

Faith, friendships and feuds on the forums: cohesion and incivility in religion discussions on Usenet's early online social networks

A thesis submitted by Timothea Kinnear for the award of Doctor of Philosophy

Faculty of Education, Theology, and the Arts

St Mary's University, London

© Timothea Kinnear 2025

Abstract

Faith, friendships and feuds on the forums: cohesion and incivility in religion discussions on Usenet's early online social networks

This research uses computational methods to analyse historical interactions about religion on Usenet, an early online social networking service that prefigured modern social media. It uses an archive of messages from c.1981-1991 and later archives from c.2003-2015. I define and analyse how cohesion and incivility manifest in the interactions, the extents to which they occur, and factors which are associated with these social processes.

The research is designed at a multidisciplinary intersection of the study of religions, Internet history, and computational methods. It is foremost religious studies research with a particular focus on methodology and digital networks. The study of religions provides the concepts of community and authority, to which I relate my findings, along with insights about how digital interactions are part of broader personal and social lives through which religion is expressed and experienced. Internet history provides technical detail about early computer networks and the social and scientific contexts in which they emerged. This 'digital substrate', as I term it, shapes how participants engage on Usenet. Internet history also provides methodological ideas and critical evaluation about how archives of online interactions can be used for research. Computational methods — here, the language R — enable me to form ethically sensitive datasets from archives, identify occurrences of social processes, and illuminate associations between facets of the dataset. For example, this includes identifying differing patterns of participation across online social networks (discussion groups) within the sites researched.

I find that computational methods are useful for large scale analyses so long as researchers understand how the computer code in those methods functions and they mitigate for imprecision. The methods are suited to identifying general trends, patterns, and associations, which can then be triangulated through manual analyses. I specify ways in which cohesion and incivility do indeed manifest, some factors associated with them, and how these inform academic understandings of community and authority when people connect online to discuss or practise religion.

I conclude that the sort of methodology used and the ethical approach embed within it are suited to analysing a broader range of social processes in the same and other similar historical online contexts. I also argue that while Usenet has been surpassed by other social media, its archives remain relevant since (1) they capture interactions spanning four decades of history, and (2) the affordances of historical online social networking, and the ways in which people behaved, are relatable to current online text-based social networking environments.

Timothea Kinnear, 23 September 2025

Submitted for the award of Doctor of Philosophy, St Mary's University, London

Acknowledgements

The PhD is a personal and professional development programme, with highs and lows, and my supervisors and friends know this is true. I am grateful to those who supervised me, Prof Stephen Bullivant, Prof Jacob Phillips and Dr Helen Costigane. I am also thankful to colleagues who have supported me in various ways, including Dr Giulia Evolvi, Dr Alana Harris and Prof James McGrath. I am likewise grateful to The Ven Dr William Jacob and friends Jon B, Christine B, James B, Lauren E, Alison H, Mark H, Lynn H, Britain K, Bernadette M, Emma M, Michelle K, Emsy P, Rachel R, Jordana R, Diana S, Zach S, and Vaughan T. I research social networks and deeply appreciate how one's flourishing depends on the networks in which they are embedded.

Contents

ABSTRACT	2
ACKNOWLEDGEMENTS	3
CONTENTS	4
LIST OF TABLES	9
LIST OF FIGURES	10
LIST OF ABBREVIATIONS	12
INTRODUCTION	13
0.1 Situating the research	16
0.2 Overview of chapters	23
CHAPTER ONE — LAYING THE FOUNDATION	27
1.1 How the early internet affected online social networks	27
1.1.1 Key facets of ARPANET	28
1.2 Usenet	30
1.2.1 The development of Usenet	30
1.2.2 The three facets of digital substrate within Usenet	32
1.2.3 Usenet’s affordances and their influence on Usenet OSNs	34
1.2.4 Usenet’s growth, popularity and change over time	38
1.2.5 Behavioural norms on Usenet	41
1.2.6 OSNs form part of social lives	44
1.2.7 Religion on Usenet	46
1.2.8 Selecting my research sites	61
1.3 Theoretical concepts	64
1.3.1 The choice of community and authority as theoretical concepts	64
1.3.2 The religious-social shaping of technology	67

1.4 Community in religion online	68
1.4.1 Relationship and commitment indicate community	70
1.4.2 Participants sense community	72
1.4.3 Sets of characteristics of community online	73
1.4.4 Communities and OSNs as networks of people	77
1.5 Authority in religion online	79
1.5.1 Affirming existing authority online	81
1.5.2 Challenges to affirming existing religious authority online	83
1.5.3 The emergence of new authorities online	84
1.6 Conclusion	85
1.6.1 Research questions	86
CHAPTER TWO — FASTER THAN WE CAN RUN	89
2.1 Using computational big data methods to study religion	89
2.1.1 How big data computational outputs represent reality	90
2.1.2 How big data computational methods fit into the academy’s methodological toolkit	91
2.1.3 Using big data to study religion	93
2.1.4 ...and to study Usenet	95
2.1.5 The strengths of using a programming language	97
2.2 Internet archives and platformisation	99
2.3 Extracting and organising messages from the archives	101
2.3.1 Downloading, extracting, and organising archived OSN messages	102
2.3.2 Selecting messages from UTZoo	104
2.3.3 Choosing the data to extract from the large text archives	106
2.3.4 The methodology used for extracting analytic descriptors from message content	108
2.4 Conclusion — summary of processes undertaken in assembling my datasets	113
CHAPTER THREE — NETWORKS AND SOCIAL PROCESSES	116
3.1 Religions are networks	116
3.1.1 Methodological approach to social network analysis	118
3.2 Computational/machine learning processes for identifying cohesion and incivility	120
3.2.1 Computational analyses to identify incivility	121
3.2.2 Computational analyses to identify cohesion	123

3.3 Machine learning to identify associations between social processes that occur on the OSN	126
3.4 Triangulating and enriching my research through manual analyses	129
3.4.1 Selecting data for the manual analyses	130
3.4.2 The manual analysis process	131
3.4.3 Ways in which manual analysis affirmed and challenged the computational methodology	133
3.4.4 Uses for the manual analysis findings	135
3.5 Mitigating challenges to the methodology and affirming its strengths	136
3.5.1 Imperfection in computational analyses	136
3.5.2 Factors not visible in my datasets	137
3.5.3 Limitations in how archives represent social realities	138
3.6 The ethical principles embedded in my research	140
3.6.1 Embedding ethical safeguards by design	143
3.6.2 The impacts of the researcher	145
3.7 Setting out the research design	146
3.8 Conclusion	147
CHAPTER FOUR — ARCHAEOLOGY OF THE DIGITAL NETWORK	149
4.1 The emergence and development of religion on early online social networks	149
4.2 Themes and subjects in early religion-focused interactions	154
4.3 Interaction	165
4.3.1 net.religion OSNs	166
4.3.2 talk.religion OSNs	170
4.3.3 soc.religion and soc.culture groups	173
4.4 Cohesion and incivility	177
4.4.1 Cohesion	178
4.4.2 Incivility	181
4.5 Applied case study — religion and science on the early Usenet	185
4.5.1 Cohesion in science and religion interactions	186
4.5.2 Incivility in early science and religion interactions	190
4.6 Conclusion	192
CHAPTER FIVE — DIGGING DEEPER INTO DATA	195

5.1 Abrahamic faith OSNs	196
5.2 Computational analyses to illuminate online social lives — characteristics of Abrahamic faith OSNs	202
5.3 Social network analyses and crossposting	213
5.4 Machine learning insights	217
5.4.1 First time participation	218
5.4.2 Incivility and longer-standing participants	219
5.4.3 Incivility and newer participants	221
5.4.4 Longer-standing participants' contributions to inter-group discussions	222
5.4.5 Inter-group discussions attracting more participation	223
5.4.6 In summary — what insights has machine learning delivered alongside the manual review of messages?	223
5.5 Applied case study — cohesion and incivility in a tight-knit OSN focusing on Judaism	224
5.5.1 Overview of exploratory analyses	225
5.5.2 Connections with other OSNs	228
5.5.3 Machine learning insights	229
5.6 Conclusion	230
CHAPTER SIX — CONCEPTUAL CONCLUSIONS AND NEXT STEPS	234
6.1 Digital substrate	235
6.2 Usenet's affordances	238
6.3 The persistence of shared aims and/or behavioural norms	240
6.4 Cohesion	241
6.4.1 Shared (religious) aims	241
6.4.2 Cooperative discussions	243
6.4.3 High signal to noise ratio	244
6.5 Incivility	244
6.5.1 Antagonistic discourse	244
6.5.2 Flooding undermining social structure	246
6.6 Major findings in relation to community and authority	247
6.6.1 Regarding community	248
6.6.2 Regarding authority	250

6.7 Significance of findings for future research	251
6.7.1 Using the methodology to analyse various social processes	252
6.7.2 Researching other online discussion group platforms alongside Usenet and Google Groups	252
6.7.3 Researching other subjects	253
6.7.4 Collaborating with others	253
6.8 Conclusion	254
REFLECTIONS IN CONCLUSION	255
7.1 Deploying computational methodology and ethical safeguards in my research context	255
7.2 Metareflexivity - coding to analyse digital artefacts	257
7.3 Interdisciplinary work in fast-moving fields	258
TECHNICAL APPENDICES	261
TA 1.1 Common facets and challenges of ML NLP, with reference to a dominant approach	261
TA 1.1.1 Clustering documents	266
TA 1.1.2 Producing word embeddings	267
TA 1.2 Association Rules Mining	269
TA 2.1 Anonymised extract of 1,000 Thread Dataset	273
TA 2.2 Sample analysed to evaluate the religion discussion classifier	273
TA 3 Extracts of the Message Dataframe	279
TA 4 Extracting data from the archives and organising them into datasets	286
BIBLIOGRAPHY	291
APPENDIX – EVIDENCE OF ETHICAL APPROVAL	314

Word count excluding bibliography and appendices: 93,104

List of tables

Table 1: approximate number of communications sent over Usenet per day in select years	39
Table 2: estimated statistics of Usenet newsgroup/OSN activity in September 1987	154
Table 3: the number of messages within my Message Dataframe from 1981-1993 which did and did not receive replies, so far as is possible to infer	178
Table 4: top ten dates on which unanswered messages were posted 1981-1993, as reflected in the Message Dataframe	184
Table 5: summary data regarding messages in select archives of soc.religion and talk.religion OSNs incorporated into the dataset	198
Table 6: summary data regarding participants and messages posted across years captured within the 2000s and 2010s	203
Table 7: numbers of crossposted messages, and the proportion in discussion threads that include first time participants	205
Table 8: average numbers of messages in threads from the 2000s and 2010s in the dataset, including figures for threads including first time participants	207
Table 9: proportions of messages posted by longer-standing participants and proportions including/excluding words associated with having shared religious aims	209
Table 10: the proportion of messages including and excluding keywords indicating shared religious aims which were crossposted	211
Table 11: average numbers of first-time participants in each discussion thread across the OSNs analysed	212
Table 12: the number of messages crossposted to multiple groups including SCJM within my dataset, excluding patterns occurring fewer than ten times	228

List of figures

Figure 1: illustrations of (left) a more-centralised network and (right) a less-centralised network with multiple potential paths between some nodes	28
Figure 3: connections established between net.religion and other OSNs from July to September 1983 as captured in my dataset.....	167
Figure 4: connections between net.religion and other OSNs from October to December 1983	167
Figure 5: connections between net.religion groups including net.religion.jewish and others from January to March 1985	168
Figure 6: connections between net.religion and other OSNs between January and March 1983	169
Figure 7: connections between net.religion groups and others in January 1986	169
Figure 8: discrete clusters formed by messages crossposted between talk.religion OSNs and others from April to June 1987	171
Figure 9: interactions between talk.religion groups and others in January to March 1991	172
Figure 10: connections between talk.religion OSNs and others from October to December 1987	173
Figure 11: connections between soc.religion OSNs and others from January to March 1991	174
Figure 12: connections between soc.culture.jewish and other OSNs from January to March 1987	175
Figure 13: connections between soc.culture.jewish and other OSNs from January to March 1988	175
Figure 14: connections between soc.culture.jewish and other OSNs from January to March 1989	176
Figure 15: connections between soc.religion.islam and other OSNs from January to March 1990	177
Figure 16: connections formed soc.culture.jewish and others between October and December 1987.....	180
Figure 17: connections formed soc.religion groups and others between July and September 1990	180
Figure 18: connections formed between soc.religion OSNs and others between April and June 1991.....	181

Figure 19: social network graphs showing connections between soc.religion.islam and others from July to December 2005	216
Figure 20: social network graphs showing connections between soc.religion.islam and others from January to June 2006.....	216
Figure 2: a screenshot of dataset column headings indicating the dataset structure.....	286

List of abbreviations

AOL	America Online
ARM	Association Rules Mining
ARPA	Advanced Research Projects Agency
ARPANET	Advanced Research Projects Agency Network
BERT	Bidirectional Encoder Representations from Transformers
FAQ	Frequently Asked Questions
GloVe	Global Vectors for Word Representation
INWG	International Network Working Group
IP	Internet Protocol
IPTO	Information Processing Techniques Office
LDA	Latent Dirichlet Allocation
ML	Machine Learning
NLP	Natural Language Processing
NNTP	Network News Transfer Protocol
OSN	Frequently Asked Questions
SCJM	soc.culture.jewish.moderated
SNA	Social Network Analysis
SRC	soc.religion.christian
TCP	Transmission Control Protocol
TRS	Theology and Religious Studies
UHC	Usenet Historical Collection
UTZoo	Usenet archive compiled at the University of Toronto Zoology Department
UUCP	Unix-to-Unix Copy Program
VPN	Virtual Private Network

Introduction

For many years I volunteered at a Christian outreach charity in London which, across three decades, has brought together hundreds of people from diverse backgrounds to learn from one another and to build Christian fellowship. We used online platforms to attract new members, organise social events, discuss Christianity and help members remain connected. Participation led people to find and deepen their religious faith, build lasting friendships, and increase understanding across divides. However, the diversity of participants presented challenges such as ensuring that people felt welcome and that disagreements were managed with civility. These dual potentials of community and conflict are not exclusive to traditional, 'offline' networks. They are also widespread in online social networks (OSNs). As Pope Francis put it in his 2019 message for World Communications Day, '[in] the best cases, these virtual communities are able to demonstrate cohesion and solidarity', while he also observed that online 'we define ourselves starting with what divides us rather than with what unites us, giving rise to suspicion and to the venting of every kind of prejudice' (Francis 2019).

Religion has been discussed and practised online since at least 1982 (Rheingold 2000, 136-138; Helland 2007) and this grew substantially in the mid-1990s when companies including America Online and Microsoft facilitated internet access for the general public. Consequently, OSN services such as discussion boards, instant messaging and mailing lists have afforded communication regarding religion around the world for more than four decades; these services have been adopted widely by those wishing to connect with others to promote and discuss religion (Vogt 2011). The phenomenon is vast and growing and these communications are crucial to understanding, through deep and longitudinal analysis, how online relationships, arguments, anger, prayers, and debates about religion have brought people together, drawn boundaries, and sparked conflict.

While many people are aware of services such as The Wayback Machine for viewing historical webpages, they may conclude that the transient, ephemeral messages sent across historical OSNs are lost to time. However, this is not necessarily the case. Many millions of these messages have been archived since at least 1981, though researchers who wish to mine them for insights are faced with significant barriers. Early archives are fragmentary and were recorded on magnetic media, limited by storage capacities and curated by a small number of enthusiasts. In the 1990s, commercial and not-for-profit organisations started collating archives and some social network service providers archived

messages sent across their services, creating broader and fuller repositories. These archives are enormous and pose serious technical challenges to the researcher who wishes to analyse them. However, removing barriers to accessing and analysing archived OSN messages, while maintaining appropriate ethical safeguards, would unlock myriad opportunities for researchers to study online discussions across many years or decades and identify how engaging in these social networks impacted participants' online social lives.

An approach drawing on three disciplines is required to transform these archived OSN messages into a useful resource for research and to derive from them robust insights into the impacts of religion online. First, concepts of community and authority from digital religion scholarship provide valuable perspectives for understanding people's engagement within OSNs. Second, the field of internet histories documents the technological developments that underly internet infrastructure and which affect the social networks that develop on it. In addition, some scholars in this field are developing tools to increase access to archived internet data. Third, computer programming including machine learning is required to undertake large-scale research into the size of the OSNs, the connections people formed within them, the substance of the messages sent over the networks and how these characteristics affect engagement between participants.

My doctoral research utilises a methodology which uses computer programming, supplemented by manual analyses, to achieve all key steps for mining insights from select types of message archive. Specifically, this innovative methodology:

- Extracts and organises messages from OSN archives
- Visualises the social connections formed between people as they exchange messages
- Infers some subjects of discussion in the messages
- Identifies and analyses
 - facets of social cohesion and potential indicators of incivility within the networks
 - patterns of participation in the networks
- and it indicates whether factors such as certain topics of discussion or network size are associated with cohesion, incivility and differing patterns of participation

The methodology has immediate applications and future potential. The insights I obtain through applying it to case studies of OSNs could help religious groups cultivate constructive dialogue online and will be of interest to others who are concerned with the positive and harmful aspects of social networks. In future, the methodological approach could be incorporated into software that enables other researchers to analyse historical OSNs that focus on many different areas of theology and religion. This could provide pertinent data for research into religious groups and discourse in the late twentieth and early twenty-first centuries, complementing other research methods. I am therefore motivated to develop this methodology both for my own research and — ultimately — to help unlock the means of analysing historical OSNs for other researchers, without them needing to employ specialist computer programming.

This dissertation focuses foremost on explaining and defending this methodology to theology and religious studies (TRS) scholars while also presenting detailed analyses of selected archives of religion discussion online. It forms a substantial 'proof of concept' that this methodology is effective and delivers datasets and insights that are valuable to TRS scholars and others. This focus on methodology is necessary for three reasons. First, so readers can follow the methodological steps that afford the empirical analyses presented in chapters four and five and have confidence in the conclusions reached. Second, the methodology and the insights that it produces represent a doctoral level contribution to TRS as, to quote the requirements for the PhD, I 'conceptualise, design and implement a project for the generation of new knowledge' which is 'at the forefront of an academic discipline', using advanced techniques (St Mary's University 2020, 6). Third, a detailed understanding of the methodology is necessary to appreciate how it provides a foundation for the further research opportunities discussed in chapter six. The remainder of this introduction briefly introduces me as a researcher, situates the research within its interdisciplinary context, and charts the path ahead by outlining the chapters of this dissertation.

My passion for this project was sparked by my interests in computer programming and the study of religions. I cultivated my passion for TRS through my master's degree in the study of religions at Lampeter, during which I attended events organised by the charity INFORM, then based at the London School of Economics. INFORM's events bring together members of minority religions, their families, academics, and others whose work engages with religious groups. Through talking with people at these events I saw the potency of religion to shape lives profoundly by connecting people and catalysing fellowship and support for one another, but also erecting boundaries and sparking conflict. I knew that many such

interactions have occurred online since the 1980s, which led me to identify that applying computer programming to archives of these interactions can enable analysis of religion online at a scale hitherto impossible. This project has required me to augment my coding skills by learning the programming language R and deepen my understanding of the technological structure of the computer networks that afford online communication. However, it is my background that equipped me to spot the opportunity for this research and to work at this interdisciplinary intersection.

0.1 Situating the research

This research is foremost a contribution to digital religion studies within TRS. This is affirmed in several ways. First, the dissertation ultimately relates its findings to the concepts of community and authority, discussed in sections 1.4 and 1.5, which are longstanding and well-developed theoretical concepts in the study of digital religion. Further, my understanding of the theoretical perspective of the religious-social shaping of technology (pp.67-68), developed within digital religion studies, framed my research by emphasising how religious individuals and groups can have agency to shape technology to accommodate their religious practices, beliefs and aims. In addition, the academic community had observed four waves of digital religion research in the years prior to when I developed my research design. These are (1) descriptive research, in which scholars explored, documented and described the religion being expressed and taking place in (then) novel online spaces; (2) categorial research which seeks to identify commonalities between expressions of religion online across diverse technology platforms (ranging from very simple text interfaces to three dimensional graphical spaces), including the use of categories such as 'authority', 'identity', 'community' and 'ritual' to conceptualise these new expressions/manifestations of religion; (3) theoretical research, in which these concepts were developed further, impacts of religion online were assessed, combined online and offline expressions of religion (such online streaming) were researched, and theoretical approaches to (for example) digital ethnography are further developed; and (4) the integrated/convergent, in which further methodological models and typologies to categorise and interpret data are developed, and a need for longitudinal research is emphasised to identify how religious individuals participate online and shape online contexts, and impacts of this on religions and wider society (Campbell and Altenhofen 2016, 3-11). Each subsequent wave expanded on research design that came before, rather than replace it, and my research relates to waves one and four. It relates to the first wave as it is descriptive in its empirical analysis of the development and substance of religion-focused OSNs on

Usenet. Regarding the fourth, framed by my understanding of RSST, it shows how religious individuals and groups joined Usenet and then used its technologies to shape/adapt the OSNs they created or used, to enable discussion relevant to their religious interests or aims.

Together, the concepts of community and authority to which my work relates, the theoretical perspective of RSST that framed my thinking in designing the research, and the ways in which my research relates to two of the above four waves, help affirm the positioning of my research as a study in digital religion.

In addition, to a far lesser degree, the research also engages with communication studies and media studies, as indicated by my research being positioned in the communications stream of a research colloquium at St Mary's University and my successful application to present at a doctoral colloquium for the International Society for Media, Religion, and Culture, both in 2023. The methodology is also relevant to sociology, as affirmed by my presentation of research using the methodology in the sociology of religion stream at the American Academy of Religion Annual Meeting in 2023. It also engages with internet history research and computer programming, which I specify in more detail in the chapters that follow.

Digital methods are multifaceted, including referring to the use of digital hardware/software when undertaking otherwise non-digital research (such as using software to transcribe interviews recorded via a laptop); research in digital environments such as social media networks; and research using coding or other digital techniques to analyse 'born digital' or digitised research objects (Campbell and Altenhofen 2016, 11). My research fits into the two latter of these categories. Further, my research is situated within the Faculty of Education, Theology, and the Arts at St Mary's University. Digital theology is its own multifaceted and developing discipline. Phillips, Schiefelbein-Guerrero, and Kurlberg propose four levels or waves of digital theology and my research is relatable to the second of these, namely '[theological] research enabled by digitality or digital culture', in which 'research now includes the analysis of (big) data, distant reading of multiple texts, online religious practice, and the visualization of data' (2019, 38). While my research is not foremost *theological*, I analyse online theological discussions using exactly the types of approach specified for this level/wave so it may be relevant/of interest to digital theology scholars. I am encouraged by Phillips' (2023) exploration of how Anselm's definition of theology incorporating faith seeking understanding could include seeking understanding of

data and of the embeddedness and impacts of digital devices (Phillips 2023, 771). My work may support this.

Two definitions are required to situate my research within digital religion studies more specifically: the first, to demarcate the type of online communications in focus here and the second to provide a term that captures the breadth of insights the research will generate regarding cohesion, incivility, and participation in OSNs.

For the purpose of this research, an *online social network* is formed when:

- Contact between two or more people is mediated via an online communications service such as a forum, social networking site or mailing list
- The service affords asynchronous communication between participants and maintains a (semi-)permanent record of messages sent
- The networks of people that form on the service are potentially open to new participants, with or without an approval process, and
- The service allows those networks to be named (and thus differentiated)

This definition of OSN excludes voice and video phone calls but includes services such as online message boards/forums including Reddit, Google Groups, and mailing lists (like those used by academic associations). Service providers have many OSNs operating in parallel on their services. For example, in 2021 there were more than 1.8 billion groups on Facebook Groups (Facebook 2021).

OSNs therefore have a network *structure*, which can be visualised as a map of individuals who are connected to others in the network by virtue of sharing communications with them. The methodological approach of my research elucidates and analyses network structures because a person's social networks can significantly influence their behaviour and social experiences, including in the sphere of religion (Everton 2018). The *substance* of discussion within networks may be on any topic but my research is relevant to networks that are used to discuss some aspect of theology or religion. This includes networks formed to connect members of a particular church or congregation, ex-members of the same, diasporic members of a religion, or networks set up to discuss religion by anyone who wishes to learn, debate, or antagonise. Therefore, participants need not all be members of a particular (or any) religion, nor need all discussions be constructive and civil. This dual focus on structure

and substance affords valuable insights into participation, cohesion, and incivility in discussions of religion online, as will be shown.

Within the range of possible applications of the methodology I utilise, this thesis focuses on select online religion discussion groups mediated via Usenet. Usenet is an online communications service formed in 1979, via which people connect and post messages to discuss myriad topics with others around the world and which grew to millions of users (Hauben and Hauben 1997, 161) as access to internet services expanded in the 1990s. Usenet was an important component of the early internet and a prominent platform. More than 150 million messages were sent over Usenet during the year 2000 (Smith 2003, 53), it remained active during the 2000s and 2010s and is still in use today, often for sharing files. The first Usenet OSNs focusing on religion were formed in 1983 (Helland 2007). These networks proliferated into the 1990s and, as my research shows, continued to attract sizeable numbers of participants into the 2000s and 2010s. Initially, participants used dedicated Usenet software and network connections but many later users from the 1990s and 2000s onwards accessed Usenet OSNs more easily via software such as Microsoft Outlook and Web-based services such as Google Groups from their normal PC, Mac, or mobile device. Chapter one discusses in more detail how services including Google Groups connected with Usenet to broaden access to its social networks. My focus on Usenet is justified by its historical importance, lasting impacts and because the insights gleaned from its social networks are relatable to other online contexts such as Twitter and Reddit; chapter one explains these justifications in detail.

As noted above, it is also necessary to define a term for the insights the methodology delivers into some of the social phenomena that occur within OSNs. I refer to these insights as *measures of engagement*, which capture:

- Cohesion within the network, reflected in the social connections and affinity that participants build with one another
- Incivility within the network, reflected in the language people use when talking with one another and which may impact their social connections
- Patterns of participation in discussions on the network

These measures reflect impacts that engaging in OSNs can have on participants' lives — building relationships, receiving positive and negative (even harmful) messages from others, and incorporating participation into one's social routines. I identified these impacts

from my literature review, then draw on extant literature to define them and set out how they manifest in religion on Usenet (pp.59-61). Further, these measures can be derived from the archived messages sent across the networks. Therefore, along with being established in other literature, they are amenable to study at large scale, potentially producing new insights through the sort of methodology I employed. This provided a suggestive starting point that cohesion and incivility would yield insights relevant to the study of digital religion.

Specifically, I find that insights gained from studying cohesion and incivility are of direct relevance to research that studies digital religion through the lenses of *community* and *authority*. I present conceptual conclusions that specify links between them in chapter six. Chapter one considers community and authority in detail, but the concept of community may be illuminated here with reference to Heidi Campbell's pioneering (2005) research, focused on digital religion, which considers sociological concepts of community, drawing on Tönnies and Parsons among others. A prominent theme among these is that community relationships entail interdependence beyond a narrow transactional exchange found in, for example, the relationship between a consumer and a producer of goods. Community relationships may inform a person's identity, provide emotional support, and allow people to know each other deeply (Campbell 2005). Consequently, in light of the above definition of OSNs, not all OSNs become communities but potential is there. Further, digital religion literature uses the term community to refer to a broad range of online interactions including text-based forums, online churches in which users interact with others, social media services such as Facebook and more (cf. Campbell and Vitullo 2016). This literature researches characteristics of these interactions such as the maintenance of boundaries, behavioural norms, social cohesion, and incivility and conflict. Even if an OSN does not become community it will still exhibit many of these characteristics, as I explore in this dissertation. Therefore, the insights delivered by my research benefit from, and contribute to, literature regarding digital religion communities.

Digital religion scholars also research how pre-existing sources of authority manifest in online communities and how new authorities develop within, or are imposed upon, these communities (cf. Cheong 2014; Kołodziejaska 2018). My work utilises and contributes to this research. Chapter one explores how social networks can afford people authority, such as a person who acts as a moderator of messages sent across a network, a gatekeeper who permits or blocks new members, or a person who has a high number of connections to others within the network and acquires respect as an authority within the group. The chapter also explores how authority can manifest in discussions on the network, for example how

religious texts or leaders are affirmed online and how people challenge them. Facets of authority are then explored through the case studies I present and discuss in later chapters.

While my research relates foremost to digital religion scholarship it also draws on research from the field of internet histories to understand how early decisions about how computers would connect, and how people would use them to interact, would shape later OSNs. This infrastructure has ongoing impacts and relevance as evidenced in Schneider's (2021) analysis of how early system operators occupied roles of significant power over the connections made and the discussions held on their networks, with similar power now held by the controllers of social media platforms. I refer to this hardware and software infrastructure as the 'digital substrate', for while it does not dictate all facets of the social networks that develop upon it, it may nevertheless influence them. Chapter one analyses relevant aspects of foundational internet infrastructure to show how they shape the OSNs studied in my research.

Since the late 1970s, OSNs have developed on computer networks that are in the lineage of the modern internet. In that time, many individuals and service providers have been custodians of the data that travel across those networks. Today, many social networks are formed on proprietary platforms such as Facebook and X (formerly Twitter) and researchers' access to these data is restricted, whereas others form on open platforms that afford researchers more access to data. However, in all cases the amount of data, in the form of messages sent over the networks, is vast and requires computer programming methods if large-scale analyses are to be undertaken. These barriers to accessing and analysing data are reflected in Possamai-Inesedy and Nixon's (2017) discussion of the need for collaborative research that is alert to the power relations between academic researchers and large social media companies; they observe that the uptake of computational methods has been slow in the social sciences and, citing Hutchings, in digital humanities. Relatedly, Burrows and Savage (2014) note that much empirical sociological analysis of social media takes place outside the academy by those who have access to the data. These barriers to researchers accessing data and the concomitant need for computer coding skills hamper large-scale research of OSNs within digital religion studies.

My use of archives of OSN messages mitigates the problem of data access for my research, but scholars in the field of internet histories remark that working with archives is difficult and presents barriers to entry for those who do not possess the necessary skillset (Winters 2017; Brügger and Milligan 2019). Writing in 2016, internet historian Ian Milligan noted the

lack of tools available for analysing archived World Wide Web data (Milligan 2016); since then, his work with collaborators at Archives Unleashed has made substantial strides in affording researchers access to such tools. Notwithstanding this, archives containing many millions of messages sent over OSNs, which are distinct from archives of Web pages, remain in a state inaccessible to most researchers due to several reasons. These include (a) the diverse computer file formats used, some of which are obsolete; (b) the differing ways information such as the message sender and date are recorded; and (c) the very cumbersome experience researchers face when trying to read and manipulate an archive on their computer, if it can be opened at all due to its size. The consequences of this are expressed neatly in Paloque-Bergès' (2015) paper which documents her use of the Google Groups website to access messages sent across social networks on Usenet. The website required her to search for and read messages in the Google interface and afforded no easy access to the underlying archives, which in practice is highly beneficial for large-scale analyses.

Specifically, my research necessitates coding skills to extract messages from the archives and manipulate them, visualise the connections formed between OSNs, analyse the substance of participants' messages, and capture the measures of engagement summarised above. There is longstanding recognition among scholars of the need for large-scale digital religion research, as in Lövheim and Linderman's (2005) chapter which advocates for longitudinal studies to analyse how online communities manage the tensions between remaining congenial while accommodating people's differing views, and Lövheim and Campbell's (2017) paper which perceives the need for more large-scale quantitative methods for studying digital religion. Cooper, Kolog and Sutinen (2020) undertook interdisciplinary collaboration (as promoted by Possamai-Inesedy and Nixon (2017)) to analyse 1,004 Twitter messages regarding church in London and assign the messages into eight topic categories. This research illustrates what may be achieved using computer programming methods and indicates the potential for such analyses to illuminate facets of churchgoing in the city. However, to my knowledge no extant research incorporates the full range of coding tasks that are necessary to study fully the structure of historical OSNs, the discussions held on them, and the factors that influence the measures of engagement I outline above. My research utilises such a methodology by, in summary, drawing on the following fields:

- Digital religion research into community and authority in online groups

- Internet histories research to understand the impacts of foundational internet technologies on OSNs
- Related internet archiving work that seeks to organise and analyse historical internet data and make it available to researchers, and
- Machine learning methodologies, which enable the analysis of large datasets of messages

This interdisciplinary approach allows me to make an original contribution to the field of digital religion studies through the empirical insights regarding religion on early OSNs from 1983-1993 and select later networks as presented in chapters four and five. These findings are made possible using internet archives and the particular methodology deployed.

The chapters of this dissertation present the project chronologically. First, the foundation is laid by locating Usenet in the historical lineage of the internet and social media. This elucidates how internet infrastructure influences social networks that developed on these services. I then situate the research through analysis of select scholarship on digital religion and social media more broadly, with particular attention to the concepts of community and authority. This culminates in research questions that focus the development of the methodology and the selection of research sites. I then explain and justify the methodology for (a) accessing and extracting messages from the selected archives and (b) diving deeper into the data to understand participation in the social networks and then identifying the measures of engagement. The dissertation then presents analyses from two complementary case studies that apply my methodological approach, before interpreting and discussing those results to identify how they contribute to digital religion scholarship and to indicate how my methodological approach may be used in future.

0.2 Overview of chapters

Chapter one starts by analysing technical and social developments of the internet that culminated in Usenet. This is essential groundwork for my subsequent discussions of the structure of the social networks that formed on Usenet and the behavioural norms participants adopted, with early decisions having lasting impacts as people joined Usenet social networks into the 1990s and 2000s. The chapter then critically evaluates digital religion research and scholarship from the wider field of social media research to identify how concepts of community and authority illuminate the impacts OSNs have on participants' social lives, and how my measures of engagement contribute to measuring and

understanding these impacts. These discussions of Usenet's history and digital religion scholarship allow me to justify Usenet OSNs as a focus for my research, since they span four decades and include technological and social developments that remain relevant as they find close analogues in topic-specific discussion platforms such as Reddit which are popular today. Consequently, Usenet OSNs provide *massive* and *relevant* datasets that illuminate the social dynamics that occur when people connect to discuss religion online. The chapter culminates in the formulation of research questions that specify exact research sites within Usenet and guide the research to make valuable methodological and empirical contributions to knowledge.

Chapter two focuses on the challenges of internet archiving, how I accessed large datasets of historical OSN messages and then extracted, organised, and applied analytic descriptors to the messages therein. The chapter opens with a critical review of how computer programming 'big data' methods and the outputs they generate benefit the study of religion. I then discuss internet archiving research. This fast-moving field seeks to store and organise archived internet data and make it available to researchers, which will be highly useful to present and future generations who seek to understand the world in the late twentieth and twenty-first centuries (Schroeder and Brügger 2017, 1). However, internet archiving presents many challenges. These include the large size and diverse format of datasets, ethical and legal considerations, organising and extracting data and providing interfaces so researchers can analyse data without advanced technical skills. The chapter explains how the work in this field informs the methodology and indicates how my research has the potential to contribute to this field. The challenges of internet archiving lead me to justify the need for a computational methodology using machine learning in order to analyse archives at scale and deliver insights that contribute to digital religion research. The chapter then explains the components of the methodology that extract, organise, and summarise data that are *explicit* in the archived messages. These include the date the message was sent, the subject line, the social network(s) to which the message was posted, analytic descriptors indicating discussion subjects and language that indicates cohesion, and an anonymised version of the sender's name. This entails reflection on the epistemological trickiness of using algorithms to study digital networks.

Chapter three explains the components of the methodology that extract *implicit* data from the archived messages. It starts by explaining processes for quantifying and visualising the social network connections that are formed when people send messages to one another across networks. This entails discussion of social network analysis and the strengths and

weaknesses of different algorithms in light of the present research questions. The chapter then explores how the measures of engagement can be inferred from the messages. This requires careful consideration of how the data include proxies that indicate the presence of cohesion, some potential sites of incivility, and patterns of participation and therefore affirm the relevance of the research sites for understanding them. The chapter explains the machine learning used to identify associations between these measures of engagement and other factors. For example, perhaps different sizes of social networks are associated with differing patterns of participation. These insights are made possible by the preceding steps of the methodology and are valuable for generating insights that are valuable to studies of community and authority in digital religion research. Finally, I argue for the strengths of my approach in light of challenges that may be made to it and I discuss the novel ethical challenges of large-scale research using archived social media data.

I then apply the methodology to two complementary research sites. The first, in chapter four, uses a dataset of messages posted to Usenet between 1981 and 1993, drawn from an archive compiled at the University of Toronto. These are some of the earliest online discussions regarding religion on networks that are in the lineage of modern social media. The dataset is far from complete due to the very early years it captures and the archivist's informal approach; this research site is akin to an archaeological analysis that has access to limited numbers of artefacts representing a bygone world. Nonetheless, it illuminates the growth of early religion-focused social networks, some interaction between those networks and others, and the subjects of their discussions. A case study is undertaken within the research site, focusing on interactions about religion and science (pp.185-192). This very early archive also places today's social media in historical context, which benefits analysis of the second research site.

Chapter five analyses the second research site, which is a subset of OSNs that formed via Usenet during the 2000s and 2010s and which focus on the Abrahamic religions. Fuller datasets are available for these later networks than for the first research site. The datasets therefore afford detailed analyses of the social network connections that form as people message one another, interactions between participants and the measures of engagement, as well as the application of machine learning for identifying patterns and associations within the data. A case study is undertaken to identify factors that are associated with cohesion and conflict in interactions on a moderated Judaism discussion group (and its connections with other groups), providing a worked example that contributes to answering my research questions.

Chapter six interprets and discusses the two preceding chapters' analyses to affirm answers to my research questions. This discussion also specifies how the research contributes to academic understanding of community and authority in digital religion, from the specific context of Usenet OSNs, and argues that insights from deep, longitudinal analyses of archived datasets are useful for understanding online interactions today. In this chapter, I make a case that further work could be undertaken to build a platform on which researchers can apply the methodology in an accessible way to analyse a wide range of archives of OSN messages related to religion. This may be useful as analyses of these archives could complement other TRS research, including projects that would benefit from understanding the dynamics in online social groups that formed around, or discussed, any one of hundreds of facets of theology and religion over the past four decades. Finally, I offer a short conclusion to the thesis. This affirms my original empirical contributions to knowledge and includes my reflections on the 'metareflexivity' of using digital lenses to analyse born-digital artifacts, and the challenges of interdisciplinary research in fast-moving fields.

At the start of this introduction, I referred to a Christian outreach charity in London. We can now see how applying the methodology to archives of this group's (or others') online social network interactions could help us understand the social dynamics that cultivated cohesion among participants. We could identify whether the varying size of the OSN over time led to more or less participation, or whether certain topics of discussion appeared to build affinity between participants. My hope is that the methodological approach I utilise can help others within and outside the academy to analyse online discussions regarding religion and deepen their understanding of the factors associated with constructive and adverse outcomes. These insights become ever more important in our increasingly connected digital world.

Chapter One – Laying the Foundation

This chapter lays the foundation for developing the methodology and applying it to specific research sites. First, I review certain early internet technologies that profoundly impacted early online social networks (OSNs). This leads me to evaluate Usenet, including a review of its affordances and the behavioural norms that developed on it, following which I review studies of religion on Usenet. This enables me to specify and operationalise definitions of cohesion and incivility. I then critically evaluate literature regarding community and authority in religion online, discussing how these relate to my ‘measures of engagement’. This provides important perspectives for understanding the social dynamics that occur within religion focused OSNs, while illustrating how large-scale computational methodologies can contribute to understanding these dynamics. The chapter closes with specific research questions that steer my project toward making such a contribution.

1.1 How the early internet affected online social networks

In the introduction, I use the term ‘digital substrate’ to refer to the technologies that comprise digital networks, upon which OSNs form. This section narrates some key historical components of that substrate, first regarding ARPANET and then Usenet. The history of the internet spans continents, decades, many networks, and many contributors. Much literature focuses on the importance of ARPANET in America, steered by work published in the 1990s. ARPANET first connected computers in 1969 (Hafner and Lyon 1996, 151-154) and developed through the 1970s and 1980s, foremost for government and university researchers before being available to others (Hauben and Hauben 1997, 120; Bay 2018, 150-151). ARPANET was used for transferring files and for research and social communications. Its pioneers developed and/or deployed some key technologies that, in idea or substance, remain a part of today’s internet.

I note that focusing on ARPANET to the exclusion of other early developments would be too narrow if writing a history of the internet, since the internet is a network of networks (Weber 2017; Driscoll and Paloque-Bergès 2017). Scholarship now analyses early networks in many countries, documenting important social or technological developments. The journal *Internet Histories* provides a forum for this work. Notwithstanding this, I focus on ARPANET for two reasons. First, ARPANET provided tools and a template for the development of other networks (Bay 2019b). It therefore influenced the digital substrate of the Usenet OSNs that I research and illuminating that substrate helps me understand the

influence of the medium on the messages it carries (Hipps 2009, 27-28). Second, ARPANET access was restricted and this helped spark the development of Usenet, reflecting how digital networks are afforded by both advances in technology and users' social needs and desires. Usenet provides my research sites and is lesser studied than some other networks (Haigh, Russell and Dutton 2015, 153). My selective approach to focus on ARPANET here is therefore justifiable as it informed the development of Usenet.

1.1.1 Key facets of ARPANET

ARPANET was a project of the United States' Advance Research Projects Agency (ARPA). ARPA was formed due to concerns about Soviet scientific developments following the launch of the Sputnik 1 satellite (Bermejo 2007, 57; Bay 2018, 145). Some ARPA projects provided university research groups with computers so they could advance technology in specific areas (Bay 2018, 145). Since the groups' focuses differed, they benefitted from connecting to one another's computers, for example the University of Utah had advanced graphics capabilities (Bay 2018, 145). The need for a network of ARPA computers was born and, alongside this, research into the development of networks was taking place (cf. Bay 2018, 141; Naughton 2015, 85).

ARPANET was sparked by the social networks of people who needed to connect, desires to develop a robust computer network, and the imperative to maximise the use of computers dispersed across institutions. I now summarise three key technologies that were developed for, or deployed on, ARPANET. These are important components of the digital substrate that influence the OSNs that emerge on computer networks, thereby informing my later analyses.

The first is the principle of the *decentralised network*. Networks are centralised to varying degrees, with fully distributed networks being the least centralised:

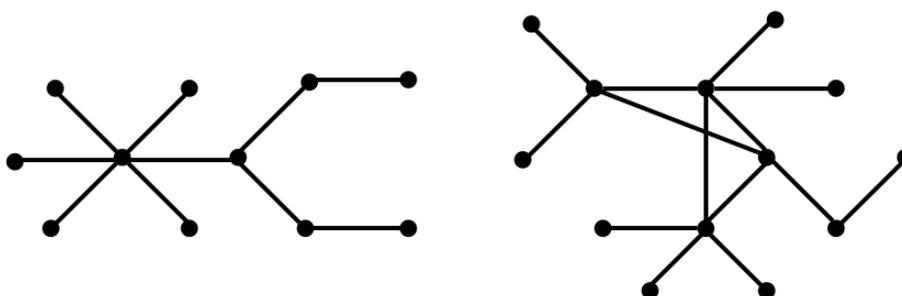


Figure 1: illustrations of (left) a more-centralised network and (right) a less-centralised network with multiple potential paths between some nodes

Each ARPANET node was typically connected to two others (Hafner and Lyon 1996, 113). On less-centralised systems such as this, it is efficient for the connections between nodes to carry multiple communications between different senders and recipients at (nearly) the same time. However, this entails the problem of conveying multiple communications across the network when each has a different optimal path of nodes to reach the correct recipient. ARPANET's solution was *packet switching*, which influenced its later widespread use across the internet (Bay 2019b, 1-2). This is the second key network technology I identify that influences OSNs. Packet switching divides a sender's communication into small components, conveying them across the network and recombining them for the recipient when delivered. Each packet, or 'message block' in Baran's (1964) paper, is stored for a minimal time at each receiving node, which seeks to forward it to another node along an efficient path to the recipient, using alternate paths if the optimal node is busy or unavailable. This requires routing tables, which indicate the next node on the most efficient network path to the packet's destination (Fidler 2019, 5).

A clear distinction between the picture portrayed so far and the global reach of the OSNs that I research is that ARPANET was bounded to certain computers and controlled centrally. The connections *between* networks must therefore come into focus. Indeed the emergence of multiple networks led inevitably to the challenge of how they could be connected (Fidler 2019). Gateways were formed, in which certain computers would receive data from one network, send it to another specified network and *vice versa* (Hauben and Hauben 1997, 50). For example, a gateway was established between ARPANET and Usenet via the University of California at Berkeley since the university had connections to both networks, allowing access to ARPANET mailing lists via Usenet (Hauben and Hauben 1997, 52). Nonetheless it is inefficient to use bespoke gateways to connect diverse networks on a case-by-case basis, overcoming language barriers between networks each time. Instead a common protocol for such 'inter-networking' was required. The work of the International Network Working Group (INWG) of experts on different networks led to the development of Transmission Control Program (TCP) and, having parted from INWG, IPTO researchers split TCP into Transmission Control Program and Internet Protocol (Fidler 2019, 12-13). TCP affords transmission of data between computers and IP handles, among other tasks, the addresses of computers across the networks, bringing standardisation and efficiency to transmitting data between computers on different networks (Fidler 2019, 13). ARPANET adopted these protocols in 1982 and, by 1989, most networks used them (Naughton 2015, 167; Hafner and Lyon 1996, 254). Known commonly as TCP/IP, updated forms of these

protocols remain in use and were crucial for helping create a network of networks that underpinned today's internet (Quarterman and Hoskins 1986, 944; Hafner and Lyon 1996, 227, 240-244).

This indicates a third component of the digital substrate that shapes the OSNs on which my research focuses, namely *gateway machines and gatekeepers* who enable, permit, or block the passage of communications. This can happen at different levels of a network. At a macro level, gateway machines connected entire networks which is important since inter-networking affords the global reach of OSNs. On the early Usenet, at a meso level, individual computers could permit or block the transmission of messages for particular OSNs, including those relating to religion. And today, at a micro level, individuals with administrator rights over a particular OSN may permit or block an individual from accessing or contributing to it. In short, gatekeepers can affect access to OSNs in ways that impact participants' online social lives.

In summary, I identify three technologies that shaped the digital substrate of networks that underpinned the early internet. The *decentralised network structure* and *packet switching* method of transmitting communications are important for enabling large and spatially distributed *social* networks of people to collaborate or discuss common interests in near real time across computer networks at larger scale than using, for example, traditional and more-centralised telephone systems. In addition, *gateway machines and gatekeepers* can permit or restrict an individual's access to, and participation in, a network. Together, these technologies help afford and shape the social networks that emerge on computer networks. They influence Usenet and continue to impact OSN participants' experiences today. I refer to these influences in my later analyses.

1.2 Usenet

1.2.1 The development of Usenet

As mentioned above, Usenet had a gateway connection to ARPANET, a network which Usenet would outlast and outgrow. Usenet's genesis lies in the Unix computer operating system, which was commonly used in computer science departments in the 1970s and remains in widespread use. Today, computers' operating systems are updated over the internet, but in the 1970s maintaining Unix proved troublesome. This led graduate students Tom Truscott, Jim Ellis, and colleagues to use components of Unix to enable online

discussion groups. In summary, they included autodial modems that connected computers via telephone lines, files of electronic messages on specific topics to which users could add, a 'Find' program to identify which files had been updated on networked computers since they were last checked and Unix-to-Unix Copy program (UUCP) which copied files from one computer to another (Pfaffenberger 1996, 366-367). This facilitated discussion groups on varied topics across multiple computers by updating and distributing the files containing messages when computers connected to one another, thereby propagating the latest messages across the network. Participants contributed to discussions, as they still do, by posting messages to one or more discussion groups, termed newsgroups, each on a specific theme. Others read messages and reply to the same newsgroup, so discussion ensues. Participants choose which groups to read, motivated by their areas of interest. For definitional clarity:

- *Newsgroups* are the repositories of messages stored and copied between computers that access and propagate them.
- A newsgroup becomes an *online social network* (OSN) on my definition (p.18), as soon as two or more people use the newsgroup to interact and consequently form a social network. In this dissertation I use the terms newsgroup (or discussion group) and OSN, depending on the context.

This network for Unix users was particularly relevant to those not connected to ARPANET, which was bounded and intended to support ARPA research projects. While first developed in 1979, the Usenet discussion group software, termed NetNews, was distributed following a Unix conference in 1980 and was intended for all Unix system users (Hauben and Hauben 1997, 39-41, 60). The first version was called A News, later succeeded by B News which could handle a larger number of messages (Fristrup 1994, 361). The network grew and by 1981 approximately 150 sites were connected. By December 1982, twelve European sites were connected, with perhaps the first being at the University of Kent at Canterbury by September 1982 (Paloque-Bergès 2017, 162). Around 1983, Kilnam Chon created Usenet connections (AsiaNet) within Japan, Indonesia, and other countries, though costs prohibited receiving news from America, so messages were sent via tape and then distributed via the network locally (Paloque-Bergès 2021, 383-384). This indicates the international demand for Usenet discussions. By 1981, Usenet had established its gateway to ARPANET via the University of California at Berkeley, enabling Usenet users to access many discussions on ARPANET; these included discussions of engineering, computing, political science, science fiction, and weaponry (Hauben and Hauben 1997, 42-43, 191-192). Later, Usenet would

use the ARPANET and then the broader internet, i.e. a network of global networks, to convey users' communications, significantly expanding its reach (Naughton 2015, 179). This was made possible by the development of the Network News Transport Protocol (NNTP) in 1985 and an updated version of NetNews software in 1987 called C News, which allowed the transmission of, and access to, newsgroup messages from sites on the internet (Pfaffenberger, 1996, 377-378; Frstrup, 1994, 361). Before NNTP, Usenet comprised a more linear network of computers feeding updates to other computers in line using UUCP, whereas conveying messages via the much larger and more de-centralised internet afforded users greater opportunities for accessing Usenet OSNs. The growth of Usenet is also fuelled by the implementation of binary newsgroups (Turner *et al.* 2005) which, rather than text discussion, allowed users to share files with one another. Further statistics regarding Usenet's growth are discussed later, but Altopia's report of Usenet receiving 171 million daily posts of files and messages in January 2021 (Altopia 2021) indicates the scope and impact of the network.

Usenet's early growth was spurred by its openness to new users, relatively low financial and administrative barriers to entry (compared with the ARPANET), and how it met a clear need for discussion between a community of professionals. It formed a physical computer network on which OSNs of *people* could form via newsgroups, including those seeking to discuss religion, and Usenet OSNs experienced massive growth when they became accessible via the internet. The three influential facets of the digital substrate (discussed above) influenced Usenet OSNs in fascinating ways and contextualise my later analysis of participants' interactions.

1.2.2 The three facets of digital substrate within Usenet

When using UUCP to connect a smaller and less de-centralised Usenet from c. 1980 to 1987, one's site (i.e. location) could be upstream or downstream from the central sites in the network. Those that were more central or were a single point of access to Usenet within, for example, a particular country, wielded power and were termed backbone sites (Pfaffenberger 2003, 29). A person's ability to receive new messages posted to their chosen newsgroups at their location is contingent on being connected to another site and that site passing on the latest messages. But why may a location upstream from one's own restrict the flow of messages and consequently impede participation in discussions at a site less central on the network? The answer lies in Usenet's openness. There were no restrictions on speech and anyone with a little technical knowledge (and latterly without even that) could

create a new newsgroup to host discussions of any topic. This differs from many current social media platforms on which terms of use are dictated by the platform owner. Newsgroups are now organised into hierarchies, which were introduced c.1986-7, known colloquially as The Great Renaming, since the previous titling system of 'net.' plus the group's name as a suffix, such as 'net.religion', became unwieldy due to the large number of newsgroups (Pfaffenberger 2003, 29-30). System administrators created seven top level hierarchies within which newsgroups could be established. These were titled comp (computer-related), misc, news, rec (recreation), sci (science), soc (social), and talk (Pfaffenberger 2003, 30). Later, many more hierarchies would be created as Usenet expanded and local administrators created hierarchies tailored to their audience. These included regional hierarchies such as uk (United Kingdom) and relcom (Russian language), along with those for demographics as diverse as k12 for American K-12 educators and HEPnet for high-energy nuclear physics (Fristrup 1994, 212-251). The implementation of the initial set of seven hierarchies was accompanied by a voting system that, *prima facie*, allowed users to influence which newsgroups would be created and which would not (Pfaffenberger 2003, 29-32). However, in practice it granted greater control to system administrators as they could vote down or ignore requests for newsgroups to be created within the hierarchy or refuse to carry the newsgroups' messages even if a group they dislike was created, which meant locations 'downstream' from those backbone sites could not access them anyway (Pfaffenberger 2003, 30-31). Backbone sites were initially powerful gatekeepers due to their central locations within the network, but their power diminished from 1987 with the rise of C News software and NNTP (Pfaffenberger 2003, 31-32). And even when the backbone wielded power, Usenet's lack of a central, formal authority allowed John Gilmore to co-found the 'alt' hierarchy, in which anyone could set up a newsgroup (Grossman 1997, 13-14). Gilmore also later co-founded the Electronic Frontier Foundation which defends privacy and freedom of speech online. Newsgroups within the alt hierarchy became extremely diverse, in part because group creation was not subject to the oversight of those managing the 'big eight' hierarchies (control which has persisted (Schneider 2022)), which included soc groups. Along with the talk hierarchy, many sites would not carry alt newsgroups' messages (Pfaffenberger 2003, 30; Grossman 1997, 14) due to the perceived undesirability of their (sometimes adult) content or the bandwidth they required and associated transmission costs. Again, by using the internet (via NNTP) to send and receive messages to/from Usenet newsgroups, users could access groups with few restrictions or even use a connection that specialised in carrying, for instance, adult discussion (Wang and Bullfrog 2004, 43).

This summary shows how the three components of the digital substrate that I identified above — decentralised network, packet-switching, and gatekeepers — can impact people’s experiences of online social networks on Usenet. As noted, until around 1987, Usenet relied on more linear network connections served by backbone sites, restricting participants’ access to newsgroups depending on decisions made by gatekeepers at their location and at locations upstream from them. Some early participants may have benefitted greatly from being able to access the newsgroup talk.religion but, depending on where and how they accessed Usenet, they may not have even known it existed. This shows the power of gateway machines and gatekeepers for influencing a person’s online social life and, as noted above, gatekeeping can occur at various levels of a network. The later widespread adoption of NNTP allowed newsgroup messages to be conveyed over the internet, expanding the opportunity for online *social* network connections on Usenet over a more decentralised network using packet switching to route communications optimally. If our hypothetical user was previously unable to access talk.religion directly from their Usenet service provider, then during the 1990s they may have been able to access it via one of many internet service providers or Web-based applications which incorporated Usenet discussion group feeds from myriad internet locations. The foundational impact of the early Usenet’s technology is affirmed in Miller, Paloque-Bergès and Dame-Griff’s (2022, 3-4) paper, which states that ‘In a considerable number of countries in Europe, Asia, and Africa, the Usenet “social” service ... and its underlying protocol, UUCP, provided the first digital routes, pathfinders, and map onto which the Internet would graft itself’. With this understanding of Usenet’s development and underlying structure, I can now explore how its technology empowered participants to act and interact in certain ways, through analysis of Usenet’s affordances.

1.2.3 Usenet’s affordances and their influence on Usenet OSNs

Usenet OSNs enable participants to exhibit and experience various social behaviours and actions. While Usenet does not direct how participants use their agency it enables, encourages, and restricts it in certain ways. My focus in this section is on two of Usenet’s key affordances which are relevant to my research. Hutchby describes affordances as ‘functional and relational aspects which frame, while not determining, the possibilities for agentic action in relation to an object’ (Hutchby 2001, 444). He presents this definition in his discussion about how technologies are shaped by how they are used, while they also shape the practices for which they are used. This need for the users of a technology to realise its affordances through action leads Costa to propose the term affordances-in-

practice (Costa 2018, 3651), which I mention here as it emphasises how affordances are realised and shaped when they are enacted.

I will show how Usenet technology empowered participants in ways that shaped their interactions within OSNs in relation to my measures of engagement, namely cohesion, incivility, and patterns of participation. This will provide important perspectives for understanding the social interactions represented in the message archives analysed in chapters four to six. The two affordances I analyse are (1) asynchronous communications within and between groups, and (2) anonymous or pseudonymous participation. I identified these specific affordances for analysis by reading social media literature critically and considering how these technologies' affordances relate to Usenet, in light of my understanding of Usenet's digital substrate, its growth over time, and my measures of engagement.

As discussed above, participants form OSNs on Usenet by posting messages to newsgroups, which are created within one of multiple hierarchies and focus on discussions of a particular topic. Many newsgroups attract participants worldwide. Consequently, Usenet participation crosses time zones so the asynchronous nature of Usenet communications is key to affording participation across international spatial divides. This intersects with the concept of the 'long tail' of the internet (Anderson 2004), which identifies that while there is *some* demand for online communications or services regarding perhaps every conceivable topic, the audience for niche topics may be small in most locations and consequently highly dispersed. Asynchronous communications across space and time zones allow a large OSN focused on such topics to form, as illustrated by the proliferation of Usenet OSNs for new religious movements which has included, among many others, Scientology (Urban 2011, 187-188) and the Unification Movement (Horsfall 1999-2000). Relatedly, Usenet allows participants to post messages to multiple newsgroups simultaneously which, if people then reply, potentially brings diverse OSNs into contact with one another (Smith 2003, 71-75). Whittaker *et al.* (2003) found this happens frequently, with an average of 34% of the more than 2.15 million messages they studied over six months being posted to two or more groups. This broadens the reach of those posters' communications but potentially sparks incivility should the OSNs disagree with one another.

The asynchronous nature of Usenet communications also allows for longer and more considered messages than may be seen in 'real time' messenger services which focus on conversation in staccato bursts. At best this facilitates, for example, detailed religious

discussions, with participants learning from one another through dialogue and critique. Nonetheless the possibility of fiery impulsive responses remains, as does the potential to proliferate nuisance spam messages, since communications can be sent to numerous groups simultaneously. Asynchronous communication also affords varied patterns of participation. Usenet OSNs are often comprised of a small number of participants who post messages frequently, a greater number of infrequent participants, and an unquantified number of people who read messages but do not respond (Whittaker *et al.* 2003; Dame-Griff 2019). This diversity of participation reflects how Usenet's asynchronous format allows participants to be as passive or active as they wish, reading messages some days, months, or years after they were posted. Over time, Usenet service providers archived messages for longer since the cost of disk space decreased. Alongside this, while messages must still travel over the network, the delays are short enough to afford near-synchronous discussions for the most active participants.

In summary, while text-based asynchronous communication may limit the extent to which participants can come to know people they have never met and who are different from themselves (Dawson 2004, 77), it may nonetheless aid the formation of cohesive OSNs by connecting participants who are passionate about specific topics and are dispersed across the world, while also letting participants choose how and when they engage. On the other hand, by bringing diverse people into contact cohesion may be undermined through uncivil disagreement due to their differing interests, motivations, and values. This potency indicates the urgency of understanding the factors that give rise to cohesion and incivility within OSNs. As Pope John Paul II (2002) stated, 'how can we ensure that this wondrous instrument first conceived in the context of military operations can now serve the cause of peace?' These factors, to the extent they are represented in OSN message archives that I access, will be analysed through concrete examples in chapters four, five, and six.

The other key affordance that I identify is how participants may present themselves in anonymous or pseudonymous ways when participating in Usenet's OSNs. Multiple aspects of Usenet's functionality combine to empower participants in this way. These are centred around Usenet's text-based format, which preceded OSN platforms that require users to have a profile, sometimes including a photograph and/or the user's real name. While Usenet participants post messages from their email address, this may of course not include their name or location. Once the technology was developed, Usenet logged participants' IP addresses (Wang and Bullfrog 2004, 57-58) which indicate their location, similar to how a house number and postcode indicate an address. However, participants with sufficient

knowledge could use services that remove their email address and/or IP before posting their messages (Wang and Bullfrog 2004, 57-58; Donath 1999, 37-38), though Kinney (1995, 767) observed that this was less common in religion discussions than in OSNs focusing on some other topics.

This is not to say that presenting anonymously or pseudonymously necessarily entails *misrepresenting* oneself. Literature regarding online interactions from a time when Usenet was more prominent indicates that while scholars may have anticipated people using anonymity to experiment with the active construction of fictional identities, in practice this was less common than expected (Lövheim 2013, 42-43; Dawson 2000, 33). Indeed, if a religious person participates in an OSN that reflects their faith, they may bring their experiences from their offline religious communities to bear on their discussions online, providing some degree of continuity to their identity across their social networks (Campbell 2012, 686-688). In particular, Linderman and Lövheim's study of a Christian newsgroup on Usenet found that sharing common religious values may have served to strengthen participants' existing religious identities (Linderman and Lövheim 2003, 237). A participant's identity online may then be shaped further by the discourse within the OSN, which may allow participants to adopt one or more positions within the group, in relation to the subject of discussion (Lövheim 2004, 70-71). For example a person could adopt one of a range of 'insider' or 'outsider' positions when discussing a minority religion within an OSN. Relatedly, Litt notes that people develop a perception of their audience when they do not know who is reading their online posts and that they tailor their presentation of themselves to this audience (Litt 2012 cited in Hocquet and Wieber 2018, 314). This applies to Usenet OSNs since they can be large, dispersed and at least partly invisible to participants, though prospective participants can read others' messages before contributing themselves to develop their perception of the people with whom they are about to engage. This modest sample of studies indicates how participants have agency over how they present themselves in text-based discussions online, but their identities may be shaped, even co-constructed, through their participation in multiple social networks online and offline, whether religious networks or otherwise.

Contributing to Usenet anonymously or pseudonymously can have positive and negative outcomes, which can impact cohesion, incivility, and participation. Considering positive outcomes, the affordance of anonymity or pseudonymity can benefit participants who wish to discuss sensitive topics. Walther and Boyd (2003, cited in Fox, Johnson and Rosser 2006) found that flexible patterns of participation and anonymity were valued by women

connecting on Usenet to discuss sensitive matters around pregnancy. In another case, some participants in discussions on the alt.religion.scientology newsgroup sought anonymity to post and discuss materials they wished to make public (Peckham 1998, 323-326). Anonymity provided different benefits in these examples; in the first, anonymity protected the sensitive personal circumstances of the participants and in the second it reduced the risk of consequence to those sharing the materials. Common to both is that the affordance facilitated participation and the potential for valued social connection by allowing people to speak up when they otherwise may not, potentially increasing participation and/or cohesion within OSNs. Masur, Bazarova and DiFranzo's (2023) research finds that social norms within online groups influence participants' self-disclosure. It may be the case, then, that participants' behaviour is enabled by the affordance and encouraged/normalised by others. Considering negative outcomes, anonymity also helps proliferate spam messages (Fisher and Lueg 2003, 257-258), which have been posted in high volumes to many newsgroups since at least the mid-1990s (Grossman 1997, 24-25; Wang and Bullfrog 2004, 59). Undesirable content can impact an OSN negatively since the ratio of noise to signal can worsen to the extent users have to work through messages involving spam, 'trolling', scams, or even malicious software while trying to engage in meaningful discussion (Grossman 1997, 24-25; Wang and Bullfrog 2004, 58-59; Graham 2019, 3-4). Such content was often posted to early religion discussions on Usenet (Baker 1995, 113-119).

In summary, Usenet's affordances facilitated actions that support cohesion within social networks by connecting dispersed people who share interests or common identities, including regarding religion, and removing barriers to participation by allowing participants to self-disclose to the extent that they feel comfortable. On the other hand, these affordances also create an environment in which cohesion can be undermined through incivility, such as when OSNs that hold disparate values, beliefs or behavioural norms are connected, or spam or trolling messages are propagated widely. Usenet's affordances are shaped by its technology and how participants use it, such as how services emerged to send messages stripped of personal information. These insights provide context for understanding the social interactions in my analysis of message archives in chapters four to six.

1.2.4 Usenet's growth, popularity and change over time

This short section builds on my above discussion of Usenet's foundational structure and technologies by exploring its substantial growth and popularity into the 1990s and beyond,

while noting some key changes that occurred over time. These changes impacted users' interactions, which is pertinent to my analyses of my research sites, presented in chapters four to six.

Soon after Usenet's conception in 1979, the first connections were formed between computers at Duke University, Duke Medical School, and the University of North Carolina at Chapel Hill. By summer 1980, eight sites were connected including two of AT&T's Bell Labs sites, Reed College, the University of Oklahoma, and the University of California at Berkeley. This mix of industrial and academic institutions is typical of the early Usenet. By summer 1981, 150 sites had joined and by December 1982 twelve European sites were connected. Additional commercial companies were also connected by 1982, including Intel and Microsoft. In 1984, Usenet comprised almost 1,000 sites and this exceeded 10,000 by 1988. (Statistics from Hauben and Hauben 1997, 39-40, 48, 168, 175-176; Paloque-Bergès 2017, 162; Pfaffenber 1996, 25, 370)

Usenet continued to grow substantially during the 1990s. A key catalyst for this was increased use of the internet to access and convey Usenet messages (discussed above), allowing more people to access many Usenet OSNs, connecting from their home computers rather than using those at specialist sites such as universities. By 1992, around 60% of Usenet communications were sent over the internet (Hauben and Hauben 1997, 54). We can chart the growth in the number of communications sent across Usenet around this time:

Year	Communications per day
1980	10
1985	375
1988	1,800
1993	26,000
2000	417,281
c. 2005	2,000,000

Table 1: approximate number of communications sent over Usenet per day in select years (Statistics from Hauben and Hauben 1997, 44; Naughton 2015, 179; Smith 2003, 53; usenet.com, cited in Krüger 2005, 8)

The number per day fluctuates over the year. Microsoft's Netscan Usenet analysis software (since discontinued) recorded 1,023,397 messages on Christmas Day in 2000, which is a significant spike given the 151,655,377 messages Netscan recorded for the entire year

(Smith 2003, 53). The global distribution of participants reflects Usenet's international reach and its roots in the United States. A review of Usenet participants' email addresses indicates that circa 1999 a large minority were based in the United States but that messages were posted from email addresses from 205 other countries (Smith 1999, 197). Further, only a minority of users may contribute on any given day. A message for new Usenet participants from 1993 indicated that around 3,000,000 people may read Usenet (Rospach and Spafford 1993, reproduced in Fristrup 1994, 288), contrasted with the approximately 26,000 communications sent each day at this time.

Alongside increased access via internet connections, Usenet's growth was also fuelled by the proliferation of widely available software for reading and posting messages. This dissertation has insufficient space to document the growth of popular internet access during the 1990s but a few examples are illustrative. America Online (AOL) is chief among these. AOL integrated Usenet access into its software from 1993 to 2005 (Dame-Griff 2019, 5-6; Reid and Gray 2007). At this time, Usenet access was still limited in some places such as in France where messages could be read but sending messages was more restricted (Paloque-Bergès, Peuvrelle and Schafer 2018, 317-318). In 1994, AOL had one million subscribers (Grossman 1997, 11) which represented a significant injection of new people into Usenet's OSNs. Alongside this, other internet users accessed Usenet via dedicated software. This is reflected in Fristrup's (1994) guide to accessing Usenet, which lists twenty-two programs for various computing platforms including PC and Mac (Fristrup 1994, 361-363). The software has different characteristics to reflect users' needs and capabilities, such as applications that focus on presenting a simple interface or others that help the user identify relevant discussions from subject lines. Other companies also incorporated Usenet access into their applications, including Microsoft's Outlook Express email software and Web browsers such as Opera (Wang and Bullfrog 2004, 45). This proliferation of software fuelled Usenet's growth as applications suited to less experienced users raised their awareness of Usenet OSNs and facilitated their participation.

In parallel, the demographics of Usenet participants evolved as access broadened from computing specialists to others within academic and research settings and then to the wider internet-enabled public in the 1990s and 2000s, when internet access became cheaper and more reliable. Newsgroups were set up to focus on topics of increasing granularity, such as those for many Christian denominations and sects within the soc.religion and talk.religion hierarchies. I review this later, but here we can note this includes Judaism, Islam, Catholicism, many mainline protestant denominations, Unitarianism, Mormonism, and

Jehovah's Witnesses among others. Beyond religion, Netscan data for the year 2000 indicates that the twenty newsgroups with the highest number of unique contributors include topics on computing, music, sex, television, motor vehicles, and sports among others (Smith 2003, 70). However, considering the broadening demographics and discussion, a tension arose around socialising new participants into the behavioural norms that developed when Usenet was founded. These norms matter since they influenced the social contexts in which later participants experienced cohesion and incivility.

1.2.5 Behavioural norms on Usenet

Usenet's roots in research and computer science departments, along with its openness, supported the development of values and behavioural norms that Usenet participants exhibited, advocated, and discouraged. To some degree, these values and behaviours persisted for many years. They reflect the hacker ethic, which includes focusing on freedom, avoiding centralised authority, using computers to do good, and advocating enthusiastic engagement with, and development of, technology (Coleman 2013, 16-20). However, while Hauben and Hauben (1997, 62-64) indicate how Usenet can have a positive, helpful, democratic, bottom-up approach to social networks this is balanced, at times outweighed, by the incivility and spam messages that the same freedoms and autonomy allow. This tension is seen in the following extracts of two messages that were frequently posted for the attention of new users. The messages come close to defining Usenet, its diversity, behavioural norms, and values:

Usenet is a large collection of computers that share data with each other. It is the people on these computers that make Usenet worth the effort to read and maintain, and for Usenet to function properly those people must be able to interact in productive ways. ... The easiest way to learn how to use Usenet is to watch others use it ... After a couple of weeks you will start understanding why certain things are done and what things shouldn't be done. ... Please remember that people all over the world are reading your words. Do not attack people if you cannot persuade them with your presentation of the facts. ... Be familiar with the group you are posting to before you post! You shouldn't post to groups you do not read, or post to groups you've only read a few articles from — you may not be familiar with the ongoing conversations and themes of the group. One normally does not join a conversation by just walking up and talking. Instead, you listen first then join in if you have something pertinent to contribute.

(Rospach and Spafford 1993, reproduced in Frstrup 1994, 286-287, 289, ellipses added)

Usenet is the set of people who exchange articles tagged with one or more universally-recognised labels, called “newsgroups” ... It is almost impossible to generalize over all Usenet sites in any non-trivial way. Usenet encompasses government agencies, large universities, high schools, businesses of all sizes, home computers of all descriptions, etc., etc. (Salzenberg and Spafford 1993, reproduced in Frstrup 1994, 281-282, ellipses added)

These messages indicate the freedoms that participants have to contribute to Usenet in line with the aims of each newsgroup, but which they could haplessly or intentionally abuse. Further, by indicating Usenet’s global reach and the types of locations from which professional users participate, the messages indicate the potency of participants’ undesirable actions to disrupt others’ social activities. These messages were necessary since while the earliest participants’ shared passions and professional community held them together, this cohesion did not endure as access broadened (Feenberg and Bakardjieva 2004, 8). Bryan Pfaffenberger (1996, 365-366) studied freedom of speech on Usenet, which he recognised was tolerated so long as it did not disrupt Usenet and thereby prevent others’ opportunities to speak. This permitted immense freedom, particularly since later participants could create their own newsgroups and connect multiple newsgroups by posting messages between them, connecting large and diverse audiences. Consequences of these freedoms are reflected in Judy Anderson’s observation (quoted in Grossman 1997, 105) that ‘Usenet, while it can be nasty, acerbic, uncaring and unsympathetic is a truly nondiscriminatory society’.

Nonetheless, the persistence of values on Usenet is multifaceted. The case for arguing they can be overwhelmed by new members is illustrated in the event known popularly as Eternal September. As noted above, in 1993 AOL afforded its members access to Usenet. This allowed participation in Usenet OSNs without socialisation into their behavioural norms. Incorporating humour and frustration, the title Eternal September reflects how each September a new cohort of university students would join the network and require socialisation, but now the influx was never-ending and was comprised of people who were not socialised within the academic or other technical communities from which Usenet previously attracted participants (cf. Paloque-Bergès 2019, 116; Dame-Griff 2019, 5-6). Further, the keen new participants may have accessed Usenet when paying by the hour at

home and using a simple interface with limited ability to search for other posts, reducing the likelihood of them observing others before taking part themselves (Grossman 1997, 33-35).

That said, Usenet's early values endured to some extent, since later participants were influenced by decisions made in the years before they joined (Fisher 2003, 10). These include the newsgroups' hierarchical naming structure. Since anyone could create a newsgroup within the 'alt' hierarchy and at the time alt was founded the 'soc' hierarchy was more controlled, OSNs formed via alt may have experienced the values of freedom and technological creativity that underpinned Usenet. Many alt OSNs discussed minority or stigmatised topics, or topics at the boundary of, or beyond, mainstream acceptability. Also, some socialisation of new participants into OSNs' behavioural norms was achieved through introductory FAQs (Wang and Bullfrog 2004, 58-59). The hacker values of freedom and individual responsibility and an absence of centralised control have also remained. At the time of writing, Usenet is used predominantly for file-sharing and less for OSNs. Its emphasis on privacy and open sharing creates tension when people air controversial views and share copyrighted materials. Usenet service providers afford those freedoms while negotiating requests from media companies to take down their intellectual property (Henderson 2012). Users still choose from various Usenet access providers and take responsibility by protecting their anonymity using services that strip their message of some personally identifiable content and/or software that encrypts messages which then require a password (Henderson 2012).

Understanding Usenet's early values and behavioural norms illuminates some of the social context in which later Usenet OSNs, captured in message archives, were formed. Indeed the lasting impact of online social networks' foundational norms is recognised by leaders in the field, as in a discussion between Reid Hoffman, co-founder of LinkedIn, and Caterina Fake, co-founder of Flickr which pre-figured later social media platforms. Fake's message to founders of online communities is '[you] are the framer of the constitution in this world that you are building' (*Masters of Scale* 2018). Affirming this, Hoffman argues that 'every founder of an online community has to shape the culture from day one, because the tone you set is the tone you're gonna keep', recognising that participants' behaviour cannot be controlled but that guardrails can be put in place (*Masters of Scale* 2018). This analogy applies to Usenet since the hierarchical naming structure for newsgroups and the freedoms normalised within, for example, the alt hierarchy, shape the types of behaviour that take place within those OSNs.

I will now take stock of my key points. The underlying infrastructure of a digital network, which may be partly invisible to end users, can shape the social networks that emerge upon it. I term this the *digital substrate*, seeking to express how it shapes but does not directly control the OSNs that it affords; both the substrate and those using the platforms have agency which structure the OSN. I identify three components of the digital substrate that are particularly relevant to studying OSNs, including those related to religion, on Usenet and similar platforms: (1) The de-centralised network structure which enables spatially distributed social networks to form. (2) The packet-switching method of transmitting communications which enables multiple people to connect simultaneously across those networks efficiently and in near real time. (3) The gatekeepers that may control the parts of a network with which a person can connect and engage. This substrate and the technical design of Usenet give rise to the *affordances* of asynchronous communication within and between groups, and anonymous or pseudonymous participation. Together, these technologies enable the formation of cohesive, international networks that form part of a person's social life through discussion of subjects that are important to them. The technologies also allow uncivil or spam messages to be distributed widely and with relative impunity, given Usenet's principles of freedom, individual responsibility, and an absence of centralised control, as encapsulated by the hacker ethic. Participating in Usenet OSNs can, in summary, impact a person's life through positioning them in new social networks and exposing them to valued and unhelpful discussions, contacts, practices, and beliefs. Together, these insights provide context for understanding the cohesion, incivility, and patterns of participation in OSNs that I analyse in chapters four to six.

1.2.6 OSNs form part of social lives

OSNs are mediated by network infrastructure installed in the physical localities in which participants' other social networks are formed and sustained. This is important since it indicates how OSN participation is embedded in people's broader social lives, opening the possibility that engaging regarding religion online could influence a person's experiences of religion offline and *vice versa*. Indeed digital religion literature recognises that religion online and offline can influence each other and serve complementary purposes. This occurs across various forms of religion online. Gelfgren (2014, 45-46) notes that graphical representations of churches in Second Life (popular at the time) bear similarities to bricks and mortar counterparts within their tradition and some are reproductions of those churches. Before Covid 19, some churches used digital media to convey services online, thereby consolidating community within their large or dispersed congregations (Vitulo 2019), and

these media became essential for maintaining religious congregations' worship and forming social connections during the pandemic (Campbell, 2020). Further, religion online may reflect faith handed down through practices, teaching, and other forms of socialisation offline. Regarding Christianity, Bennett concludes that online communities can be church insofar as they are connected with the historical tradition and practices the church has received and passed on (Bennett 2012, 129-130). This resonates with Dawson and Cowan's (2004) understanding that the internet reflects and acts as a shadow of the wider world, since much of the world is represented online and a lot of online content is sparked by something offline.

Given this online-offline continuity, a person's behaviour when interacting with religion online can be influenced by the behaviours they exhibit when interacting with that religion offline. Lövheim considers this regarding young people constructing religious identity in a Swedish discussion group. Participants' past experiences appear to influence how they engage with the group but they are then influenced by the dominant ways in which others approach religion within the group's discussions (Lövheim 2004). Relatedly, the function of a religion focused OSN can be shaped by what participants consider is lacking or requires change within that religion offline. Marta Kołodziejka's (2018, 126) study of Catholic discussion forums finds that some participants had rich dialogue with people knowledgeable about Catholicism and those outside the faith to an extent that was not present in their offline church or educational settings. The power of the digital substrate and Usenet-like affordances shape these forums, since the forums connect large numbers of diverse and spatially distributed people who have common interests or are wishing to engage with people different from themselves. This is achieved through asynchronous communication with participants deciding the degree to which they share personal information, affording freedom of expression which is then enabled or curtailed in accordance with the behavioural norms and expectations established within the OSNs (Kołodziejka 2018, 62-63). These similarities to Usenet are important as they indicate the relevance of Usenet for understanding some other contexts in which religion is discussed online. This relatability is reflected in my definition of OSN (p.18), which indicates that my research on Usenet will be relevant to a range of contexts that share certain characteristics, including Kołodziejka's research sites.

The macro perspective afforded by analysing survey and panel data also affirms that a person's engagement online is part of their whole social and religious lives rather than a separate sphere. Nyland and Near's (2007) survey of 184 people found that religious

respondents were likely to use social networking sites to maintain existing relationships. Although a small study, this indicates how computer-mediated communication could form part of one's existing social life even when smartphone use was less widespread and the internet was more typically accessed through desktop and laptop computers, as for much of Usenet's history. More recently, McClure's (2016) analysis of National Study of Youth Religion (NSYR) data finds that social networking site use can undermine plausibility structures and be associated with increased tolerance of syncretic religious beliefs. Regarding plausibility structures, the digital substrate can increase the breadth and diversity of social contact across divides via social networking sites, which reflects McClure's findings. Regarding syncretism, McClure identifies that NSYR data predict increased tolerance of syncretism with social networking site use, though he finds that factors such as parental religiosity interact with social networking site participation to influence a person's acceptance of syncretism. This research affirms how a person may be influenced by both the online and offline networks which comprise their full social life.

These qualitative and quantitative studies affirm the value of OSN message archives as a research site for theology and religious studies, since they are digital traces of social networks that may affect participants' lives online and offline. This is because OSN participation is one part of a person's whole social life and may influence participants' religious beliefs, practices, and contacts. This happens now and was possible via desktop and laptop computers before the advent of smartphones. With these points affirmed, I will now explore literature regarding religion on Usenet before critically evaluating the pertinent concepts of community and authority that are developed in broader digital religion literature.

1.2.7 Religion on Usenet

While religion has been discussed on Usenet since at least 1982, the first newsgroup dedicated to it, net.religion, was created in February 1983 (Helland 2007, 958-959). Messages were often adversarial. Early discussions included debates or arguments regarding creation, religious scriptures, and the existence of God among other topics (Helland 2007, 959-960). Usenet's open nature and freedom of expression both led to the formation of this OSN and enabled discussion in which very little cooperation or cohesion was present.

The freedoms and discord in discussions of religion on Usenet continued into the 1990s, with *Time* magazine remarking that alt.fan.jesus-christ and alt.religion.scientology were

‘among the busiest — and most contentious — of the nearly 20,000 discussion groups carried on Usenet’ (Ramo and Burke, 1996). Both were created in the alt hierarchy, in which principles of freedom were, as discussed above, a catalyst for creating the hierarchy in which any participant can create a newsgroup, leading to unbridled discourse on myriad topics. Some other, more cohesive newsgroups had already emerged, such as net.religion.judaism which was established in February 1984 and connected the Jewish diaspora (Helland 2007, 961). It was intended specifically for discussions of Judaism and, at least initially (Chua 2009, 242), the OSN was effective notwithstanding the diversity of Jewish traditions represented (Helland 2007, 961).

The newsgroup net.religion.christian was formed by 1985 as was net.lang.India, which included discussion on Indian religions (Helland 2007, 962). The number of Christian OSNs grew substantially by the 1990s. This is reflected in the results of Robert Glenn Howard’s (1997, 299) search for Christian newsgroups circa 1995, which identified the following among others:

- alt.christnet
- alt.christnet.bible
- Others within the alt.christnet sub-hierarchy, including:
 - .ethics
 - .sex
 - .theology
- rec.music.christian
- soc.religion.christian
- alt.fan.jesus-christ

OSNs for various Christian traditions, and other traditions rooted in Christianity, proliferated into the 2000s. This is indicated in the following sample of the newsgroups for which I accessed archives within the ‘talk’ hierarchy, which was originally set up to accommodate highly diverse subjects including some which are controversial and unrelated to Usenet’s original focus on computer-related topics and research (Pfaffenberger 2003, 29-30):

- talk.religion.christian
- talk.religion.christian.anglican
- Others within the talk.religion.christian sub-hierarchy, including:
 - .charismatic

- .lutheran
- .mennonite
- .presbyterian
- .quaker
- .roman-catholic
- talk.religion.jewish.messianic

Various aspects of Usenet that I discuss above created an environment in which many religion OSNs could form and fulfil varied demands for religious connection and discussion. First, the hierarchical structure established at Usenet's foundation, augmented and liberated through The Great Renaming and the creation of the 'alt' hierarchy, enabled people to form many OSNs for discussing religion via denomination-specific newsgroups such as those within the talk.christian hierarchy above. Second, since by 1987 messages were increasingly conveyed over the internet using NNTP rather than the more bounded network of specialist sites that preceded this, more people could access Usenet and connect with a wider range of its OSNs. Third, the wide availability of software such as Microsoft's Outlook Express and services such as AOL for accessing Usenet OSNs allowed a wider demographic to join than comprised Usenet's earlier participants. This is likely to have brought with it increased demand for diverse religion discussion. Fourth, the value placed on freedom of speech on Usenet enabled OSNs to form in which people openly discussed religious groups that were at times controversial, such as Jehovah's Witnesses and Scientology. Fifth, the discussion of controversial topics in religion, or sensitive matters such as abortion, is for some people afforded by Usenet enabling them to remain anonymous or pseudonymous, removing some potential barriers to these discussions. Together, these insights indicate how the digital substrate, Usenet's key affordances and its behavioural norms are useful for understanding the growth and nature of religion focused OSNs on Usenet.

Supplementing this, religion elsewhere on the internet was generally less interactive than on Usenet at this time. In the year 2000, a Pew Internet and American Life Project survey of 1,309 websites maintained by Jewish, Christian and Unitarian Universalist congregations found that only 3% of those websites enabled online discussion (Larsen *et al.* 2000, 4, 11). This was despite the same survey recording that 29% of respondents use the internet themselves to discuss religion; respondents included rabbis, church ministers, other officials, and those who ran the websites (Larsen *et al.* 2000, 19). Christopher Helland also observed that, around the turn of the millennium, many religious groups were reluctant to

enable interactivity in their presences on the Web (Helland 2005, 1). Consequently, Usenet and similar services filled a demand for two-way connection and discussion that was often unfulfilled by 'Web 1.0' era websites, which were more static and focused on one-way conveyance of information to users. Indeed, I argue that religion on Usenet at this time prefigured the religious connections people made with each other in the later Web 2.0 era. Popularised by computing educational company O'Reilly Media, Web 2.0 emphasises the use of the World Wide Web as a platform to which people can contribute their knowledge with focuses on user agency, collaboration, and dynamic content (O'Reilly 2005). These are reflected in Usenet and, while it lacked other sophistications of Web 2.0 (see O'Reilly 2005), this indicates why Usenet gained such popularity in the early 2000s.

I now analyse literature that examines examples of religion focused OSNs on Usenet, with the aim of acquiring insights into how cohesion, incivility, and/or conflict manifest. This allows me to flesh out my understanding of these concepts and subsequently define them in more detail, helping me operationalise them in my research. In turn, this allows me to specify computational processes for identifying these social processes and this focuses my analyses of message archives in later chapters. I begin with studies that focus on the fostering of cohesion on Usenet and end with a focus on incivility and conflict, though these processes are visible to varying degrees throughout the literature that I analyse.

Herring (2010) studied two years of activity from the newsgroup uk.religion.christian. She surveyed participants, with 80% considering that the group provided some form of community for them. Herring analysed how Usenet's text-based medium, the discussion of religion and prayer contributed to this. This includes some respondents who only read others' messages, rather than contributing themselves, indicating the value of OSNs for this less-visible audience. Herring identifies that within the network, participants know one another through identities constructed via the messages they post since they have mostly not met face-to-face, and that perceptions can require correction when a participant's self-revelation leads to a change in understanding regarding, for instance, their religious faith (Herring 2010, 37-38). Alongside this, given the breadth of discussion, Herring (2010, 38) finds that participants may disagree with each other on one matter while agreeing on another, and still support each other's requests for prayer or support regarding difficult personal circumstances. She uses the term *perichoresis* — a borrowing from Christian trinitarian theology — to describe the form of community she observed, which exhibited fluidity of movement within a bounded space (Herring 2010, 39-40). This reflects how the network persisted over time and participants reported feeling community despite their

evolving alliances with one another. In turn, this indicates the potential for religious belief and identity, enacted through discussion and practice, to foster cohesion at times when people may be otherwise divided. For Herring, participants' messages may be interpreted as them taking action, such as the purposeful welcoming of those who are marginal within the community, engaging with other participants as individuals, the common purpose of discussing Christianity despite participants' diversity, and the 'praxis of expression' in which people commit their views regarding Christianity to a public and archived forum. This indicates how OSN discussion can impact a person's social life through grafting them into a community which exposes them to new people and ideas, and makes public their views on topics which may then be discovered by others. I consider Herring's perspective on perichoretic community to be useful since it illuminates how supportive and argumentative relations between participants may be fluid within the finite space of an OSN. This fluidity is made visible by the methodology I develop in chapters two and three. However, Usenet's affordances of crossposting messages to multiple newsgroups simultaneously can remove these boundaries and impact cohesion by connecting diverse OSNs who focus on dissimilar topics, as I shall soon discuss.

Helland's (2005) analysis of the distinction between religion online and online religion leads me to understand further how practising religion via OSNs can enable or support cohesion including, in some cases, when disagreement occurs. Helland's concepts create a spectrum in which religion online is akin to the conveyance of information regarding religion and online religion is the performance of religion. He notes the epistemological complexities around knowing whether a particular action online is religious, since this may be apparent only from the perspective of the actor, though the researcher can seek to gauge this (Helland 2005, 5-6). One of Helland's case studies for this process of discernment examines a discussion of prayer requests shared via the large Christian newsgroup soc.religion.christian. Helland considers that the participant requesting prayer was sincere, as was a person who expressed concern about one aspect of the request, and that consequently both people were participating in the public expression of religion, as were later contributors to the discussion who further explored and reconciled the disagreement using biblical scripture (Helland 2005, 8-9). In this case, disagreement occurred within a shared and apparently sincere search for truth through the practising of religious prayer and discernment, so it supported cohesion rather than leading to incivility. I consider this indicates how religious activity can foster cohesion within an OSN, as can disagreement in some circumstances when it occurs in the pursuit of shared aims that reflect the OSN's purpose. Sometimes, however, disagreement can undermine or prevent cohesion and the creation of new OSNs

may be an effective solution. The following two papers provide complementary case studies of this. In the first (Chua 2009) incivility toward existing members of a Jewish OSN led to the creation of a new, more tightly controlled newsgroup via which an alternative OSN could form. In the second (Howard 2011) Christian fundamentalists sought to connect with mainstream Christian OSNs but were rejected so formed their own.

Chua cites literature to affirm the premise that some form of governance is necessary for online 'virtual community' to succeed, considering why such communities regulate speech and whether this benefits the group's success in terms of its persistence, the number of messages posted and participants' assessments of the group's success (Chua 2009, 236-237). Chua studied six newsgroups, of which two are religious, namely soc.culture.jewish and soc.culture.jewish.moderated. The former allowed people to post messages freely, whereas the other utilised one or more moderators to filter messages and block those that were considered harmful, including messages about certain aspects of politics, the promotion of other religions and anti-Semitism (Chua 2009, 241). Chua's (2009, 242-243) summary of messages posted to the unmoderated group on 22nd February 1999 (its moderated counterpart was formed in July 2000) indicates a mix of antisemitism, spam, and other messages that would not pass moderation, among some messages that aligned with the network's purpose. This high proportion of unwanted content continued after the creation of the moderated newsgroup and various attempts to improve messages posted to the unmoderated one were unsuccessful. These included creating alternative newsgroups for connecting those wishing to discuss Jewish and non-Jewish issues that did not fit with the original group's intended aims, the development of an FAQ to guide behaviour, and the use of technology to filter messages (Chua 2009, 245-246). In this case, Usenet's decentralised structure and freedoms of expression and participation connected a large, diverse, and to some degree anonymised group of people, which afforded incivility and conflict in messages on the unmoderated Jewish newsgroup. Further, frequent crossposting between the unmoderated Jewish newsgroup and others (Chua 2009, 245) served to decontextualise discussion from the stated aims of the Jewish group's participants since the messages connected multiple OSNs which may not socialise people into the same behavioural norms. At worst, it is possible that the posting of uncivil/offensive content may have shored up the partisan position(s) of those opposed to the Jewish OSN, reflecting Lyu's (2023) research that incivility towards others can affirm one's own identity in contrast the target group. The consequences of this are seen in Chua's analysis that 3.9% of the messages posted to the moderated newsgroup across a ten-day period in 2003 were anti-Jewish culture, compared with 31.7% on the unmoderated one. In this case, moderation

was associated with reduced incivility and conflict. Complementing this, Robert Glenn Howard's (2011) study of a 'virtual ekklesia' of fundamentalist evangelical Christians online includes analysis of end times discussion on Usenet, illustrating the benefit to fundamentalists of creating their own OSN after they were rejected by other Christians and remained unconnected with each other via public Usenet Christian newsgroups.

Howard (2011, 47) observed that mainline and evangelical Christians who connected via popular Christian newsgroups in the early-1990s were not receptive to fundamentalists who used proof-texting to argue for certain interpretations of biblical eschatology. It would not be surprising if Usenet's early demographic of government and academic researchers created an environment unaccepting of biblical proof-texting as a means to ascertain knowledge. Howard (2011, 46-47) posted a message seeking discussion about end times to five newsgroups in 1994 and received five replies, all from one group, though he received fourteen responses direct by email which, as a more private medium, enabled the sender to avoid criticism from people unsympathetic to this discussion. Along with Howard's (2011, 50-51) illustration that soc.religion.christian was dismissive of end times discussion, this indicates the presence of fundamentalist Christians accessing Usenet for whom the behavioural norms within established Christian Usenet OSNs meant they were not connected to one another via these networks. This changed in 1995 with the creation of the newsgroup alt.bible.prophecy, via which an OSN formed in which end times discussion was fostered through participants' 'ritual deliberation' of interpreting biblical scripture and discussing others' interpretations (Howard 2011, 58-59). Howard (2011, 50-52, 58) identifies an increase in end times discussion on Usenet from 181 archived messages mentioning 'Endtimes' or 'End Times' between 1981 and 1st January 1992, 269 in 1992 and 3,081 in 1996, of which 1,166 were posted to alt.bible.prophecy. This increase also reflects the influence of the *Left Behind* series of novels, the first of which was published in 1995 (Howard 2011, 58). Usenet participation in general grew substantially from 1992 to 1996 (see the trajectory indicated by table 1 above) and the incomplete nature of archives, the de-centralised structure of Usenet (meaning not all messages are available everywhere) and participants' diverse vocabulary render it very challenging to establish definitively how many messages are sent regarding specific topics (see Technical Appendix 1.1 on some challenges of computational natural language processing). Nonetheless these figures, supported by Howard's analysis, indicate how forming a newsgroup for discussions of a particular religious subject, intended for participants who are like-minded since they share religious epistemology, can help an OSN to form and grow when its members may not be able to connect to other OSNs that rejects their discussions.

The formation of a new newsgroup can also help connect people who share an interest at the intersection of their religion and a different topic. Jay Howard and John Streck's (1999) analysis of contemporary Christian music illustrates that this occurred on Usenet for Christian fans of various music genres. In this case, people connected via the newsgroup `rec.music.christian` to discover bands that reconciled their secular musical preferences with their religion by finding 'the Christian' alternative to secular bands (Howard and Streck 1999, 99-100). The hierarchical naming structure for Usenet newsgroups allowed a dedicated Christian music newsgroup to be established within 'rec.music' and this enabled people to contact others and form an OSN sympathetic to this intersection. However, messages at intersections of two or more distinct topics may also be posted to newsgroups that are not intended for such an intersection, sometimes leading to rejection of those who instigated the discussion. I now discuss two examples of this in extant literature on Usenet, one focusing on the new religious movement Heaven's Gate and the other on Christian Identity militia.

The religious group that came to be known as Heaven's Gate sought to promote their apocalyptic beliefs, which were rooted in Christianity, New Age, and popular UFO culture, by posting messages to newsgroups that each focused on one of many topics including computing, science fiction, philosophy, drugs, consciousness, artificial intelligence, and near-death experiences (Zeller 2014, 189). The group's wider internet proselytization resulted in only one new member (Zeller 2014, 198-199) and the group's engagement on Usenet yielded rejection and sarcastic and dismissive responses. They were intensely serious about their religious mission and, for at least one former member, this failure to convey their message effectively on Usenet indicated that they could no longer communicate with the public (Zeller 2014, 196). For several reasons, coalescing around hopes invested in the Hale-Bopp Comet, thirty-nine Heaven's Gate members took their own lives in March 1997.

The group exhibited a different strategy from some other small religions that communicated via Usenet. They sought to connect with OSNs that had established discursive boundaries that did not provide fertile ground for the message of Heaven's Gate, rather than forming their own newsgroup within, say, the `alt.religion` hierarchy and seeking to develop an OSN of their own. Heaven's Gate had some success in attracting members at meetings in the 1970s but group leaders and some members were part of the New Age milieu (Wessinger 2000, 233-234) rather than occupying the discursive world of, for example, the artificial

intelligence Usenet OSNs they contacted in 1995. Therefore, in this case the incivility Heaven's Gate experienced on Usenet from connecting with OSNs that exhibit topics of discussion, behavioural norms, and an epistemology markedly different from their own was foreseeable and contributed to the group's rejection and the ending of their public engagement.

Some members of American militia fared better when seeking to proliferate discussion rooted in minority and controversial religious beliefs. Militia discussion was substantial on Usenet in the 1990s via newsgroups focused on discussing guns, survivalism, and the US constitution among others, and via the later dedicated newsgroup misc.activism.militia (Weeber and Rodeheaver 2003, 190, 197). In this case, the religious component is the Christian Identity view that underpinned a large minority of participants' motivations for militia activities, indicated by Weeber and Rodeheaver's (2003, 199-200) analysis of Usenet messages posted to three militia-focused newsgroups in which such motivations were indicated. The Christian Identity view entails particularly controversial racial beliefs rooted in religion and the anticipation of an apocalyptic battle. Weeber and Rodeheaver (2003, 198-199) analysed militia-focused discussions on talk.politics.guns, misc.survivalism and misc.activism.militia, some of which were sympathetic to militia and others were uncivil or were messages crossposted to varied newsgroups simultaneously. Messages sympathetic to militia indicate compatibility between those OSNs' other discussions (politics, guns, survivalism) and militia beliefs and practices, unlike Heaven's Gate's messages when they contacted artificial intelligence focused OSNs. Given the proliferation of conflict in historical and current OSNs, it would be helpful to see how militia discussion became grafted into OSNs that focused on related but distinct topics such as guns and survivalism in a way that would indicate how militia participants were able to broaden discussion within those established newsgroups and cohabit them. The research I present in chapters four to six indicates such broadening where religion focused OSNs discuss non-religious topics or become connected with non-religious OSNs via crossposting.

Weeber and Rodeheaver's paper is also useful for showing how sociological theory can be tested through analysis of Usenet data, in this case Neil Smelser's theory that people join radical movements because they experience strain in their life. Weeber and Rodeheaver (2003, 193-196) identify 125 Usenet messages from militia participants who experience a range of strains including fear of federal government, rapid social change, or loss of employment. Relatedly, the paper identifies 335 messages which indicate events that precipitated the militia movement, including the tragedies at Waco and Ruby Ridge among

others. This indicates the integration of OSN participation into a person's whole social life by showing that 'offline' events or strain fuel militia belief or activity, which leads to OSN participation that in turn connects a person to others and stimulates discussion. This illustrates the utility of historical OSNs, including those on Usenet, for analysing social phenomena.

The paper also prefigures discussions about the proliferation of 'fake news' and the dissemination of minority perspectives via social media, since it shows Usenet to be an alternative means of accessing information to mainstream news (Weeber and Rodeheaver 2003, 196-198). The paper notes the lack of censorship on Usenet which allows people to discuss controversial topics and promote different perspectives on events from those reported in mainstream media. This is enabled by Usenet's values of freedom and not having centralised control, particularly once Usenet messages were conveyed primarily over the internet. However, these freedoms permit chaotic events to occur which undermine cohesion in terms of participants being able to engage without disruption and avoid disruptive incivility. Chief among these was the struggle to dominate discourse in alt.religion.scientology.

The newsgroup alt.religion.scientology was created in July 1991, partly as a vehicle for challenging the Church of Scientology (Grossman 1995). The OSN that formed via the newsgroup comprised Scientology members and critics who held their disagreements in tension while cohabiting the network (Grossman 1995). This was not to last, following a letter being posted to the newsgroup which outlined an alleged unofficial strategy to dominate discussions via an influx of participants posting pro-Scientology messages. Contestation also emerged around the quotation and posting of copyrighted Scientology materials to the newsgroup, leading to legal action being threatened or taken against some participants and Usenet service providers. Many messages were also removed by third parties, which disrupted discussion (Grossman 1995). Peckham's (1998, 318) analysis is pertinent to my above discussion of affordances since he considers these events from the perspective of resource mobilisation theory, in which social movements are understood to focus their activity on the acquisition and deployment of various resources that sustain the group and enable them to affect change. For Peckham (1998, 322), the key resources deployed in this conflict are the 'virtual resources' of Usenet bandwidth and anonymity. I argue that these are not purely virtual since capacity to send and receive messages on Usenet relies on physical network infrastructure which has an ultimate maximum capacity,

and anonymity is provided via computer software that affects the information shared when messages are sent across, and stored on, that infrastructure.

Nonetheless, resource mobilisation theory illuminates how Usenet's affordances shape the conflict that occurred. Regarding anonymity, participants' ability to hide their identity supported their posting of controversial materials and views by reducing fear of legal repercussion (Peckham 1998, 339). This heightened and perpetuated the conflict among participants in the OSN, since it encouraged actions that critics of anti-Scientology messages took to reduce the visibility of those messages, including posting a high volume of messages in support of Scientology and seeking legal redress to prevent dissemination. Those actions attracted new people to the newsgroup who sought to protect the freedom of speech online, irrespective of their views on Scientology (Peckham 1998, 333). This increased activity brings the issue of bandwidth to the fore. The de-centralised network and asynchronous communication allowed the exploitation of Usenet, since the messages could be sent to the newsgroup from any international location that could access it and be read by others whenever they next connected. In this case the early participants' initial fragile cohesion, which could be conceptualised as two sub-groups for and against Scientology each exhibiting their own like-mindedness, was undermined by threats, flooding the network with messages, and the cancellation of select messages. For Peckham, this approach was about increasing the proportion of the communications that were positive stories about Scientology. From my perspectives of cohesion and incivility on social networks, I see this persistent posting of positive stories as decreasing cohesion as established participants opposed the influx and their ability to be heard and connect with other participants in their OSN was threatened.

The Scientology-focused conflict is different from previous examples that I have reviewed. This illuminates two distinct ways in which conflict and incivility can occur and undermine cohesion: *discursive* and *structural*. Regarding Heaven's Gate and Christian Identity militia, the groups' messages were sometimes met with hostile responses when they connected with other OSNs via newsgroups that provided infertile ground. In the case of the unmoderated Jewish newsgroup, there was incongruence between the Jewish participants' views and those who challenged or were hostile toward them. In these cases, incivility occurred due to the topics of discussion and participants' responses to others' views; it was rooted in antagonism expressed *discursively*. By contrast, the conflict and loss of cohesion on alt.religion.scientology was caused by the significant increase in pro-Scientology messages and the cancelling of select other messages. Engagement among and between

the Scientologists and critics who connected via the newsgroup was reduced since their messages formed a smaller proportion of the messages posted. The attacks reduced their ability to sustain their network connections with one another and expand their OSN. In other words, the network was disrupted *structurally*, impacting the functioning of the network infrastructure and the social connections between participants within the groups.

These studies also shape how I perceive cohesion in my research, leading me to operationalise it in ways that reflect how it manifests in OSNs and is visible in participants' digital traces captured in archives. I specify this as cohesion may be operationalised in ways particular to the study at hand, even if definitions remain similar. For example, Lane's (2021) research focuses on individuals' bonds to each other and to their group, 'rooted in varying levels of consensus in large groups', and he focuses particularly on individual group members (Lane 2021, loc 91.1-97.1). This reflects Lane's use of cognitive science in building artificial intelligence computer models of social groups. The above studies I reviewed illuminate three ways in which cohesion can be expressed and identified. Each are indicated in more than one of the above studies, but I pick out the most salient examples here.

First, cohesion manifests through the persistence of the OSN and stability of discussion topics that relate to a *shared (religious) aim* that is valued by group participants. This is visible in Howard's (2011) study of end times discussion, since the network of those interested in such discussion could thrive when it took place via a newsgroup dedicated to that topic, in which participants shared boundaries about the role of subjectivity in hermeneutics by agreeing on the inerrancy of biblical texts and the potential errancy of participants' interpretations. This shared religious aim, aligned with participants' theological passions, afforded cohesion that helped the OSN flourish.

Second, cohesion may be expressed in *cooperative discussions* in which participants are aligned to the shared aim, even if disagreeing in the course of pursuing it. Herring's (2010) study of perichoretic community in uk.religion.christian and Helland's (2005) analysis of soc.religion.christian indicate this, since they show how agreement occurred alongside disagreement which could galvanise the group where it occurs in the pursuit of the networks' aims, thereby supporting cohesion.

Third, a *high signal to noise ratio* may be a prerequisite for cohesion within an OSN. For this research, a high signal to noise ratio is present when: (a) there is a higher density of

social connections between participants by virtue of them replying to each other's messages, with a high proportion of messages receiving replies, indicating people are engaging with one another; and (b), a low proportion of messages that are spam, posted by bots or by third parties flooding the network and disrupting its functioning. This is to say that the 'signal' reflects sustained social connections and engagement, and the noise is disruptive. Table 6 shows different degrees of engagement across OSNs with more sustained engagement between participants in the soc.culture.jewish.moderated (SCJM) Usenet OSN, in terms of duration of participation and average number of messages per discussion thread. Similarly, the more focused soc.religion.christian.bible-study saw participants contribute to the group for longer than average of other OSNs I researched. By contrast, the looser – in terms of participation and focus of discussion – talk.religion groups, saw lower levels of engagement. They had less signal for participants looking for social relationship and relevant discussion. Therefore, data indicating signal are relevant to cohesion as they indicate sustained social relationships through discussion over time, and with greater intensity where discussion threads are longer. This would be undermined by high levels of spam, or 'noise', which I consider an indicator of incivility and seek to identify in my research sites.

Chua's (2009) study of moderated and unmoderated Jewish newsgroups and Peckham's (1998) analysis of disruption on alt.religion.scientology are instructive here. The high proportion of undesirable messages posted to the unmoderated Jewish newsgroup sparked the creation of the more tightly controlled moderated one and, in doing so, enabled discussion on Jewish topics to flourish. The success is indicated by the migration of participants from the unmoderated newsgroup to the moderated one. Regarding alt.religion.scientology, the 'flooding' of messages and cancelling of critical messages imperilled the bipartite nature of the network, in which supporters and critics cohabited. While this particularly impacted the visibility of critics' perspectives since they became less prominent, the higher number of messages undermined cohesion throughout the OSN by rendering it difficult for participants to interact and thereby sustain their connections with one another.

The above analyses allow me to specify in more detail the concepts of cohesion and incivility that I stated in the introduction, and which comprise my measures of engagement (along with patterns of participation), informed by concrete examples of how they have manifested on Usenet OSNs that involve religion. These concepts capture important facets of how OSN participation influences people's social lives, which affirms the value of my

research. I identify that cohesion and incivility are indicated through the presence and enactment of:

Cohesion:

- The *shared (religious) aims* of participants, often expressed in the group title and behavioural norms
- *Cooperative discussions* among participants toward those aims, notwithstanding disagreement
- A *high signal to noise ratio* within messages on the network, reflected in
 - Participants connecting with several other members by replying to their messages (indicating high network density)
 - A low proportion of messages being spam or other automated communications by third parties

Incivility:

- Antagonistic discourse among participants
- The 'flooding' of a network with messages that *undermine the social structure* of the network as participants' engagement with one another is disrupted

Relatedly, the interactions explored in this literature suggest that incivility is potentially found in discussions which are higher participation, cross-posted, and involve first-time participants, among elsewhere.

Pulling this together, I draw on the studies discussed presently to define cohesion for my research as: 'the presence of social network ties, formed by social interactions, remaining persistent for a duration longer than average for the OSNs analysed, and characterised by shared aims, not disrupted by antagonism or spam.'

This has facets that are found in the operationalisation of cohesion as a concept in other research, For example, Chiovaro *et al.* (2021) consider 'linguistic cohesiveness' in discussion themes on Twitter (now X), and how they reflect real world events at the time of the Arab Spring. While not the same as my 'shared aims' it recognises that cohesion can be present through the development and sustaining of discussion themes, as occurs on Usenet OSNs that focus on certain topics. Recognising cohesion through social network ties formed by interactions is intuitive, finding its roots in the decades-long history of studying social networks (see Granovetter 1973), and I recognise persistence as indicating

cohesion. In a recent example from the study of religions, Stein *et al.* (2020) researched the presence and persistence of social intra-congregational familial ties in Amish communities, and the impact of congregational splits on this cohesion. They find that such a network perspective is useful for understanding retention. I consider this affirms my understanding of cohesion manifesting, in part, through the persistence of ties, since perceiving social network ties at simply a snapshot in time says nothing of their endurance, which shapes social groups and participants' experiences.

With the same approach, I can define incivility as: 'messages which undermine the structure of an online social network through their antagonistic language, or by overwhelming social connections by posting high volumes of posts.'

'Antagonistic discourse' is compatible with various definitions of incivility that Rega, Marchetti and Stanziano (2023, 2-4) and Novotná, Macková and Rossini (2023, 2-3) identify when reviewing extant literature. Further, uncivil discourse may be represented by the derogatory or profane language present in a substantial minority of online media (Weber *et al.* 2020, 1) and within a religious OSN contemptuous language or insulting names may be used toward those with whom a participant disagrees, as Mullan (2015, 113) found in her study of a Catholic social media site. Particularly, uncivil language may be directed toward minority or out-groups, as found in Ortiz's (2020) research about trolling online, which identified that abuse is often racist or sexist in content. Religious groups may be the subject of such abuse, as documented in Kesvani's (2019, 98-101) discussion of abuse received by a female Muslim taking part in a counter-protest. These characteristics of uncivil online discourse are reflected in Trifiro *et al.*'s (2021) research which, though focusing on American political discourse on Twitter, reviewed extant literature that described the properties of incivility and then articulated characteristics by developing that work through their own big data analysis. Such characteristics can be identified by analysing language.

Incivility through the flooding of the network is more novel and relevant to Usenet. Specifying these concepts enables me, in chapters two and three, to determine computational processes that measures facets of cohesion and incivility, and factors associated with them, when analysing the message archives that comprise my research sites. Using a computational methodology is valuable here as it enables me to process and analyse a high amount of OSN data, which is desirable since the growth of religion on Usenet produced a great deal of religion focused discussion that may be analysed. However, before proceeding to this methodology, the final section of this chapter evaluates

the concepts of community and authority which are widely used lenses in the study of religion online and incorporate facets of cohesion and incivility in their analyses. This affirms the relevance of my research to digital religion studies by showing how my perspectives of cohesion and incivility contribute to understanding community and authority online. These evaluations are not exhaustive; they focus on select literature that helps me achieve this aim.

1.2.8 Selecting my research sites

Having explored religion on Usenet, I was in a position to select a subset of the available Usenet OSN message archives as research sites. I analyse these archives to answer my research questions. I also undertake discrete case studies of certain religious groups and interactions within the research sites to show how the methodology can be applied to analyse, for example, a specific religious group or event. These case studies nonetheless contribute to answering my broader research questions. To make my selection, I first surveyed the archives that are available. I became aware of primary sources of archives through the research I cite above and in chapter one. Some sources republish Usenet messages with limited capacity for users to search, such as Google's Groups service, but I consider these secondary sources as they rely on underlying archives. There are three key primary sources of archives: the UTZoo archive of more than two million messages from 1981-1993; the archives made available by the Giganews Usenet service provider; and the Usenet Historical Collection (UHC) of data from an anonymous source, perhaps derived from Google Groups. UTZoo is comprised of a nested series of folders each containing multiple files which typically record single messages, save for a relatively small number of 'digest' messages, such as digests containing sets of messages sent to certain OSNs on ARPANET with which Usenet connected. As noted above, Giganews and UHC archives are in MBox format, in which messages are presented sequentially with some limited structure.

My first decision was to include religion discussion from UTZoo as a research site. I have four reasons for this. First, UTZoo is unique in providing perhaps the earliest extant dataset of online discussion of religion that is in the lineage of modern social media. While far from complete, manipulating and extracting the archive is akin to a digital archeological exercise in which fragments of data indicate social networks and activities from a previous time period. In fact, some later developments on Usenet which I analyse in chapter five are illuminated by analysis of UTZoo, such as the prominence of the soc.culture.jewish group

which is likely supported by early establishment of net.religion.jewish, as seen in the UTZoo archive. Therefore this archive provides historical context for the Usenet inherited by a wider audience in the 1990s and 2000s.

Second, it spans 1981-1993 so traverses the divide between Usenet running primarily through a limited number of more centralised sites which retained administrative control and the latter, more open Usenet running over the internet (pp.30-34). This may illuminate some impacts of central network control over online religion discussion. Indeed, the archive illuminates varying degrees of density between religion groups and their connections to others, with net.religion as a stand out example in its early establishment, growth and social connections. The talk.religion group also indicates openness and breadth although, as discussed above (p.33), access to 'talk' groups was more limited during Usenet's earlier years and this will be reflected in my social network analyses and discussion of talk.religion groups from UTZoo.

Third, Usenet's early demographic of scientists, government researchers, and elite enthusiasts differed substantially from the later Usenet which was widely accessible and popular. UTZoo may suggest the role of religion discussion among a more elite demographic in the 1980s and early 1990s, which is relevant to literature on religion, science, and universities (cf. Ecklund and Scheitle 2018; Marsden 1994).

Fourth, considering my research questions, elements of cohesion are visible in UTZoo such as the development of dialogue through discussion threads, rather than standalone messages, with around 24,569 messages in my extract of the UTZoo dataset being replies to existing threads. Further, complementing this, the potential of flooding with spam can be investigated and particular ways of managing conflict/incivility (thereby supporting cohesion within those developing discussion threads) are present. This latter observation relates to my point in the preceding paragraph, as participants approach of expressing measured disagreement, noted in my analyses (pp.181-184), is becoming of the professional researcher/scientist demographic of Usenet's early years.

Fifth, UTZoo presents a particularly difficult methodological challenge to researchers organising and extracting data. Hence deriving ethically sensitive datasets and insights from UTZoo provides a particular service to the wider academic community.

The UTZoo dataset comprises a fraction of Usenet messages from 1981-1993, due in part to the cost and limitations of storage space at these times and the archivist's informal approach, but analysing it illuminates some discussions from early online religion OSNs, some interaction between religion groups and others, and some of the topics discussed. It helps place today's social media in historical context and therefore benefits my analyses of later archives and will be citable in other research. Nonetheless, later archives are fuller and afford 'proof of concept' for testing and refining the computational methodology that I have assembled. Consequently, my second decision was to include among my research sites a subset of later archives of specific OSNs relating to the Abrahamic faiths from the Giganews archives.

Rather than representing an amendment to my research questions (pp.86-88) by focusing on specific religions, it is instead a decision that allows me to address my research questions most fully, since my first question includes specifying a focus on a substantial sample of *prominent* religion-focused historical Usenet OSNs. I sought to identify OSNs that reached a large scale (so provide a substantial sample) and prominence. With this in mind, I had five reasons for selecting archives of Abrahamic religion groups.

First, the larger scale and breadth of these OSNs led me to include a OSNs with a range of characteristics. This was important for enabling comparison and contrast between OSNs, since it enables me to identify cohesion and incivility to *relative* degrees across the dataset. This, in turn, helps me to answer my research questions. These characteristics include the range of religions being analysed, the numbers of participants, the level of participation as measured by the numbers of messages and lengths of discussion threads, the Usenet hierarchy in which the group is situated (such as 'soc' or 'talk'), whether a moderation policy is in force, and the proportion of first-time participants that take part in discussions on average.

Second, the archives provide substantial datasets as discussions of these large religions was popular. Third, and relatedly, the breadth of datasets includes discussions of religion at the intersection of other topics highly relevant to cohesion and incivility, such as politics, science, minority religions, and LGBT+ lives. Fourth, focusing on different faiths within the Abrahamic religions shows how the methodology can be applied to different groups. Fifth, they enable contrast with religion discussion in UTZoo (so far as is useful since UTZoo is fragmentary) as my UTZoo analyses refer principally to the Abrahamic religions. Sixth, analyses of these provides a foundation for future potential projects regarding online

discussion of these faiths using additional datasets to answer further research questions; focusing on the Abrahamic faiths, compared with very small religions, broadens my options for such research.

Dame-Griff's (2019) analysis of Usenet archives for transgender discussion (p.96) helped my consideration of sources. Dame-Griff used select archives from the UHC, whereas for my later research sites I chose to focus on the Giganews collection. UHC is a defensible choice but I preferred Giganews as, like UHC, it provides very large archives from a broad range of OSNs relevant to my research but, unlike UHC, Giganews' provenance is perhaps clearer; the data were provided by a leading Usenet access provider (Henderson 2012) to the Internet Archive. As Giganews provides access to Usenet, messages flow through and are retained by its servers. The resulting archives have good integrity (notwithstanding limitations that I discuss below), helping to assure integrity of my research.

1.3 Theoretical concepts

1.3.1 The choice of community and authority as theoretical concepts

My review of literature identified concepts developed and deployed frequently in digital religion studies. Prominent among these, on my reading, were community, authority and identity. I discuss scholarship on community and authority subsequently in sections 1.4 and 1.5. Religious practice also figures, such as in Hutchings' reflections and research on immersive digital three-dimensional church services which include prayer (Hutchings 2015; 2010). The concept of 'network religion' in the 2010s also incorporated these concepts as in the five traits Campbell (2012, 68-83) presents: 'networked community, storied identities, shifting authority, convergent practice, and a multisite reality'. This nuances these concepts by contextualising them to the affordances of social worlds that blend offline and online elements into one social whole. This also indicates their salience to Usenet which is, perhaps more than some online discussion systems, multi-faceted in the way networks develop within and across groups. The rootedness and ongoing salience of community, identity and authority are further reflected in Campbell and Evolvi (2019), discussing how scholarship understands community forming in new ways, authority potentially challenged, and the performance and negotiation of identity in relation to religion via digital technologies.

As well-established concepts by the time I formulated my project in the late 2010s, I considered each of these as potential lenses through which to analyse Usenet. Indeed all are relevant. While not including ritual as a theoretical concept in my research I note that religious practice, particularly prayer, does occur on Usenet and ritual can be a topic of discussion across Usenet's myriad OSNs. Identity, which I ultimately chose not to include, is perhaps more salient and I considered focusing on this. Usenet affords multiple markers of identity, which I explore indirectly in this thesis. These include the varying degrees of self-disclosure afforded through pseudonymity; the professional status of scientist, academic or researcher of Usenet's early participants and signalled through their signatures; and ways in which behavioural norms inform how people express through writing, such as the use of certain language, emoji and more. In the absence of later social media affordances such as profile pictures and avatars, these text signifiers influence the identity participants perceive other participants to have. Indeed these aspects of identity interplay with community and authority on Usenet. Regarding community, identity may influence the development of relationships, with compatible identities potentially supporting longer-standing participation and the pursuit of shared aims and practices online together. Regarding authority, some identity positions may clash with others, such as in religion and science in which scientific methodology or findings may clash with certain readings of religious texts. Here, those who align themselves with identities of scientist or Bible scholar may find identity influences authority.

Nonetheless, on reflection I chose to exclude identity as an explicit theoretical concept in my research as I consider that community and authority are the two most salient. This is because they intersect effectively with my measures of engagement on Usenet — cohesion, incivility and patterns of participation — so I identified that researching Usenet could make a contribution to understanding community and authority within digital religion studies. Chapter six presents my conceptual conclusions and then links the themes of my major findings back to what we learn about community and authority (pp.247-250). In short, cohesion can enable and reflect some social behaviours that certain scholars identify as indicative of community:

- Sustained interaction
- A relative stability of a core of active members
- Established behavioural norms
- Distinct group identity
- Management of conflict

- Shared faith (when referring to community in religious contexts)

Section 1.4.3 further explores select scholars' identification of markers of community in online spaces.

Further, authority can enable this by maintaining boundaries around discourse and practice through, for example, moderation, establishing behavioural norms through introductory FAQ documents, excluding those who do not conform to behavioural norms (through formal banning or informal social exclusion of not replying to their messages).

On the other hand, incivility can undermine cohesion through discourse and other online behaviour. Regarding discourse, the use of vulgar language or tone, or discussion of offensive subjects 'off-topic' for the OSN at hand, can dissuade others' ongoing participation by creating a space less conducive to supportive and collaborative discussion. Further, spamming an OSN with off-topic, adult or scam messages, or simply a very high number of messages that present a certain view on a topic, can undermine others' ability to converse in groups and form relationships through sustaining collaborative on-topic interactions via the OSN. I explore this in section 1.2.7 (pp.57-59), referring to a 'high signal to noise' ratio as supportive of cohesion in an OSN. While cohesion and incivility may co-exist, incivility can be an opposing and undermining force, so researching it adds an additional complementary dimension to understanding online social interactions.

In sum, for the present research I selected community and authority as these are salient lenses for understanding social interactions regarding religion on Usenet. My particular perspective, and contribution to digital religion scholarship, is in how cohesion and incivility affect these social interactions and relate to community and authority. Community is indicated through the presence of certain group characteristics, summarised above, explored further in section 1.4.3 (pp.73-77) and reflected upon critically after I concluded my research in sec. 6.6.1 (pp.248-250). Authority is represented in discussions through (1) technical roles such as gatekeeping and service provision which may choose to permit or deny access to certain groups, as (2) performance of the authority of texts, religions, or people.

My conceptions of community and authority are related to other models in that they are drawn from scholarship which conceptualise and engage with community and authority in various ways, which I discuss in sections 1.4 (pp.68-79) and 1.5 (pp.79-85). I focus on

elements of community and authority that are visible in my analyses through the lenses of cohesion and incivility. These elements are sufficiently broad and well-rounded that my findings contribute to understanding community and authority online, as I discuss in section 6.6, made possible by the computational methodology supplemented by manual analysis.

1.3.2 The religious-social shaping of technology

The theoretical perspective of the religious-social shaping of technology (RSST) ran as an undercurrent as I undertook this research, helping me to see how religion shaped interactions on Usenet. Developed by Campbell (2010a), it draws on the concept of the social shaping of technology (SST), which recognises that the development of technology takes place in social contexts, that its development and use are negotiable by its users, and indeed SSC emphasises users' agency alongside that of those in control of technologies' production (Campbell 2010a, 50, 53). Here, technology is not developed in a vacuum but in developers' and users' social worlds. This is salient to Usenet, particularly (though not exclusively) in its earlier years, when users' actions shaped the structure of Usenet's OSNs through their naming conventions, the technology that connected different OSNs, and in how they voted pseudo-democratically on OSNs' creation. Further, users' actions shape the content of Usenet OSNs and what is defined as acceptable use, which can then become reaffirmed through socialisation into an OSN, FAQ documents for new participants and through moderation. SST balances narratives of the power of 'big tech' and relates well to Usenet with its open and distributed nature lacking centralised control.

However, when religion is involved, it goes further. Campbell's religious-social shaping of technology 'seeks to explore in more detail how spiritual, moral, and theological codes of practice guide technological negotiation' (Campbell 2010a, 59). RSST recognises how religious groups, and their individual members, negotiate their use of technology in light of their religious identity, beliefs and practices, accepting and rejecting elements, and shaping its use in potentially innovative ways to suit their preferences and needs (Campbell 2010a, 60). We see this on Usenet and here, I note three ways in which participants guided by religion influence Usenet.

First, Usenet's relative openness to the creation of new discussions groups (OSNs) made space early for elements of religious community, as in net.religion.christian (explored in chapter four). This naming of OSNs to demarcate who might use them, and for what, shapes Usenet's landscape and made space for religion discussion which influenced the social

world of Usenet that later participants inherited and shaped further. Further, participants connect religion-focused OSNs with others through crossposting, potentially increasing others' exposure to religion.

Related to this, some religious participants may feel the imperative of proselytisation. While this certainly occurs among, for example, Christian participants, the group known as Heaven's Gate provide a small scale though clear and high-profile example of this. As noted above, the group engaged with Usenet to share their message, ultimately unsuccessfully, but their religious imperative was served by posting and crossposting their messages in OSNs that reflected the syncretic nature of their message, which infused Christianity, New Age and aesthetics of science fiction. This Usenet functionality (and the technology of packet switching) enabled this dissemination and, taken together, in doing this, religious users' actions shape the connections formed between Usenet OSNs and influence discourse.

Indeed religious participants can further shape topics of discussion. As noted above, the use of FAQs and moderation can determine parameters of discussion, influencing others' behaviours on Usenet. Additionally, longer-standing participants can shape discussions in a way that accommodates/reflects their religious perspective since they often participate with higher numbers of messages and engage with new participants, thereby deploying a form of authority. Taken together, this malleability enabled the development of elements of (at times diasporic) community

In sum, RSST frames how Usenet's religious or religion-focused participants shaped the technology they use, in how they influence OSN structure and content, which in turn affects the cohesion made possible and consequently the elements of community present. They use Usenet to affirm and enact religious authorities (explored later in this thesis) and their interactions also give rise to the possibility of incivility. RSST is therefore an undercurrent through my analyses.

1.4 Community in religion online

'Offline' religion is often practised in community. Extant literature notes how practising religion online has limitations, which indicate how religious community may be limited online when it forms around religious practice. First, literature notes the limitations of mediating religious practice via a screen and text or video, which may impede participants' abilities to

experience the 'otherness' of those with whom they engage, flattening the experience of encountering other people to intellectual interactions with reduced commitment (Holmes 1997 cited in Dawson and Hennebry 1999, 32-33). Dawson (2005) makes a similar point regarding ritual online, which may feel detached and disembodied due to audio-visual limitations. Of course these sources reflect the technology used when the authors wrote, though while modern technologies afford a visually and aurally richer experience, interactions still necessarily lack the embodiment afforded when religion is practised by people present in the same physical space. This is a second type of limitation. O'Leary (2005, 40, 45) notes that online Pagan ritual participants cannot hold hands and doubts that the Catholic church would accept the validity of online sacraments. Similarly, Foltz and Foltz (2003, 324) indicate that some elements of Christianity require close presence and they note that practising Christianity within local communities negates digital divides. Tim Hutchings (2015, 152-154) also acknowledges that for some theologians, online community can fulfil a relational element of religious community, while identifying that for the Catholic church online engagement may complement embodied religious community but not replace it adequately. Hutchings quotes the Pontifical Council for Social Communications publication *The Church and Internet* which affirms that 'virtual communities' can lead people to connect and support one another, but that online engagement cannot replace 'real interpersonal community, the incarnational reality of the sacraments and the liturgy, or the immediate and direct proclamation of the gospel' (Foley 2002). This distinction became very much more live when, in some countries, churches closed at the start of the Covid 19 pandemic. Tan's (2020, 81-82) reflections on the embodiment of communion indicate the limitations of digital media during the pandemic since such media cannot replace embodied communion, while Tan also emphasises how digital media afford engagement with religion in other ways such as via live-streamed services which may still support religious experience and connection. I argue that this supports my focus on OSNs, since they can facilitate online discussion and some facets of community (which I discuss presently) effectively, allowing me to focus on these and not facets of religion that some scholars consider cannot be mediated online.

Now I have briefly summarised these limitations, I can evaluate literature which considers definitions or, less formally, facets of community and how these manifest online. My starting position that OSNs *can* facilitate facets of community is affirmed by Wellman and Gulia's (1999) analysis of computer networks as social networks. This challenges the dichotomy between offline close-knit ties and online dispersed networks by identifying that many people's community ties are already mediated via communication technology and are

specialised, not based on geographical proximity (Wellman and Gulia 1999, 17). If we consider that loss of spatial proximity is not a barrier to forming community, this supports my view that *facets* of community can emerge within religion oriented OSNs so long as we exclude community formed around activities that require spatial proximity between community members. Indeed, Herring (2005) uses the framework developed in Wellman and Gulia's paper to affirm that the Usenet OSN uk.religion.christian has the features of a community.

The term community, in relation to online social groups, escapes neat definition (Coco 2008). Nonetheless definitions exist and these offer overlapping and often complementary perspectives. This section analyses select literature to articulate three approaches to identifying community online. I then explain how studying groups from the perspectives of networks illuminates facets of community. In doing so, I identify how the methodology contributes to the study of religious community online.

1.4.1 Relationship and commitment indicate community

The first approach I consider emphasises relationship and commitment as markers of community online. I start here as these threads also run through the other approaches I review. Hutchings' (2015, 151) analysis of what people mean by online community opens by noting that users of early online communication systems, of which Usenet is one, referred to their groups as community since they supported one another, built a sense of shared identity and established friendships through ongoing conversations. Hutchings has studied online expressions of church extensively, including research for his (2010) PhD, which incorporates a study of the Church of Fools. In 2004, the Church of Fools ran as a well populated three-dimensional representation of church in which people used avatars to participate and interact (Hutchings 2015, 156-157). Hutchings evaluates Church of Fools against Rheingold's oft-cited definition that communities 'emerge from the Net when enough people carry on public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace' (Rheingold 1993 quoted in Hutchings 2015, 152). In the revised edition of Rheingold's book, he states he could call such communities 'online social networks' (Rheingold 2000, 359), indicating the contestation that has emerged around the term community and prefiguring my discussion (below) on how the perspective of networks relates to conceptions of online community. Hutchings also quotes Meredith Gould's (2013) perspective that '[T]ime and energy put into quality interaction is what makes a group of individuals who share interests and concerns become a community no matter

where it's located' (Gould 2013, quoted in Hutchings 2015, 152). Hutchings (2015) argues that Rheingold's definition indicates Church of Fools was a community, while at times experiencing disruption from antagonists. This is affirmed by Etzioni's definition of community, which is not drawn from online religion, but adds useful specificity. For Etzioni, communities have two elements. The first is 'a web of affect-laden relationships among a group of individuals, relationships that often crisscross and reinforce one another (rather than merely one-on-one or chainlike individual relationships)' (Etzioni 2004, 225). This reflects the dense social relations that can occur within church, since people often know many other members of their congregation. I argue that since the digital substrate of Usenet and other online networks affords many-to-many communication, such relations can occur online. Second, Etzioni states that communities have 'a measure of commitment to a set of shared values, norms, and meanings, and a shared history and identity — in short, to a particular culture' (Etzioni 2004, 225). This indicates the relevance of my measures of engagement to the study of community when focusing on relationships and commitment, since I identify cohesion as indicated by the presence of shared aims within the group, cooperative discussions toward those shared aims, and a high density of social relations between participants. I identify incivility through antagonistic discourse and the 'flooding' of a network with messages that undermine the social structure that enables participants to sustain relationships with one another — with enough antagonistic 'griefers', this could have happened at the Church of Fools.

This understanding of community is borne out in Neumaier's (2019) survey of participants in online Christian discussion boards, which adds nuance by identifying specific characteristics. Neumaier asked respondents whether, and how frequently, they engage with each other outside the discussion board and sought their level of agreement regarding which of sixteen characteristics of online discussion boards (which reflect my definition of OSNs) would indicate community. The facets with highest mean average agreement among the 842 respondents include participants recognising each other through, for example, the opinions they express, an intention to carry on participating, and the presence of insider references or knowledge gained through participation (Neumaier 2019, 29-33). I note that the OSN having shared aims or goals was lower down the list of facets recognised as indicating community, which Neumaier (2019, 32-33) suggests may indicate users experience belonging through shared history and supporting one another, rather than sharing religious (or other) convictions. Nonetheless, the mean average response of 2.74 for this characteristic out of a scale from 1 (agree not at all) to 5 (agree absolutely) indicates that many participants did not disagree that shared aims indicate community. Indeed

Neumaier's paper (2019, 33-37) includes analysis of qualitative interviews with discussion board participants, in which three types of community on the board are identified. The first two are 'siblings in faith' and 'board family', indicating the role of relationship and support among group members for one another. The third is 'combat zone', in which participants from various religious backgrounds engage in sophisticated and intellectual discussion, not always agreeing but engaging sympathetically. Neumaier defends this as an expression of community and I consider it illustrates cohesion through the pursuit of shared aims and cooperation insofar as participants were aligned through their discourse in the respectful pursuit of knowledge and discourse.

1.4.2 Participants sense community

Neumaier's study also shows that community can emerge subjectively from participants' experiences. This is the second approach to identifying online community that I evaluate. This approach often involves defining criteria that indicate community, while emphasising an informed but subjective assessment of its presence. Blanchard and Markus' (2004) paper is useful here and is cited by Campbell (2013) as a key text for understanding online community. Blanchard and Markus (2004) identify from literature that there is no consensus definition of community, but that objective measures of participation in online discussion groups (which I define as OSNs) such as continued participation, the number of messages, and the proportion of members who are active, are alone insufficient to identify community. The paper illustrates through quotations from their study of a Usenet OSN focused on sports training that participants considered the group to be a community. The OSN had approximately 17,000 daily readers and received approximately 100 messages each day. Blanchard and Markus (2004) identify, from participants' statements, multiple reasons for their perception that the OSN is a community. The reasons prove familiar in light of the research I cite above. They include recognising other participants and identifying them through the substance of their messages, supporting one another, developing personal relationships and, for fewer participants, feeling an emotional connection or obligation to the OSN. The authors identify that participants' behaviour reflected processes of community construction in line with these characteristics.

While Blanchard and Markus' paper is not focused on religion, other scholars writing about religion online agree with Neumaier in reporting that they or others sense community through their participation. In an early example, David Lochhead (1997, 67) writes from a Christian perspective that 'virtual community' is a product of digital networks and their global

reach which unites people of common interest. Referring to discussions of Kierkegaard online, Lochhead (1997, 67) asserts that '[if] I become an active participant in a group like this, I soon have a sense of being part of a very real, if unseen, community'. Later, Bennett (2012, 117) argues that online communities exist, not least because participants often perceive themselves to be part of a community. She quotes Campbell's (2005) research into online community, which I discuss below, to note that Campbell's participants identified religious community due to their relationships or shared ideology. This has some overlap, particularly in terms of relationships, with how Neumaier's and Blanchard and Markus' respondents identified community, and both relationships and shared ideology reflect how I identify cohesion within my measures of engagement. I argue that having specific criteria for identifying community is important, lest the term could be applied too broadly or be criticised by those unaware of how various facets of community can be mediated online effectively. I agree, therefore, with Lorne Dawson (2004, 81) who stated that Rheingold's definition of religion that I quote above is plausible, but that specific indicators are required to measure community. Having such criteria, which have been affirmed from emic and etic perspectives, allows us to identify the presence of community online and compare and contrast between different online communities / social networks or within one group over time. The criteria may not gain scholarly consensus, but in practice there is significant overlap between characteristics identified in research. The development of such criteria leads me to disagree in a limited way with Shane Hipps (2009, 114) who, writing on the influence of technology on religious faith, perceived that online community may be like 'cotton candy', providing some immediate satisfaction but not sustainable nutrition. Hipps states that community requires intimacy, permanence, and proximity, of which, in general, only intimacy is experienced online and to only a limited degree. I accept Hipps' statement regarding proximity, but permanence may feel ephemeral in many community settings as people come and go over time – some change may be inevitable – and indeed Usenet connections can last to some degree, as I establish in my quantitative analyses of the average duration of OSN participation, in which longer-standing participants emerge among those who join and depart quickly. Neumaier's respondents and other research that I discuss presently indicate that some concrete facets of community are mediated online and that participation can be persistent.

1.4.3 Sets of characteristics of community online

I will now contrast three scholars' highly pertinent sets of characteristics of community online and, in doing so, affirm the usefulness of my measures of engagement for capturing

important facets of online community and contributing to research into religious community online. This is the third approach that I evaluate for identifying community, namely elucidating sets of characteristics of community. It therefore relates to the approach of participants sensing community, as that sometimes results in specifying characteristics. The three sets of characteristics are from Dawson (2004), Herring (2004) and Campbell (2005). Herring was not writing specifically regarding religion, but Kołodziejska (2018) operationalises Herring's characteristics in her analysis of Catholic discussion forums, indicating their pertinence to the study of religious community online. All three were published in 2004 or 2005, reflecting the focus on community in computer-mediated communication scholarship of the time and, in turn, following earlier studies such as Rheingold ([1993] 2000), Baym (1995), and Jones (1995). These 1990s studies were not focused on religion but refer in part to Usenet and are cited in literature I read for this review, indicating their influence.

I chose Dawson, Herring, and Campbell's work for two reasons. The first relates to how I kept a track of the development of scholarship related to my research, and my reading of it, by compiling a literature and scholar bibliography. This refers to around 765 publications, including around 204 on digital religion. I kept notes on all texts I identified and read, and categorised them to group them with others relevant in the field. This helped me identify which scholars were prominent in academic discourse and I found that Dawson and Herring's work had prominence and was engaged with by others, and Campbell is a leading scholar in the field. This on its own, however, is not enough, as I had to consider their relevance to my research. To do this, I considered that these scholars' sets of characteristics refer to ways of engaging online that have similar functionality to Usenet. This is perhaps unsurprising given the early 2000s vintage of their sets of characteristics, in which text-based message board and instant messaging were more prominent among the mix of computer-mediated communication services available. Consequently they help create a connection between community, as can occur on Usenet, and similar services and my operationalisation of cohesion since these characteristics of community relate to ways I identify cohesion, as I set out above (pp.58-61).

Dawson reviewed earlier scholarship on online communities considering the technical affordances of computer-mediated discussions, drawing in examples of religious communities online. He identified six characteristics that indicate the presence of community:

- Interactivity between participants
- Stability of membership within the group
- The stability of participants' identity
- Netizenship and social control, which is to say social processes to guide expression and behaviour
- Personal concern for one another
- The community taking place in public

(Dawson 2004, 81-85)

On Dawson's characteristics, which he indicates are not necessarily definitive (Dawson 2004, 85), community may be identified in part through stability of the network, with a persistent core of participants who have stable identities interacting with one another, along with the application of behavioural norms that shape participation toward agreed or imposed aims. Dawson also emphasises the importance of interpersonal relationships within the group, indicated by the characteristic of personal concern for one another and the stability of participants' identities, not least so they recognise one another. Although *prima facie* similar, the six groups of characteristics identified by Herring have a slightly different focus. They are:

- The presence of active and self-sustaining participation with a core of regular participants
- The group having a shared history, purpose, culture, norms, and values
- Participants exhibiting solidarity, support, and reciprocity
- Criticism, conflict and means of conflict resolution
- The group having self-awareness as an entity distinct from other groups
- The emergence of roles, hierarchy, governance, and rituals

(Herring 2004, 355-356)

Herring categorises the first and fourth characteristics as 'sociability', the second and fifth as 'identity' and the third and six as 'support'. This overlaps broadly with Dawson but has less emphasis on stability. While I appreciate how stability can help foster community, at the same time communities are inherently dynamic. Indeed a strength of Kołodziejska's (2018) approach when operationalising Herring in her analysis of Catholic discussion forums is that she emphasises how community is always in the process of being enacted.

This recognises how online communities or OSNs inherently evolve over time, since the network of people is reshaped through the relationships they form and dialogue they share with one another. The substance of discourse also evolves, bringing with it changes in the degree to which characteristics of community are present and, as I identify, changes in cohesion, incivility, and patterns of participation. In short, users influence and reaffirm the structure and substance of the community or OSN through their participation. Therefore, individual participants' actions can have ramifications on others' social lives. This instability is clearly visible when analysing a group over time, as I illustrate in chapters five and six. I find that the methodology resonates with Herring and Kołodziejska's approach to the analysis of community dynamics.

Lastly, Campbell identified the following characteristics from her observations and interviews with participants in three online communities which she terms Community of Prophecy, Online Church, and Anglican Communion Online, analysis of which is rooted in Campbell's broader understanding of the nature of community online, as recorded in her (2005) monograph and (2001) PhD dissertation. They are:

- Forming close and supportive relationships
- Demonstrating care for one another
- Feeling valued as an individual
- A sense of connection fostered by frequent participation
- Cultivating intimate and transparent conversation
- Sharing religious faith, which informs the group's practices, expected behaviours, topics of discussion and provides common language

(Campbell 2005, 181-186)

These overlap with characteristics identified by Dawson and Herring and those that Neumaier's (2019) participants identified as indicating community, in terms of members forming relationships, supporting one another and the continuing to participate over time. This does not incorporate the more dynamic perspective that Kołodziejska elucidates, but Campbell's much earlier work helped lay the foundation for later scholarship.

Collectively, this literature indicates how religion can be an active force in galvanising and directing an online community or social network. I capture facets of this through my analysis of group cohesion, all elements of which (shared religious aims, cooperative discussions,

and high density of social relations with a low proportion of disruptive contributions) indicate community and show how religion can be a force that fosters group cohesion. Advances in machine learning natural language processing allow researchers to make progress toward identifying these characteristics in OSN message archives at a large scale. This is challenging and requires careful evaluation. It can be implemented through computer code that analyses the substance of messages, including their topics of discussion and indicators of co-operation, which may be combined with other measures such as persistence of participation and high numbers of intra-group personal relationships to infer elements of community. I capture such elements in my measures of engagement and computational methodology, which I discuss in chapters two and three. This also indicates the relevance of the methodology to studying OSNs that form around phenomena which function similarly to religion in social groups, as well as those which are substantively religious. Functionally similar groups could include high-demand political, activist, or millennial groups, since these may also foster shared aims, cultivate identity, persistent commitment and shape participants' discussion and behaviour within their group. I explore this later in the context of considering avenues for future research (pp.251-254).

1.4.4 Communities and OSNs as networks of people

My preceding analyses show how my objects of study are networks of people interacting in ways that scholars call community at some times and in some places. Perceiving online groups as *networks* of people with ties to one another, expressed through patterns of participation, is valuable since certain properties of network size and structure may be associated with cohesion, incivility, and topics of discussion within OSNs, as I discuss in chapters five and six. All of these factors influence community (given the indicators of community discussed in the preceding subsection) so the perspective of networks is valuable for studying community online. Here, my work is informed by Cox (2008) who identifies that the presence of dense, multiplex ties between people within networks is often associated with the presence of community online. Further, Porter, Onnela and Mucha (2009), whose work I discuss later (p.120), and Lancichinetti *et al.* (2011) identify community as reflected in densely connected networks of people, i.e. people having social connections with many others within their group and fewer connections to those in other groups. Relatedly, Preece and Maloney-Krichmar (2005) argue that focusing on the strength and nature of people's online relationships is more useful than physical proximity for identifying community. Strength can manifest in the duration of positive engagement with one another online and, as discussed above, religion can be a force to galvanise such engagement.

Marc Smith (1999), then a researcher at Microsoft, applied social network analysis to Usenet and identified, among other insights, that many Usenet OSNs are populated by few people and contain little dialogue, instead conveying spam, scams, and uncivil comment. Noting this, Fayard and DeSanctis (2005) cite a range of scholars who conclude that many OSNs fail to function as communities. Indeed the density of people's relations within networks, in terms of how many other people each participant connects with through communication, may be impacted by cohesion and incivility over time, affecting the possibility of community. In practice, we may see less dense ties when participants conflict and disengage, though a short-lived high density of relations may also reflect incivility as participants engage in a cohesion reducing argument. Looking at a social network from the perspective of ties between participants indicates the dynamic nature of communities (Campbell 2013, 65-66) and their potential instability. Informatively, Lundby (2011) makes the case that patterns of belonging are a tool that link community and network since the social bonds that form through belonging to a group, including bonds between those who share religion, can form community. From this perspective, social networks *require* some stability of social ties and for members to share a 'common focal point of identification' (Lundby 2011, 1221). Here, then, the agency of the participant to choose whether or not they continue to engage, along with any forces exerted by their religious bond, can support the emergence, continuation, or fragmentation of community.

Networks and community are therefore distinct, as argued more recently by Neumaier (2019). For Neumaier, communities can be 'emotionally or symbolically charged' social entities and be seen as communities by participants, whereas a network indicates relational ties between people and does not have, from participants' perspectives, shared aims or behavioural norms (Neumaier 2019, 27-28). I agree with Neumaier that networks need not be communities and that, as discussed above, community may be a subjective quality perceived by its constituents. Of course we must not see network and community as a dichotomy since community, considering the indicators discussed above (pp.73-77), can emerge and dissipate on networks, reflecting how networks are constantly forming, evolving, and fragmenting by virtue of participants' actions and the forces, religious or otherwise, that bind them.

The dynamic nature of networks is also indicated in Kołodziejaska's (2018) research (discussed above), in which she observes that focusing on the ongoing process of community creation is particularly important in online contexts since they may change

rapidly. Kołodziejska rightly identifies analysis over time as crucial for understanding the formation of communities and how they are sustained and evolve. This affirms the relevance of analysing OSN message archives since they often span many years and afford a quasi-longitudinal approach in which I can analyse participation over time. In chapter three, I discuss how I use computational social network analysis to glean insights regarding network structure and patterns of participation. I recognise that not all facets of religious community are captured when perceiving community through the lens of networks (Campbell 2005), though the network perspective contributes valuably to understanding community when: (1) analysis considers the structure of a network through patterns of participation *alongside* the substance of the messages people send on the network, as the methodology does, and; (2) research focuses on OSNs that discuss religion but which are not necessarily *religious* communities. This is the case in my research and in many of the examples of religion on Usenet discussed above.

In conclusion, while the measures of engagement that I identify through my research methodology do not capture all facets of religious community described in literature that I review above, particularly those subjective aspects internal to participants such as feeling valued or simply *sensing* community, the methodology does capture:

- The development and persistence of relationships between participants as represented through their discourse
- The stability of core participants' membership
- The identification of shared group aims, represented in cooperative discussions and ongoing participation
- The accommodation of disagreement
- The apparent presence of shared faith, indicated by group practice such as prayer, shared language, and topics of discussion

This indicates how my research contributes to literature that uses the lens of community to study religion online. Before drawing this chapter to a conclusion, I now critically review scholarship on authority in religion online and indicate how my measures of engagement relate to it.

1.5 Authority in religion online

What is meant by authority in online religious contexts? Campbell (2007, 2010) develops four components which are a useful starting point since they broadly reflect the types of authority considered in digital religion literature and some scholars, such as Cheong (2013), cite Campbell's categories directly. They are:

- Roles (such as saints or church leaders)
- Structures (of community, religious practice, or hierarchy)
- Beliefs/ideologies
- Texts

(Campbell 2010b, 257)

They reflect how Weber's ideal types of authority — traditional, charismatic, and rational-legal — manifest when authority is influenced by media and its material structure, in this case the dynamics of online networks (Campbell 2010b, 254; Feldt and Høgel 2018). This, in turn, indicates how the type(s) of authority present in religion online may depend on the context (Cheong 2013). Complementing this, Kołodziejska (2018) also identifies forms and sources of authority that may emerge online. To summarise these very briefly, authorities may be top-down official religious authorities or emerge bottom up — and the two may influence each other. Authority may be situated in a person or be impersonal, such as a church or scripture, and it may be constructed and referenced within the online context, and/or participants may refer to existing sources of authority from offline contexts (Kołodziejska 2018, 60-61).

How authority manifests may therefore depend on the religion being discussed or practised. Campbell (2007, 1048-1054) compares interviews with Christians, Muslims, and Jews, noting their concepts of authority differ. Many of the OSNs that I analyse focus on Christianity and consequently discuss Christian authorities — such as biblical exegesis and the church — though these align with Campbell's focuses on structures, beliefs, and texts, suggesting the insights I produce are relatable to other religious contexts. Also, my measures of cohesion, incivility, and patterns of participation are not dependent on any one religious tradition, though the ways in which authority may *impact* those measures of engagement could depend on the substance of the religion being analysed.

A hypothetical scenario will illustrate the importance of understanding the religious context being analysed. OSNs that discuss a form of Christianity which holds the historical church

in suspicion may be unlikely to draw knowingly on the Catholic church's tradition and theology. Indeed doing so to legitimise one's arguments may, in this context, be rejected by other participants. Within a Catholic OSN the position may still be complicated since while drawing on, for example, the catechism is likely to be considered an appropriate authority, its interpretation or application in a certain context may still spur disagreement. In some discussions of religion, the catechism will aid cohesion through shared pursuit of religious truth and in other cases it will not. Understanding this requires sensitivity to the religious context and affirms the value of my interdisciplinary approach which combines knowledge of theology and religious studies with machine learning to capture factors that reflect the ways in which cohesion and incivility manifest in the discussion or practice of *religion* online.

It serves my research to focus in this brief section on (1) ways in which authority from other contexts is affirmed online, (2) the challenges of doing so and (3) the emergence of new authorities online. I am guided by Campbell (2007) and Kołodziejska (2018), discussed above, in identifying facets of authority online. These three processes capture how online religious contexts emerge against and are influenced by existing religious landscapes. In contrast, Blondheim and Rosenberg (2017, 50) state that '[the] experience emerging from the breakdown and collapse of all entrenched conventions and narratives in the digital world, and the opening of a chaotic abyss can thus serve as a prelude to a fresh new theological start'. This may be the case in some contexts which seek to be detached from authority offline. As Neumaier (2020, 17) identifies, inter-religious discussion online can occur in more neutral contexts without the inequalities that may exist when dialogue occurs in a space formed and dominated by one religious tradition. Nonetheless, online dialogue can be shaped by the ongoing impacts of offline authority as in the cases I review presently.

1.5.1 Affirming existing authority online

Discussion in the 1990s from postmodernist perspectives suggested that religious institutions that principally communicated via 'print and literacy' would be seriously disrupted by online communications (Ess, Kawabata and Kurosaki 2007, 941). However, textual and audio-visual media can be conveyed effectively online, allowing the authority of religious institutions to be expressed, challenged, and transformed through OSNs. Pope Francis' X (then Twitter) account is among the uses of social media that are less disruptive to established religious authority. Genovese (2018) argues that through (then) Twitter, the Pope contributes to shaping discussion of global issues. However, the Pope engages in one-to-many communication, which is different from the dialogical nature of many OSNs in

which people form a dense network of social connections by responding to one another and social hierarchy is, to varying extents, flattened with participants co-enacting the religious authority to which they may be subject (Cheong 2017, 27-28). Therefore, for religious leaders, online engagement may be valuable for sculpting discourse toward forms that they recognise as valid expressions of the religion they lead. This aligns with the concern expressed by the Pontifical Council for Social Communications that 'it is confusing, to say the least, not to distinguish eccentric doctrinal interpretations, idiosyncratic devotional practices, and ideological advocacy bearing a 'Catholic' label from the authentic positions of the Church' (Foley 2002).

Religious leaders' online influence is evidenced in Cheong's (2014) analysis of the Twitter messages of Pastor Kong Hee, leader of an Asian megachurch. Hee's followers form social networks that generally, though not always, affirm their pastor's authority (Cheong 2014). This indicates how traditional authority can be reproduced and affirmed online, co-enacted by the social media communications of religious leaders and members, which support cohesion. Such co-enactment is relational, as Faimau and Behrens (2016) identify regarding the Facebook posts they analysed from the prophetic leader of the Gospel of God's Grace Ministries church in Botswana. Followers' posts that affirm their religious leader uphold the leader's authority, showing how a leader's influence may be enabled by their followers' actions. In a slightly different context, Hutchings' (2013, 168) research of an online church video stream with accompanying chat facility notes that the pastor dissuaded atheistic discussion in the chat but that participants offered to engage with an atheist participant via email instead. In this case, the pastor's authority defined appropriate use of the online chat facility and the participants brought that to pass. In these examples, religious leaders' authority supported cohesion and shared religious aims in their followers' online discussions.

The established authority of religious texts can also be affirmed online by OSN participants, as seen in Howard's (2011) observation of fundamentalist 'proof texting' (p.163) and Abdulla's (2007) analysis of messages regarding the September 11 terrorist attacks on the United States from three Arabic language discussion boards. Abdulla identified that 43% of the messages analysed condemned the attacks and that 30% indicated some justification, mostly with reference to American foreign policy. At the same time, 30% of the messages emphasised how the attacks are contrary to the teachings of Islam, referring to established sources of authority such as the Quran and the ruling of Sheikh Yusuf Al Qaradawi to justify condemnation (Abdulla 2007, 1072-1077). Abdulla finds that on each of the three message

boards analysed, messages seeking to justify the attacks were often posted within the first four days following September 11 and became fewer over time, with messages condemning the attacks being more detailed and posted throughout the period analysed. Textual and other authorities were used to condemn the attacks and counter ideas that they were justified.

1.5.2 Challenges to affirming existing religious authority online

The diversity of perspectives expressed on the message boards that Abdulla analysed indicates how reproducing established religious authority online can be challenging. This was also seen in the unmoderated Jewish discussion group discussed above (Chua 2009, see above pp.51-52), which exhibited incivility since those disrupting the group did not recognise, or did not respect, the Jewish sources of authority that were affirmed by the group's established participants. The discourse that flowed from this fractured the group, undermining cohesion. Indeed the challenge of re-establishing religious authority in online contexts may be heightened in environments such as the early Usenet, where behavioural norms unsympathetic to religion permeate much of the platform. Given this, and my preceding summary of how religious authority is co-enacted online between followers and leaders, online religious authority may be seen as fragile and challenging to establish and maintain. Where successful, such authority provides a basis on which shared aims may be pursued and cohesion cultivated, with disagreement managed constructively within epistemological boundaries. The discussion of texts can provide a starting point to seek such common ground. Bennett (2012, 92) notes Catholics' use of papal encyclicals, the catechism and canon law when participating in online discussions, indicating a strategy to discern within a plurality of opinion by leveraging the authority of church teaching. Indeed, Catholics are encouraged to share their faith online (Mullan 2015, 98-99) and this indicates how they may draw on existing authorities to do so.

In addition to the challenge of (re)establishing authority in new social contexts, a second and related challenge to religions' authority online is the proliferation of incongruent information. This can occur in at least two ways. First, as expressed in the Pontifical Council's concern cited above, an individual or group may use online contexts to extend Church teaching and practice in ways that the Church does not recognise as valid. This may be seen in Brasher's (2001, 170-176) summary of Marian devotion online which refers to, among other sources, the Cult of Mary and visions of Veronica Leuken, the dissemination of which, one website states, some Bishops sought to ban. This affirms

Dawson and Hennebry (1999, 34) who note that a move towards grass roots or, to use Kołodziejska's perspective, bottom up, authority that sidesteps church authority is afforded when communicating via decentralised digital media. In part, this is enabled where boundaries between leaders and other members become indistinct, as Gelfgren (2011, 11-14) notes regarding the formation of online churches which may not recreate the traditional relationships between leaders and lay participants, or may be formed without the official approval of the denomination they represent. Second, the sharing of information online regarding a religion may serve to undermine its plausibility structure, as Eileen Barker (2005, 74-77) identifies regarding websites or OSNs that share information regarding a new religious movement that the NRM would find uncomfortable, leading members to doubt. This may have been a concern regarding materials posted to the newsgroup alt.religion.scientology, discussed above (pp.55-57).

1.5.3 The emergence of new authorities online

The authority of a religion may be challenged online not just by OSN participants but by new online authorities such as network administrators or OSN moderators. These roles are formed within the specific online contexts that enable them (Lövheim and Hjarvard 2019, 216) which, in the case of Usenet, is the digital substrate that allows people to form and (particularly in earlier days) control access to OSNs. Campbell and Golan (2011) explore such online authority roles in relation to Orthodox Judaism. They argue that those in gatekeeper roles may seek to link themselves with offline religious authorities for justification or may develop their own autonomous authority (Campbell and Golan 2011, 716-717). The online authority and power of those in gatekeeper roles may be unrestrained when religious leaders established in offline contexts do not engage online. Kong (2001, 407-408) notes that religious leaders often would not enter Usenet discussions since their authority may not be recognised or they would be challenged and rejected by sceptics. In this situation, new online authority figures have freedom to shape discourse and participation on their OSN, but without the influence of established offline authorities the OSN may become detached from the wider religious community that they represent.

The influence that online authorities such as OSN founders or moderators have on discourse (see p.43 regarding Hoffman and Fake on *Masters of Scale* 2018) can impact new participants through the diffusion of behavioural norms or a discourse style that new participants must adopt to be socialised into the network. Bennett (2012, 101-102) notes that such adoption may be necessary before established members engage with new

participants. I identify this in terms of cooperative discussions toward shared aims, in which people who contribute in a way unaligned with the behavioural norms and aims of the group may be marginalised. This in turn impacts cohesion since those aspiring participants are not socialised into the network effectively and have fewer relationships with others. Consequently a form of authority may be diffused between established members who safeguard the integrity of the group. Similarly, if a religious leader from an offline context enters an established online platform or OSN, then s/he may need to adhere to the behavioural norms of that environment. Narbona (2016, 94) identifies that these include the appropriateness of the language used, the timeliness of the message, participation in the platform's 'technical specializations' such as Twitter hashtags and the formation of relationships that engage followers and position himself/herself as a leader. Here, the authority of the platform administrators or owners may be in tension with the authority of the religious leader, who may seek to engage or influence followers in ways that are not congruent with platform norms. Similarly, Lövheim and Hjarvard (2019, 216) identify that in order to affirm their religious authority, a person must be competent in negotiating the challenges to their authority that the environment can present in the form of, for example, negative comments, while also being sensitive to their followers.

In conclusion, a common thread in my analysis is that online authority often manifests through influence rather than direct control, save for the power of OSN administrators and gatekeepers. Authority that seeks to influence people online may be fragile, changeable and require co-operation from others within the online context. The methodology can infer the influence of such authority in OSNs in the following ways, thereby delivering insights that contribute to analysing authority online:

- Network structure, with persons in authority occupying central roles within a network of dense relationships, or acting as a bridge between multiple clusters of people who have relationships with one another (as may be the case with Pope Francis)
- People perceived as authorities receiving a high number of responses to their messages
- Positive discussion of religious texts or other canonical sources of teaching
- The topics of discussion promoted by authority figures becoming dominant and shaping discourse for others within the OSN

1.6 Conclusion

My research journey started with my interests in internet history, theology and religion, and computational methodologies including machine learning. My study of these areas developed in parallel and, from them, I identified how a machine learning methodology could afford large scale analyses of potentially impactful religion-focused online social networks. Researching and writing for this chapter then enabled me to specify the foundation that enables and authorises my research, as follows:

- Usenet's digital substrate, affordances and behavioural norms show that how people connect with one another, and what they say, are influenced by the networks over which they hold their discussions and form relationships.
- Extant case studies of religion on Usenet indicate religion is a force that can result in positive and/or negative online social interactions. In the present research, I use the lenses of cohesion and incivility, which are measurable in ways I specify, to analyse some such positivity and negativity in online social interactions.
- The substantial presence of religion on Usenet indicates that it is a promising site for undertaking such research at a large scale.
- Given the potential large scale of the research, a computational methodology including machine learning is highly beneficial for undertaking analysis, identifying macro level patterns and social behaviours, as captured in my measures of engagement.
- Such research is relevant to digital religion studies, since it provides insights into specific aspects of community and authority online, which are well established lenses in this field.

This leads me to propose the following research questions, which license and catalyse my research.

1.6.1 Research questions

1. In what ways, and to what extent, do cohesion and incivility manifest in a substantial sample of prominent religion-focused historical Usenet OSNs?
2. Across the research sites analysed, which of the following factors of these discussions are associated with cohesion and incivility? Whether the message:
 - a. Was a reply to an existing discussion thread
 - b. Part of a discussion with higher than average participation
 - c. A participant's first message within the dataset

- d. Crossposted to multiple OSNs
- e. Posted by a longer-standing participant
- f. Indicated a discussion focused on religion

For RQ1, my preceding review of literature regarding religion on Usenet indicates potential ways in which cohesion and incivility may be present in the research sites I analyse (pp.58-61). My research for RQ1 seeks to identify their presence and the extents to which they occur through my quantitative and manual analyses. For RQ2, factors are identified through the computational method of association rules mining, explored in chapter three (pp.127-128 and Technical Appendix 1.2).

In more detail, I will answer those questions by fulfilling these objectives:

1. Identify and access archives of select prominent Usenet OSNs (research sites)
2. From each message, extract the following data that are *explicit* in the message archives:
 - The date the message was sent
 - The subject line
 - The OSNs to which it was sent
 - Analytic descriptors indicating message themes, including indicators of cohesion and factors suggestive of incivility
 - An anonymised version of the sender's name
3. Regarding OSNs for which archives are sufficiently abundant, extract pertinent features that are *implicit* in the messages:
 - The social networks that form as people send messages to one another
 - Measures of engagement:
 - Indicators of cohesion
 - Factors suggestive of incivility
 - Patterns of participation in OSN discussions
4. Identify associations between the measures of engagement and other characteristics of the messages and social networks that participants formed, to identify which factors impact cohesion, incivility, and participation in religion focused OSNs
5. Apply the analytical approach to two case studies within the research sites, focusing on topics pertinent to theology and religious studies

6. Analyse 1,000 message threads manually to triangulate the results of the computational methodology
7. Compile the results into visualisations and critically evaluate them to answer the research questions

These research questions and objectives position my project as inductive research and enable me to develop the precise computational methodology that I will use for this project, as well as specifying my exact research sites within the vast archives that are available. I address these tasks in chapters two and three.

Chapter Two — Faster Than We Can Run

This chapter explains and justifies my choices regarding the parts of the methodology that extract, organise, and analyse facets of data that are *explicit* in online social network (OSN) message archives. I discuss internet archiving, specify my selection of sources, and explain why they are relevant. I then focus on natural language processing, discussing my methodological decisions regarding computer programming and the algorithms I use. First, however, I take one step back to explore in more detail why a computational methodology, including machine learning, is necessary for answering my research questions.

2.1 Using computational big data methods to study religion

In chapter one, I explained that a computational approach affords the analysis of archives at scale. Such archives are a form of ‘big data’. There are various facets of big data and understanding these provides a foundation for my present discussion of how big data and computational methods can benefit research. The following key facets are synthesised from Salganik’s (2018) prominent and detailed guide to digital research, along with Eagle and Greene’s earlier (2014) summary and Gahar *et al.*’s (2018) citation of the characteristics of data handled by big data computer systems. In summary, datasets may be very large, including hundreds of thousand, millions, or billions of data points; they may be broad by virtue of including many categories or facets, covering long time periods, or holding data on one narrow research focus in extreme detail; they are often used without the subjects they represent sensing the researcher, unlike for example in a laboratory setting in which researchers monitor certain behaviour; they may be cumbersome and require manipulation before analysis and visualisation; and they may require careful analysis to deal with sensitivity of the data, biases within it, incompleteness, and other extraneous data present that are not of interest, all in order to generate reliable outputs and insights.

These characteristics influence the tools we use and this requires something of the ‘hacker ethic’ to select and adapt established methods, and design new ones, specific to the data at hand and the aims of the research. Bainbridge (2020, 20) indicates how researchers must get to know the digital data they use in their research and the computational methods at their disposal, sometimes improvising a novel approach rather than following one prescribed. Scholars in digital humanities may learn how to program computational tools and discover potential data sources in parallel, identifying ways to answer research

questions through their learning of these two components. This way of learning resulted in the methodological approach I utilise.

By applying their tools to the data, researchers generate datasets and other outputs that are representations of the social realities to which their research relates. In my research, these representations are datasets that capture people's interactions on OSNs and indicators of cohesion, incivility, and patterns of participation. The methodology is the pipeline between my raw data, which are the digital traces of interactions compiled into the archives I use, and my datasets and findings, such as factors that influence participation. The pipeline must preserve relevant details of the social interactions that the archives reflect, while transporting and transforming the data into forms analysable at very large scale and incorporating ethical safeguards.

2.1.1 How big data computational outputs represent reality

My datasets must reflect the OSNs from which they are drawn as accurately as possible, since I rely on the datasets to represent the social reality of participating in the OSN. The present chapter considers this in relation to (1) how I use computational methods to extract analytic descriptors regarding messages sent across the OSNs I analyse and (2) the extent to which the archives I use are reliable records of those OSNs. For now, following Davis and Love (2019, 638), I note that '[all] data are imperfect proxies for social life', but my datasets and findings need to preserve and reflect the social interactions present in the archives. Second, I must ensure my computational methods for extracting analytic descriptors of message content and identifying cohesion and incivility are indeed identifying those social phenomena. Computer models that represent and/or simulate some aspect of reality are simplifications (Luke 2015) but must specify the relationship between each facet of the model and the corresponding facet of social reality, and relationships between different facets of the model (Lane and Shults 2020), such as between certain topics of discussion and cohesion in my research. Done well, the resulting findings and datasets enable predictions to be made based on prior occurrences. In my research this could involve insights around the presence of cohesion likely in social networks of certain sizes, or whether certain topics of discussion will encourage people to participate for the first time.

My approach to my research design, which functions on the basis that computational approaches can identify social processes, is to first consider that the messages captured in the archives are digital traces of interactions. From here, I consider that certain

characteristics of those social interactions, reflected in the messages and metadata, indicate social processes – especially where archives allow analysis of interactions over time. This includes cohesion and incivility as I operationalise in the present research, since I specify ways to analyse them (pp.58-61) which are amenable to computational analysis and are checked and enriched through manual review. In identifying those characteristics computationally, these methods suggest the presence of these social processes happening in the ‘real world’ social interactions captured in the archives.

2.1.2 How big data computational methods fit into the academy’s methodological toolkit

This conceptual understanding of how big data methodologies relate to, and aid understanding of, reality, enables consideration of how they complement other well-established historical, textual, quantitative, and ethnographic research methods. For some, they add the possibilities of new insights without superseding any other methods and can create macro level data which sparks closer analysis of certain cases or texts (Beuving 2019; Webster in Brügger and Milligan 2019). Indeed in mixed methods approaches, computational methods afford otherwise inaccessible analysis at very large scale and researchers then use well-established non-computational textual or ethnographic methods (or others) to verify and deepen the computational analyses. For example Šimunjak and Caliandro (2019) use computational methods to obtain samples of Twitter data, extract relevant tweets (in this case Twitter users mentioned by Donald Trump), and document the interactions between those people. They then undertake manual content analysis on a small subset of tweets. In support of this approach, the researchers cite Lewis, Zamith and Hermida’s (2013) research which employed manual methods of coding journalists’ tweets during the Arab Spring alongside using computational methods to organise and extract their dataset. However, computational methods can take us further than in these papers by using algorithms for extracting analytic descriptors of messages and identifying social processes, as discussed in the present chapter and chapter three.

In addition, researchers must understand the wider world in which the subject under analysis is embedded, so we can interpret what the algorithmically produced insights are telling us about that reality (Roth and Luczak-Roesch 2018). Chapter one’s summary of Usenet history and how it evolved helps me understand the digital world that participants inhabited, which contextualises my reading of their messages (Polonski 2017, 114-116). I also manually analyse a substantial subset of discussions to affirm and deepen my computational analyses, as I will discuss. This can then enable a cycle of methodology

decision-making in which a researcher's detailed understanding of the social context and computational methods allows them to design an appropriate computational approach for their research questions and avoid selecting a tool that is, to the researcher, black-boxed because they don't understand it. This combination of methods can then provide a large-scale overview via the quantitative methods and a rich, micro level perspective from, for example, ethnography. Such an approach steers clear of generalising from detailed individual case studies of, for example, religious practice that may not represent broader patterns of behaviour (Bruce 2018). Broad quantitative analysis also helps ensure the academy does not rely on out-of-date understandings about religious phenomena, a point made by James Lewis (2014) who showed that new religious movement (NRM) scholars' qualitative focuses meant they missed large-scale changes in the demographic profiles of new NRM members. While Lewis' approach is not computational big data, the point holds that quantitative methods deliver important insights because of the large scale of analysis they afford.

Before proceeding, it is important to note the limitations of the *learning* in machine learning as I discuss and apply it in my research. This addresses any unease about the use of the term learning with reference to mathematical algorithms which scholars, perhaps particularly in the humanities, may feel. I justify my use with reference to Davison's (2021, 259) discussion of how analogical links between the capacities of humans and capacities of computers may be conceptualised in a weak manner by use of metaphor. I use the term learning in this manner, as a metaphor for how I would learn if undertaking such analyses as a researcher, since aspects of the methodology I have assembled 'learn' latent characteristics of messages and social relations within OSNs, then apply that learning to understand associations between, for example, certain characteristics of discussions or groups, iterating to build knowledge presented in algorithms' outputs. This understanding of learning contrasts with strong analogical comparison which may imply metaphysical claims about computers' sentience or understanding, as seen, for example, in popular depictions of machine learning algorithms as anthropomorphised robots.

I now critically evaluate selected scholarship, identified through my literature review, that uses big data approaches to study religion. This enables me to evaluate critically the contributions these methods can enable and how my methodological approach contributes. I then set out the methodology in detail within the remainder of this chapter and chapter three.

2.1.3 Using big data to study religion

While it remains a small proportion of scholarship on religion, several published papers use computational methods to study religion. I survey a select sample here to illustrate the approaches that have been used and some research sites analysed. This helps to situate my own work and the original contribution it makes to the field.

Some extant research uses computational methods to extract data from websites and organise it into an analysable form. Such a process is documented in Svensson's (2020) collation of articles from a Swedish website that provides Islamic texts. Svensson used the computer programming language Python to assemble a data frame, which organises data into rows and columns as I discuss in more detail later (see Technical Appendix 4). This structure allowed Svensson to identify Islamic authorities cited frequently in the website's articles. Data frames, and research objects like them, also allow researchers to run computer code that identifies associations and/or correlations between data, as in Thomas, Al Shehhi and Grey's (2019) research which finds that obscenity on social media has a negative association with religiosity. This entails first specifying the phenomena of obscenity and religiosity, identifying them within the text and then analysing associations between them. Similarly, Davidson *et al.* (2019) study political left wing and Islamic discussion forums, identifying and operationalising seven different roles that participants could play in the discussions, such as collaborators or leaders. In addition to careful definition of terms, operationalising social concepts such as obscenity, religiosity and collaborator successfully requires assurance that the dataset accurately reflects the social phenomena that the researcher seeks to capture.

In addition to extracting, organising, and identifying associations, research also uses computational methods for natural language processing (NLP) to infer the topics of discussion within messages posted to social media or in other texts. Elwert, Tabti and Pfahler (2020) demonstrate this successfully. They use latent Dirichlet allocation (LDA), a popular method for inferring the topics present within a corpus of text. LDA requires detailed understanding of its assumptions and limitations, which I present later. For Elwert, Tabti and Pfahler it provided some nuanced understanding of what participants in their research sites, Evangelical and Salfist discussion forums, were discussing and enabled their research findings about the role of interreligious discussion in the groups' social standing and identity formation. Reflecting its popularity, Sanders and Ferré (2020) also employ LDA as part of a methodology to group together and identify topics of discussion within news

articles and Reddit comments about Answers in Genesis' Ark Encounter. This enabled research into whether news articles might influence online discussion and suggests the relevance of NLP to my research since Reddit's affordances overlap with Usenet's, to which Reddit is a spiritual successor. Further, Sharon (2021) applied an LDA analysis to a dataset of Twitter messages including some relating to religious nationalism, followed by a close reading of different types of nationalism having inferred topics. This differs from the methodology since I don't use LDA (pp.108-113) and my research incorporates other social processes and analyses, but it indicates that applying NLP followed by a close reading of messages about religion at intersections of other topics, as I do, can generate insights into social lives.

Such NLP models can also identify the presence of religion discussion even when religion is not the primary focus, as in Corvo and De Caro's (2019) research into social media messages that discuss loss and grief. Some participants indicated how religion and spirituality supported those who have experienced bereavement. Apart from online discussion, NLP is applied to religious texts such as Thomas' (2018) analysis of Seventh Day Adventist Ellen White's *Testimonies for the Church*. Thomas identified characteristics such as readability, the positivity, neutrality, or negativity of the sentiments White expressed and the formality of the text, with non-native readers of English in mind. Reflecting the growth of such research, Nakissa (2021) provides an overview of studies that use computational methods to analyse Islamic texts including documents in Arabic. This indicates the language-neutral nature of such methodologies, once developed and configured appropriately.

While many studies research OSNs, other focuses include Phillips' (2019) research to identify historically popular Bible passages and those popular online more recently. Phillips used Google Ngrams to search Google's corpus of print books and he used datasets obtained from popular Bible reading and study tools YouVersion and Bible Gateway, which indicated passages that users engaged with by sharing, highlighting, or marking as favourites. This research has two particular strengths. First, it shows how archived historical data can be used alongside more recent data to answer research questions. This requires operationalising research concepts so that both recent and historical datasets can contribute to the aims despite the data's heterogeneity. Second, it determines change over time, which is made possible by the combination of datasets and methodology used. Phillips' work illuminates a trend toward the popularity of therapeutic Bible passages over those focused on propositional beliefs. Balazka, Houtman and Lepri (2021) discuss the

possibility of using historical data for the study of (non)religion but note challenges to accessing, digitising (if necessary), extracting and organising such data. Phillips utilised the potential of Google's Ngrams tool to access and extract data from the historical texts analysed, but many sources do not have such an interface available, as is the case in my research.

Together, this field of research affirms the feasibility of my use of computational methods to organise my data into data frames and mine it to analyse social phenomena, having configured and applied NLP algorithms to extract relevant details from messages, such as facets of the language used and themes within discussions. The combination and configuration of these methods for use in my research is challenging and the data I use are not structured homogeneously as data may be if using, for example, commercial platforms for extracting data from social media websites. How, then, do my data compare to those used in other computational large-scale analyses of historical textual data, particularly scholarship focusing on Usenet? Engaging with the literature in this emerging research area has helped me to develop and affirm the methodology.

2.1.4 ...and to study Usenet

Some of the above studies combine multiple computational methods and some previous research on Usenet does the same. Two of the strongest studies, in terms of using computational methods to study at detail and depth are, first, Turner *et al.*'s (2005) research which analyses hierarchies of newsgroups, interactions between them and variations in participants' contributions, focusing on the 1,000 most active newsgroups from 2003-2004. Reflecting the hacker ethic, they used software designed for analysing how computer hard drive space is being used to visualise Usenet hierarchies, alongside more bespoke software for visualising participants' behaviour, inferring different roles that participants may take on. Second, and similarly, Smith's (2003) research used Microsoft's Netscan software to examine newsgroup popularity, interconnections between newsgroups, and the data available for analysing user behaviour. Both studies illuminate participants' micro-level behaviour while presenting a macro-level summary of their online environments. Like my work, they analyse broad context and specific behaviour, though do not utilise machine learning algorithms to identify what is associated with certain characteristics of social networks or behaviour.

These publications accessed Usenet data closer to the time it was generated, rather than using the longitudinal archives that I analyse. Some research has used such archives, however. Avery Dame-Griff's (2019) study of transgender OSNs on Usenet is a good example. Dame-Griff makes a methodological contribution by using the programming language Python to extract and organise messages from archives, and visualise patterns of participation and the social networks that form. Dame-Griff did not present results from NLP or use machine learning algorithms to deliver insights about factors that help explain certain behaviours but, in addition to the methodological contribution, the study illuminates the development of the term cisgender and supports the researcher's broader project of enabling easier access to historical messages discussing trans lives and concerns on Usenet. This work to improve access to archived data, to which my own project may also contribute, is important as demonstrated in Paloque-Bergès' (cf. 2015; 2017) use of Usenet archives. Paloque-Bergès discussed the access to archived Usenet messages made available via Web interfaces, such as the now-superseded Deja News and ongoing Google Groups service. She notes that researchers have difficulty finding and extracting the messages they wish to analyse and that the large standalone files (such as those that Dame-Griff and I use) have complexities that pose barriers to entry and use. Quite apart from her empirical findings, for example in relation to early French online services, Paloque-Bergès' research is helpful for its critique of the format of Usenet archives and the barriers these formats present to researchers. Critical awareness of archives' limitations is very important. For example, in Churchill's (2009) study of the origins of the militia movement, Google's Usenet archives are used to identify an increasing prominence of the Second Amendment in discussions. Churchill indicates that Google's archive includes all messages from 1981, which is not accurate since Usenet was not archived systematically by service providers at that time. This needn't undermine Churchill's overarching point but indicates the importance of knowing how Usenet was preserved in its earlier years and the nature of the resulting archives.

Milligan's (2013) and Schmidt's (2015) blog posts also make methodological contributions to analysing Usenet archives. Both explore extracting data from the UTZoo Usenet archive, which I discuss later (pp.102-104) since I use it in my research. Milligan applies NLP to infer topics of discussion in Canadian Usenet newsgroups; Schmidt creates visualisations such as the most popular newsgroups and takes a particularly critical look at the scope of the archive and its apparent gaps, while also noting that his methodology may require refinement. Nonetheless neither of these studies, nor any research discussed in this section, incorporate all components of the methodology I assembled for my research. This

is not a criticism; it reflects my above discussion that researchers can program computational methods to meet the specific needs of their research site and/or question. However it does support the idea that computational methods can be used to analyse the research sites I study to answer my research questions. My research is unique in its particular combination and application of research methods, and the computer code I write to apply them to large archives spanning decades. The remainder of the present chapter and chapter three discuss how I configure and apply methods used successfully in previous research while making my original contribution.

2.1.5 The strengths of using a programming language

Researchers may use tools, platforms and/or languages for computational analysis of large datasets. Tools may be defined as discrete pieces of software, such as code to filter data. Platforms enable the creation of tools within their software environments, such as macros within Microsoft Excel (among other bespoke platforms for advanced analytics). Computer programming languages allow highly flexible creation, combination, and deployment of new tools and platforms. I use the programming language R on Mac to combine, apply, and configure data analysis tools for the following tasks:

Extracting:

- Individual messages from the text files that comprise archives
- Message metadata including the date the message was sent, the sender, the OSNs to which it was sent, and the subject line
- Analytical descriptors that indicate topics of discussion and language used within the dataset

Organising:

- The data into data frames, in which each row refers to one message and each column includes one category of data

Swapping:

- Message senders' names with a mix of numbers and letters to pseudonymise participants

Identifying:

- The social processes of cohesion that I analyse, and facets potentially associated with incivility, using the criteria I have identified to operationalise them
- Patterns of participation in the OSNs I analyse

Mining:

- For associations between categories of data, such as whether certain subjects of discussion are associated with cohesion

Visualising:

- The prominence of topics of discussion and relations between them
- The social networks that emerge as participants interact by sending messages
- Patterns of participation

A strength of using a language is that save for the earliest steps of downloading messages and some initial formatting of those files (see pp.102-104), I can use it for all the above tasks and write code that undertakes tasks in sequence, rather than manually executing one after the another in different software packages. Ali *et al.*'s (2022) research indicates this by using R to organise data and then run a sentiment analysis (I discuss sentiment analysis below), though my research incorporates additional processes as summarised above. Code can also be customised and transferred, sometimes with small amendments, to undertake the same tasks with new data efficiently. R is well-established for large scale data analysis including machine learning. See, for example, Luke (2015, 4-6) who argues that R is effective for social network analysis as, compared with many other software packages, it is powerful for importing and transferring very large datasets and it has a broad range of programming functionality which has been contributed by R's users. Also see Lee (2019) who discusses R methodology for NLP, then applies it to identify co-occurring nouns in Korean news reporting regarding immigrant workers. Alternatively, I could have used the language Python as Schmidt (2015) and Dame-Griff (2019) used for analysing Usenet, though R is strong for all the types of analysis I undertake so provides an environment in which to program my entire workflow.

Nonetheless, using a language to undertake the whole analysis workflow holistically presents a lot to learn. I undertook two short courses via UCL, two online courses, read textbooks, journal articles, used online forums for data analysis professionals, and developed proofs of concept for conference presentations. Through this I learned the syntax of the R language, how R manages system memory, means of importing and exporting

data, methods for visualising analyses, how to use and configure others' algorithms and develop my own, principles for writing efficient code, and I developed my knowledge of object-oriented programming (in which data are held in objects and code is written to interact with them). Using a language such as R was necessary for undertaking the full breadth of work required to answer my research questions, but it also conveys the additional advantage that, through my learning, I came to understand how the algorithms I use and configure work at a mathematical level. They are not 'black boxes', in which the researcher is not privy to the workings of the system. This aids my selection, configuration, and evaluation of algorithms, ensuring my work is reliable. In sum, my new learning in computer programming, in combination with my previous postgraduate education in theology and religious studies, ensures that I have sufficient knowledge of both fields to undertake this kind of interdisciplinary research.

2.2 Internet archives and platformisation

I will now briefly discuss the field of internet archiving and how using archives, where appropriate to the research question, may avoid the problem of 'platformisation' which can impede academic research.

Future historians will likely make extensive use of archives of the internet to understand life in the early twenty-first century (Schroeder and Brügger 2017). Archives typically focus on the World Wide Web, rather than messages sent across services such as Usenet. While there is no stable definition, internet archives include 'born digital' (rather than *digitised*) artefacts such as webpages and are collected intentionally with the aims of building a collection focusing on a specialist subject, area or indeed collecting everything within defined parameters, which can then be accessed and analysed (Dougherty *et al.* 2010, 7). It is a developing field in terms of methodology, the community of researchers participating, and the small but growing uses of archives (Schroeder and Brügger 2017). The *Internet Histories* journal (launched in 2017) provides a forum for this work. Internet archives are important for understanding social and political events (Dougherty *et al.* 2010) and, in 2016, archived internet data were used in a USA trial regarding trademark infringement (Bychowski cited in Winters 2017, 176). While researchers have developed user-friendly interfaces for reading archived webpages, such as The Wayback Machine at the Internet Archive and the SHINE interface implemented at the British Library, accessing the archived Web is increasingly challenging with areas of the Web being inaccessible. This is due in part to commercial companies using proprietary protocols that keep their services beyond

archivists' reach. Or, when such data are made available, they may be curated in a way that protects certain commercial interests (Acker and Donovan 2019, 3-4). In response, researchers innovate to scrape data from public-facing interfaces to ensure critical research can still be done (Bainotti, Caliandro and Gandini 2020), or focus on more readily accessible data such as X/Twitter hashtags, which represent lower hanging fruit (Özkula, Reilly and Hayes 2022). OSN message archives, large and cumbersome as they are, are an under-utilised source.

During the 2010s, some attention was paid to the UTZoo archive of Usenet posts, compiled at the University of Toronto Department of Zoology between 1981 and 1993, containing over two million Usenet messages. Systematic Usenet archiving may have started in 1995 with the service Deja News, whose archives Google acquired in 2001 (Reid and Gray 2007). Deja News retained, organised, and made Usenet messages available via a user-friendly interface, deleting messages where participants did not want them archived and helping Usenet become a resource for accessing information on myriad topics (Lyman and Kahle 1998). Google grafted the UTZoo and Deja News archives into their Google Groups online social networking service. However, search functions provide researchers with limited access to read through search results online, rather than providing downloadable datasets for offline analysis. This limits the use of Google Groups to small scale research of Usenet. And while Deja News was large-scale, the decentralised (pp.28-29) nature of Usenet means that it's not practicable to archive the entire network, though the datasets that *are* archived can be substantial and sufficient for a range of research questions (Driscoll and Paloque-Bergès 2017, 55). Many substantial archives are available from the Internet Archive which, in addition to hosting the UTZoo archive, also holds the Usenet Historical Collection (UHC) and Giganews archives. The UHC was donated by an anonymous contributor and appears to have been drawn from Google Groups. Giganews is a Usenet service provider and have provided their own archive of Usenet messages to the Internet Archive. Both sources include large message archives from OSNs focusing on many topics including religion. In the UTZoo archive, each message is typically stored in an individual text file nested within a structure of folders which partly reflects the magnetic tapes from which the archived messages were later digitised. The UHC and Giganews archives use MBox formats, which result in very large files (such as those larger than a gigabyte) potentially containing hundreds of thousands or more messages presented sequentially as text and with very limited structure. These archives are valuable, but not at all user-friendly (Paloque-Bergès 2017); some computers or applications will simply fail to open them.

However, these archives make the investment in manipulating, extracting, and organising them worthwhile. The rewards in terms of insights into social lives are evidenced in chapters four to six but here I'll mention the additional benefit that such archives, so long as they suit one's research questions, can help avoid the problem of platformisation to which I allude above. This occurs when a social networking service provider uses the internet to convey users' messages but does so via their own proprietary applications and protocols, so the messages remain inaccessible to researchers except via certain routes authorised by the provider. While this can help private conversations to remain so, it can mean that even data shared via OSNs that participants know are public and are read by strangers, are nonetheless beyond the researcher's grasp for conducting ethically sensitive and valuable studies. Platform-authorized access can be partial (potentially in both senses of the word) when service providers curate the data that researchers access (Acker and Donovan 2019, 4) and acquisition may be prohibitively expensive, relegating academic researchers behind commercial data analytics companies. Such restrictions have become more widespread following the events surrounding Cambridge Analytica (Loutfi *et al.* 2019). Application programming interfaces (APIs) are available for accessing some platforms and can provide tools to researchers for extracting and analysing data (Mackenzie 2018). Archives don't afford such tools, but neither do they have the opacity that a platform can acquire by restricting access to a proprietary interface.

Usenet was popularised at a time when messages shared on public OSNs were more accessible to interested others, including researchers. Although current social media services have enhanced functionality such as video and animation, many of the ways people behave, and the connections and discourse that social network services afford, remain stable over time (Polonski 2017, 310). Historical archives therefore have the potential to provide insights into social behaviour online and present rich data that may not be available from other platforms due to platformisation. Therefore, I argue that researchers could make greater use of historical OSN archives, but the limited use so far reflects barriers to entry and the nascent nature of this research area.

2.3 Extracting and organising messages from the archives

Data from archives require extracting and organising due to archives' limited and heterogeneous structure, and to enable my subsequent analyses. Many large-scale data analysis projects require an 'extract, transform, load' (ETL) process to convert data into a

useable structure that contains the features pertinent to the research (cf. Wu *et al.* 2019). For this, researchers must choose which features of the data to extract with their particular research aims in mind. This process of feature engineering (Salganik 2018, 238) may transform the data into formats bespoke to the researcher's aims, including introducing ethical safeguards which I discuss later (pp.140-146). This results in archives of data derived from the primary archives and the structure that researchers impose when creating them results in a format suited to the researchers' bespoke analyses. This step must therefore be undertaken carefully, as omitting certain features during the ETL/feature engineering stages may enable or prohibit the resulting datasets from being used to answer certain questions (Dougherty *et al.* 2010, 17, 20). In doing this, researchers create the objects of their research which are one step removed from the primary sources, so researchers need to ensure those objects reflect the reality they are seeking to analyse (see sec 2.1.1). Archives, whether compiled by a digital service provider or derived by researchers, are 'reborn digital material' (Bebeil in Brügger and Milligan 2019) This term reflects how the digital material has undergone a transformative process, bringing to mind Bourdieu's (1992) observation that in sociology the problems raised in our research may be related to the objects that we, as academics, have created. For example, the datasets I prepare influence the social network analyses that are possible (Bainbridge 2020, 9-10) and which I describe in chapter three (pp.118-120), so I had to make design decisions that ensured the dataset format afforded the analyses I require. Specifically, I analyse trends in interactions within and between OSNs so I opted for my datasets to include one row per message and a column (or vector, in the language of data frames) containing the OSN(s) to which messages were posted, separated by a comma to demarcate different OSNs, along with including the date each message was posted to ensure social network analyses could reflect time periods I would specify. Messages are my units of analysis, so my datasets are structured to give messages primacy rather than, for example, individual participants, which helps afford these social network analysis insights. I now summarise the steps for extracting and organising my data, explaining the decisions I have made about what to include in my datasets in order that I can answer my research questions.

2.3.1 Downloading, extracting, and organising archived OSN messages

I downloaded all the archive files I used in compressed formats from the Internet Archive. I then reviewed the file formats that had been used within the archives and identified that UTZoo, my first research site, would require a different approach from the later Giganews archives.

As I mention above, the archives assembled at the University of Toronto Zoology Department (UTZoo) were stored on magnetic tapes. These were later digitised and made available to the research community via compressed files, each typically representing one of the 141 tapes that were digitised. I downloaded and decompressed the file for each tape. The files were saved in .tar format, a Unix format for saving multiple uncompressed files together. I was able to decompress these files on Windows and Mac computers. Each digitised tape contains multiple folders, initially numbered sequentially and later according to the specific newsgroups to which the archived messages were posted. At times those folders are nested within other folders reflecting the part of the Usenet hierarchy in which that group was set up, such as the net or soc hierarchies (see pp.32-33). These changes over time in how the archive folders and subfolders are organised reflects the evolution of Usenet in which newsgroups were setup within an increasingly complex hierarchy to accommodate Usenet's growth and the breadth of topics discussed. Subfolders generally contain many individual files, each containing one message posted by a participant of that newsgroup, though some files are 'digests' of multiple messages. As I note above, these are typically messages posted to ARPANET discussion groups with which Usenet connected (I note an example on p.29), but to which Usenet participants could not post. In later years, messages written in response to an earlier post may contain all or part of the earlier post. This reflects how Usenet software became more sophisticated and messages became 'threaded' as people conversed by commenting on others' messages. This functionality helps enable social interaction.

This structure of folders and subfolders makes it challenging to open and navigate the archive, though it's remarkable and commendable that the tapes were digitised at all. The file format of each message appeared unspecified, so I used an automated process to append each file name with '.txt', which prompts computers' operating systems to attempt to open the file within default text editing software, such as Notepad on Windows and TextEdit on macOS. This made opening files more efficient since I did not have to specify an app/program when opening each file.

The format within each message varies over time but they consistently contain some header information and the substance of the message, bearing similarity to the structure of an email. An early message header typically contains the poster's name and/or username, the path through the network that the message had travelled, a message date (though it is not always clear if this was the date sent or received), the newsgroup to which the message

was posted and a subject line. This information is not presented consistently across time due to variations in software and the need to establish conventions, so early messages often do not precede each piece of header information with a standard piece of text, such as 'Date:' or 'Subject:'. Date formats are also highly heterogeneous, for example an early message may use the format 'XXX MMM DDD HH:MM:SS YYYY' (where X is the day of the week) and a later message 'DD MMM YYYY HH:MM:SS GMT' and some messages indicate both a date and time the message was posted and when it was received. The header standardises in later years as software for posting messages to Usenet matured. Some messages end with a signature of varying lengths, containing custom information as with emails today.

2.3.2 Selecting messages from UTZoo

Since the UTZoo archive contains more than two million messages, I sought to extract the messages most likely relating to religion. I achieved this by developing a small ontology of words that are likely to occur in messages discussing religion and less likely to occur elsewhere, and then extracted only those messages in which words in the ontology appear. This is an imprecise process. If casting the net wide with a broad ontology, I would reduce the risk of omitting religion-related discussion but increase the risk of including messages unrelated to religion (see below regarding the term Christian). My approach to this task was iterative. I first extracted messages from the entire archive that use the term religion. While *prima facie* this is a blunt tool, early Usenet OSNs that discussed religion formed via the newsgroup net.religion and newsgroups within that hierarchy, such as net.religion.judaism. Once Usenet's hierarchy expanded, prominent groups for religion discussion were founded with the talk.religion and soc.religion hierarchies, such as talk.religion.christian and talk.religion.jewish. Extracting messages including the term religion includes all messages containing the word religion and those sent to newsgroups within the religion sub-hierarchies, since the search includes the 'header' metadata to which I refer above. I then focused on the subset of messages that were remaining, extracting messages including the terms Jewish, Judaism, Islam, Islamic and Muslim, reflecting my focuses on Abrahamic faiths. The term Jew proved too imprecise. Messages that I checked which contained the term Jew but not Jewish/Judaism were irrelevant, as were messages I checked that contain the term Bible but not religion. For example, messages referred to reference publications as a 'bible'. The term Christian yielded some messages irrelevant to my research as it's also a first name, so while I included messages that matched the term Christian, I subsequently removed those sent only to newsgroups wholly unrelated to religion, such as

those focused on computer programming. Any remaining off-topic messages are managed via my analytic descriptors which suggested messages that related to religion discussion.

I also decided to exclude messages that were posted to Usenet as part of 'digests' since, as noted above, I found these were typically messages compiled from ARPANET discussion groups rather than Usenet proper. I found they lack some crucial 'header' information so my automated methods (see TA 4 below) presented null returns in place of, for example, the sender or date. This made it easy to exclude these messages from my resulting dataset since I automatically excluded messages for which this information was missing. I then identified words that frequently co-occurred in messages focusing on Christianity, Islam and/or Judaism. For example, I found that the word Muslim co-occurred in around one in eight messages that include the word Moslem including messages which were 'threaded' to include earlier messages within the discussion, perhaps indicating the historical or regional usage of this spelling. I identified six such co-occurring words which were likely to be distinct to Abrahamic religion discussion — Qur'an, synagogue, Muslem, Moslem, theology, Yahweh — and created a subset dataset of messages which included *any* of these words and *none* of my initial keywords. However, on inspection this small subset of approximately 163 messages did not focus on religion discussion. For example some were lists of available newsgroups, which included the names of groups on myriad topics. I chose to exclude this small subset.

This fairly simple ontology helps to safeguard the general integrity of my dataset and is relatable to the approach taken in de Wildt and Auper's (2019) research which successfully used a relatively narrow ontology to identify discussion of religion in video game discussion forums. This nonetheless risks including messages that are irrelevant to my research but use the term religion in an unrelated context, for example 'football is my religion'. My analytic descriptors helped me identify such messages as they had few occurrences of other words used commonly when discussing religion and had co-occurrences of other words unrelated to religion. This enabled me to remove certain messages from my dataset. In addition, as I explain below, I affirmed the methodology and deepened my analyses by manually analysing 1,000 discussions. This allowed me to infer that I had included only a small number of false positives. These steps provided me with a subset of the overall UTZoo archive containing 40,146 messages posted to newsgroups between years 1981 and 1993. I compiled these messages into one large text file ready to be extracted and organised into a data frame within R.

I did not take this approach with the later archives from Giganews that comprise my second research site, as the Giganews archives I used specifically contain messages posted to specified newsgroups, so capture the OSNs that formed within that newsgroup and the connections that the newsgroup had to others where messages were posted to multiple groups simultaneously. These archives' MBox format includes each message sequentially in a text file but with more structured and consistent header data than in UTZoo. Nonetheless, such MBox files can be greater than a gigabyte in size and require computational processes to extract and organise their data. Opening them within basic text editing software can render them very hard to read as the software may not format the messages, presenting them as one exceptionally long string of text.

2.3.3 Choosing the data to extract from the large text archives

This resulted in multiple, often very large text files each containing messages representing many thousands of discussions sequentially. I had one for UTZoo, which combined all extracted messages from across the archive from 1981-1993, irrespective of the newsgroup to which the message was posted. I also had 48 archives of later messages, one or more for each newsgroup on which I focused; in sum, these later archives included discussions of religion from 2002-2015 and totalled approximately 792,393 messages. This neatly reflects my two research sites: the more fragmented but historically important early time period and the later fuller archives closer in lineage to current social media.

I identified that my datasets would require the following information:

- A short random mix of numbers and letters (an alphanumeric string) in place of the message sender's name, using the same string for each sender so their personal information is excluded (see ethical considerations discussed on pp.140-146)
- The newsgroup(s) to which the message was posted
- The message date
- The subject line
- Analytic descriptors that I identified in the message, indicating participants' topic(s) of discussion

I later appended the results of statistical analyses that underpin findings I present in chapters four and five (summarised in TA 4 below), which are useful for identifying some potential sites of incivility. With reference to my above discussion and my measures of

engagement (pp.19-20, 58-61), I now explain how my datasets containing this information enable me to answer my research questions and represent the underlying social reality captured in the original archives so far as is feasible, while not including personal information. Following this, I describe the process I used to extract these data.

Including a substitute set of alphanumeric characters in place of the sender's name for each message allows me to identify the density of social networks, which is to say the extent to which participants are replying to each other and thereby connecting socially, rather than groups merely containing sequences of messages to which nobody replies. This is possible because the unique alphanumeric string for each participant is the same throughout the dataset. I can also then identify patterns of participation, for example whether uncivil discourse tends to be in patterns of one-to-one interaction as Ziembowicz *et al.* (2021) found regarding conflict on Wikipedia, and the sorts of factors that are associated with first time participation. My analyses, and my outputs from them, will focus on broad patterns, not investigating any individual's participation, with the exception of using the identifier to potentially weed out any automated spambot participants that I identify.

Knowing the newsgroups to which messages were posted is essential context for understanding the sorts of discussions I might expect to see. For example, messages posted to one group that focused on a particular religious denomination may have a different character from messages crossposted simultaneously to that group and other groups that focus on, say, science or politics. I can then identify whether a discussion is in line with the aims of the group(s) to which it was posted. This information also enables me to visualise connections between OSNs by identifying where messages were posted between multiple groups and identify if any factors are associated with such connections, such as certain topics of discussion, increased or decreased participation, or incivility.

An indication of the message date is important for quasi-longitudinal analyses, which suggest how OSNs' discussions may evolve over time in terms of network size and topics of discussion. It also potentially allows associations to be identified between discussions and events occurring contemporaneously in the world. For example, one might expect increased discussion of the papacy around the time of Pope Benedict XVI's visit to England and Scotland in 2010. However, specifying the exact date of a message is difficult for two reasons. First, as noted above, date formatting in the UTZoo archive varied greatly so it's challenging to extract a message date; I explain my resolution for this below. Second, while messages may be sent on one date, they may be received in different locations on different

dates due to time differences and, more significantly, because Usenet's decentralised nature means it may have taken some time to convey messages across the network. This issue diminishes over time as messages posted to newsgroups were increasingly sent over internet connections.

The subject line of the message allows me to group together messages that form one discussion. This is not perfect since participants can change subject lines, though such a change may also indicate that the discussion has moved on and is distinct from previous messages. This, along with the other information extracted, affords insights into how many participants discussions had on average when filtered for other factors such as topic of discussion. The subject also contributes useful data when extracting analytic descriptors.

Those analytic descriptors indicate: subjects that are the focus of conversations between participants; language used in the messages prominently or distinctively. Without these descriptors, one could identify that a discussion between certain OSNs includes a large number of participants but have little context regarding the substance of the discussion. Consequently, the thematic summaries that these analytic descriptors provide add significant depth to the analyses that are possible when combined with the other data I extract, though deriving them is challenging (see pp.108-133, Technical Appendix 1.1).

2.3.4 The methodology used for extracting analytic descriptors from message content

When analysing smaller datasets it is feasible to examine the substance of discussions manually, as in Narbona's (2016, 98-99) analysis of the leadership expressed in Pope Francis' Twitter messages and Chua's (2009) research of Jewish Usenet OSNs. However, datasets of hundreds of thousands or millions of messages require, in practice, computational machine learning (ML) processes. As Sir Demis Hassabis, co-founder of the artificial intelligence company DeepMind has put it:

For me artificial intelligence and computers, in some way they're an extension of our mental capabilities. I view computers in the same way as a car allows us to move faster than we can run. (Hassabis on *Across the Board*, 2014)

Here, I discuss in some detail the algorithmic approaches for NLP that I have configured to derive analytic descriptors of message content. I spent time learning, selecting, configuring, and evaluating the methodological approach I have taken to this task. Machine learning

entails a series of computational processes, the outcomes of which are, to varying degrees, interpretable in light of the mathematics. These steps need to be explained and I see this done to differing extents in extant literature, which sometimes omits or presents oversimplified summaries. The explanations also inform my decisions regarding which algorithms to use and help the reader see the relationship between the data I use and my conclusions, building confidence in my work.

The internal workings of ML processes are comprehensible to varying degrees so this understanding of exactly how an algorithm arrives at its output becomes more difficult with less interpretable and more 'black boxed' approaches. As Kolkman's (2020) research illuminates, perhaps only those closest to the development and deployment of ML models in an organisation may be able to assess their credibility and, to them, the inner workings may remain black boxes, posing even greater challenges for others in the organisation who are non-experts. A danger here for academic researchers is employing an algorithm without sufficient understanding of its logic — does it operate according to assumptions and logic that actually deliver what the researcher intends (Nanni in Brügger and Milligan 2019, 7), and might the application of a poorly understood methodology appear robust simply because it is applied to 'big data' (Markham, Tiidenberg and Herman 2018)?

The ML processes I configure are comprehensible, allowing me to explain the detail and show a clear line from my research aims to my datasets, analyses, and findings. This enables me to take responsibility for the algorithms I use, which is an important ethical principle given the agency that ML algorithms can have in research (McQuillan 2018, 4) to process data, deliver insights, and apply classifications. They may even develop narratives through 'algorithmic emplotment' as Jacobsen (2020) identifies, in which algorithms render data meaningful in a way that may provide some certain inevitable narrative about a person or their language and images. The algorithms I configure can therefore be seen to provide a methodological substrate (Markham, Tiidenberg and Herman 2018, 5) as their outcomes influence the recommendations I make in chapter six regarding OSN participation and future research. Importantly, and helping to avoid any inaccurate narratives resulting from algorithmic emplotment, I do not have to make post-hoc interpretations of how the algorithms I configure have arrived at their insights. Although this can occur with less interpretable approaches such as deep neural networks (McQuillan 2018, 2-3), I explain plainly and in some detail how the processes work. This enables comprehension of *how* I arrive at my results and clear inference regarding *why* I obtain the results I do, which is a strength of using a programming language since writing, configuring, and reviewing my

code myself enables my depth of understanding. Showing my ‘working out’ in this way, without publishing the underlying dataset in alignment with my ethical approach, also allows me to test the results through comparing the process with the insights gained — do they reflect what we would expect to see? This explanation and sense checking process reflects the ways in which models can be tested by those who may utilise and are familiar with them but are not experts in the technical aspects (Kolkman 2020, 7-8), thereby providing further assurance to the reader.

In light of the strengths and limitations of methods for computational natural language processing (see technical appendix one for discussion of these), and those like them, I considered how to identify subjects/themes of discussion and other relevant linguistic expressions from messages in my datasets, doing so in a manner that I could then apply labels to messages in my dataset to meet my research aims. The approach I used delivers three sets of results: (1) it identifies words which are representative of each message; (2) it labels messages that include language likely to be associated with incivility; (3) it generates a word embedding that captures semantic relationships between words with messages. I now summarise and justify my approach to each in turn.

First, I wrote code to apply R functionality that assembles a term-document matrix, in which each message that occurs within the dataset is represented in a unique column, and each term which occurs within the dataset occupies a row. I applied two stages of pre-processing to vocabulary of terms included in the term-document matrix. First, I excluded a small subset of terms which are unlikely to indicate themes/subjects within a message, such as conjunctions ‘so’ and ‘or’. Second, I applied a stemming algorithm to reduce words to a root, as summarised in my discussion of LDA (see Technical Appendix 1.1). This means that, for example, religion and religious are considered together. Having assembled the matrix, I identified the six most frequent terms that occur within each message and saved that analysis.

I then applied ‘term frequency — inverse document frequency’ (TF-IDF) weighting to the term-document matrix. TF-IDF is a longstanding approach for identifying words that are characteristic of texts and is often used in combination with other algorithms such as LDA within R and in other languages. It may be considered a form of key word/phrase extraction, overlapping with NLP research into text summarisation, an NLP sub-field which seeks to extract or produce phrases/sentences which are representative of a text being analysed. (See Allahyari *et al.* (2017) and El-Kassas *et al.* (2021) for summaries of NLP approaches

to text summarisation.) It works by applying weights to words within a corpus in accordance with how frequently they appear in a document multiplied by the inverse of how frequently they appear across a corpus of documents; the term frequency multiplied by the inverse document frequency (Silge and Robinson 2017, 31-32). Consequently, words that occur infrequently and may be limited to a small number of documents in the corpus will receive higher weights than words that occur frequently across the corpus. The infrequently occurring words are likely to be characteristic of the documents in which they *do* appear so the highest weighted words within a text may summarise important features. In religion research, Svensson (2020) used TF-IDF to identify words that are characteristics of certain Islamic authorities' writings, finding that such terms helped identify online articles referring to that authority, and Kołodziejka and Neumaier (2017) used TF-IDF to identify words that characterised the discussions they were analysing from online German and Polish groups, then selecting a subset that related to community.

TF-IDF does not require an initial random distribution of words to topics, so provides a clean line of sight between the original corpus and the results of the algorithm. It may also be used without removing stop words, since those words are likely to occur frequently and receive a low weighting, not appearing when the researcher identifies the terms most associated with each document. I required both analyses — term-frequency *and* TF-IDF — since the strength of TF-IDF is in identifying terms that are distinctive to a single message or small group of them. Messages may use commonly occurring words that are low-weighted but highly indicative of a subject of discussion, so the term-frequency count from my initial unweighted term-document matrix was also required. Having run both analyses, I appended the top six words that occur in either analysis to each row (i.e. each message) of the dataset as short analytic descriptors, complementing the message subject lines which I retain in my datasets. This affords the possibility of insights about subjects/themes that cross across multiple years of analysis, OSNs and participants.

In contrast with LDA, this approach removes the needs to run the clustering analyses multiple times with different initial random distributions and it removes the need to specify a number of 'topics'/clusters prior to analysis. In addition, this approach can identify relevant terms in short texts since, so long as the message has *some* words that indicate its themes/subjects, they will be weighted accordingly in the term-frequency or TF-IDF analyses. These strengths address the challenges specified in the bullet points above for the purposes of my research.

In addition to identifying message subjects/themes via the above methodology, I also tested using the Syuzhet library in R to identify messages containing language that indicates incivility. Syuzhet parses each sentence in each message and, with reference to an underlying coded dictionary, identifies whether each sentence indicates any one or more of eight qualities including ‘anger’, for example picking up on the use of profane language. I wrote code that deploys this functionality and then evaluates the number of sentences indicating anger in each message, *excluding* ‘threaded’ messages where a participant has replied to a previous message and included that message or quoted from it in their reply. This is important, since the message being analysed may be a short, civil response to an angry uncivil message, and I want the label to reflect that. Having tested various parameters, on a small sample dataset I found that where more than 85% of *negative* sentiment sentences within a message indicated anger *and* the message includes more negative than positive sentences (or at least one sentence indicated anger within messages of four sentences or fewer), these parameters provide a broad high-level indicator of incivility within the message. However, when applied at scale by writing code to label such messages with an analytic descriptor of incivility across the entire Message Dataframe, given the diverse topics, behavioural norms and the heterogeneity of the datasets (such as varying message formats), I found that such measures were too imprecise. This analysis failed to provide an indicator from which high level trends regarding incivility in the dataset can emerge. Instead, as noted following my review of literature regarding religion on Usenet, and my indicators of incivility, I found that discussions that were higher participation, crossposted to multiple OSNs, and involved first time participants provide a heuristic for identifying some potential sites of incivility, which I then consider in my manual review to identify times when incivility was present (see pp.121-122).

Third, to afford deeper analyses of the terms associated with different themes/subjects within the messages, I used the GloVe algorithm to create a word embedding which, as in my experimentation using BERT, embeds words within the corpus in a high-dimensional space so that those which, it is inferred, are semantically similar are in relative proximity to each other. Key to this approach is learning the contexts in which each word appears within the messages. To do so, the algorithm considers each word alongside a set number of preceding and subsequent words — as if applying a sliding window over a text to capture the words either side of the word in the centre being analysed at that moment. GloVe thereby accounts for the proximity of words within individual messages. The algorithm then uses a logistic regression technique to plot the words in the matrix into the high dimensional space, ensuring that words which share proximity within messages are relatively spatially

close. This resulting word embedding is typically (and from my experience) faster and less computationally expensive to train than using transformers-based models such as BERT, with which I had experimented. It allowed me to identify terms associated with prominent themes/subjects within messages. Specifically, I identified language associated with religion discussion, starting with the terms I used to identify religion focused messages in the UTZoo dataset (pp.104-105) and identifying words associated with those as captured in the GloVe algorithm. The first five of these are, in alphabetical order, Allah, Aquinas, Aramaic, baptism, and biblical. I then wrote R code to identify whether any of these words occur in the keywords extracted via TF-IDF for each message and labelled the message to indicate this.

Having critically evaluated these approaches through applying them to my messages, I combined TF-IDF and my word embedding (the first and third methodological tools discussed above) to form my analytic descriptors of messages. Recall that my dataset in R, which I now term my Message Dataframe, includes the analytic descriptors in place of full text messages using one row to recall details for each message. These include descriptors attributed through my above term-frequency and TF-IDF analyses and an indicator of whether religion focused keywords featured in the message distinctively or prominently. As with my other analyses, I triangulated my findings when analysing 1,000 discussion threads manually. Together, these steps provided basic high level thematic summaries of messages.

2.4 Conclusion – summary of processes undertaken in assembling my datasets

In summary, I applied this process to form my dataset of messages from select archives, which I term my Message Dataframe:

Downloading and extracting archives held online:

- Download and decompress selected archives
- Extract religion-related messages from 1981-1993 archive
- Compile those messages into one sub-archive

Importing archived messages into R:

- Import those archives into R, splitting them into individual messages

- For each message, replace the sender details with an alphanumeric string
- Then extract the...
 - OSNs to which the message was sent
 - Message date
 - Subject line

Extracting analytic descriptors of subjects and language in the datasets' messages:

- Pre-processing to create a corpus
 - Tokenise the documents into a corpus of individual words
 - Remove from the corpus
 - Punctuation
 - Numbers
 - Email and web addresses
 - Convert upper case to lower case
- Identify the words which are representative of each message within the dataset
 - Create a term-document matrix, which captures the number of occurrences of each term within each message and the co-occurrence of all terms over the whole dataset
 - Create a term frequency — inverse document frequency (TF-IDF) version of the term-document matrix to identify terms which are representative of a message or group of messages, but which occur less frequently across the dataset as a whole.
 - Apply the top six terms from both of these analyses as analytic descriptors to each message within the dataset.
- Identify message characteristics which may act as a heuristic for messages that exhibit incivility:
 - Discussion threads that may account for a proportion of the incivility that occurs include those which are
 - higher participation
 - crossposted, and
 - involve first-time participants
- Identify the presence of religion focused keywords in each message
 - Use the GloVe algorithm to identify words associated with religion in the dataset

- Apply a descriptor to each message stating whether or not any of those keywords occurred distinctively or prominently in the message. i.e. whether any of those words were extracted via TF-IDF (see above)

Drawing in threads from the beginning of this chapter, the above process delivers a dataset that contains everything I need to discern my measures of engagement, namely cohesion, incivility, and patterns of participation, and thereby answer my research questions (pp.86-88), using datasets that cross four decades but which aren't subject to the privatisation of social media data due to increased platformisation. The methodology described in the present chapter results in datasets of what is *explicit* in the archived messages (notwithstanding some subjectivities around identifying social processes), seeking to represent faithfully the social interactions captured within the archives by extracting, transforming, and loading data from the messages and using NLP to apply analytic descriptors of their contents. The dataset embeds ethical principles by including a string of characters as an identifier for message senders while excluding both full messages (notwithstanding that I retain a smaller set of redacted messages for triangulation, as discussed in chapter three) and personal information such as their names and email addresses where these appear. Chapter three specifies the methodology for what is *implicit* in the messages. Specifically, identifying the measures of engagement themselves and associations between these measures of engagement and other factors such as subjects of discussion and the size of the social networks that form as people connect via OSNs.

Chapter Three — Networks and Social Processes

In this chapter, I focus on the methodology for elucidating insights that are *implicit* in datasets of online social network (OSN) messages. First, I present an approach to social network analysis (SNA) that illuminates the OSNs that form when people connect online to discuss religion. I then discuss how network analysis and other computational methods including machine learning, supplemented by manual review, may be used to identify my measures of engagement, namely cohesion, incivility, and patterns of participation. In turn, these measures inform research into religious community and authority online (see pp.68-85, pp.247-251). I also evaluate how association rules mining, a form of machine learning, can be used to discern associations between facets of the social networks, subjects of discussion, cohesion, patterns of participation and, in turn, identify factors potentially suggestive of incivility, that I have elucidated and analysed. These approaches are all suited to the datasets that I formed, affirming the methodology for extracting and organising messages from the raw archives and applying analytic descriptors to them. I then summarise how I identified a subset of 1,000 discussions from my datasets of archived messages to deepen and triangulate my computational analyses. Following this, I explain how I embed ethical principles in my research by incorporating privacy by design, thereby mitigating ethical concerns while unlocking insights which, once fully developed, may be of interest. Finally, I argue for the strengths of the methodology considering challenges that may be put to it. This concludes the methodological section of my dissertation and provides the foundation for the analyses of my research sites and specific case studies, which are presented in chapters four to six.

3.1 Religions are networks

Networks – groups of entities and the connections between them – are central to religions (see Everton 2018). Among other forms of religious network, the entities may be buildings, organisations or people, and the connections may be physical proximity, governance, or social relationships. For example, church hierarchies of bishop, priest, and deacon enable a networked hierarchy of leadership. The Alpha course, connecting people to discuss the Christian faith, begins with a shared meal which creates social connections between participants who may be meeting for the first time (Heard 2008, 129-132). And scholars of religion recognise that religious practitioners' commitment and behaviour may reflect the type of in-person religious services that they attend with others, such as whether a group demands high commitment and regulates behaviour or tolerates less exacting behaviour

and lesser contributions to the religious group's activities (Stark and Finke 2000, 146-150). This indicates the power of social networks to influence religious practice, so studying online social networks helps us understand how they influence aspects of online interactions, such as my measures of engagement.

The graphical visualisation of networks emerged in the early twentieth century, as in Moreno and Jennings' (1938) statistical technique for studying relationships between groups of people. Since networks afford access to healthy and harmful relationships, ideas and other resources, and individuals may adopt the beliefs and practices of a religious group because a person in their social network is already a member (O'Leary and Brasher 1996, 3, 6), visualising social networks can aid understanding of why an individual or group may display certain behaviours. In a recent example of (offline) network analysis that relates to religion, Fincham and Burton (2020) identify Quaker economic networks that were dense insofar as members were connected with many other members while remaining open to new people. In this case, being part of the social network conferred commercial advantages, so any study of the participants' business activities is enriched by understanding their location within a network motivated by religion.

The structure of relations within a religious social network can impact the behavioural norms that emerge, such as if stable relations emerge between certain religious and non-religious groups. Lövheim observed this in her (2004) study of a Swedish OSN in which the behavioural norms that influenced how religion could be discussed were formed through give and take between the participants' previous experiences and knowledge, and the dominant ways in which religion was already being discussed within the group. The authority of certain members as more experienced participants conferred to them more influence over religion within the OSN. The influence of participants' authority within the network could be made apparent through SNA which illuminates reciprocity and transitivity. Reciprocity, for my research, is the extent to which others respond to a participant's message and transitivity is the property that two network participants who share a mutual connection with a third person will, themselves, become connected (so forming a triad of relations). These properties indicate influential members who develop dense social networks. As Sun and Shen (2021, 3) note regarding authority online, 'those who have more ties and are more connected in the network have better access to information and resources, which may help them solidify influence in the community'.

Intuitively, we may postulate that the structure of social networks can impact the prominence of certain themes or subjects of discussion within an OSN, based on the religious beliefs and practices (or absence thereof) of certain well-connected participants, or the reproduction of dominant ways of behaving by a densely connected cluster of long-standing participants. This suggests the relevance of combining SNA with machine learning and natural language processing (ML and NLP) algorithms (see technical appendix 1.1) supplemented by manual review to infer subjects of discussion and linguistic expressions indicating cohesion, and message characteristics potentially suggestive of incivility, within my datasets of messages. The relevance of this combination — SNA, ML and NLP — is affirmed by extant literature. Kong's (2001, 406-407) perspectives on how ideas come to dominate within groups through struggle between groups and ideas indicate the relevance of SNA for contributing to understanding the location of, and interaction between, ideas online, and how this can impact group formation and boundaries since SNA along with machine learning analysis of messages can show the 'location' of ideas within a visualisation of a social network. Subsequently, SNA can indicate the reach of certain topics within and between discussion groups/OSNs, as demonstrated by Boy, Uitermark and Wiersma's (2018) analysis of discussions of hijab fashion on Instagram. The limits in reach of topics or participants may indicate the boundaries of a cohesive group, as in Elwert, Tabti and Pfahler's (2020) investigation of inter-religious discussions in conservative OSNs in which outsider participants' contributions to religious groups' discussions reflected the boundaries of who was considered a member of the religious in-group. Relatedly, but looking within an OSN rather than in relations between religious and non-religious groups, certain themes or subjects of discussion may be associated with strong or weak ties between participants, since strong and weak ties provide different types of support (Granovetter 1973). And while it may have previously been considered that relationships mediated principally or entirely online would result in weaker ties than 'offline' relationships, OSNs may afford the development of strong ties which provide valued social relations and support online (Krämer, Sauer and Ellison 2021). This positions my use of SNA within the theoretical framework of graph theory (Dervan 2015, 30-32), as I use sociograms — visualisations of social networks — to illustrate connections between OSNs and then consider how the attributes of these networks may be associated with cohesion, incivility, and subjects of discussion. By combining SNA with ML NLP, I analyse the propagation of discussion across social networks (Al-Garadi *et al.* 2018, 19).

3.1.1 Methodological approach to social network analysis

The social connections that I analyse are formed when people post messages to one or more OSNs. These messages form discussions, through which social connections are developed. I analyse interactions across my datasets to derive overall patterns of participation in OSN discussions, which are one of my measures of engagement. These interactions cause connections to form between OSNs when people crosspost messages to multiple OSNs (which may be focused on different subjects) and I visualise these interactions. My SNA uses adjacency matrices, in which each OSN appears on one row and in one column of the matrix, and the intersecting cells of the matrix state the number of times those two OSNs occur within the same discussion, indicating social connections formed between them as participants posted messages to groups across a defined period of time.

I calculated the adjacency matrices by using the message subject, date, and newsgroup data that I extracted from message archives and loaded into my datasets (see Technical Appendix 4). I then assembled the matrices in R and selected an algorithm to turn the matrices into graphs that visualise the social connections. These graphs are secondary to the data contained within the matrices, which are themselves derived from my core datasets, though the visualisations illustrate connections in a manner more comprehensible than my datasets or adjacency matrices. Therefore, my SNA visualisations play a supporting role in my analysis. I use them in chapters four and five to illustrate the structure of social networks that occurred at a macro level within my research sites, and at more micro level when undertaking my case studies, showing how different structures of social network were associated with interactions between certain OSNs, indicators of incivility or subjects of discussion, supporting how I evidence my measures of engagement and what appears to influence them. While I present static visualisations, they nonetheless illustrate the dynamic nature of OSNs that I analyse since the visualisations reflect change over time (Kolaczyk and Csárdi 2014, 179) due to patterns of participation and how participants interacted. SNA matters since a person's position within an OSN may provide them with authority to do and say certain things (Golder and Donath 2004, 1), in line with the behavioural norms established within a group (as discussed in chapter one). Further, if a person is central to the group as they are particularly well-connected and potentially act as a bridge between less-connected participants, that person may have influence by virtue of the reach of their discourse (Downing and Dron 2019). Consequently, it is valuable to consider the OSNs within which a discussion takes place to identify its potential impacts on participants.

This makes the choice of algorithm important as a variety of mathematical approaches may be used and these may produce significantly different visualisations. I consulted literature on algorithms commonly used and, considering that my measures of engagement include cohesion and incivility, and my work relates to the study of communities online, I chose the Fruchterman-Reingold algorithm. This algorithm is force-directed. It simulates forces that push and pull nodes away/toward each other iteratively, based on the frequency of their adjacency in the matrix that is being visualised; it can maximise modularity (clustering) in the visualisation which, while imperfect so potentially missing some clusters, for my purposes can help visualise distinct clusters of OSNs that form through the sending of messages between them (Miller 2015, 78-80; Porter, Anneal and Mucha 2009; Weber in Brügger and Milligan 2019, 8-9). This can help illustrate a facet of cohesion within religion focused OSNs, namely the density of social relations among discrete groups of people, which can indicate community (cf. pp.77-78 and Porter, Onnela and Mucha 2009; Lancichinetti *et al.* 2011). Indeed applications of SNA to Usenet have previously identified different types of OSN, such as 'question and answer' groups and social support groups, which have different network structures in terms of how dense social connections are between participants, with the former having fewer social connections between participants than the latter (Fisher, Smith and Welser 2006).

Using SNA to study relations *between* OSNs is also illuminating, as inter-group connections may be important for understanding the reach of religion discussion and the impacts of connecting religion OSNs with those focused on other subjects. As Bainbridge (2020, 34-35) notes regarding Facebook, studying 'interlocking' groups can elucidate the structure of an online subculture, and Olson and Neal (2015) illustrate this by identifying network connections between different subreddits (discussion groups) on Reddit by virtue of them sharing a statistically significant number of participants. Previous research has shown that Usenet OSNs are not silos in isolation (Smith 2003, 72-75) so people outside a religious OSN may encounter posts from the religious group and vice versa, where they are crossposted, and building close relational ties based on religion discussion may be associated with a person then remaining within the religion OSN over time. The possibility of this is indicated in Schafer's (2018) research which, while focusing on offline social networks, found religion discussion was a strong predictor of maintaining a social tie over time.

3.2 Computational/machine learning processes for identifying cohesion and incivility

In the preceding section, I mention how SNA can indicate cohesion and community by illustrating the density of social connections between a network of people. Here, I elaborate on this and consider how other computational methods, with a particular focus on NLP and ML algorithms, supplemented by manual analyses (see technical appendix 1.1 and pp.129-136), can help identify cohesion and incivility within large datasets of OSN messages. This contributes significantly to enabling me to identify my measures of engagement (pp.19-20, 58-61) and answer my research questions (pp.86-88). Here, I operationalise cohesion and incivility in line with my definitions in my measures of engagement. While researchers' and participants' perceptions of whether an OSN is cohesive or uncivil may vary, this enables me to measure their presence with some objectivity. And although I discuss incivility and cohesion separately here, the two concepts are not mutually exclusive and facets of both may be present at different, or even overlapping time periods. This is partly due to the interlocking nature of social networks including OSNs, with some people within sub-networks exhibiting different social processes from the OSN as a whole, showing how cohesion can depend on the boundaries of one's analysis. It also reflects how disagreement can tread a fine line between sparking conflict on the one hand, and cohesion-increasing discussion or resolution of differences on the other (Simmel cited in O'Leary and Brasher 1996, 15-16; Coleman cited in Hocquet and Wieber 2018, 44-45), as well as an opportunity to practise faith-based accommodation of diverging opinions by 'disagreeing in love' (Hipps 2009, 127-129). Nonetheless, incivility and cohesion have distinct properties and can be identified and analysed discretely.

3.2.1 Computational analyses to identify incivility

Incivility online is widespread and a growing concern (Trifiro *et al.* 2021), so identifying incivility and factors associated with it could improve social lives. To identify incivility, I use supervised computational approaches, supplemented by manual analysis. The supervised element of my method is comprised of certain classifiers I determined and applied to identify characteristics of discussion threads, specifically those messages which are (1) high participation, (2) crossposted between (at times disparate) OSNs, and (3) involved first time participants. Higher participation discussions can indicate spikes in discussion due to heated interactions, 'flames', or the flooding of an OSN with messages (as in alt.religion.scientology, pp.55-57). Crossposting can bring incivility where groups holding highly disparate views connect (as on soc.culture.jewish, pp.51-52). First time participation can flow from such crossposting and such participants may not be socialised into group behaviour or, again, may hold highly differing views on contentious subjects from an already

established core of longer-standing participants (noting Eternal September, see pp.42-43). Therefore, these characteristics can act as a heuristic for identifying some of the discussions within my dataset in which incivility *may potentially* occur. I then review the 1,000 Thread Dataset to identify where it does occur in such discussions, and ways in which it manifests. To be clear, these characteristics are binary classifiers applied to each message (Bramer 2007, 79-80) — they are deemed present or absent.

As specified in my measures of engagement (pp.19-20, 58-61), language is a factor indicating incivility and can contribute to incivility in multiple ways. My analytic descriptors also indicate the terms which are prominent and distinctive within messages and some are indicative of uncivil language, as are some message subject lines. As discussed above (p.112), I tested the use of sentiment analysis via the Syuzhet package in R, seeking to use this to indicate uncivil language. In this form of sentiment and emotion analysis, comparing the corpus of a document to a dictionary of terms in which each is labelled as expressing a negative, neutral, or positive polarity of sentiment, and/or emotions such as happy, angry, or anxious. This can be undertaken across languages and lexicons, so long as a suitably labelled dictionary base has been established (cf. Refaee's 2016 PhD dissertation on sentiment and emotion analysis in Arabic). It can be useful for identifying broad contours of emotional expression across a set of documents. Indeed sentiment analysis has been used in the study of civility in online religion discussion, such as in Kimmons *et al.*'s (2017) research which uses it alongside specify keywords to identify civility in Latter-day Saint Twitter accounts. However, I found it beset by challenges using my data. First, it is challenging to apply sentiments or emotions to individual words as this may not handle negation — 'I'd hardly say I love this artwork' — or sarcasm — 'yeah, the best party ever', though researchers continue to innovate in this subfield of language analysis; see Can and Alatas' (2019) review for relevant examples. I also found great imprecision in the accuracy when applied at scale, having (1) tested the method with some success on a data sample, and (2) applied parameters (p.112) to reduce the likelihood that the sentiment and emotion analyses would erroneously flag a message as uncivil when words were negated, by only flagging as message as uncivil via this route where 85% or more of the messages' sentences indicate anger and there is a greater proportion of negative sentiment sentences in the message compared with positive sentiment, or at least one sentence indicates anger in messages of four sentences or fewer. However, this was insufficient in light of the second challenge, namely the heterogeneity of the data within the archives in terms of its structure and metadata, and behavioural norms within individual OSNs, meaning that incivility manifests differently even if potential factors associated with its presence in some

circumstances (namely spikes in participation, crossposting, and first-time participation), may be consistent. Because of this, it was not possible to apply sentiment and emotion analysis to derive the findings I required.

The indicators of incivility that I did extract, summarised above, can be triangulated through (1) manual review of messages including subject lines (see pp.129-136) (and may be suggested by the language prominent or distinctive in messages identified via TF-IDF) and (2) by analysing patterns of interactions between OSN participants. This is important given the challenges of imprecision of identifying incivility computationally, indicated in this dissertation and in Stockinger, Schäfer and Lecheler's (2023) research with content moderators. Dame-Griff's (2019, 18) analysis of transgender discussion on Usenet found that certain flashpoint posts changed the tone of discussion significantly. This is relatable to the term 'flame war' in which the number of messages conveyed in online discussions spikes momentarily due to heated discussion and can therefore be measured quantitatively (Hocquet and Wieber 2018, 47).

In summary, and to add technical specificity to the facets of incivility that I specify in chapter one, I identify a range of manifestations of incivility in my dataset by first using the following labels I apply computationally, namely (1) high participation discussion threads which may represent spikes, or flashpoints, in the number of messages, (2) the message being posted to multiple OSNs simultaneously, and (3) messages which are from first-time participants. Taken together these may represent spikes or flashpoints in discussions, between OSNs that focus on disparate subjects or have different behavioural norms, and involve people potentially not socialised into the behavioural norms of the group to which they are posting. Incivility may occur in other ways, but identifying such discussion threads acts as a heuristic for interactions that may have more potential for incivility.

I then triangulate this algorithmic approach via my manual review of a subset of messages that have these characteristics to identify whether the presence of these characteristics is associated with incivility and to identify the ways in which incivility manifests. I discuss the findings of my manual reviews in chapters four and five alongside computational findings. This also then enables me to discuss the impacts of incivility on community and authority in online religion discussion in chapter six.

3.2.2 Computational analyses to identify cohesion

As I summarise above, if social network participants are well-connected with each other, this may indicate the presence of cohesion, in line with research summarised by Tulin, Pollet and Lehmann-Willenbrock (2018, 162) and reflected in Wang, Yang and Thorson's (2021) research into climate discussions on social media which found a tight knit community evidenced by the structure of their social network. In addition to participants being well-connected by sharing messages, we can also consider how they continue to participate over time. Persistence of participation in a dense network may indicate cohesion, since ongoing participation in social media can cultivate bonding social capital (Williams 2019, 10-11). Wertheim (1999, 302) uses the term 'network of responsibility', in which the networks that participants create require their ongoing participation to sustain them. Without ongoing participation, the density of social relations on the OSN may reduce as people don't know each other so don't interact and the behavioural norms that were formed at the start are not reproduced.

There is, however, likely to be an upper bound to the number of participants that a network can sustain in which those participants are densely socially connected to one another and are focused on the aims of the OSN, such as the discussion or practice of religion. In larger networks it may be more difficult to maintain relations and 'know' people. Tulin, Pollet and Lehmann-Willenbrock's (2018, 162-163) research found that larger groups on Facebook tend to have lower density due to the number of social connections that would be required for high density (expressed as the average number of other members to which each member is connected, as a proportion of the total number of members). Perhaps consequently, large groups tend to lose participants as well as gain them and active participation becomes concentrated among a subset of those connected to the group, as Panek *et al.* (2018, 3) found on the OSN platform Reddit. Indeed this has been identified historically on Usenet, with Whittaker *et al.*'s (2003, 83-88) research of 500 Usenet OSNs finding that 27% of participants posted only one message and many discussions involved a small number of participants who formed cliques. This concentration of participation among a small, particularly active subset of OSN members is not necessarily negative, as Davidson *et al.* (2019) observe, since it allows for different roles within OSN communities from those listening to those talking and those leading, but such roles are not mutually exclusive — those who talk can also listen and consequently reduce the volume of traffic across the OSN. Of course the substance of discussion may affect whether such an equilibrium may be achieved, since the flame wars discussed above may lead people to participate where an issue leads them to feel they must contribute and be heard (Johnen, Jungblut and Ziegele 2018, 3148-3150). And while they may not contribute actively to

cohesion, more marginal members may remain connected to the OSN by reading others' messages and posting occasionally, as this connection retains their access to people and discourse that they otherwise would not have (Lőrincz *et al.* 2019).

Considering further the idea that dense social network connections may indicate cohesion, the proportion of posts that receive replies may also be associated with high or low cohesion, reflecting the related SNA principle of reciprocity. Since groups have different purposes with some focusing more on community and others on simply sharing information (as noted above), where messages to the OSN receive replies and discussion ensues the higher engagement may, depending on its subject and civility, foster cohesion, whereas messages without replies may indicate the marginalised or rejected status of the person posting the message within the OSN (cf. Pomson 2008; Stadtfeld, Takács and Vörös 2019).

Connections *between* social networks may be a factor in supporting or undermining cohesion in the longer term and have complicated effects. On the one hand, inter-OSN connections may undermine cohesion within any one OSN, where the connection of diverse groups leads to uncivil discourse as discussed above. However, if social networks experience repulsive as well as attractive forces, insofar as attraction to others within the network leads people to continue engaging within their OSN and repulsive forces dissuade them from interacting with other groups due to a lack of affinity or the risk of incivility, an OSN may become densely connected, bounded and cohesive internally with few external connections to other groups. This role of repulsive forces on social network formation is modelled in Stadtfeld, Takács and Vörös' (2019) research. However, in such a situation cohesion could be associated with echo chambers that foreground and normalise radical views among a homogenous group (Recuero, Zago and Soares 2019). Consequently, cohesion in a bounded group may not always be desirable, even though it may reflect facets of community and stable sources of authority. I evaluate this in the case studies I analyse in chapters five and six, since my methodological approach enables analysis of cohesion alongside the subjects and themes present in the participants' discourse.

These elements of OSN participation that contribute to indicating cohesion – the proportion of messages receiving replies, the proportion of new participants, the degree to which discussions are crossposted or 'contained' within one OSN, overall numbers of participants and the duration of their participation (suggesting persistence and 'churn' among membership) are visible through the methodology I used for this research.

The ways in which I identify cohesion and incivility computationally relate directly and accurately to the facets of cohesion and incivility that I specify in my measures of engagement (pp.19-20, 58-61), which in turn arose from my critical engagement with extant literature regarding digital religion and Usenet OSNs. As I establish in chapter one, these social processes reflect aspects of community and authority, which have themselves been key focuses in digital religion studies. These datasets, derived largely from my computer programming methodology and including the insights regarding these social processes, equip me to answer my first research question: to what extents do cohesion and incivility manifest in a substantial sample of prominent religion-focused historical Usenet OSNs? The next section of this chapter explains the methodology by which I answer my second research question: which factors are associated with cohesion and incivility?

3.3 Machine learning to identify associations between social processes that occur on the OSN

To recap, my Message Dataframe provides the following information about messages sent over the OSNs I am researching, as captured in my selected archives:

- A random mix of numbers and letters representing the sender
- Message dates
- Subject lines
- Newsgroups (OSNs) to which messages was sent
- The size of the OSNs involved in discussions in the dataset and their network structure, formed by participants engaging with one another by posting messages
- Analytic descriptors of the language used in messages

I later appended to this the results of certain statistical analyses, indicating whether messages were:

- The first posted by a participant to the OSNs I am analysing
- From a longer-standing participant (two months or more)
- Part of a longer than average discussion thread

(See empirical analyses in chapter five, particularly tables 7 and 10.) These data, and my theoretical concepts that underpin them — online social network, cohesion, and incivility — represent the ontology for my research, that is to say what I have conceptualised and what

I model in relation to my research subject (Gahar *et al.* 2018, 226). In chapter two and the present chapter I explain how the components of my datasets including my word embedding (which represents the language used in the corpus of texts) capture the underlying realities of the social networks, discussions, cohesion, and incivility represented in the archives. To answer my second research question, I must apply machine learning to my datasets to identify factors that are associated with cohesion and incivility across the breadth of social interactions that I analyse, which enables me to present my findings in chapters five and six. In turn, this will indicate some factors that influence community and authority in OSNs, which I explore in chapter six. For this task, I use association rules mining (ARM) (Hahslers *et al.* 2023), which is an aspect of association analysis. I draw in part on Bramer (2007) and Lesmeister (2019) in the following summary.

ARM is a computationally intensive form of machine learning that searches for associations between categorical data within datasets. It is commonly applied to analyse customer behaviour. For example, an online film streaming service may analyse a dataset that specifies the films that subscribers watched along with subscribers' demographic data, to identify sets of films commonly watched by the same subscribers and those popular with certain demographics. In its simplest form, ARM identifies rules that state 'if this then that' with a degree of probability, based on the dataset analysed. Other applications include healthcare (Lesmeister 2019, 244) and decision-making in manufacturing contexts (Lin *et al.* 2019). ARM has been used to analyse social media profiles, with Si *et al.*'s (2019) research using ARM to identify associations between activities that LinkedIn members stated as their interests. These insights, such as, from a healthcare context, what proportion of patients exhibit certain symptoms having had a particular treatment, are termed association rules. Rules are accurate to varying degrees. The level of a rule's 'interestingness' (as termed in the literature) can be specified by three measurements — support, confidence, and lift — which I will summarise in Technical Appendix 1.2 (TA 1.2) and indicate how they apply to my datasets.

Together, these ARM insights indicate social phenomena that occur together frequently, indeed more frequently than is statistically likely if their co-occurrence is not related, and which I can then analyse to understand some social impacts of OSN participation. While manual, micro analysis of a research site may indicate that certain social phenomena such as incivility occur when certain other conditions being present, the likelihood of such social phenomena occurring together can only be evidenced using a method such as ARM, which applies machine learning at a larger scale than could feasibly be analysed manually. In

addition, ARM may illuminate other associations that occur less frequently (low support) but are strongly associated (high confidence) and so provide valuable insights that may be missed through manual analysis. ARM suggests leads to follow up, which I then investigate through manual textual analysis, enabling me to answer my second research question.

I applied an ARM algorithm to mine for interesting rules across all messages in my Message Dataframe from 1994 onwards. This excludes the UTZoo dataset, which is too fragmentary and so cannot be used reliably to identify rules. I configured the algorithm to search across combinations of the following 'itemsets', which are sets of phenomena that may potentially co-occur. I tested for associations between indicators of cohesion, indicators of incivility, and the following features of each discussion captured in the dataset:

- The OSN(s) in which the discussion was held
- Whether the message...
 - Indicated religion discussion
 - Was crossposted to more than one OSN simultaneously
 - Was part of a longer than average discussion thread
 - Starting a new discussion thread or replying to an existing thread
 - Was a participant's first message
 - Was posted by a longer-standing participant in these religion focused OSNs

This delivers analyses spanning years or decades so I can identify whether the relationship between factors varies over time, avoiding the weakness of computer models that do not capture this and may assume invariance (Lőrincz *et al.* 2019, 48). And when considering the association between these factors, I am not inferring *causation* but instead association. So I am not assuming whether, in a hypothetical example, a certain length of discussion thread tends to *lead* to cohesion, or whether cohesion tends to give rise to longer discussion threads. My caution arises because of: (a) the complexity of the systems I am analysing since behaviour may be influenced by some third factor (Kolaczyk and Csárdi 2014, 122) that I would not naturally associate but has been influential in the participants' mind, such as a previous discussion on a different OSN or offline; and (b) because I do not have formal experimental data, in which I can control for other influences and isolate one social phenomenon that I can then adjust to identify its impact (Taris 2000, 2-4). My data are non-experimental historical digital traces. While I must note that in some digital social scientific research non-experimental data can indicate causality tentatively, such as when a phenomenon has affected a random subset of a group of people (and its influence on other

things can then be inferred) or in some cases when a scenario has been repeated with only one variable changed (Salganik 2018, 50-70; Broström 2012 147-157). There may be some instances where this has occurred in message archives, but I am retaining the more modest goal of association, which still aids explanation of facets of the social interactions I observe. Can and Alatas (2019, 23-24) review social network analysis research into causality detection using OSN data but even where the term causality is used research may be identifying associations, since the complexity of the system renders causality of social behaviour problematic, particularly when linking social media to wider world events (such as stock markets) as in some of the research that Can and Alatas discuss.

This concludes the summary of my computational methodology that started in chapter two. I have summarised how I: accessed OSN message archives; extracted and transformed pertinent metadata such as message dates; loaded it into organised datasets; extract analytic descriptors and undertake network analysis, which together allow me to infer social processes of cohesion and incivility; and finally use machine learning (ML) to identify factors that are associated with cohesion and suggestive of the potential for incivility across the OSNs that comprise my research sites and the specific case studies I analyse within them. In the remainder of this chapter, I delineate how and why I analysed a subset of OSN messages manually, defend the methodology in light of the limitations of computational analyses of OSN data that I have not addressed hitherto, and argue that the ethical principles I embed in my research promote a just approach to ML research using OSN data that can unlock benefits for the common good.

3.4 Triangulating and enriching my research through manual analyses

Studies analysing very large potential datasets, such as OSNs, may focus on analysing a subset of the data, as in Squirrell's (2109) ethnographic research on Reddit. Squirrell did not use machine learning methods and found that manually coding every message was impractical, so analysed a subset for which he had identified selection criteria including certain historical messages and new messages as they were posted. In doing so he reached a thematic saturation (Squirrell 2019, 1914). This suggests that analysing a subset of messages can provide insights regarding a larger corpus. While this does not negate my need for machine learning approaches to answer my research questions (and the benefits of which I have discussed), manual analysis helps affirm my insights through a substantial detailed check and ensures my analyses accurately convey the events taking place. The value of such a qualitative component alongside large scale quantitative work is indicated

by Crooks and Currie's (2021) research which illuminates how research methodologies and outputs can treat minority groups in just and unjust ways — we must ensure our methodologies tell minority groups' stories accurately. With this in mind, I manually reviewed 1,000 discussions from the archives I accessed to triangulate my computational analyses, ensuring my ML approaches were reliable and enriching my insights.

3.4.1 Selecting data for the manual analyses

This foregrounds the need to select the 1,000 discussions carefully. My criteria were to include a range of messages to ensure the following are captured within the subset:

- Every OSN on which I focus in chapters four and five
- Every broad analytic descriptor, i.e. my computational labels that help me identify characteristics of messages in the datasets (see technical appendix 3)
- Flashpoints in participation (p.123)
- Different sizes of social network (by virtue of the number of participants posting)

I extracted these messages by using code in R to isolate a subset of 1,000 redacted discussions as part of my initial work to compile the datasets for my research, following which I excluded all other full text messages from my data. I ensured the discussions extracted fulfil my above criteria and I configured the code to extract an even split of 500 discussions from UTZoo (for use in my analyses for chapter four) and 500 from later, fuller archives (for chapter five). The discussion threads were selected at random within these parameters. I now term this second dataset my 1,000 Thread Dataset (the first dataset being my Message Dataframe).

As a representative example of how I ensured the 1,000 threads met my criteria, I used R to count the messages by subject lines and identified discussions that had large numbers of both messages and unique senders, as indicated by the random unique identifiers that represent senders in my datasets. This illustrated how discussions occur within various sizes of social networks and how some threads become flashpoints. I then ensured that my subset of messages for manual analysis also included flashpoints (in which participation spiked).

Having identified my subset of messages for all the parameters in the above bullet points, I applied an ML algorithm to remove last names and email addresses from these messages,

so I could analyse them in a redacted form and help maintain OSN participants' privacy. As discussed above, my core datasets do not contain the original messages, instead incorporating analytic descriptors of messages, and are designed to aid privacy.

3.4.2 The manual analysis process

Equipped with my dataset, I then manually reviewed the messages with two purposes in mind. The first was to identify whether any computationally derived analytic descriptors applied to them were fair reflections of the subjects being discussed and/or indications of cohesion or incivility, depending on the descriptor. The second was to apply short descriptive labels. For example, I could apply a note to a message indicating the thread focuses on creationism and appears to indicate cooperative discussion (i.e. an element of cohesion), or focuses on Judaism though is antisemitic. Notes refer to context of the thread, so while distinctive aspects of messages are reflected in the notes, they capture the tenor of the thread generally. This is suitable as I then reviewed entire threads again when later analysing interactions that reflected certain characteristics, in order to write narrative analyses in support of various computational findings in chapters four and five. The content of my review of religion on Usenet and the elements of cohesion and incivility that I identify act as sensitising concepts for writing these labels (pp.49-61). These elements are:

- Antagonism, including adversarial and contentious discussion
- Messages flooding a network, including spam
- Discussion of religious denominations and new religious movements
- Sensitive matters including abortion and antisemitism
- Cooperative discussions around shared aims/interests including religious doctrine, eschatology, millennialism
- Religious practice including prayer, prayer requests, evangelism
- Gatekeeping, including the effects of moderation

The labels are not a formal part of my analysis – I do not rely upon them directly for findings – but when later weaving discussion of OSN interactions into chapters four and five, to bring my quantitative analyses to life for the reader and provide a flavour of life within these social networks, I used these labels as a shortcut for identifying pertinent discussion threads (rather than individual messages) to read additionally closely in light of the quantitative analysis at hand and incorporate details of them as more narrative elements of chapters

four and five (cf. pp.181-184, 219-222). A sample of the message subjects and the notes I applied is presented in technical appendix 2.1 to indicate the approach I took.

In addition to manually reading threads and then incorporating details of them alongside my quantitative analyses in chapters four and five, the other facet of my manual analysis was to triangulate the computationally derived descriptors applied to messages. I consider there are two distinct types. The first is characteristics of messages relating to the participants' social network, specifically whether a discussion is higher participation, crossposted to multiple OSNs, is a message the first from a participant, and whether that participant is persistent, i.e. whether they participated in Usenet discussions across the OSNs I analysed for two months or more (as captured within the archives I used). It was straightforward to check that the computer code had flagged these characteristics of messages accurately. Since my 1,000 Thread Dataset includes all messages from the 1,000 threads selected, I could see that threads flagged higher participation had a greater number of messages included in the dataset whereas messages with no replies or only one were not flagged high participation. It was likewise simple to confirm that messages identified as crossposted ('interOSN' on my 1,000 Thread Dataset) were sent to two or more OSNs, as metadata for the messages appear on the same row of the dataset as 'interOSN' and other identifiers so can easily be cross-referenced. Finally, checking whether a message flagged as a participant's first took one additional step but was still feasible, and confirmed accuracy. For this, I sorted the rows of the Message Dataframe by the unique alphanumeric combination that represents each participant anonymously and then by date. This was simple as the dataset is manipulable within R Studio software, akin to sorting a large spreadsheet. Sorting the data allowed me to see that the first message from each participant in the dataset had been flagged accordingly and other messages had not been flagged. Since I had iteratively written, tested and improved the code for extracting the dates messages were sent and recording these in the dataset (see p.104), I had confidence in this aspect of the methodology. Sorting the dataset this way also enabled me to confirm that the methodology accurately flagged persistent participants, since I could easily observe the dates of the first and last messages that they had posted.

The second type of computationally derived descriptor entails more subjectivity, is probabilistic, and does not have a binary accuracy. I initially had two such descriptors: one for flagging potential incivility and a second for indicating religion discussion, so far as will be identified by searching for the distinctive or prominent presence of keywords associated with religion in my dataset (see technical appendix 4). I have since excluded the former

from my analyses. Regarding the latter, I knew the breadth of religion discussion that takes on Usenet from my review of previous research (pp.49-58), and focused my datasets on groups relating to Abrahamic faiths, so I sought to identify messages that have the characteristics of religion discussion within those broad faith groups. These messages are likely to be focused on the shared aims of religion discussion within the OSNs I analyse. For this, I referred to my Message Dataframe to identify whether messages across the dataset, which featured the keywords prominently or distinctively, had been flagged as indicating religion discussion, and cross-referenced the subject line which is a reasonable indicator of the subject of the message.

These analyses amounted to a substantial detailed check, in line with Eagle and Greene's (2014, 151-152) recommendation that researchers must have a means of double-checking the results of the computer models they assemble from big data sources. Since my computational and manual analyses span long time periods, I naturally incorporate triangulation of phenomena that I observe as any anomalous findings regarding, for example, a specific OSN in a particular month, will stand out against the broader landscape of what I observe. This is a little like Pomson's (2008) research in which OSN messages from 2005 comprised the core research site but messages from two years preceding this were analysed to check for representativeness. The subset of redacted messages is also highly useful for my analyses of certain case studies within the datasets, which I present in chapters four and five.

3.4.3 Ways in which manual analysis affirmed and challenged the computational methodology

Through these analyses, I found that the extraction of metadata and messages from message archives had a good level of accuracy. Specifically, the methodology extracted an initial 50,666 messages from the UTZoo dataset for use in my research. This reduced to 40,785 when removing duplicate messages, indicating that this element of the methodology worked. When then searching for, with a view to removing, messages for which a username was not captured, no such messages were identified. Since a username will always be present in message metadata, unless there was an error capturing this in the archive, the presence of a username in all UTZoo messages (the more difficult of the messages to parse into a dataset, given their diverse formats and metadata) affirms the accuracy of the methodology for this initial data extraction and organisation into a dataframe.

This good level of accuracy regarding data extraction was unsurprising since I improved my code iteratively during development on identifying any errors in how it extracted, for example, message subject lines or sent dates. The analytic descriptors (frequent and distinctive words, a binary classifier identifying whether such words indicate religion discussion, and an indicator of incivility) were reliable indicators to varying degrees as specified presently when considered at a high level across a dataset.

I found that messages which *had* been flagged as indicating religion had the hallmarks of religion discussion with a notable exception where the personal name Christian appears in the participant's message, rather than it referring to members of the Christian faith (though this is mitigated since some such messages nonetheless are about religion). A second limitation of the keyword approach is the sheer breadth of religion discussion that may occur – keywords need to maximise accuracy and mitigate against including messages *not* relating to religion discussion. As a result, messages relating to religion that do not include any of the keywords prominently or distinctively are not included. However, some such religion discussion is rightly excluded for not focusing on the Abrahamic faiths, which are the focuses of the OSNs I analyse.

I analysed data for a random sample of 359 posts across 100 threads to identify how the classifier performed. See technical appendix 2.2 for a list of those subjects of messages classified as indicating religion discussion and those that were not, and my views on the tentative usefulness of this classifier.

My critical review of this classifier affirms the importance of supplementing the findings of the computational analyses (such as indicating which factors of messages or the social network are associated with religion discussion) with manual review, to enrich our understanding of what aspects of religion are being discussed and to confirm the findings *suggested* by the computational analyses. I do this in chapter five specifically, when adding qualitative detail to my quantitative findings.

As I note above (pp.122-123), sentiment analysis was not reliable when applied at scale. In light of this, there are two related *potential* indicators of incivility. The first is a spike due to spam messages or flammable 'flashpoints' in discussions. The manual analyses identified some spam messages (adult content). My analyses did not identify any *clusters* of spam messages indicative of flooding an OSN with spam – having reviewed the 1,000 Thread Dataset and paying attention to messages flagged by my computer code as not indicating

religion discussion – but did identify incivility occurring in longer discussions. At times, discussion degenerates. While I analysed messages in my 1,000 Thread Dataset manually I found these spikes, which occur amongst higher participation discussion threads, to be a useful focus for my analysis when seeking to identify incivility. Second, refining this, messages that are crossposted, in higher participation threads, and are participants' messages, may reflect the presence of people not socialised in the OSN's behavioural norms, and perhaps holding views disparate to those of core members, which may be potential sites of incivility. This is affirmed by my review of literature regarding religion on Usenet and review of social networks formed when OSNs connect – they are at times diverse and potentially opposed to one another (pp.49-58, pp.165-177, pp.213-217). Therefore, the threads my computer code flags as higher participation, crossposted and involving new participants are a useful starting point here. Nonetheless, *both* possible manifestations of incivility are affirmed by the literature, which suggests the suitability of these measures as indicators of *potential* leads to follow up. Therefore, to identify incivility, I used the findings of my manual reviews honed by the above criteria, while observing the findings of the sentiment analysis but not relying upon it in any way.

3.4.4 Uses for the manual analysis findings

The manual analyses affirmed strengths and limitations in the data. I found two broad categories of robustness. First, some measures are categorical – whether the message was a participant's first, was crossposted to multiple OSNs, was part of a higher participation discussion thread, and whether it was sent by a longer-standing participant. I found these analyses to be accurate by following the checks detailed above. Second, some measures are probabilistic and their usefulness is more limited. I found the indicator of whether a message relates to religion discussion, within the set of keywords, flagged a small proportion of false positives. The messages it did flag were mostly relating to religion, but it left messages more broadly focused on religion (beyond the keywords) unflagged. Knowing this allows me to bear it in mind in chapters four and five, where I discuss interactions qualitatively to enrich my quantitative findings – I'll know what types of messages were indeed flagged as relating to religion discussion. As summarised above, I found sentiment analysis unreliable when applied at scale, so while I note its findings when reviewing my 1,000 Thread Dataset I do not rely on it in any way.

In short, the categorical computational analyses, and to a lesser degree the probabilistic analysis that flags religion discussion, helped me identify patterns that are plausible and

make sense. Importantly, I consider that they provide leads to follow up through the manual review rather than findings that stand alone, since they benefit from checking and enriching through incorporating a review of the discussion threads themes, bringing the quantitative findings to life.

As natural language processing is probabilistic and imperfect, these descriptors therefore served useful purposes of indicating some broad-brush trends and providing leads to follow up through closer analysis, including manual triangulation, in chapters four and five.

3.5 Mitigating challenges to the methodology and affirming its strengths

I now discuss challenges that could be made to my research methodology and, in doing so, I both acknowledge its limitations and affirm its strengths. These challenges are drawn from the literature I read while developing my knowledge of the field, along with others that came to mind as I defined my research questions and research sites and designed the methodology. I have already discussed some — specifically, the needs for the researcher to have expertise in their algorithms *and* the field to which they are applied, and how I am not inferring causality between cohesion and incivility and other factors within my datasets, instead identifying association. Here, I discuss others.

3.5.1 Imperfection in computational analyses

I must accept some imperfection in order to use computational methods for big data analysis — in my case when deriving analytic descriptors from messages that indicate subjects of discussion, and when inferring cohesion and incivility. I discuss one aspect of this above regarding the imperfection of computational processes for analysing natural language. I minimise imperfection and avoid it undermining my results by: (1) undertaking substantial checks through manual analysis to ensure general reliability and check for misclassification; (2) accepting association rules that have sufficient confidence, support and lift (see above) and identifying them via the exhaustive Apriori algorithm so that, taken together, the phenomena that they identify are likely to be present in a significant number of discussions and consequently if misclassification was present in a small number of cases it will not undermine the rules' integrity more generally; and (3) focusing my analysis on these general trends, not one-off individual threads/messages in which a misclassification could have an outsized contribution to the data analysed. This also mitigates the impact of any noise in the dataset in terms of outlier messages that are less well-fitted to the analytic descriptors

I derive and/or that contradict strong association rules I have mined. I spot outliers through simple visualisations of the labels I applied algorithmically to messages. Nonetheless, outliers don't undermine the relevance of my stronger findings *in general*. And besides, human coders may classify differently due to their subjective interpretation, particularly with marginal cases, and large-scale manual analysis may present other challenges in terms of identifying messages that fall within the parameters of the research and then analysing them. Abdulla (2007, 1069-1070) notes this in relation to manual processes for identifying, summarising, and analysing content for research on Arabic language discussion boards after the September 11 terrorist attacks.

A different type of human subjectivity potentially influences algorithmic processes for identifying themes/subjects within text data and I avoid this through the (relatively) straightforward approach I use for identifying and applying analytic descriptors. In Technical Appendix 1.1.2 (p.267-269), I summarised how deep learning language models (such as BERT) are trained on very large corpuses of text written in the English language, in BERT's case by masking 15% of input words and iteratively improving its prediction of missing words based on the context around them (Devlin *et al.* 2018). Consequently, biases that inform how English language is used are reflected in such models. This is evidenced by Hugging Face's tests in which BERT predicted a missing job role, which was masked in a sentence, in stereotypical ways based on the gender of the jobholder as indicated elsewhere in the sentence (Hugging Face, no date). Such biases potentially influence the word embeddings that researchers generate when using these increasingly popular algorithms. The methodology is not subject to this.

3.5.2 Factors not visible in my datasets

The contents of my datasets allow me to answer my research questions (see above). Nonetheless, other factors that I do not observe may also influence cohesion, incivility, and patterns of participation on the religion focused OSNs that comprise my research sites. This reflects the epistemological limitations of my work, in the same manner that an ethnographer has a limited vista onto the social world that they are researching in terms of the places they visit and the people they interview or alongside whom they participate. Literature on social research and digital religion indicates some of these factors. Age and gender may influence the use of social networking sites for religion (Nyland and Near 2007, 20) and historically Kong (2001, 407) found that men dominated Usenet OSN discussions. Later, Mullan (2015, 109) found that women starting discussions on, for example, modesty

or makeup, on a Catholic social media site were trolled. While the methodology (by design) does not capture participants' personal characteristics, so I cannot identify if certain characteristics influence subjects of discussion, participation, cohesion, or incivility, I nonetheless make a contribution to knowledge using the factors I *do* consider, ensuring the associations that I identify are significant (see discussion of ARM, pp.127-128 and TA 1.2).

Relatedly, religious events from other (online or offline) contexts may influence the factors that I measure. For example a significant world event may impact topics of discussion for a certain time on a particular OSN and Robinson (2005) found that online discussions of the September 11 terrorist attacks reflected, to some degree, perspectives that tended to predominate in the participants' national contexts. This reflects how OSNs are embedded in broader social lives, they are not closed systems, and my thematic summaries could be relevant to understanding popular religious discussion of such events. Or, as discussed, religious culture into which people are socialised offline may be brought online as in Lövheim's (2004) study discussed above, so prior socialisation may influence how participants behave. This indicates the need for knowledge of the broader (offline) belief and practice to which a religion focused OSN relates (such as, for example, British Catholicism or Methodism) and I seek to present such knowledge during my case studies in chapters four and five.

3.5.3 Limitations in how archives represent social realities

In chapter two, I considered how the archives that I use represent the social realities of the OSNs from which they are drawn. I elaborate here on how I mitigated concerns about this, since any deficiencies may find their way into my datasets (Weber in Brügger and Milligan 2019, 8-9). First, I considered whether the archives could be missing messages in a way that undermines analysis. This is less likely in later datasets since I use those provided by Giganews, a major Usenet service provider (Henderson 2012) who facilitate traffic of messages to and from recipients. As Usenet is a decentralised network, Giganews wouldn't have received all messages sent across all Usenet connections but this issue is not unique to Giganews. Dame-Griff (2019) noted discursive gaps in Usenet Historical Collection message archives most likely drawn from Google data; these omissions became 'null' nodes in the resulting social network analyses. Nonetheless, as a major hub, Giganews are in a stronger position to preserve and archive messages as they flowed through the system than, say, a third party who uses a computer program to 'crawl' the internet and save webpages to which they would only otherwise have access as an end user. Indeed the use

of these archives provides more direct access than accessing messages via the Google Groups interface which, while more immediately user-friendly (cf. Petit 2022), adds an additional layer of mediation and does not afford systematic access to large archives suited to computational analysis.

The UTZoo archive is more limited but is valuable (see pp.61-63) and I mitigate its incompleteness by limiting the types of analysis I undertake. Regarding later datasets, I have no reason to think messages would be systematically removed or removed in a particular pattern, save for participants intentionally having messages excluded as they don't want them archived (which is beneficial from an ethical perspective) and what I *do* find is not undermined by the possibility of incompleteness. In other words, I make associations between factors that are present, based on the messages in my dataset, and I don't make inferences regarding what I don't have, for example by arguing that a certain percentage of OSN discussions (more broadly than my dataset) exhibit certain characteristics. Had I identified that records are missing systematically, I had researched means of reducing their effect on analyses (Bramer 2007, 17-18; Taris 2000, 31-37) and, in any case, sound computational analysis of large archives enables researchers to work with more data than ethnographers or historians may have regarding other social milieux they may research.

It is also possible that OSN participants connect via other means, such as private email, whereas I assess messages posted to public groups. This is common to most research in social settings, in which participants may (or in many cases, certainly *do*) have communications not observed/recorded by the researcher. It does not undermine the relevance of what I do observe, and what I observe has a larger reach than private communications as messages are available to everyone who accesses the OSN, not just one person. The main possibility regarding the use of private messages is that a discussion takes place in public between two people and then continues in private — it's then still one-to-one though, not large group discussion, so of limited social impact in terms of number of people participating in the discussion. This is decreasingly a problem as times progressed because people could choose to mask their email addresses if they wished, bringing anonymity (see pp.36-38) which may increase willingness to post messages publicly.

I also considered how the OSNs I analyse reflect the social groups and religions from which they draw members, which relates to how certain demographics (such as men, as noted above) may be overrepresented at some times in certain OSNs. If we know that some groups are unlikely to be represented, we can reflect this in our analyses (Schroeder 2019,

10). This indicates the need to know Usenet's social context, which I do. As access broadened (see pp.31-32) the diversity of participants increased, though in *any* research using online contexts there are still digital divides in which skills, anxieties, and access to devices restricts some from using online services (van Deursen and van Dijk 2019). Weeber and Rodeheaver (2003) considered that the socioeconomic status required for internet access did not undermine their research at that time. Further, the more limited demographic profiles of early participants makes the UTZoo dataset amenable to the case study I undertake regarding discussions of religion and science online (pp.185-192). Knowledge of the narrower demographic then becomes a strength for my research.

In summary, the data I use inevitably do not contain all factors that may influence OSN discourse and participation. This epistemological limitation is common to all historical documents, ethnographic and interview studies; researchers have some limits to what they can access regarding their research subjects. Nonetheless, the methodology I've presented delivers datasets of large scale and targeted detail. They relate directly to my specified measures of engagement which, in turn, provide insights into important factors that impact OSNs and relate to digital religion studies' broader research focuses of community and authority online. Therefore, the existence of other factors that I cannot observe does not undermine the value of what I *can* observe. And in any case, I have some knowledge of the religious and social contexts to which I apply my analyses in later case studies (a strength, as I note above, of developing abilities in both computational methods and TRS), so I can be sensitive to some elements of participants' wider social worlds.

3.6 The ethical principles embedded in my research

Archiving takes considerable resources and, where it takes place, ethically sensitive research can reward that investment by showing the value of research using archived OSN data. Otherwise the opportunity to undertake valuable research (with appropriate safeguards) may be missed, as Dougherty *et al.* (2010, 11) note regarding Facebook data about the Iranian Green Movement around the 2009 election not being archived as, while its future usefulness was recognised, funds were not available. Here, academic researchers can balance the commercialisation of OSN data, which may not always be in consumers' interests (Mellor 2017, 240; Feenberg and Bakardjieva 2004, 21-22), with research that aims squarely at the common good.

Modern social media platforms require users to sign up to terms of service, which may include references to data being used for research, and this is not the case with all Usenet participants. Usenet newsgroups were international and public from their early years, with messages saved to files for each newsgroup and at each host's location (see pp.30-31), although retention periods differed and were often short in the earlier years due to the cost of disk space (Pfaffenberger 1996, 370-371). Indeed newsgroups remained public, unless retained as private such as on a local network that was not carried to hosts in other network locations and, consequently, are not within the major archives compiled. Usenet's public nature, established by its structure which I've discussed, is reflected in a proposal for a Usenet policy posted to the network in 1981. This stated that Usenet 'is a public access network. Any User is allowed to post to any newsgroup ... All users are to be given access to all newsgroups except that private newsgroups can be created which are protected' (Horton 1981 quoted in Hauben and Hauben 1997, 175). My work focuses on larger (or larger clusters of), public newsgroups carried across many hosts on the network over large geographic areas. The public nature of Usenet data simplifies ethical issues around its use (see Herring 2005, 149) while not eliminating the needs for protection, since public data may be used for unethical or even criminal acts (Tromble 2021, 4-5). Given the decades spanned by the interactions in my datasets, I also note that some participants in Usenet discussions may have since deceased. On some people's views, social media participants who have deceased should retain the same or similar rights to privacy as those still alive (Morse and Birnhack 2022).

I view the needs to handle public data carefully in light of Nissenbaum's (cf. 2011) oft-cited concept of contextual integrity, in which one factor in considering the appropriateness of processing others' information is whether it is reasonable in light of the context in which it was shared or created. Taking data from one context to another may entail 'context collapse' in which data are read, perhaps misunderstood, by other audiences (Misra 2017, 14). While noting that the nature of Usenet made participants' public messages available to strangers around the world, I reflect this by avoiding any additional sharing of participants' messages through my outputs, showing sensitivity to the message content through pseudonymising and removing personal details in my datasets and, through literature I read for this dissertation, having an awareness of the Usenet's social norms at the times the messages were shared along with my understanding of religion through my TRS studies. Dame-Griff (2019, 10) recommends awareness of social norms and message sensitivity in his paper discussing research using archives of transgender Usenet newsgroups, a subject that can be equally or more sensitive than religion.

Usenet's earliest participants are unlikely to have foreseen twenty-first century computational analysis research projects, though computational data analysis including artificial intelligence, for example for playing chess, was certainly a subject of interest since Usenet's early days, and the launch in the 1990s of commercial Usenet archiving and public search services such as Deja News enhanced users' awareness of archiving (Grossman 1997, 168). These services also helped ensure the data I now use have been public, in various places, for many years and consequently present fewer ethical concerns than private data. Neuhaus and Webmoor (2012, 51-52) note reduced ethical concerns when researchers use data that are already public. Alongside this, as discussed in chapter one, some participants used services to afford anonymity (Grossman 1997, 78-79), similar to how people use virtual private network (VPN) services to mask their internet location when browsing websites. However, poor knowledge of research undertaken using OSN data may persist today, since while users may sign up to terms of service for social media platforms, research indicates that few users read them in detail and may spend only seconds reading the complicated legal text they're asked to approve (Obar and Oeldorf-Hirsch 2018; Pangrazio and Selwyn 2018, 5). Consequently, meaningful informed consent is not obtained.

It is therefore beholden on all researchers, whether using historical or up-to-the-minute data, to implement appropriate safeguards even where users have signed up to terms of service that allow their data to be used for research. Increased awareness of research using OSN data, when ethical safeguards are in place, may reduce OSN users' concerns about such research taking place, particularly where informed consent is not practicable (Gilbert, Vitak and Shilton 2021, 11). Indeed, informed consent is generally impracticable when using OSN archives due to the age of some data, the impossibility of contacting historical participants and because such contact may itself be perceived as intrusive. The principle of informed consent is often embedded in institutional review boards' approaches to ethics but may be impracticable when conducting very large scale digital social research (Neuhaus and Webmoor 2012, 57-59). Instead, my research embeds safeguards by design and balances any risk by seeking to answer research questions that deliver recommendations in the common good — improving cohesion and reducing incivility that can come with 'toxic techno-cultures' that the affordances of social media networks can enable (Bannerman 2018). My approach reflects some OSN participants' interest in how data are used, such as them disapproving of commercialisation that works against the user by focusing on personalising advertising and surveilling (Allmer *et al.* 2014, 57-59), rather than improving

user behaviour (Gilbert, Vitak and Shilton 2021, 5-8). The machine learning I use relies on human input captured in the archives; it rests on participants' contributions and seeks to derive beneficial outcomes without any further investment from OSN users. This relates to Mü's (2019) reflections on human-aided artificial intelligence — that ML insights draw on aggregated human input — and, although a different use of the data from my research, Cachia, Compañó and Da Costa (2007, 1195-2000) illustrate this in their analysis of the use of OSN data to glean large cohorts of people's views about, for example, foreign policy decisions.

3.6.1 Embedding ethical safeguards by design

I state that Usenet was founded as a public platform, which is a useful perspective as it encapsulates certain principles that are foundational to how it works, such as those in Horton's statement and its decentralised network in which messages are copied to files and made available at hosts' locations. Nonetheless, broader definitions about what constitutes public and private communications may be misleading when considering public OSNs, as people may nonetheless share personal information that they do not want reproduced elsewhere. Zimmer (2018) considers this point in relation to research that released online dating profile information. While already 'public' in one context (people searching online profiles of others who seek romantic contact), the public release of a compiled dataset of such information may be considered inappropriate.

The 20 Newsgroups dataset is an interesting case. It is a series of text files containing well-organised messages from 20 Usenet OSNs including three related to (non)religion. The dataset is used in the ML research community to test NLP algorithms that researchers develop (Kaggle 2017; Xia and Doss 2021). This shows that in some contexts it may be uncontroversial to share these public messages and, perhaps more importantly, indicates that ML research on Usenet may have relevance to other OSN services, since many researchers using 20 Newsgroups to cut their ML models' teeth will be employing their model elsewhere later.

Research has found that software engineers, while recognising the importance of privacy, also struggle to define what it means when embedding it in software that acquires or uses personal information and so does not always provide absolute privacy to such information (Bednar, Spiekermann and Langheinrich 2019). Other research participants also have different, complementary, or contrasting views of what privacy means to them, such as

having some concerns around who may use communications and for what, and differing ideas around the barriers to access and controls that should be put in place. This is reflected in Pangrazio's and Selwyn's (2018) research into young people's views of privacy, Dobber *et al.*'s (2019) research which showed that younger people tended to have fewer privacy concerns than their older survey respondents, and Dumbleton's (2016) PhD dissertation which indicates students' diverse opinions about what constitutes privacy. And finally, OSN research I read for this dissertation takes different approaches to privacy, with some reproducing full text of messages, some doing so while obscuring authors' names, and some paraphrasing. The diversity of views about privacy and differing degrees of privacy afforded in research indicate that researchers undertaking social media analysis with potential benefits in the public interest need not be stifled by ethical concerns but show pragmatism (Richterich 2018) and reflexivity about how such research can embed ethical safeguards throughout its design, execution, and publication, thereby enabling valuable research to take place.

With this in mind, I apply higher standards of privacy and pseudonymity than are present in many other contexts where public Usenet data are held, whether on Usenet servers themselves, public facing archives or print or online research publications. This mitigates the risk of shining a spotlight on any one OSN participant whose historical data, while public, could be undesirably unearthed and showcased in full text (Driscoll and Paloque-Berges 2017, 53). I apply different and proportionate approaches to my handling of (1) the raw archives (which are drawn from the public domain), (2) my derived datasets, and (3) my outputs, including this dissertation. I now summarise these and explain how the granularity of the data in the object I'm discussing — raw archive, dataset, or output — correlates with the degree of protection I give it. Raw data are guarded closely, my datasets are protected, and public outputs focus on broad analysis of trends, reflecting internet historian Ian Milligan's (2016) perspective that writing at a high/broad scale entails few privacy concerns.

My *raw archives* were stored in encrypted files, available to myself and my supervisors. I protected these files to a greater degree than the copies available online. I have assembled *datasets* from the OSN archives using the methodology summarised in chapter two and the present chapter. My Message Dataframe, which includes metadata and descriptors about messages, is my first and primary dataset. Original messages are excluded in place of my analytic descriptors and I replace senders' email addresses with a random mix of alphanumeric characters. This approach to pseudonymisation reflects Mancosu and Vegetti's (2020) approach to creating an anonymous hash of Facebook users' unique web

addresses (URLs) when scraping data from Facebook, which they consider supports the undertaking of digital social research while preventing exposure of social media users' sensitive information. Above, I explain how I identified a subset of 1,000 OSN discussions for manual analysis. Together, these discussions comprise my 1,000 Thread Dataset. Since these discussions *do* include original messages, I applied regular expression ('regex', see TA 4) to remove email addresses, and last names from the dataset. I store my datasets in encrypted files. My word embedding, which captures semantic relations between words used in the archives, based on their co-occurrence in documents within the corpus (pp.112-113), does not contain any messages in their original structure. I store it in an encrypted file and it has not been published. Lastly, the *outputs* that I produce from my research exclude any direct quotations from messages, instead including brief paraphrasing from my 1,000 Thread Dataset to summarise points where necessary, and the outputs focus on broad trends as reflected in my word embedding and the association rules that I identify. My written analyses are aided by quantitative visualisations that present aggregated data, thereby protecting anonymity while still evidencing the insights necessary to answer my research questions. In taking these steps, I applied more stringent protection than apply to copies of the same underlying full text data that are available on the internet and my approach reflects elements of the methodologies used in other large scale OSN research, using recent or archival data (cf. Proferes *et al.* 2021; Bainotti, Caliandro and Gandini 2020; Mancosu and Vegetti 2020; Dame-Griff 2019).

3.6.2 The impacts of the researcher

I must also account for the impacts that OSN researchers potentially have on current and future OSN participation. As awareness of big data social media research grows, the perception of surveillance could potentially impact OSN participants' behaviour if they feel increasingly surveilled (cf. Zuboff 2019). Relatedly, perceptions of government surveillance may apply chilling effects to online discourse, as evidenced by Stoycheff *et al.*'s (2018) research into engagement in political discussion by Muslims in the United States. This panopticon effect ought to apply less to academic research that is clear about its data sources and handling, algorithmic processes, aims and outputs. The researchers' aims and how they process and present data are therefore key to addressing OSN participants' potential concerns and thereby minimising the researchers' impacts. In the longer term, an ideal situation would be that research in the field of digital religion (to which my work is a contribution) fosters constructive online discourse through knowledge exchange with participants, platforms, and those with responsibility for governance. This could lead to

changes in the datasets gathered in future, which may reflect the impacts of earlier research. Beer and Burrows (2013, 67) note that such recursivity in data reporting on social trends, such as music charts or companies' sales data, has always existed, but that the pace has accelerated and that digital and archive tools and services have created new ways for researchers and other individuals to interact with the culture that produce the data. This indicates the potency of social media research and the importance of reflexivity in defining research aims and articulating outputs. I embed this reflexivity in my research questions, which aim toward the common good, my operationalisation of terms such as cohesion and impact, which allow me to answer those questions using datasets that require no direct engagement with individuals, and my outputs which further protect individuals through reporting on rigorously evidenced high level trends. I also note that there is a risk to *not* doing research (Salganik 2018, 317-321) with OSN archives and communicating the outcomes, since we would then miss out on potential benefits that could reduce adverse outcomes from future OSN participation if participants, platforms, or moderators (as appropriate) take recommendations on board.

3.7 Setting out the research design

I now take stock by setting out the research design end to end. While chapters two and three cover individual elements in more detail, this summary aims to provide a high-level view and show the overall cohesiveness of the design.

The research design is a mixed methods approach. It begins by drawing on advances in internet archiving, which afford access to large scale archives of messages posted to religion-focused Usenet online social networks. The heterogenous format necessitated incorporating computer programming (here, in the language R) to extract and organise those messages into a dataframe, which is an analysable format. The programming then applies descriptive labels to each message, indicating whether the message was: a reply to an existing discussion thread, the participant's first message, crossposted to multiple OSNs, within a higher participation discussion, posted by a longer-standing participant, indicative of religion discussion. This part of the research design relies, in part, on natural language processing. I apply well-established methods to my specific context; these methods pre-date later AI innovations, which rely (to varying degrees) on processing on third party servers and are more probabilistic and have limited comprehensibility.

Such factors contribute to identifying elements of cohesion (see pp.58-61 for how I operationalise the concept), and the dataframe as a whole helps me identify behaviours

indicative of incivility (flooding a network with messages, spam messages, or antagonistic content). Further, the programming extracted 1,000 discussion threads at random – 500 from the UTZoo archive of messages from the years 1981 to c. 1993 and 500 from the Giganews archives, covering c. 2002 to 2015. The research design then incorporates further computational and manual analysis.

The computational analysis applies social network analysis (SNA) and association rules mining (ARM). SNA visualises their connections to other Usenet OSNs, i.e. the wider context in which religion OSNs are found. This shows how different OSNs established different patterns of posting over time and provides some context for why OSNs exhibited differing patterns of behaviour, as identified by the statistical computational analyses. For example, the early founding of OSNs focused on Judaism discussion, and soc.religion OSNs, may have supported their later flourishing as they were established and well-connected by the time access to online services broadened in the 1990s.

ARM is a form of exploratory analysis which identifies factors of messages that are associated with one another. I pay particular attention to associations between factors that occur more frequently than would be expected if those factors were unrelated. One prominent example in this research, and other case studies I have undertaken, is that crossposting messages to multiple groups is associated with people going on to participate in discussion for the first time. ARM helps me identify behaviours that may be associated with, or challenge, cohesion.

For the manual analysis element of the research design, I review the 1,000 Thread Dataset with sensitising concepts and message characteristics in mind for inferring the presence of cohesion and incivility (pp.131-132) in discussion threads. This provides richer detail to the computational analysis indicating, for example, some ways in which disagreement was handled within OSNs, topics of discussions, the content of spam messages that were posted, and the focuses of antagonistic content.

I then draw these findings together in chapter six to present conceptual conclusions, before relating my research back to concepts of community and authority in digital religion (as explored in chapter one) to confirm what has been learned regarding those concepts from the specific context of religion-focused Usenet OSNs.

3.8 Conclusion

This chapter has presented the remaining components of the methodology. I referred to the structure of Usenet and how Usenet OSNs function (discussed in chapter one) to explain how social network analysis can help illuminate cohesion and incivility as defined in my measures of engagement, along with patterns of participation. My datasets and these supplementary analyses enable me to answer my first research question: ‘in what ways, and to what extents, do cohesion and incivility manifest in a substantial sample of prominent religion-focused historical Usenet OSNs?’ I then explained how association rules mining of my datasets including the analytic descriptors derived computationally (see Technical Appendix 4) can illuminate factors that are associated with my measures of engagement. These analyses enable me to answer my second research question: ‘which factors are associated with cohesion and incivility?’

Following this, I discussed challenges to the methodology and the ethical principles that I embed in my research to assure readers of its general reliability and suitability for large scale data analysis, while acknowledging its limitations. I then close with a summary and this concludes the methodological chapters of my dissertation. I now apply this methodology to my research sites, UTZoo and select Giganews archives and, within each of them, a specific case study. This shows how the computational methodology works in practice and delivers insights of interest to TRS scholars.

Chapter four — Archaeology of the Digital Network

This chapter has five sections, each the result of applying elements of my computational methodology to the UTZoo archive of early Usenet messages from c.1981 to 1993. First, I review the growth of religion discussion within these early online social networks (OSNs), illuminating how the network structure afforded critical and, later, faith-based discussion of religion online. I primarily draw on my own primary analyses of the formation and evolution of these OSNs, though I cite other scholarship where useful. Second, I present prominent themes and subjects that emerged in these interactions, indicating what participants talked about when Usenet's demographic was primarily government and academic researchers, followed by technology enthusiasts. Third, I present a sample of interactions between religion focused OSNs and those set up to discuss other topics. This sets up the fourth section, which analyses how my concepts of cohesion and incivility, as developed in chapter one, manifest in discussions of religion on the early Usenet. The chapter concludes with a case study that develops from this context to illuminate cohesion and incivility within discussions about religion and science. In taking this approach, the chapter moves from analysing some high-level structure of the early Usenet and users' interactions, before moving to closer analyses of interactions within that structure and then focusing specifically on the social processes drawn from the conceptual framework for my research. This enables me to draw on broad results from my computational methodology and distil down to specific insights, first presenting factual findings followed by interpretations using my concepts. While UTZoo's incompleteness restricts me from applying the full scope of the methodology (saved for chapter five), these insights make a significant contribution to answering my research questions.

4.1 The emergence and development of religion on early online social networks

I applied the methodology developed in chapter two to extract and organise messages from the UTZoo archive and the Giganews archives of later OSNs analysed in chapter five into two datasets, incorporating ethical safeguards. The first — my Message Dataframe — contains summary information regarding every message and the second — my 1,000 Thread Dataset — is a redacted sample of full text messages from 1,000 threads including 500 from UTZoo (see pp.130-131). This allows me to identify discussions of religion and the OSNs within which they took place. In addition, contrasting the full text messages of my 1,000 Thread Dataset against the computational findings of my Message Dataframe allowed me to affirm the general reliability, strengths, and limitations of my coding for

extracting, organising and computationally analysing messages. Two key limitations of UTZoo are that I cannot infer the length of discussion threads, in other words how many responses were received for each opening message on a subject line, and the dataset does not contain even one message from all religion focused discussions. These limitations are due to the fragmentary nature of how UTZoo was compiled, though it is a credit to early participants and later archivists that any data from the 1980s has survived.

Prior to the restructuring of the top-level hierarchies within which Usenet OSNs could be formed ('The Great Renaming', see pp.32-33), groups to discuss religion were formed within the 'net' hierarchy. The first of these was net.religion, following participants' deliberations about whether to form it in 1982 (stated in Helland 2007 and identified in my own analyses). A concern was the fiery tone of many discussions about religion and whether space should be made for this topic or whether it should instead be discussed in net.flame, where such a tone was typical. Others argued for net.religion so they could ignore religion discussion as it would then be contained to some degree, especially with storage space at a premium on early computers used for research. Therefore, the creation of net.religion was a response to, not the start of, online engagement about religion. Creating net.religion provided space for religion discussion to continue and be open to those who desired it, in the same manner as other topics such as politics within net.politics which was formed by 1982. My analyses identify discussion or mentions of religion as early as 1981. Among the earliest are humorous posts and others' responses to them on net.jokes. These often relied on stereotypes, prefiguring content that would (very quickly) come to be considered flaming and, in today's language, trolling.

A proposal was also made in the OSN net.general during February 1982 to start a group for discussing religions separate from, or preceding, Christianity, which would be titled as a Paganism OSN. This was not successful at the time, though discussion of Pagan religions (and people identifying as Pagan) continued and grew within OSNs focused on religion and other topics. Again, the fragmentary nature of the UTZoo dataset prohibits any reliable statistics, but at least a few thousands mentions of Paganism were made during the 1980s. This prefigures the growth of Paganism online in the 1990s, captured in Davis' (1995) *Wired* article, which was one of the first articles on digital religion. Helland (2007, 961-962) notes that members of other religious groups sought to have OSNs created and dedicated to those religious traditions around this time, such as Buddhism and Wicca, but similarly to Paganism these did not receive enough backing.

Indeed this reflects a growing problem that emerged. Increasingly diverse discussions of religion occurred within net.religion, connecting people of diverse (non)religious identities and motivations for participating, as is enabled by Usenet's decentralised network structure (p.28). My analyses found that proposals were made in June 1983 to retain net.religion but provide 'subgroups' within a net.religion sub-hierarchy for those who are participating from the perspective of a certain religion. The aim was to reduce some heat and noise, improving the usefulness and tone of discourse. This reflects how community and authority can benefit cohesion (which I explore further in chapter six), since engagement within OSNs focused on a specific religion are more likely to enable shared religious or non-religious aims for the group and affirm shared sources of authority whether, for example, a religious leader, text, or church. They could also improve the signal to noise ratio, benefiting cohesion since the latter requires participants to be able to engage effectively and not be frequently crowded out by uncivil, off-topic, trolling and flaming messages.

This preceded the formation of OSN net.religion.jewish, which was proposed in January 1984 on net.news.group. This was before groups for other major faiths had been created and the echoes of this group's early creation may be seen in later decades, since discussion of Jewish life and faith comprise a larger proportion of discussion than other religious groups within the archives I used for later analyses, presented in chapter five. Messages to net.religion.jewish appear from February 1984, when the group was founded, and diverse discussions and connections with other OSNs ensued (discussed below). The first dedicated group for discussing Christianity was net.religion.christian, founded in November 1984, intended for Christians and those interested in Christian religion. This distinguished the group from net.religion by making it more focused, regarding both the subjects that could be discussed and the expected demographic comprising a proportion of Christian participants. Although there was no proscription on debating the truth of Christianity, some people expected the group not to focus on this, as indicated by discussion within the group in February 1985.

As explained in chapter one (pp.32-33), in c.1986-7 a set of seven top level hierarchies were created, within which individual OSNs could be formed. This enabled both more breadth and more nuance around the subjects/groups on which an OSN could focus and expectations around the behavioural norms of the group. This included the 'talk' hierarchy, which proliferated and made space for diverse, casual discussions. Some persisted into the 2020s, accessible via Usenet software or, perhaps more commonly, the Google Groups interface. Religion discussion developed here with talk.religion formed in August 1986 and,

in September 1986, talk.religion.misc. By 1987, if not before, talk.religion.newage had also been formed. While this may appear niche now, the appeal of New Age discussion on Usenet during the 1980s makes sense in light of at least three circumstances. First, diverse syncretic religions, often drawing on Westernised versions of Asian religious tradition, experienced growth in twentieth century America within an individualised ‘spiritual supermarket’ (Hanegraaff 2001). Second, this diversity of practice and belief broadened the potential remit for interactions within a New Age OSN. Third, the countercultural underpinnings of Silicon Valley may have led to a greater than average level of interest in New Age discussions among Usenet’s early demographic.

Access to talk.religion was likely more limited than net.religion since, before Usenet messages were sent over the internet, many locations (recall these were often government or academic research institutions) chose not to forward on such messages to recipients ‘downstream’. I discuss this in chapter one (p.33) and a desire to limit costs of transmission and storage as Usenet grew is evidenced in a UTZoo discussion from 1987, which refers to a petition to remove expensive, high participation groups including talk.religion.misc due to perceptions about its poor ratio of heated interactions to those more illuminating. This gatekeeping decreased over time as Internet access improved and talk.religion groups — including many subgroups for specific religious denominations — survived and comprise some of the archives I evaluate in later case studies.

The ‘soc’ hierarchy was also formed at this time. This focused on social topics and notable within this — for the purpose of this research — is the creation of soc.culture within which groups for certain geographic regions and demographic groups formed, including those relating to religion. Here, ‘soc.culture.jewish’ is prominent. I consider Chua’s (2009) study of this group in chapter one (p.51, p.58) and it forms part of my case study in chapter five. Formed in September 1986, soc.culture.jewish indicates the porous boundaries around what may be considered ‘religion’ online. These are encouraged by the three elements of the hierarchy in which the OSN is based — social, culture, Jewish. Rather than interactions comprising debate or the expression of views on religious traditions, history, propositions, or practices such as prayer, as comprised much (though certainly not all) discussion in net.religion groups, the society and culture focuses of soc.culture groups encouraged broader parameters. Here, discourse on different cultural groups, geographic locations, diaspora communities, cooking, schools, higher education, politics, and more was within remit. These different behavioural norms suit the discussion of Judaism and Jewish life well, given the extent to which Jewish practice may permeate one’s life and environment. These

broader parameters indicate why participation on soc.culture.jewish was high. By contrast, soc.religion.christian (discussed below and in chapter five) experienced low participation — a message from May 1987 indicated it was within the twenty least-read OSNs on Usenet (and see figures for September 1987 below). At least three factors may contribute to this: first, discussion of Judaism and Jewish life was established early on Usenet as noted above. Second, as indicated by a message in April 1987, soc.religion.christian was a moderated group. Indeed it replaced the earlier mod.religion.christian, since advances in Usenet software meant that moderated groups needn't be located within a dedicated hierarchy named 'mod'. Moderation enabled a tighter rein over the content, subject, and tone of messages posted, but compounded the (then) slow transmission of messages across the sites that carried Usenet content. Third, the narrower focus on religion rather than all facets of life from a Christian perspective more broadly.

By contrast, groups such as soc.culture.jewish broadened parameters from discussing religion directly and frequently from a distant, rationalised perspective as was becoming of Usenet's early demographic of government, academic, and commercially employed researchers, to discussing myriad topics that are influenced by religion and comprise a religious person's life. This can build cohesion within an OSN, since religion can provide the basis for a shared identity and interests, as explored in chapter one (pp.57-60). Antisemitism also abounded, limiting cohesion and introducing incivility as I explore below and analyse more deeply in my case study in chapter five, which uses later and fuller archives of soc.culture.jewish and its moderated counterpart, soc.culture.jewish.moderated, which sought to overcome these issues.

Discussion of Islam had also proliferated during the 1980s and a message from September 1989 indicated that it was being discussed in a range of Usenet OSNs, none of which were dedicated to the subject. This was similar to interactions about Judaism and Christianity before groups were setup specifically for discussing them. A proposal was made at this time for a moderated soc.religion.islam to be formed. After some contestation around the legitimacy of the voting process — concerns around counting digital votes accurately are not new — by November 1989 messages were being posted to soc.religion.islam, with the first in UTZoo presenting the Quranic names of Allah.

Moving to a higher level, two reports posted in 1987 illuminate how many messages were sent to forty different newsgroups and provided estimated statistics of how many people read and participate in 280 groups. This situates the place of religion discussion within

Usenet and indicates its prominence. For context, the most popular OSN by readership was comp.sources.unix with approximately 28,000 readers, which is unsurprising since Usenet was established to discuss and share files regarding the Unix computer operating system (pp.30-31).

At this time, soc.culture.jewish received approximately 378 messages per month and had an estimated 4,800 readers or participants — approximately one in fifty people on Usenet in 1987. It was received at 94% of sites (i.e. locations with computer terminals that sent/received Usenet messages) and had an estimated cost of \$0.99 (USD) per reader per month. Given the cost of sending and receiving messages, this high level of propagation indicates the group was valued. Statistics are provided for three other religion-focused groups, listed here:

OSN	Readers/participants	Number of messages per month	Usenet sites receiving this OSN (%)
soc.culture.jewish	4,800	378	94
talk.religion.newage	5,000	122	85
talk.religion.misc	3,800	22	87
soc.religion.christian	1,700	4	85

Table 2: estimated statistics of Usenet newsgroup/OSN activity in September 1987

4.2 Themes and subjects in early religion-focused interactions

As indicated above, the sequential development of the ‘net’ hierarchy and OSNs within it, followed by ‘talk’, ‘soc’ and others, helped to establish various behavioural norms within different Usenet OSNs. This had lasting impact, since this high-level structure persists to this day and, so long as behavioural norms are enforced or reproduced through socialisation (see pp.41-43), these norms remain to some extent and help support (or hamper) cohesion over time. It is therefore illuminative of later OSNs, which experienced higher participation among a diverse public, to explore the subjects and themes present in early discussions. In this section, I draw on subject lines and keywords extracted for my Message Dataframe and the subset of full text message threads, which comprise my 1,000 Thread Dataset, along with notes I made about them (pp.131-133). This provides me with a high-level overview of the themes and subjects of discussion across the dataset — the computational analyses are particularly useful here — and some specific threads from which I draw more granular

and nuanced insights. Therefore, while space precludes an exhaustive review, these overviews are broadly indicative of the themes, subjects, and tenor of interactions.

As the first religion focused OSN, net.religion witnessed a breadth of religion discussion without exclusive focus on one religious group or topic. Usenet's early demographic of academic and government researchers brought a rationalised approach to Usenet discussions in general — they focused initially on conveying facts and making arguments — though the prominence of religious belief and practice among Americans in the latter-twentieth century (cf. Stark and Fink 2000, 218-258) means it is unsurprising that religion was discussed early and often. Nonetheless this potent mix, along with net.religion not being moderated, sparked diverse discourse. I will use my two datasets to review that here, before considering the subjects and themes present within other prominent net.religion, talk.religion, and soc.religion groups.

The rationalised approach is reflected in net.religion interactions that emphasise reasoning and argument in the early 1980s. These span ostensibly sincere discussion about key theological issues — such as the existence of God and the nature of sin — to humorous engagement using the early Usenet demographic's 'native' languages of math and computer programming to engage with theology and religion. St Thomas Aquinas also features, perhaps because he resonates with those approaching religion from the perspective of reasoning. Indeed, Aquinas' approach to reasoning the truth of Christianity was discussed along with his reflections on distinctions between mortal and venial sin. On the other hand, Aquinas was also cited in whimsical engagement that uses mathematical notation as a form of reasoning to indicate characteristics of God. I explore interactions between science and religion at this time later in the chapter.

In addition to reasoning and argument, the human body is a site of much discussion and contestation. This becomes persistent, as I also explore it in chapter five's analyses of later Usenet OSNs, and continues today with fiery interactions online around individual autonomy over one's body in relation to abortion and LGBT+ rights. This is prefigured on the early Usenet in, for example, discussions of pregnancy, infant death, and gay people. These difficult topics spark impassioned responses, such as regarding the rights of individual participants to promote their views to others, whether on Usenet or elsewhere in their social lives. Perceived dissonance between more conservative Judeo-Christian views on, for example, abortion and the adoption of more liberal views in wider society are also explored. General trends toward secularisation may help explain this. Indeed, later

interactions about the place of religious values or symbols in public spaces and education, considered in chapter five, reflect more explicitly a distinction between what some religious participants, and those who oppose them, consider acceptable. In addition, the counterculture of 1960s America which infused the development of Silicon Valley suggests that early Usenet participants may have been more likely than the wider public to sympathise with views (and alternative religions) that counter longstanding religious morality.

Helland (2007, 959-960) notes that people would engage with others about their faith on net.religion and receive extensive criticism of their beliefs in response. I observe this highly critical tone, at times separate from the above rationalised critique of faith from scientific or mathematical perspectives, in messages that are immediately on the attack criticising those who express religious belief or practice, claiming they lack mental capacity. This can draw on tropes such as 'brainwashing' to undermine arguments based on theology or religious philosophy and adds heat without light. Later, I discuss ways in which, and extents to which, cohesion and incivility manifest on the early Usenet but I mention this here to contextualise another insight identified from my datasets, which is that net.religion was also home to sincere promotion or discussion of Abrahamic faith conveying a highly positive tone about religion. While it is difficult to tell due to the UTZoo archive's incompleteness, such discussions seem to have gone uncontested by others at times, or at other times were contested but then other participants challenged the tone of the contestation or affirmed the religious post in some way. I conclude this is present where any engagement recorded in UTZoo with a discussion highly positive and promotional about religion showed an absence or very low level of negative response on manual review of my 1,000 Thread Dataset. Topics of such discussions include religious retreats and Bible apologetics and discussions of characteristics of God, occurring within the first three years of interactions on net.religion. This timing is perhaps relevant, since by then net.religion had become the designated space for religion discussion so participants who sought not to be exposed to religion could choose not to read or participate. Consequently, those who engaged did so intentionally. In addition, the collaborative and collegiate foundation of Usenet from 1979 had persisted to this time, so overtly uncivil or antagonistic messages were less a part of Usenet's behavioural norms than later. Nonetheless, discussions often reflect the critical perspective that Helland identifies, giving rise to a need for an alternative social network in which the more positive and religiously motivated interactions, such as those above, could take the foreground. This is reflected in interactions on net.religion.christian.

In particular, net.religion.christian affords space for discussions in which for some people the existence of God, with characteristics as articulated in Christian theology, appears axiomatic. This provides a foundation for cohesion within the group (discussed below) and is seen in messages captured within UTZoo, such as those discussing participants' perceptions that God is acting in their lives. Some such interactions are aimed clearly at coreligionists and indicate a willingness to be vulnerable to sharing these experiences and embedding them within a narrative about their social or spiritual lives. A shared foundation of Christian belief or experience among some members is also indicated in the promotion of Christian texts and beliefs, in which discussions reflect a positive tone around their acceptance. The Roman Catholic Church features among these, including issues of priestly authority and Catholics engaging with and submitting to well-established Church teaching. More widely, messages around this theme also indicate religious believers' awareness of the limitations around what they know. Participants indicate limits to their epistemology. This awareness may have come to participants' minds while writing on Usenet since online discourse to explain or express religious ideas or experience, particularly through text, requires articulation of theological understanding which may resist neat expression, particularly where a person is not equipped with a sophisticated theological vocabulary. This is a stark contrast with promoting or sharing religion based around participation in, for example, an aesthetically rich Eucharistic service which engages all senses and invites a person to step into a potentially transcendent experience.

Notwithstanding the apparent shared Christian beliefs among some participants, other interactions on net.religion.christian exhibited more contestation around religious beliefs and practice. Indeed, when founded, the remit of net.religion.christian included discussion of Christianity and related matters, so although coreligionists found a supportive environment it was not without challenge. However, perhaps the more collegial early Usenet behavioural norms and the professional demographic are reflected in the tone of critical discussions which are more constructive than later flaming/trolling messages. One dialogue from the mid-1980s is sceptical of Christianity in light of human suffering and ambiguities in the Bible, calling into question whether the focus of Christians' faith has a character worthy of worship. This dialogue takes a witty or sarcastic tone at times, but also focuses substantively on the reasonableness of belief in God considering critics' perceptions of God's actions and Christians' basis for faith. It reflects a rationalised approach to knowledge, reflecting the world of scientific research that many participants from this time inhabited. Another discussion from around this time focuses on whether a person can know they have encountered God from sensory experiences, and whether the attribution of such

experiences to God is independent or occurs only/primarily when seen through a lens acquired by socialisation into religion. Discussion reflects on participants' own religious experiences, and absence of experiences, and how they are interpreted through the person's theistic or non-theistic beliefs. Some messages remark that participants can be spiteful in their responses to religious participants' discussion of their experiences of God and distinguish between those more malicious reactions and others' calmer, critical language. This is reflected in discussions summarised in this paragraph. In this way, cohesion between Christian and critical participants need not be undermined since interactions are civil, fall within the aims of the OSN, and remain within its specified aims.

As may be expected since net.religion.christian and net.religion.jewish were both formed within the broader net.religion hierarchy that had its own behavioural norms, these two OSNs' participants exhibit similar behaviours. Save for flaming/trolling and spam, participants operate within the parameters of discussing the OSNs' eponymous religions with or without their own religious commitment, and typically with civility notwithstanding disagreement. However, there is a distinctive breadth to subjects discussed within net.religion.jewish, perhaps reflecting the breadth of Jewish life within and beyond what might be considered religion. (This breadth reaches full fruition in the later soc.culture.jewish OSN.) These interactions reflect at least four motifs, which I summarise here. While distinct, at times they overlap and relate to one another. The first is how many interactions relate to defining boundaries. These include boundaries around whether certain branches or communities of Jews may be considered authentically part of Judaism. At times these interactions reflect on God's calling of Israel as a chosen people, the manner in which this happened, and how it affords a distinctive human identity within Judaism which is not available to all. Separately, boundaries are also defined around acceptable subjects of discussion with the OSN, affirming a broad focus on Judaism and rejecting discussions of Christianity and other religions, when considered apart from Judaism. This may benefit cohesion within the net.religion hierarchy on Usenet more broadly since it helps ensure interactions within OSNs remain on topic. It also reflects a form of gatekeeping (see p.30 and pp.84-85) with certain discourse permitted or not, influencing exposure of certain participants to certain topics, based on the OSNs within which they interact.

Some interactions around the subjects of the second motif, antisemitism, and the holocaust, were later subject to gatekeeping with their discussion moderated within the later soc.culture.jewish.moderated OSN, created in part in response to off-topic and offensive messages within its unmoderated counterpart (which I recognise as reducing cohesion and

increasing incivility). Within net.religion.jewish, discussions of antisemitism and the holocaust are often from a more emotionally detached point of view, based on my observations of the tone conveyed. They focus on media, politics, or scholarly work that focuses or touches on antisemitism. As with net.religion.christian above, engagement does not reflect widespread offensive flaming/trolling messages that were part of later Usenet (and other) OSN participation, though we cannot rule out the effects of how UTZoo was archived manually, reflecting the archivist's interests and messages that were accessible at his location. As noted above, participants engaged around myriad practical and material aspects of Jewish life. These are captured this third motif. Here, discussions range from the sabbath and negotiating religious obligations around work, to Jewish custom in married life. Suggestions and requests regarding kosher food, recipes, and restaurants emerge. This may partly reflect how Usenet participants could be away from their country of origin to study or research, so may actively seek connection with other Jews, perhaps themselves dispersed spatially, to ask advice on food and other such immediate material needs.

The fourth motif retains a focus on Jewish life broadly conceived but is at a more macro level, with discourse around (geo)politics, geography, Israel-Palestine relations, and secularisation. Israel and the promised land feature prominently here and this prominence continued into the later, fuller archives of Jewish OSN discussion analysed in chapter five. Within this site, the geography of the land and political interventions to address the consequences of contestation between Israel and Palestine arise. Diverse views are shared, likely indicating the varying religious affiliations, political views, and geographic locations of Usenet's early participants. In addition, prefiguring later engagement about the Iraq war, in the mid-1980s participants engaged around the strategic place of Israel and oil reserves in light of the Cold War. This indicates how some participants engage from the perspective in the United States, in which many of Usenet's initial participants were located, alongside Canada and then internationally as the 1980s progressed.

A focus on the body is common to many discussions, such as using one's body in compliance with the law and one's identity as a Jew being imbued by historical lineage. The material culture of food and cooking may also be seen to relate to the body, and God's role as an authority in Jewish lives to require certain behaviour and instil identity marking out Jewish participants as a distinct group with a strong communal shared identity. In light of my identification from Usenet research and digital religion literature that shared (religious) aims are an indicator of cohesion (pp.57-60), Jewish communities have a strong foundation

of personal, embodied identity from which to build. This is reflected in the very early emergence, persistence, and blossoming of Jewish OSNs.

Such a shared basis for identity and community is not found among participants in the talk.religion hierarchy. The talk hierarchy was intentionally less focused on directed discussion within shared aims on specified subjects. It was a place for OSNs that accommodate conversations and chat that emphasise more relaxed socialising — a place for talk. As with net.religion.jewish, the breadth of discussion led me to identify discrete, albeit overlapping, motifs that represent themes within discourse on talk.religion OSNs. The first represents how people used these OSNs to promote and affirm their own Christian belief. This was before popular level books that supported Christians engaging online (cf. Baker 1995; Lochhead 1997) and indicates that even without such guides, participants chose to try and share their faith with others. This may have been for evangelisation or to connect with coreligionists, though the presence of this content within talk.religion rather than within, say, net.religion.christian suggests some missional intention. In some such messages, participants reflect on their personal faith and perception of God's activity in their life. Others are less personal and have a more eschatological focus, seeking to educate and indicate urgency of religious adherence, and some affirm the authority of the Bible as a means of justifying their position. The reliance on a religious authority may not cultivate cohesion here, as it could in a group populated primarily by coreligionists, due to the likely diversity of (non)religious affiliation among participants in talk.religion discussions. Other messages simply present theological information and notes on important historical religious figures, which treads a line between cultivating discussion — talk — about these themes while also educating readers about a religious faith. The most zealous of messages have a tone that indicates conflict or confrontation between, on the one hand, humanity's behaviour and the world we inhabit, and on the other hand, righteous Christian conduct. Although some of these messages were presented as information, rather than explicitly seeking dialogue, they appear to have gained little traction. This reflects Howard's (2011, 47-48) later experience that messages posted about Bible prophecy were received with silence or hostility. Participants in talk.religion often didn't engage around more evangelistic subjects so such discourse did not cultivate cohesion between them on my definition.

Interactions on talk.religion included a focus on alternative spiritualities including the New Age, reflecting the broad, exploratory, light, and often playful behavioural norms of talk hierarchy groups. To provide a flavour of this second motif, interactions refer to witchcraft, satanism, alternative spiritual texts and theories of humankind's origins. Nonetheless, there

is perhaps a tension here between the countercultural milieu in Silicon Valley from which many participants were drawn and the rationalised epistemology that characterised many early interactions. This is indicated in the apparently critical responses that some messages around these (and related) subjects received and helps contextualise some Usenet sites' (locations') decisions not to receive messages sent to talk.religion groups. This tension is also exhibited in the third motif I identified, which is the prominence of argument with a critical tone about religion, often some aspect of Christianity. As seems perennial, these sometimes zero in on difficult moral aspects of religion, such as slavery and the existence and nature of hell. Here, putative inconsistencies are identified between the positive ways in which religious participants express their faith and the awkwardness of certain parts of the Bible or theological teaching. At times, these interactions become conversational and have high participation, so far as can be observed given the limitations of the UTZoo archive. Humour may be used to mock or lighten the tone and a sense of the community that would emerge in later atheism focused OSNs is seen. Indeed at times, atheism is referenced directly and people indicate atheistic views. While the dataset here is too small to draw conclusions, this candidness about atheism occurs at a time when diverse forms of religious belief and practice were prominent in America (Finke and Stark 1992) and religious participants may have experienced more criticism of religion online than in their wider social environment. Here, talk.religion may have been potent for sowing seeds of atheistic thought, since research shows later atheism focused OSNs were influential in helping people form nascent atheistic identities (Nash 2002).

At times, the fourth motif extends this criticism implicitly or explicitly in discussions around the place of religion in public life and society. These interactions include discussion of belief in creationism as opposed to natural evolutionary accounts of life's existence and diversity, and the presence of any religious influence in large companies and among their leaders. This can be seen in light of the privatisation of religion, a key element of some secularisation theories (cf. Bruce 2011, 27 fig. 2.1) which articulate how religion's role in the public sphere can change. On this trend, public and private companies and organisations' leaders may hold a religious faith but, generally speaking, religion is less likely to be incorporated into the institution's identity and ethos. Instead, faith is held privately with less overt expression. If privatisation becomes the norm, the influence of religion in the lives of public leaders and institutions may appear unusual and noteworthy. This may be particularly so if that expression of religion is perceived as deviant by some. This is the case with creationism, which receives much criticism from scientists and other researchers (cf. Brumfiel 2008; Numbers 2006), demographics that populated the early Usenet. UTZoo shows these

subjects yield several responses (archive limitations preclude figures) and indeed my fifth motif is that some talk.religion interactions become conversational in length and tone. Like creationism, some such interactions relate to subjects likely to appeal to — or otherwise animate — Usenet’s community of researchers. These include a discussion of epistemology in relation to God’s (fore)knowledge of particles’ positions on theories of quantum mechanics and I evaluate interactions at intersections between science and religion in the case study below. In addition, geopolitical discourse — such as discussions about Israel — feature here and explore intersections between religion and politics. I note below that some elements of incivility can be found in discussions of religion and science on the early Usenet and in chapter five I find that discussions of Judaism and Israel can spark incivility. Nonetheless, these talk.religion messages did not trip my analytic descriptor for incivility according to the parameters I set for the Syuzhet algorithm. As I discuss in my case study regarding religion and science (later in this chapter), I find that incivility may be less present than in later years and I offer some factors that may influence that.

Finally in this section, I explore the soc.religion.christian OSN. It poses a contrast to talk.religion groups since it is moderated. The other major site of religion discussion is the soc.culture hierarchy, which I will analyse in chapter five since soc.culture.jewish.moderated provides the case study for that chapter. Discussion on soc.religion.christian appears more focused than talk.religion and soc.culture which reflects its religion focus and moderated nature. Specifically, topics within the study of biblical languages, interpretation and critical reflections on scripture, and epistemology abound and these provide the first motif I identified. These reflect varying degrees of technicality, but many interactions show serious engagement with the subject reflecting a shared basis for discussion, whether that is a mutual faith or shared interest in Christianity. Critical views are expressed but, even then, it seems often in the pursuit of dialogue and deeper understanding. Here, discussions include what it means to say that the Bible is inspired, or is God’s Word, and participants grapple with thorny aspects of the concepts of hell and sin. Where this brings disagreement, it does not reflect the aspects of incivility that I identify in chapter one. The shared behavioural norms, focus on Christian religion and moderation are likely important here. Indeed this focus on religion as a subject makes space for the second motif — dialogue around Judeo-Christian church and Bible history, such as the council of Nicaea and festivals including Passover and Easter. These do not always indicate a participant has personal faith but the sharing of views and historical information transcend that need and afford cohesion in relation to the group’s shared aims and, to the limited

degree that can be inferred, some formation of social relationships through sustained dialogue.

Eschatology and theology more broadly also figure in discussions of Revelation and the rapture (the third motif), though without the approach of 'proof texting' that Howard (2011) identifies as less salient to participants in Christian OSNs that do not take a fundamentalist approach to faith. Discussions reflect an inquisitiveness around God's character and the substance of divine Mosaic law. Again here, there is a seriousness in engagement seen perhaps less frequently in the 'talk' groups, reflecting why 'soc' groups were more likely to be accessible at sites who retained them while dropping access to talk. This perception of quality may of course also inform behavioural norms within the group, since people approach it with expectations of more substance and less noise in their interactions.

Moving from the more cerebral to Christianity as a lived religion, the fourth motif identifies how interactions often focused on practical aspects of Christian life. These include talking about the importance of prayer, correct approaches to prayer, and encouraging others to pray. Elsewhere, dialogue focuses on women's lives in Christian families and the church. One message, part of a longer thread not captured in the dataset (recall the UTZoo archive is fragmentary), discusses whether biblical authority and the Holy Spirit might licence women to take leadership roles. Another discussion thread focuses on married women taking on work and careers. Participants wrestle with changing behavioural norms in society, the authority of biblical teaching, and their own reasoning as to what is practical and sensible. Written in the early 1990s, engagement on these topics reflects the climate of contestation (still ongoing in some contexts) around the role of women in leadership within and outside the church at this time. This shows the potential usefulness of OSN archives for accessing popular discussions around contentious issues.

Evangelisation and expressing devotion, the fifth motif, are other practical aspects of Christian life for some participants. Here, people share Christian poetry and other creative writing, which reflects their expectation that such religious expression is acceptable within the soc.religion.christian network. This is notable since it is less dialogical than the typical focus on debate or conversation in many other Usenet OSNs. It is indeed acceptable within the aims of soc.religion.christian, considering a moderator's message from 1989 which expresses how messages may focus on any aspect of Christianity, thereby licensing the publication of personal, devotional text that focuses on Christian religion. Other expressions of faith are more evangelistic, though some are presented in poetic or another creative

written form. Some carry a sense of urgency in light of the fleeting passing of time and/or the need for salvation. They encourage the living of life with intention to do good and awareness of the Christian narrative of salvation. Still other dialogue grapples with the need for evangelisation and how/whether Christians should be involved, which intersects with the extent to which Christians believe they should be active within the world. There is a slight 'meta' perspective here because when such messages are posted exclusively to soc.religion.christian, so not crossposted to other groups, those Christians are communicating within a discrete social network, albeit one that newcomers may observe and to which they may contribute, if their messages pass moderation. This reflects how some participants do not use the platform to evangelise to others, instead focusing on their social connections with other coreligionists. Others, however, are less reticent to evangelise via OSNs.

The sixth and final motif relates to participants using the OSN to send messages about (1) religion as a phenomenon/activity and (2) Christian religious groups. This reflects participants' interest in a range of Christian denominations and sects, such as Wesleyan churches and Watchman Nee, and interest in questions such as whether religion is a universal characteristic of human society. Here, the diversity of religious authorities to which participants refer exposes a way in which topics of discussion may limit cohesion within an OSN if a person writes from a confessional perspective and relies on Christian religious authorities — such as leaders and texts — which are not recognised as such by others. Since soc.religion.christian is moderated uncivil responses may be vetted, but those posting may be met with silence. The UTZoo archive's limitations make it difficult to test this, but my chapter one argues that having a shared religious aim in discussions can indicate cohesion and diversity of religious authority and interest may limit that at some times.

Behavioural norms established within a group can shape the subjects and topics of interactions. This may be enforced by moderation but also through responses (or distinct lack of responses) made to those who post messages which are outside the scope of an OSN. For example, within religious OSNs belief in God's existence (or the reasonableness of holding such a belief) or the authority of certain teaching or tradition may be axiomatic. The body recurs as a locus for such discourse about religion. This appears between coreligionists engaging about the use of their bodies for religious observance and practice, and critics challenge religious views on abortion and LGBT+ lives.

In addition, the structure of engagement — whether it is dialogical, a large group conversation, or more informational and laconic — may be influenced by the title of the OSN and its location in Usenet’s hierarchy (for example within the ‘talk’ or ‘soc’ hierarchies). Once established, new participants may be socialised into these behavioural norms and they are then perpetuated over time. However, the effects of behavioural norms on the content and structure of interactions may be tempered by the norms of participants’ wider social milieus. I observed a recurring rationalised tone, critical argument, and focus on epistemology, argument, and reason. This may be understood considering the professional research environments in which early Usenet network and access were embedded, and the strength of this influence on Usenet behaviour is likely to diminish over time. ‘Eternal September’ (pp.42-43) is an exemplar of this, in which the task of socialising annual cohorts of new students into Usenet OSNs was replaced by the infeasible task of socialising a rapid growing and diverse public user base.

This summary identifies differing structures and substance of interactions across these OSNs, so far as is possible within the UTZoo archive’s limitations. This provides context for the next section, in which I will identify patterns of interactions between religion focused OSNs and others. This will show how OSNs’ differing shared aims and behavioural norms are associated with different structure of social interactions *beyond* their boundaries.

4.3 Interaction

Here, I present social network graphs which illustrate the connections formed between certain Usenet OSNs when messages are crossposted, i.e. when participants choose to send a message to multiple groups simultaneously. This is often because users want to share information or spark discussion with or between two or more OSNs because the subject relates those groups’ focuses. There were some restrictions such as moderation (where present) and, particularly in Usenet’s early years, limited access to certain OSNs at some locations. These patterns of crossposting often reflect the subject matter focuses and behavioural norms established within groups. Looking back, these graphs add to my above analyses of subjects discussed within Usenet OSNs by showing how those focused on religion connected with others, illustrating the reach of religion discussion into other areas of discourse. Looking forward, these graphs contextualise my subsequent analyses (below) on how ways in which cohesion and incivility manifest around religion discussion.

In these social network graphs, each node (often referred to as vertices in social network analysis (SNA)) represents one OSN and is labelled with that OSN's name. The lines (or edges in SNA literature) indicate that a connection was formed between the two OSNs that the line connects within the timeframe specified in the note under the graph. The lines are in a scale from light grey to black, with darker lines appearing thicker. These darker lines indicate a higher number of messages were crossposted to those groups compared with other crossposting connections within the timeframe analysed. However there is a limitation, since the UTZoo dataset contains only fragments of the early Usenet, the darker lines indicate crossposting patterns which *are* present in the specified timeframe but mostly likely others formed which are not captured here. To be clear, I have no reason to think that messages captured in the UTZoo archive were posted to *additional* groups not captured in my dataset and graphs. The issue is that many messages were not captured in UTZoo to begin with. Therefore, my discussion of these graphs focuses on what *is* visible and do not infer that the absence of certain connections between groups within the dataset indicates the absence of crossposting at the time messages were sent. I now present extracts of my social network graphs, with each subsection focusing on connections formed between certain religion focused OSNs and others. I focus specifically on net.religion groups first, followed by talk.religion, soc.religion, and finally soc.culture. This reflects the structure of my previous section.

4.3.1 net.religion OSNs

As indicated above, net.religion was the first religion focused OSN on Usenet and initially had no subgroups. It therefore included discussion on a full breadth of topics relating to religion and this is seen in the breadth connections established to other subjects in the second half of 1983.

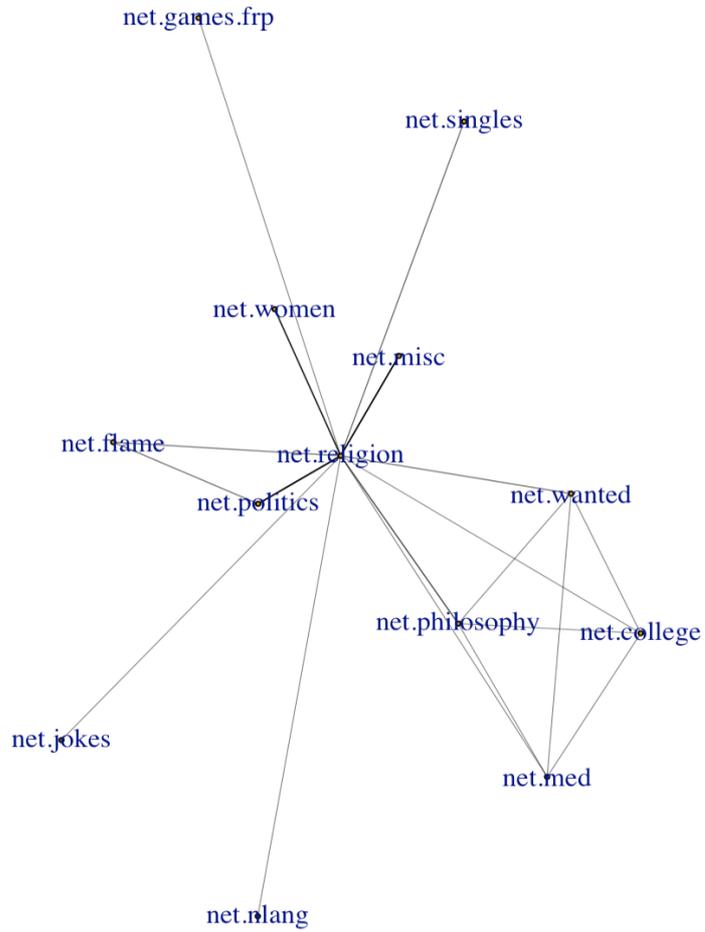


Figure 2: connections established between net.religion and other OSNs from July to September 1983 as captured in my dataset

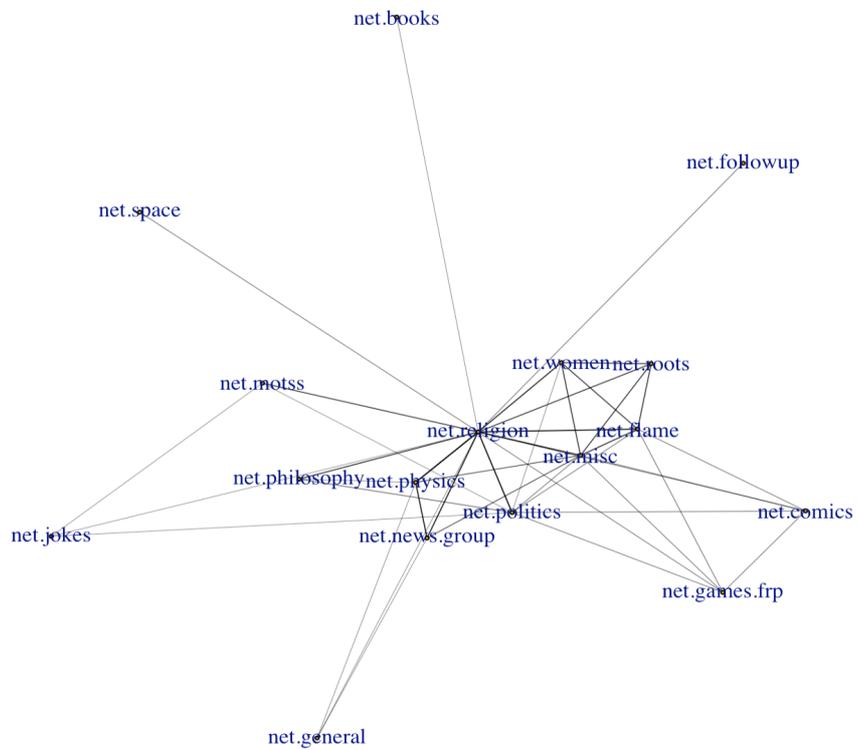


Figure 3: connections between net.religion and other OSNs from October to December 1983

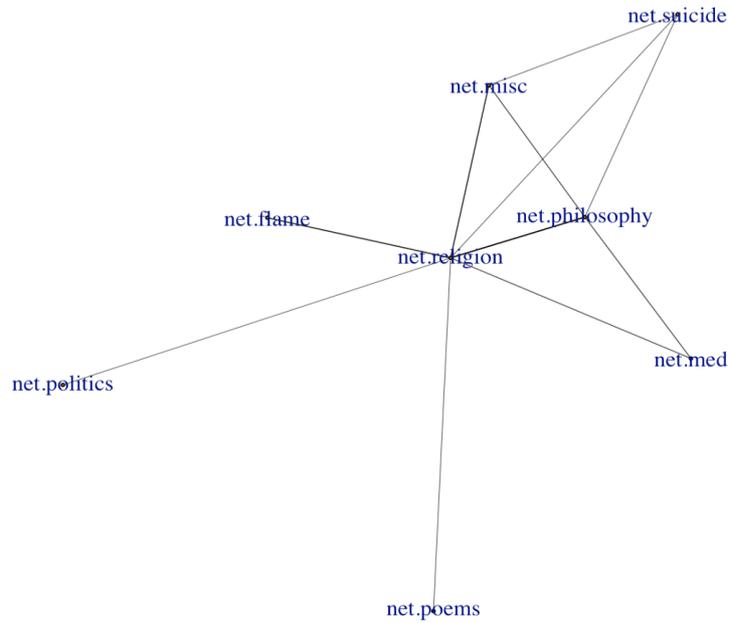


Figure 5: connections between net.religion and other OSNs between January and March 1983

As I summarise above, some interactions on net.religion.christian indicate a shared foundation for belief, whereas other interactions indicate contestation and/or scepticism. This is indicated by interactions in 1986, such as those between net.religion.christian and net.politics and net.abortion.

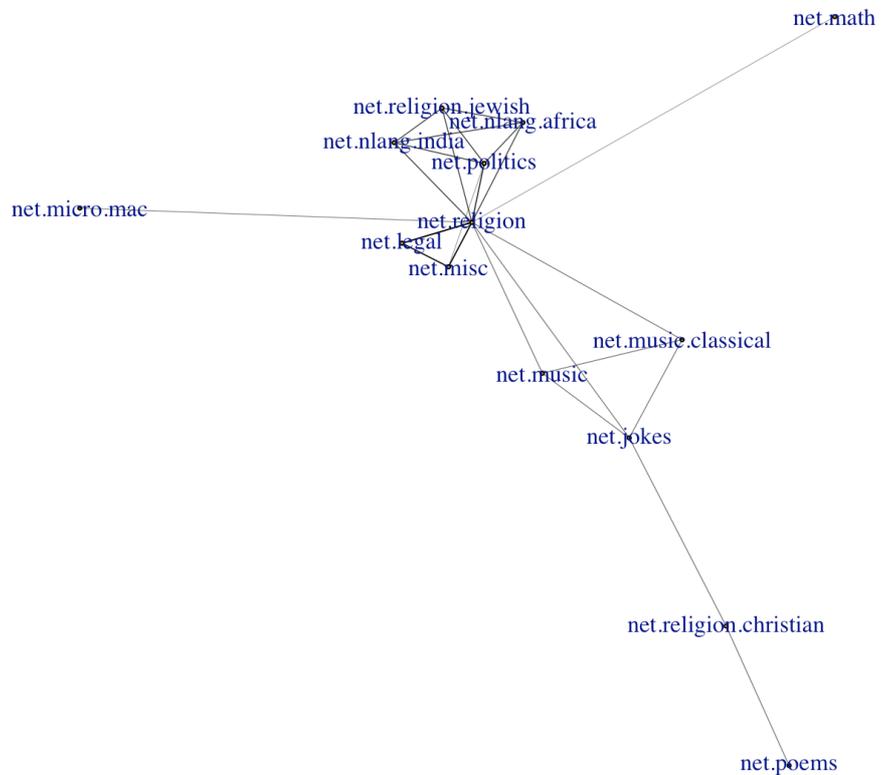


Figure 6: connections between net.religion groups and others in January 1986

The same quarter shows a breadth of connections with net.religion connected directly with myriad subjects. This started early, with 1983 showing a trend (which persists at times) of net.religion discussions intersecting with, for example, politics and physics, but those interactions taking place in separate discussions so, for example, politics and physics do not become linked. Such interactions also blossomed between groups in the talk.religion hierarchy and others, exhibiting diverse expressions of religion.

4.3.2 talk.religion OSNs

I summarised above (pp.150-152) how discussions of Paganism and other minority religions grew on the early Usenet. While a Paganism OSN was not formed when initially proposed, discussions are visible on talk.religion.newage by the first quarter of 1987. This follows the creation of the 'talk' hierarchy among others in The Great Renaming (see pp.32-33) in 1987. Nonetheless, access to these discussions was possibly limited as messages sent to groups within the talk hierarchy were not being carried across all computers connected to Usenet due to the high costs of transmission at the time (p.33). This is reflected in limited cross posting in the first quarter of 1987 though this had increased by the second quarter (and indeed third quarter), with use of internet protocol to send messages increasing the number of network routes and OSNs were available at more locations.

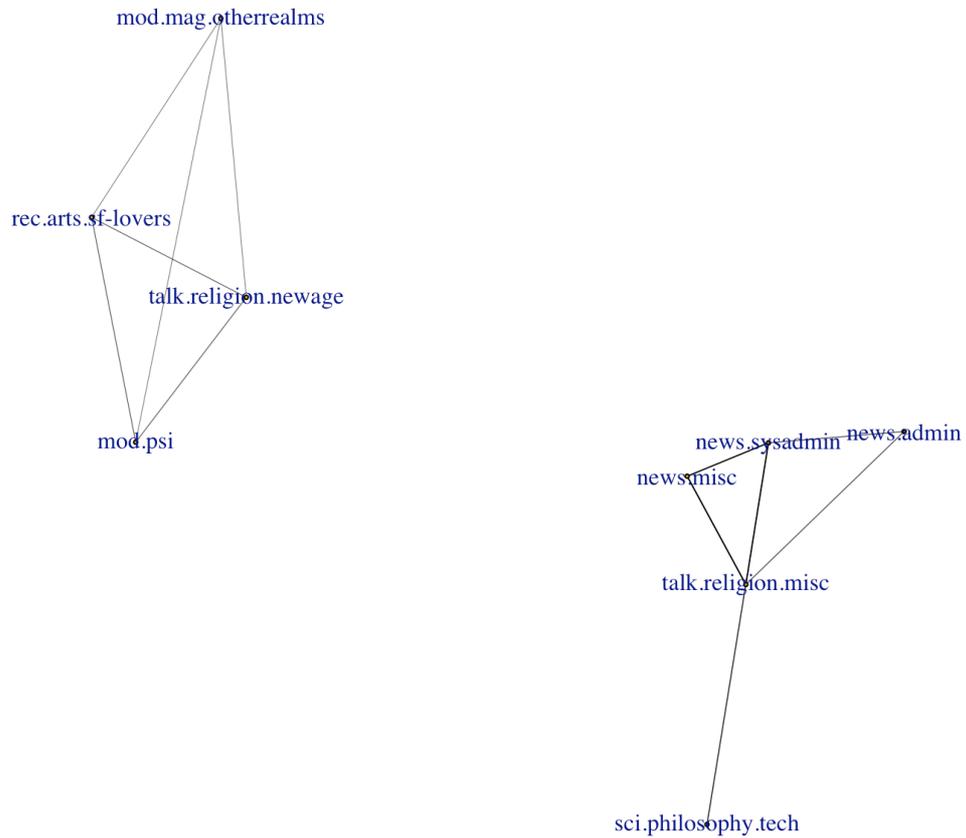


Figure 7: discrete clusters formed by messages crossposted between talk.religion OSNs and others from April to June 1987

The social network graphs for talk.religion groups also reflect the behavioural norms of more relaxed socialising compared with more focused OSNs/hierarchies. Nonetheless, in the second quarter of 1987 we see a connection to sci.philosophy.tech, which persisted into 1988, and talk.religion.misc and sci.misc had multiple messages crossposted in the third quarter of 1987. At this time, talk.religion.newage — a home for discussion reflecting the spiritual milieu cultivated within Silicon Valley and the west coast — had already established connections with other groups beyond religion reflecting its explorative and diverse scope. Indeed this dichotomy of interactions with science-focused groups on the one hand and diverse religious groups on the other, is seen in talk.religion groups' interactions with clusters of other groups in the first quarter of 1991.

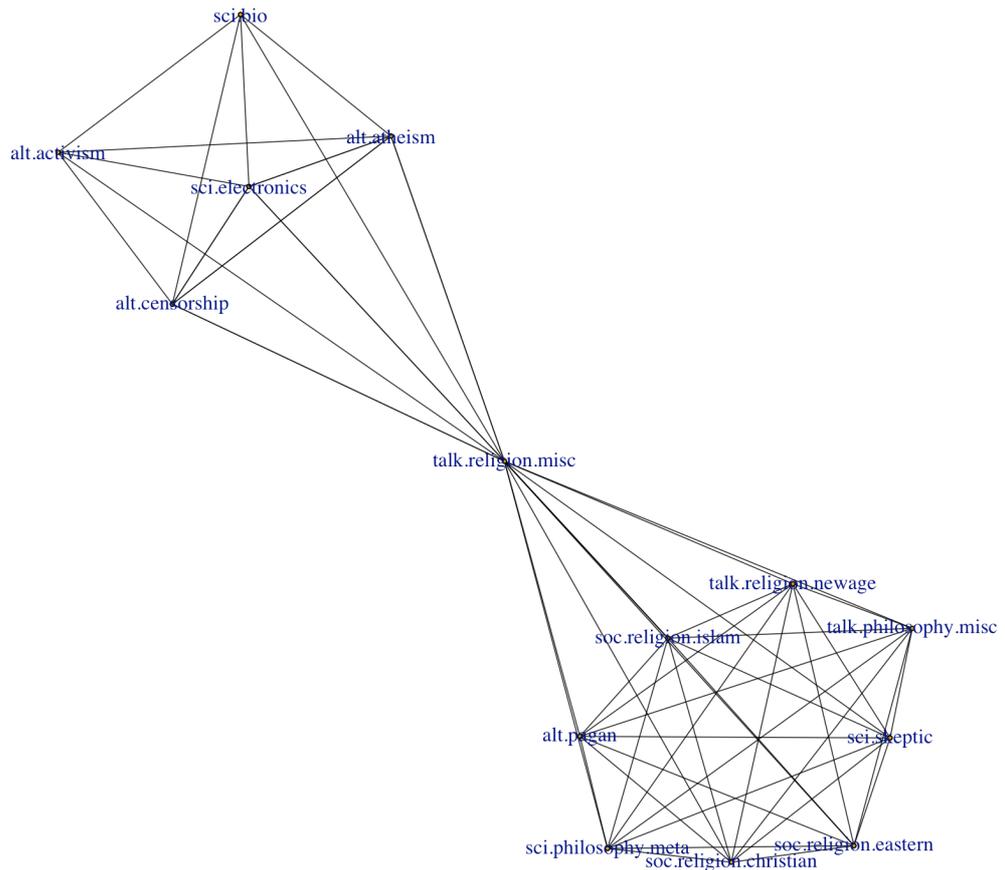


Figure 8: interactions between talk.religion groups and others in January to March 1991

Here, interactions within talk.religion did not lead to other groups of very different subjects being connected, though connections established are commensurate with broadening discussion beyond any one group's narrower focus. This befits talk.religion groups' diversity and social aims. Such discussions are not necessarily 'light' though. In my above analyses of discussions, I identified the motif of discussing religion in public life or society (pp.161-162), which can bring up topics about which people hold strong opinions and may relate to a person or religion's position and power in society. The social network graphs indicate interactions around such subjects in the connections visible between talk.religion OSNs and soc.motss (LGBT+ lives), soc.women, soc.culture.jewish and talk.origins by the fourth quarter of 1987, which was often used to discuss creation and evolution, an emotive and legally contested subject in American education (Numbers 2006; Larson 1985).

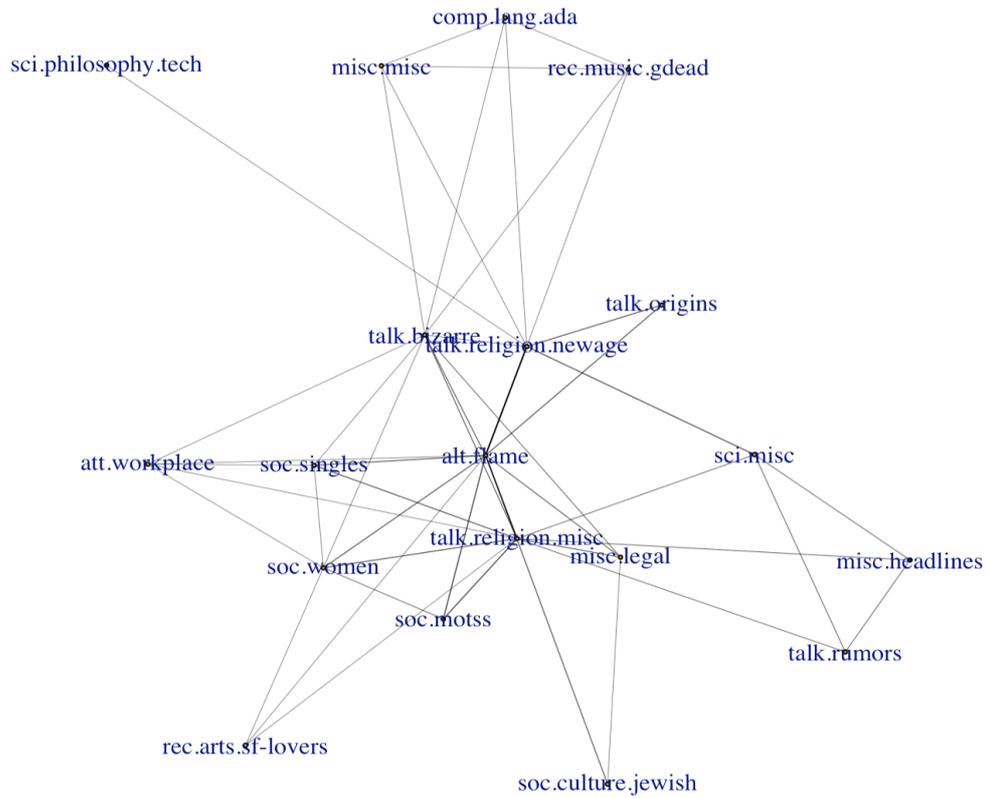


Figure 9: connections between talk.religion OSNs and others from October to December 1987

4.3.3 soc.religion and soc.culture groups

This leads to my discussion to ‘soc’ hierarchy groups, which similarly fostered deep and substantive discussions of religion at the intersections of many topics, but within more focused — and in some cases moderated — religion focused OSNs. The effects of moderation are seen within soc.religion.christian, a moderated group which saw less participation than groups focused on Judaism. I infer this from the less frequent presence of soc.religion.christian within UTZoo which may reflect Usenet more widely, since this remains the case in my fuller datasets that capture participation in later years (discussed in chapter five). This lower level of participation is visible in my social network graphs which identify no crossposting in 1990, some crossposting in the first quarter of 1991 but not in the second quarter.

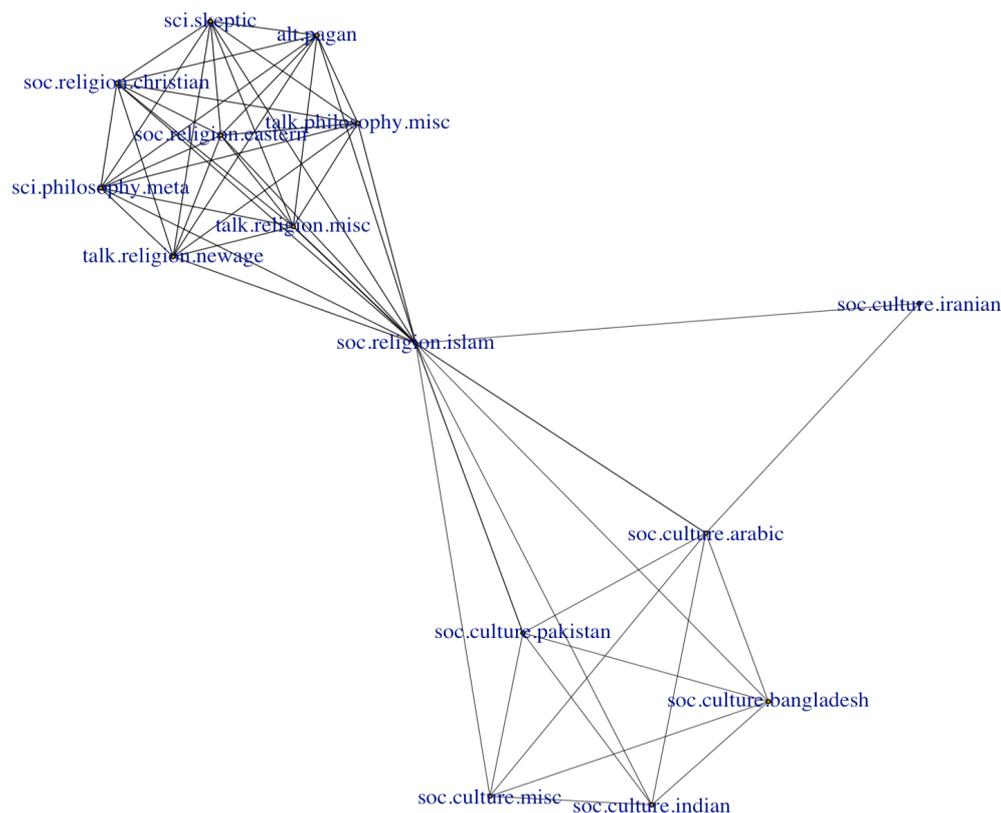


Figure 10: connections between soc.religion OSNs and others from January to March 1991

The apparently lesser participation in `soc.religion.christian` and crossposting between `soc.religion.christian` and other OSNs compared with groups such as `soc.religion.islam` and `soc.culture.jewish` may be for at least three reasons. First, space dedicated to discussing Judaism had been established early on Usenet (p.47). This may cultivate increased discourse on Judaism and visibility for the group since, as I discovered during my manual analyses for my 1,000 Thread Dataset, it is mentioned elsewhere on Usenet such as in lists of OSNs and how much they are used. This may increase the prominence of the group in a self-perpetuating cycle of increased visibility and participation. Second, unlike `soc.culture.jewish`, `soc.religion.christian` was moderated which affords the opportunity to ensure the subjects and tone of discourse reflect the group's aims and the behavioural norms established. This makes crossposting less likely since the subjects and tone may reflect other OSNs to which the message was crossposted and moderation may then reject the message. Third, the Christian remit of `soc.religion.christian` may at times be interpreted more narrowly than Judaism and Islam groups which, while also focusing on religion, embed religion discussion within broader discourse on other personal and public aspects of Jewish and Islamic life. This is reflected in the first quarter of 1991 (above). Here, we see `soc.religion.Islam` crossposted with `soc.culture.arabic` and `soc.culture.iranian` among other groups. This builds on similar connections visible in 1990. By contrast `soc.religion.christian`

crossposting in the first quarter of 1991, at least as indicated in the dataset, was concentrated around groups that focused on discussing or contesting *religion*. This reflects the subjects of discussion within soc.religion.christian which I summarised in the previous section, including an emphasis on dialogue and understanding, history, eschatology, Christian denominations and more, among some discussion of broader aspects of Christian life and societies. The latter may have nonetheless been less central than in Judaism- and Islam-focused OSNs.

Connections between soc.culture.jewish and computing OSNs were visible very early, as in this network graph for 1987, and persisted while also broadening by 1988 to include OSNs focused other subjects such politics.

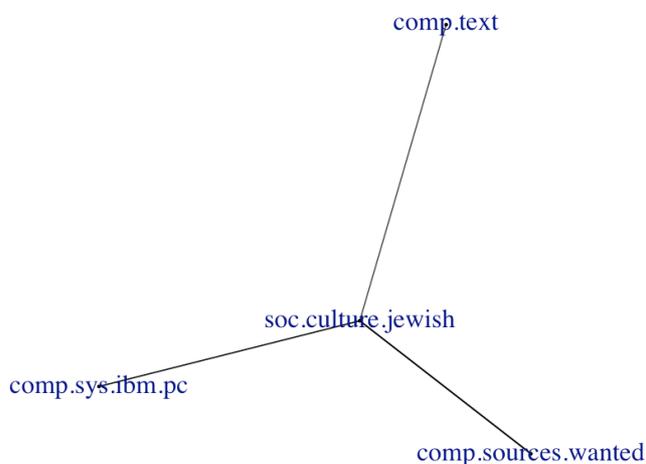


Figure 11: connections between soc.culture.jewish and other OSNs from January to March 1987

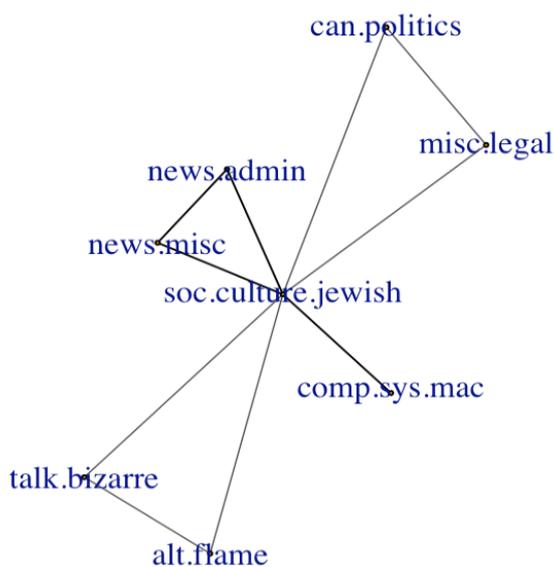


Figure 12: connections between soc.culture.jewish and other OSNs from January to March 1988

By 1989, connections between soc.culture.jewish and other soc.culture OSNs had been established.

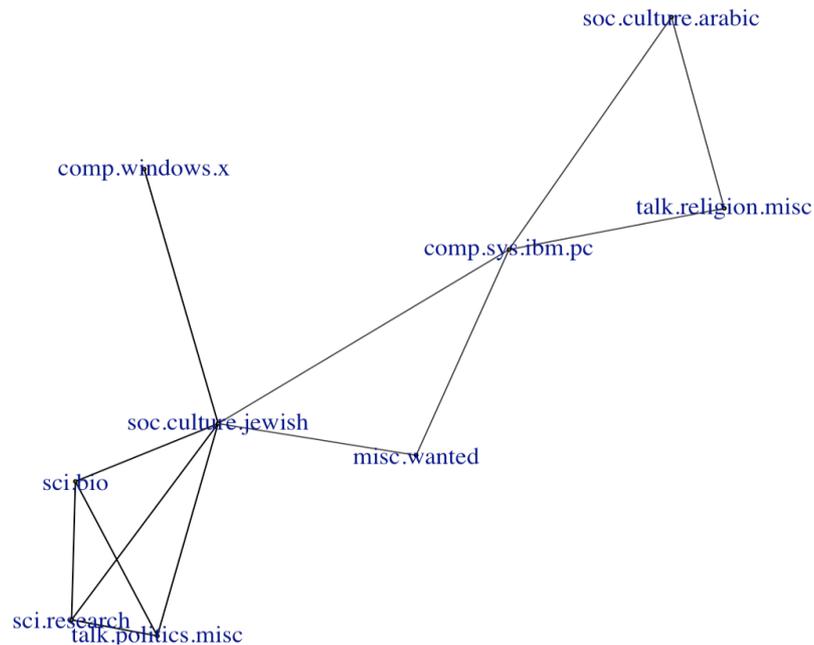


Figure 13: connections between soc.culture.jewish and other OSNs from January to March 1989

This perhaps shows soc.culture groups were better connected at this time, since soc.culture.jewish and soc.culture.arabic are now connected by comp.sys.icm.pc, which acts as a bridging node between them. These connections may have been due to participants' increased awareness of different OSNs and the breadth of discussion accepted on these OSNs making the inclusion of multiple groups in discussions acceptable. This possibility is affirmed by the inclusion of the diverse talk.religion.misc OSN in crossposting. Additional crossposting between soc.culture groups and soc.religion.islam is also visible by 1990. This further indicates how (potentially) cohesive discussions can be held across a range of OSNs with different, but overlapping, focuses and behavioural norms. The relevance of this overlap amidst the groups' diversity for enabling/sustaining crossposting is affirmed by the distinct triadic network established between soc.religion.islam and Indian and Arabic soc.culture groups.

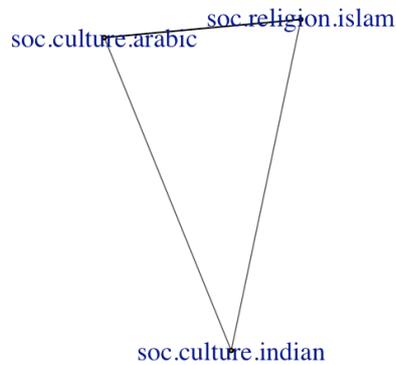


Figure 14: connections between soc.religion.islam and other OSNs from January to March 1990

In reviewing the patterns of network connections established between net, talk, soc.religion and soc.culture OSNs, we have seen that patterns of crossposting may be influenced by several factors. First, the breadth of the subject/focus of an OSN impacts the range of other OSNs to which it may usefully connect in pursuit of discussion in line with shared aims. The above example of soc.religion.Islam connecting with Indian and Arabic groups indicates this. Second, the hierarchy within which an OSN forms influences its behavioural norms, with groups in the talk hierarchy experiencing more relaxed and open discussion in myriad subjects and consequently connecting to a great variety of other groups. Third, moderation likely affects the acceptability (even the possibility) of crossposting, since moderators may manage boundaries around their OSN tightly in relation to subjects discussed and the groups with which their OSN connects. The limited crossposting that includes soc.religion.christian may reflect this. Fourth, events and discourse in the wider world may influence patterns of crossposting as online discussion is one part of participants' social lives. Crossposting which includes talk.origins is notable here, since discussion of creationism and evolution has been prominent in legal, educational and science communication in America.

4.4 Cohesion and incivility

These patterns of crossposting influence cohesion with social groups. In chapter one I argue that various factors indicate cohesion and incivility. The fragmentary nature of the UTZoo archive limits the extent to which these social processes can be observed in my dataset for Usenet's early years. Nonetheless, certain factors are visible. For cohesion, I can search for religion discussion as this reflects the religion-focused OSNs' shared aims and indeed my UTZoo dataset is a predominantly religion-focused subset of the larger UTZoo archive. I can then also see where such messages receive replies. I cannot assume messages

without replies indicate a *lack* of cohesion as those replies may simply not have been included in the dataset (my analyses of later archives in chapter five afford more robust analyses here). For incivility, I can infer the presence of uncivil messages which use antagonistic language and look for any flooding of messages that would disrupt participation.

4.4.1 Cohesion

Singleton messages	Messages in multi-message threads	Total
15,612	24,569	40,181

Table 3: the number of messages within my Message Dataframe from 1981-1993 which did and did not receive replies, so far as is possible to infer

The above table indicates that 61% of messages included within my dataset from years 1981-1993 are part of discussions, or at least threads in which at least one reply was received. There are two caveats. First, very early discussions may not have used the convention of prefixing ‘Re:’ before the subject line in replies, which I use to detect messages which are replies to existing threads. This could lead me to understate the proportion of messages in multi-message threads a little. Second, it is possible the dataset curator saved more longer threads than singleton messages since those threads would be more visible — and perhaps more interesting — to the curator. Nonetheless, this implies some density of social interaction as a proportion of religion-focused messages were interactions with preceding messages. This perhaps also reflects the early time period in which there was less spam than in later years; participants came from demographics particularly committed to engaging with — and contesting — intellectual topics. There was also less traffic online so more chance of messages being seen. This could benefit the signal to noise ratio, which is one factor that indicates cohesion (pp.57-59).

The social network graphs I produced also indicate that OSNs in different hierarchies and with different behavioural norms (sometimes informed by their hierarchy location), exhibited different degrees of density in terms of their connections to other groups. The SNA indicates that net.religion groups experienced early, persistent, and growing density of connections to other groups, so far as is visible within the UTZoo dataset. This is likely related to factors I summarise above. Specifically, these include: the breadth of discussion topics accommodated within net.religion; the number of connections to other OSNs (focused on different subjects) that this breadth affords in the course of interactions; and the early

founding of net.religion made them prominent within the (then) small but growing online social worlds, which in turn afforded them prominence. Combined, these factors provided a context in which net.religion groups' centrality in social networks — by virtue of being connected with other multiple groups over time — could grow.

The talk.religion groups similarly fostered an openness and breadth, at times irreverence, of discussion, which helps to cultivate sustained crossposting between different groups. This can indicate one facet of cohesion, insofar as the groups become embedded in a network of OSN discussions which spur interaction and, potentially, persistent participation. Unfortunately, I cannot identify any one person's persistent participation due to the limitations of the UTZoo dataset (I explore persistent participation in chapter five). Further, talk.religion groups sometimes become part of discrete clusters of OSNs in which a series of messages is crossposted to a large number of groups. Such patterns of crossposting are visible in figure 9 above. In this figure, talk.religion.misc is connected with two separate clusters of OSNs across a three-month period. These two clusters indicate distinct discussion focuses; the first reflects a more scientific and critical perspective to discussions and the second includes groups which reflect a broad range of religious traditions. In both cases, but particularly the latter, cohesion is indicated by the density of the social network formed between the OSNs by virtue of persistent posting as interactions ensued across the period. Of course *other* such interactions may also have occurred which were not visible in the dataset, but I can argue for the presence of some cohesion based on what we *do* observe.

As summarised above (p.151), net.religion.jewish was the first religion focused OSN to be created. This may help explain why soc.culture.jewish connected to computing groups early and persistently from 1987 when the group was formed. This intersection reflected the interests and demographic of Usenet's early participants. These connections continued and other discrete clusters of discussion emerged, such as between October and December 1987:

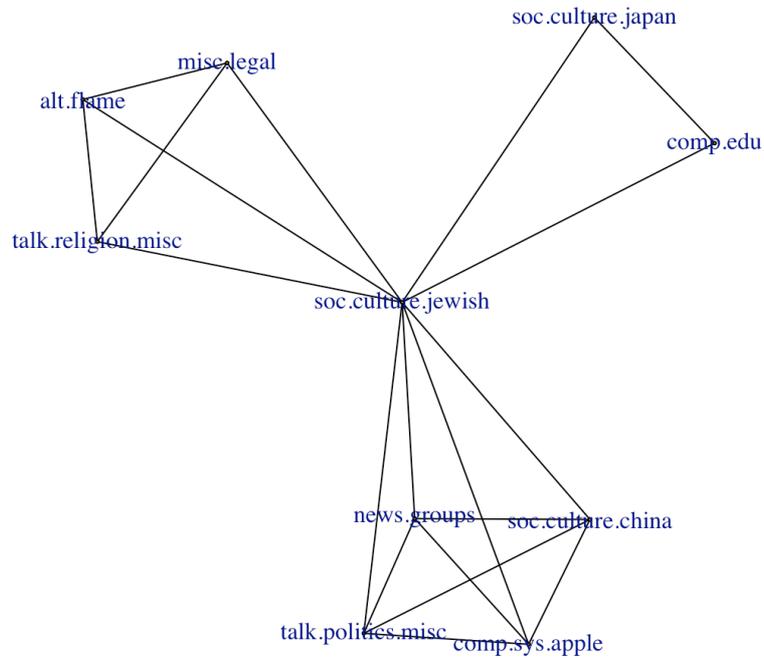


Figure 15: connections formed soc.culture.jewish and others between October and December 1987

On my observations, there is a less distinctively *religious* tenor in the connections established, which reflects the diversity of religion discussion within soc.culture groups, even where they relate to religious groups, such as soc.culture.jewish. Indeed, connections between soc.culture and soc.religion OSNs often involve groups which focus more broadly than, say, the debate of religious propositions which can animate narrower discussions of Christian religion online. Such a connection is seen in the third quarter of 1990.

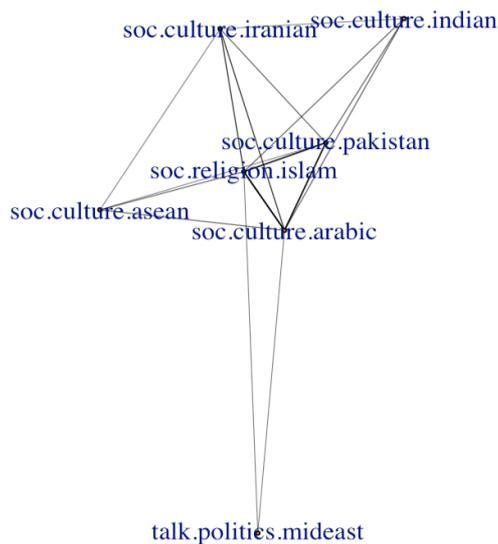


Figure 16: connections formed soc.religion groups and others between July and September 1990

These connections are denser than the modest triad formed between soc.religion.Islam and Arabic and Islamic soc.culture groups in the first quarter of 1990. However, in my observations (and noting the limitations of my UTZoo dataset), connections such growth in density does not occur with soc.religion.christian, save for that forum being involved in highly crossposted interactions in the second quarter of 1991. This may reflect how soc.religion.christian was moderated so had tighter boundaries around crossposting and acceptable content than some other groups. Also, although the group afforded discussion of all aspects of Christianity, this perhaps did not result in discussions on the breadth of geopolitics and other aspects of life that, for example, Islam may when intersecting with discussion of Near and Middle Eastern countries. Here, I have in mind the diversity of connections between soc.religion.islam and other OSNs in the second quarter of 1991.

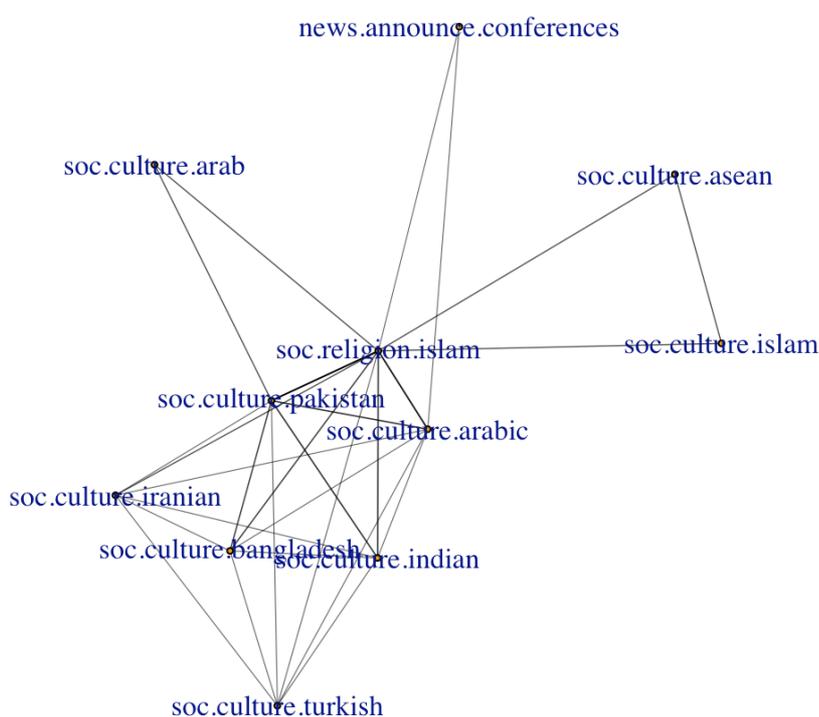


Figure 17: connections formed between soc.religion OSNs and others between April and June 1991

4.4.2 Incivility

In this subsection, I incorporate findings of my manual review and my computational analysis that identifies potential flooding (spikes in participation which are in fact high numbers of messages disrupting on-topic discussions) which may, on inspection, indicate incivility.

From my manual analyses I note that, at times, critical messages from early participants make space for disagreement by asserting counterpoints to one another on topics including theology and theologians, such as Aquinas and arguments for the existence of God, Satan and suffering, the punishment of hell, Israel and Palestine, and ethics, such as how one grounds a moral framework without God and violent historical actions of the church. This reflects Usenet's demographic of scientists and academics, often maintaining an informed and critical tenor around their discussions, with disagreements characterised by advocating reasoned points of view, even if expressed with the sentiment of deeply held conviction.

At other times, vulgar language is used in relation to, among other focuses, individual participants, gay people, religious belief, and antisemitic language is present too. Here, participants sometimes respond to this dialled up rhetoric with antagonism of their own through their language or shouting through block capitalisation. This enflames arguments and detracts, it seems, from cohesion since opposing views are no longer being articulated in a collaborative way to extol and educate, instead becoming adversarial. Here respondents might say others are 'attacking' them and respond with a vigour that reflects their perception of needing to defend themselves. At this point, cohesive accommodating of disagreement can recede and incivility through antagonism increase.

Participants as early as the first half of the 1980s anticipate the potential for inflamed responses and may ask for 'no flames' in response, setting an expectation that the discussion will remain collaborative. Indeed the counterpoint also occurs, with messages designated as 'flame' and then proceeding to be antagonistic through, for example, offensive assertions about a previous contributor and using a sarcastic tone. This bracketing out of a message as flame marks it out as distinct from the general tenor of messages. This expression of boundaries also occurs where participants assert that messages aren't suitable for a certain OSN due to their content.

Respondents nonetheless defend themselves, such as when criticised for their unthinkingness toward religious belief or behaviour, or defending their religion (as happens regarding Catholicism and Paganism, for example). While sometimes antagonistic, for some respondents it presents an opportunity to explain and frame the disagreement in a way that may defuse. This shows the power of individuals to be reflexive in how they shape the future course of discussion.

For some participants there is a skirting around a more antagonistic tone while remaining focused on criticising Christians and the perceived poverty of their grounds for belief.

One feature is common across the diversity of these discussions: the ability to reply to portions of messages at a time. When incorporating a previous message in a reply, the previous message is indented:

- > As in this text,
- > previous messages are prefaced by arrows

And participants can respond to specific elements, such as adding a line of their reply between two indented lines. This affords close and careful dissection of arguments line by line, such as theological arguments. It is also used for flaming by responding in vulgar or otherwise uncivil ways to individual portions of messages – I noted this regarding, for example, Palestine and discussions of gay people. These interactions become staccato and enable both careful disagreement and incivility.

Humour is also used to express sentiment, whether levity or vulgarity (so incivility). These posts are often short and less dialogical. Consequently they are less amenable to developing argument whether cohesive or uncivil, so are likely to be lower participation discussions and, as a result, contribute less to developing cohesion. These posts do not occur frequently, for example approximately eight times in my sample of 706 posts manually reviewed from 1981-1993, but are distinctive in having these characteristics.

I also noted the topic of discrimination in messages, which is to say participants discussed discrimination. This occurs at the intersection of topics including affirmative action, perceptions that Christians can be antisemitic, discrimination in employment and some discussion around discrimination in having Usenet accounts withdrawn. Such discussions are unlikely to display incivility in relation to their topic, since they are discussing rather than enacting discrimination. The latter occurs, of course, such as in soc.culture.jewish as I discussed (Chua 2009).

In summary, the manual review shows that Usenet participants at this time often articulated disagreement in a way that furthered discussion, but were aware of the possibility of inflaming discussions through incivility (such as vulgar language). Indeed, some discussions on hot button topics exhibited this when discussing, for instance, Judaism or

other religions and homosexuality. Therefore incivility was certainly present in these early years on Usenet, but many participants avoided it through measured disagreement on a topic within religion of shared interest, creating social norms of collaboration in discussions and perhaps even expressing their desire to avoid ‘flaming’.

Additionally, I wrote and applied R code to identify the dates on which the most singleton, in other words apparently *unanswered*, messages were posted. This can indicate the flooding of a social network with irrelevant messages, which I explore in chapter one (pp.58-59). This may be done to reduce the signal to noise ratio and make dialogue impractical for sincere participants. The following table presents the top ten as a sample:

Date	Number of singleton messages
12 Jan 1990	39
16 Jan 1990	39
12 Feb 1990	39
9 Feb 1990	38
7 Feb 1990	37
27 Dec 1990	36
17 Jan 1990	35
8 Feb 1990	35
11 Jan 1990	32
15 Jan 1990	29

Table 4: top ten dates on which unanswered messages were posted 1981-1993, as reflected in the Message Dataframe

(Note – see below regarding possible imprecision due to the archiving method and a mitigating note about this.)

I then extracted message data for the top twenty-four dates, which was enough to identify dates on which more than twenty unanswered messages were posted. There remains some imprecision in the figure due to the manual curation of the UTZoo dataset — perhaps some messages received replies which are not captured. Nonetheless, flooding would be indicated if many of these unanswered messages were clearly off-topic for the OSNs to which they were posted. With this in mind, I wrote code to extract message data for all unanswered messages from those twenty-four dates for manual review. Specifically, I reviewed the keywords that I extracted, which serve as descriptors of each message, and the message subject lines. This manual review indicated that many messages were related

to religion discussion and were not likely to be flooding. Consequently, I did *not* identify flooding as an indicator of incivility within religion discussion on the early Usenet.

Why may flooding not have been present, when I identified its occurrence through my review of extant literature about religion on Usenet (see pp.58-59) I found a pertinent insight in observing that twenty of these twenty-four dates — and indeed all the top fifteen dates — were from 1990 and 1991, rather than dates from the 1980s. At these times, Internet access was still expensive which may have dissuaded flooding (as distinct from spam, which may bring its own financial rewards) but the number of messages sent across Usenet OSNs daily had grown substantially (see table 1). Given this growth, the likelihood of messages simply being lost among the volume posted increased; it became more difficult to be visible and cultivate engagement from others. In addition, the broadening demographic of participants perhaps reduced the average level of interest and engagement with any one post. Early participants — scientists and other researchers — may have been more likely on average to engage with intellectual debate around religion subjects than some later participants. It is of course difficult to generalise though, given the breadth of discussion topics I summarised in the previous section of this chapter.

4.5 Applied case study — religion and science on the early Usenet

I now present the steps I took, and what I found, when researching interactions at intersections of religion and science within the early Usenet from 1981-1993. I use the same UTZoo component of my Message Dataframe for this as in my above analyses.

It is important to recall that the archives used for my later analyses relating to the 2000s and 2010s (as presented chapter five) are fuller archives provided by a Usenet service provider (whose servers send and receive messages across the network). I used selected archives focusing specifically on religion OSNs and the OSNs with which they are connected by crossposting. Consequently, everything in the archives relates to messages in religion focused groups. With UTZoo, I identified and used keywords to extract messages likely to relate to religion discussion (pp.104-105). This resulted in a subset of the UTZoo archive generally relating to religion. Given this, for my case study of discussions at the intersection of religion and science, I had to extract a subset of my UTZoo subset that related to science discussion.

To achieve this, I used my GloVe word embedding language model built from my Message Dataframe (see pp.112-114) to identify words associated with science. I then used R to extract messages which include a subset of these terms, specifically science, scientific, physics, mathematics, and evolution, where my analyses identified these were used prominently or distinctively (see TF-IDF on pp.110-111). I also included the terms biology and chemistry. In total, I extracted 1,369 messages from my Message Dataframe. Although biology and chemistry were not identified via my word embedding as associated with the word science, they are often used to refer directly (of course) to scientific subjects and including messages that use these terms prominently or distinctively enhanced the breadth of my dataset. I excluded the terms philosophy, research, logic, and knowledge — which were also associated with the word science — since I argue they may be more likely to be used in discussions unrelated to science so could increase the risk of I correctly including messages on topics unrelated to science. While appreciating the term evolution may be used in other contexts, since my the UTZoo component of my Message Dataframe focuses on religion discussion from the 1980s and early 1990s, I considered that the term evolution may be used frequently in relation to debates about evolution and creationism. This proved to be the case and I focus on that subject presently.

4.5.1 Cohesion in science and religion interactions

The earliest connections visible in my dataset between religion and science OSNs occurred in 1983. By the end of the year, net.religion had connected to physics and space groups and then, by 1984, a maths group. Then, net.religion.jewish had connected to net.astro and net.math by the first quarter of 1985. These early connections reflect early participants' interests, when taken together, in both (computer) science and its applications and religion. In particular, discussions here reflect the rationalised tone taken on early Usenet, often focusing on more philosophical and evidence-based approaches to religion, its history and propositions. This contrasts with the notion of lived religion, which focuses on every day and embodied experiences. Two pertinent examples illustrate this. The first is a discussion of the age of the earth and moon in relation to creationism. This challenge, at times made by young earth creationists, posits that the earth and moon are not ancient as one would expect far deeper sediment on the surface of the moon if that were so. This argument has an immediate attraction for some in the context of a pithy debate in front of a lay audience and requires rebuttal using naturalistic, perhaps uniformitarian evidence. The second relates to ontology, with an exploration of how an ontological argument expressed through mathematical notation may be used to affirm the existence of God. Both of these examples

illustrate the early intersections of science and religion on Usenet in light of its demographic's interests but are distant from the everyday experiences of lay religious people in the Abrahamic faiths to which these discussions relate. These early discussions are, then, cohesive as they reflect the shared aims and behavioural norms of the groups to which they were posted.

Connections between religion and computing groups flourished with `soc.culture.jewish` connected to `comp.sys.ibm.pc` by the first quarter of 1987. Soon after, `talk.religion.misc` and `soc.culture.jewish` connected to `comp.sys.apple`, `comp.sys.cbm` (Commodore Business Machines — a major 1980s manufacturer) and `comp.sys.mac` by the third quarter of 1987. Among these discussions are practicalities of using computers for non-Roman script, when computers on the network had been designed and setup in America for an English-speaking, writing, and coding community. Again, these conversations were cohesive as they represented an intersection between the lives of religious participants and computing OSNs. There were fewer such connections between computing and religion groups by the fourth quarter of 1987. This may reflect the changes in Usenet's hierarchy, which are relatable to online social networks today. Following the establishment of eight main Usenet hierarchies such as 'soc' and 'talk' (see pp.32-33), discussions took place across a wider range of discussion groups, with the potential for each to focus on narrower subjects or subsets of social groups. For example, within the talk hierarchy, `talk.religion.christian.jehovah-witness` was formed to focus on this narrower topic within discussions of Christianity in general. Such increasingly discrete groups perhaps limit the number of ways in which multiple groups may naturally overlap and perhaps decrease crossposting, until such a time as participation had grown to the extent that there were enough participants to make discussion viable at myriad niche intersections between subjects (as in the long tail of the internet, Anderson 2004).

This is perhaps evidenced by observing two ways in which connections *were* still established between religion groups and others around 1986 and 1987. Both reflect interests we would expect to flourish around a scientific/research community, part of which has imbibed Silicon Valley culture. The first is the connections established between `talk.religion.newage` and `sci.philosophy,tech` by the fourth quarter of 1987, and the connections between `talk.religion.newage` and `sci.misc` in the first quarter of 1988. I infer from my dataset that these interactions were playful, explorative, and speculative, at times focusing on the nature and limits of science, experience, and ontology. They fit well into the openness of Valley culture at the intersection in time between counterculture, new religious

movements from Eastern traditions and the rapid computing and concomitant science developments of the 1980s.

There are connections around similar themes involving the eclectic talk.religion groups in later years. This reflects the eclectic nature of talk.religion groups, which affords a broad range of connections with others. In the first quarter of 1989, talk.religion was connected to a large cluster of other groups including science groups — astro, physics, space, psychology — and talk.origins was connected to talk.religion.misc. This reflects the theme of creationism I mention above, since talk.origins focused particularly on scientific and religious discussions around evolution and creationism (Fristrup 1994, 196). Separately, talk.religion.misc connected to comp.ai around discussions of AI and ontology and, again separately, to comp.sys.ibm.pc in a discussion around software incorporating religious texts.

Notwithstanding talk.religion OSNs' eclecticism, discrete clusters between talk.religion groups and others exist, indicating cohesion since not all subjects were being focused on in parallel, with the potential of this introducing noise and impeding the flow of focused discussion. For example, in the first quarter of 1991 talk.religion.misc connected to sci.electronics, sci.bio and others including alt.atheism, and connected in separate discussions to various religion groups along with sci.skeptic. This formed two distinct clusters of discussion, which are visible in figure 9 above.

Building on the above, and having reviewed the subjects of discussions where religion and science groups intersected on the early Usenet, I conclude that there are two prominent themes amidst the diversity of topics within interactions, and with a lot of breadth within both of them. The first is around evidence for the truth of religions' claims and interpretation of the Bible. Here, discussions reflect the rationalisation and individualisation elements of the secularisation paradigm (Bruce 2011, 27), which may be expected considering participants' professional work and social networks within research institutions, which I explore above and in chapter one. Message threads include discussions of proof texting, prefiguring Howard's (2011) findings (p.52 above), contradictions within the Bible, considering scripture changing over time, and other critical approaches to biblical texts. Here, Ecklund, Park and Veliz' (2008) research into the decline of religious belief among elite scientists from 1969 to 2005 is illuminating. Their review of literature situates this decline in secularisation theories' connection between secularisation and increasingly rationalised approaches to knowledge claims. Many of the institutions in which Usenet was accessible formed part of this elite

scientific world. My computational analyses and manual review of a subset of messages (from my 1,000 Thread Dataset) indicate civility is retained within the broad parameters of accepting candid and outspoken discussion. This perhaps reflects the rationalised paradigm, in which religion is being critiqued empirically and logically, rather than relying on emotive arguments seen on later OSNs and social media more broadly today. I explore the limited presence of incivility below.

I mentioned the second broad theme above and expand on it here. It concerns arguments about creationism and evolution. These interactions must be seen in the wider context of American pedagogy and science, which had been impacted by the promotion of creationism at a popular level. Public-facing organisations such as Answers in Genesis, the Institute for Creation research and Creation Ministries International reflect a broader movement which gained traction in contesting the evolutionary account of origins, the age of the universe indicated by cosmological science, and using such arguments to affirm a human identity which for example, denies certain LGBT+ rights. Many vectors are used to achieve influence and promote these activities, including political actions from state to school board level, the promotion of private school curricula, hosting conferences in churches, public debates with willing scientists, and outreach through print, audio, and digital media. Consult Numbers (2006), Ruse (2005), and Scott (2004) for overviews of the science and social aspects of creationism; see Ray (2012) for analysis of how young earth creationism contrasts with other Christian approaches the science and religion of origins; and see Batten *et al.* (2014) and Ham (2013) as examples of creationist literature aimed at a popular audience. Collectively, these texts indicate why, for scientists and educators, creationism may be seen as a challenge to science pedagogy and communication which has social impacts in terms of human rights and dignity, and scientific impacts. They show differences in how the Bible is weighed as a source of evidence for early historical periods and they indicate epistemological incongruence between natural scientific approaches which allow for longer ages and a blend of uniformitarian and catastrophic change in Earth's history, and creationism which challenges how science makes claims about an unobserved past.

Certain aspects of those arguments, such as those I summarise presently, manifest on early Usenet. This is to be expected given Usenet's demographic and network roots in scientific university and research institutions, and how online interactions form part of a person's broader social and discursive worlds. These include discussion of geology and stratigraphy, which are potent since creationism challenges dominant (then and now) paradigms which perceives that sedimented layers of strata are laid down over long time periods through a

mix of uniformitarian and catastrophic processes, with the fossil record reflecting these long ages. On creationism, strata are largely attributed to the catastrophic processes of the Noahic flood. Therefore, this element of creationism particularly challenges scientific theories of evolution and the development of Earth over long time periods and attracts attention accordingly.

Further, evidence in support of creationist and evolutionary models are discussed. This reflects my above point that early Usenet interactions around religion often focused on religions' claims or propositions and this may be particularly likely at intersections of science and religion, given the focus on evaluating empirical and logical claims within science. Creationist apologetics tend to point to data as being in support of their paradigm, without asserting proofs. This limits the conclusiveness of debates on Usenet around proofs, since proofs are beyond reach. It is possible that participants engaging here operate within physical or mathematical sciences that afford certain proofs, when defined appropriately. See Villani (2015) for an engaging discussion of the development of a mathematical proof.

There is also some discussion around the law and pedagogy in relation to teaching evolution and creationism in classrooms in the United States. This reflects the political aspect of creationism in which, from school board to state level, interventions are made to limit the teaching of evolution and/or ensure the inclusion of creationism alongside it. With roots in legislation that led to the 1925 Scopes Trial, in which teacher John Scopes was convicted of illegally teaching evolution (Larson 1997, 87-110), these interventions from school board to state level frustrate the teaching of evolution as science and the exclusion of creationism. The Arkansas Balanced Treatment Act of 1981 was contemporary to the early Usenet and mandated the teaching of 'creation science' alongside evolution. Here, creation science is positioned as an alternative scientific paradigm to evolutionary science. Political actions such as this, along with discussions of biology which are of course impinged in the classroom by legislation such as this, occur on Usenet.

4.5.2 Incivility in early science and religion interactions

I present here discussion of the limited and nuanced ways in which incivility occurs in early interactions regarding science and religion, focusing on my manual review and considering one thread that sentiment analysis flagged as indicating incivility.

Many of the discussions appear collaborative, with participants sharing views and advancing discussion through affirming or disagreeing. This reflects the demographic on Usenet at this time — primarily academics, scientists and researchers. Programming is sometimes used to express a logical sequence of events in relation to theology — setting out a scenario that the poster can then engage with. This draws in some discussion of suffering and salvation — perennial topics in apologetics and objections to Christian faith.

At times, heat rises in discussions, indicated by the use of block capitals and some negative sentiment expressed toward others. Nonetheless, this does not appear antagonistic, or not strongly so. Humour is similarly used to express rejection or criticism of the Christian faith gently, such as using arguments from a scientific paradigm in discussions of the nature of heaven and hell, and it does not bear the hallmarks of the intensity of antagonism that I observed in later archives, nor in my review of literature researching or documenting religion on Usenet.

Within the broad possibilities of discussing religion and science, discussions often focused on creationism. This is understandable in light of the advance of creationist groups and their advocacy for creationism in education. For some scientists, this brings religious doctrine into contestation with scientifically derived perspectives on origins, evolution and human identity. This can lead to attempts at boundary maintenance around what is perceived as legitimately science, and what is religion. While ebbing and flowing, the 1960s, 1970s and 1980s witnessed this contestation in education and evangelism/apologetics (Numbers 2006, 312-372) and Usenet discussions among scientists of the time reflect that, while often remaining civil. In summary, taken together, these observations suggest these earlier interactions, at least within the domain of religion, were less uncivil than later participants would experience into the 1990s and beyond, garnering Usenet a reputation for, at times, flaming and hostility.

Although incivility appears less common than on later Usenet, contrasting the observations I present here with data in chapter five, one message thread tripped the algorithm I configured to infer incivility (p.112). Though I do not rely on this in my analyses, on this occasion it was pertinent to my review. This related to a discussion of creationism and presented a more emotional tone than at least some other interactions. Here, one participant expresses incredulity at another for stipulating that other religious beliefs require proof in the same manner that is demanded of creationists. Posted to two religion OSNs, including a Christian one, and no science groups, it is foreseeable that demanding proof of

religious beliefs would elicit critical responses. This remains a science-related message insofar as it focuses on creationism, but the participant may have sought to apply the behavioural (discursive) norms of science OSNs, in which empirical or logical evaluation is foregrounded, to religion discussion. This was not successful. Here, then, criticism did not occur because disparate groups were connected, but it could be argued that it is because the epistemology of science OSNs was applied within religion OSNs. Beyond contesting creationism and evolution, my manual observations indicate that other subjects indicated some elements of incivility in an argumentative tone, where focusing on criticism of biblical inerrancy and taking a critical tone toward prayer. My other indicator of incivility, flooding, was not present and I consider reasons for this above. Nonetheless, the fragmentary UTZoo dataset means I cannot infer the absence of flooding from its absence in the dataset.

A few factors may help explain this limited presence of incivility. First, the demographic of early participants on Usenet, and the professional locations in which Usenet access was often available, encouraged behavioural norms that were established on the network. These were initially collegiate, with participants taking the position of scientist, researcher, or collaborator. Later demographics were less socialised into these behaviours due to their diversity of (non-)professional backgrounds. In addition, there were challenges to socialisation indicated by quantitative changes, specifically the increasing number of new participants each year, the rapid growth in participation in the 1990s and 2000s (supported by the often-decreasing costs of internet access), and the number of groups available for potential crossposting grew. The breadth of focuses/subjects within Usenet OSNs also broadened so was less focused on subjects aligned to the early, more professional participants. Lastly, the likelihood of crossposting to disparate groups increased as later Usenet software made it easier to find groups and send messages to them, such as AOL's software and Microsoft Outlook Express (p.40). This increased the chance of sparking incivility, reflecting my concepts of cohesion and incivility (see pp.58-61), by connecting people opposed to one another. The difficulty of socialisation was captured in the vignette of Eternal September (pp.42-43). Consequently, the presence of incivility was perhaps more likely, and occurred in greater volume, in later years. I explore these later contexts in chapter five.

4.6 Conclusion

The chapter researches some of the earliest online interactions about religion, and science and religion in particular, that are in the lineage of modern mobile internet social networks.

It shows the growth of religion discussion soon after Usenet was founded, identifying how the subjects discussed and behavioural norms exhibited reflect the group within which the interactions took place. Interactions between religion focused OSNs and others reflect the interests of early Usenet participants and the behavioural norms, such as 'talk' groups having at times broader connections reflecting the conversational tone. I identified cohesion since discussions were often aligned with the OSNs' aims and the conflicts captured in later archives are less visible, perhaps reflecting the professional institutions and users that comprised the early networks. Concomitantly, incivility was less evident than in later archives, perhaps for the same reasons and because Usenet was less broad in the early years, with OSNs and participants focused on fewer subjects, diminishing (not eradicating) the likelihood of conflict-sparking interactions across OSNs' divides. I found these observations played out in the case study of religion and science interactions, which also showed how online interactions provide insights into popular discourse around topics of the day. It is unsurprising that creationism featured so prominently at this time, with this cohort of participants. Taken together, the insights represent a new empirical contribution to knowledge as far as, I have found, these data have not been analysed in religion research to generate such insights about topics of discussion among these early online participants, or the social processes present in the interactions, which I have illuminated. The case study in particular shows how computational analyses illuminate social interactions about subjects crucial to identity, social inclusion, pedagogy, and science.

The research in this chapter also affirms the usefulness of this methodology for examining discourse on myriad subjects that were salient in the 1980s-2010s (the decades for which we have archives to varying degrees), adding to other methodologies for examining social impacts of religious groups or religion-related events. These analyses are infeasible without a computational methodology to extract, organise, and analyse these particularly cumbersome archives. Further, this chapter provides regarding the structure of Usenet OSNs (the hierarchy and sub-hierarchies formed and connections between them), some popular discussion subjects and the behavioural norms established. This provides context for observing ongoing impacts, continuity, and change in chapter five's analysis of later decades.

Nonetheless, the fragmentary dataset from this time period (1981-1993) which I used in the present chapter brings three limitations. First, I cannot infer the absence of flooding, since perhaps some highly uncivil interactions were not saved. Second, I cannot identify typical patterns of participation. Third, I cannot use machine learning to infer factors associated

with cohesion and incivility — such as topic of discussion, network size. I need fuller datasets for these tasks. This leads to chapter five, in which I use fuller datasets to analyse select Usenet OSNs in the 2000s and 2010s.

Chapter five — Digging Deeper into Data

Discussions of religion on online social networks (OSNs) were pervasive by the 2000s. OSNs helped new inter-personal relationships form, while also causing conflict with adversarial discussions between people of opposing views, sometimes sparked by world events such as the September 11 terrorist attacks. Since online and offline worlds form one social whole, examining these archives shines light on social interactions that, for some, will have been valued aspects of their social lives (cf. case studies in Campbell 2005, 80-105, which indicate the value some religious participants placed on faith-based online social interactions in the 2000s).

This chapter uses the methodology to examine select interactions on Usenet OSNs from the 2000s and 2010s, using archives from Usenet service provider Giganews, which are made available for research purposes. The strength here is that Giganews' position as a prominent and large service provider enables them to capture traffic conveyed across their servers. Usenet is decentralised (see p.28) so Giganews doesn't capture all Usenet OSNs. Nonetheless, Giganews can archive data in a much more robust manner than any end user, such as the curator of the UTZoo dataset used in chapter four. This enables me to undertake the full suite of analyses developed within the methodology to answer my research questions.

I focus on Abrahamic faiths as they had high participation and, of course, these faiths are prominent in the offline world. This increases the breadth of topics in interactions, since they can be guided by offline events, and means they could be populated substantially by people who are committed to these faiths, increasing the likelihood of these interactions being valued by their religious OSN participants.

The chapter has some similarity in structure to chapter four, helping me illustrate continuity and change between the earliest OSNs in the lineage of modern online discussions and the later OSNs that followed. As in chapter four, I present factual findings and then my interpretations. First, I use the archives to outline the terrain of religion discussion relating to Abrahamic faiths on Usenet OSNs by the 2000s and 2010s. This illustrates the ongoing influence of structural decisions and some behavioural norms from the early years. I focus on groups within the core 'soc' and 'talk' hierarchies, which typically had high engagement and therefore higher potential social impact in terms of forming connections between participants for cohesive or uncivil discourse.

I then present computational analyses of selected OSNs to elucidate characteristics of their social networks, focusing on quantitative data regarding the number of participants, their interactions, and patterns of participation. I reflect on why these OSNs exhibit these characteristics considering my previous findings from literature and research. I then broaden to consider interactions between these OSNs and others where this helps to explain these characteristics. I then gather these insights together and using machine learning to identify factors associated with cohesion and incivility within the dataset, thereby answering my research questions.

The chapter concludes with a case study, focusing on the soc.culture.jewish.moderated OSN. This narrower focus affords a more detailed analysis of historical trends in social interactions within an OSN. This is important and timely, given both the global reach and size of Jewish communities, but also the potential for divisions across religious or national identity. These divisions are indicated in Nagar, Hoter and Hasler's (2021) paper studying the outcomes of online interactions between Jewish and Arab students in light of Israel-Palestine relations. 'Othering' can occur, and conflict between people positioned in opposition to one another may take place, but the online interactions paved the way for some empathetic engagement reducing, for some participants, the perceived distance between themselves and the other group (Nagar, Hoter and Hasler 2021). My present case study contributes to understanding factors that are associated with potentially harmful discourse and those associated with healthy interactions.

5.1 Abrahamic faith OSNs

In the 2000s and into the 2010s, OSNs focusing on the Abrahamic faiths bear some imprint of decisions made in Usenet's earlier days. I discuss three aspects here, the first two of which being Usenet's interrelated *structure* and *freedoms*. Although many early decision makers may have been long gone, the structure of Usenet's hierarchy — with the 'big seven' top level hierarchies including soc and talk groups — persisted in its impact by structuring later discussion groups and influencing their behavioural norms. The same is true of the alt hierarchy which, as I note in chapter one (p.33), was formed separately. One clear distinction is with the remit of soc and alt groups, with the latter having an exceptionally broad remit and flexibility to accommodate discussion — and file sharing — on any topic. The soc groups remain more focused, as will be seen here, on discussion and articles on their specified topics, notwithstanding the diversity of contributions within them. Simply put,

alt groups garnered a reputation for sometimes illegal content and often dysfunctional conversations, or ‘unabashed anarchy’ as Frstrup (1994, 198) put it. Thanks to early participants’ actions to structure Usenet, participants approaching online interactions in the 2000s, often for the first time as internet access broadened, were met with a clear way of identifying religion-focused groups. These were often well-established and longstanding, which provided a readymade online network which they could join through reading messages and participating. In addition, as I explore in chapter four, the structure allows for the creation of groups within hierarchies, creating substantial sub-hierarchies which allow further groups to be created within. This allowed regional groups and groups focused on bespoke, often narrow, topics or demographics to flourish. These freedoms varied a little, with alt groups affording particularly high levels of autonomy, though indeed strong diversity can be seen in other hierarchies such as talk, which I review here.

My analyses for this chapter focus on archived OSN interactions regarding Abrahamic faiths (and some groups rooted in Abrahamic faiths) within the soc and talk hierarchies, including many smaller groups. The following list indicates the breadth of OSNs created within the soc.religion and talk.religion hierarchies, and is simply the sample that I analyse for my research:

Online social network	Approximate number of messages	Years spanned
soc.religion.christian	83,876	2003-2014
soc.religion.christian.bible-study	19,993	2003-2013
soc.religion.islam	49,416	2002-2015
talk.religion.christian	2,198	2003-2014
talk.religion.christian.anglican	390	2003-2011
talk.religion.christian.jehovah-witness	4,494	2003-2010
talk.religion.christian.protestant.adventist	510	2003-2012
talk.religion.christian.protestant.baptist	706	2003-2012
talk.religion.christian.protestant.charismatic	84	2003-2011
talk.religion.christian.protestant.episcopal	74	2004-2011
talk.religion.christian.protestant.evangelical	43	2005-2011
talk.religion.christian.protestant.lutheran	107	2003-2011
talk.religion.christian.protestant.mennonite	80	2003-2011

talk.religion.christian.protestant.methodist	158	2003-2011
talk.religion.christian.protestant.moravian	66	2003-2010
talk.religion.christian.protestant.pentecostal	207	2003-2011
talk.religion.christian.protestant.presbyterian	136	2003-2012
talk.religion.christian.protestant.unitarian	137	2003-2012
talk.religion.christian.quaker	46	2003-2012
talk.religion.christian.roman-catholic	8,156	2003-2013
talk.religion.christian.science	414	2003-2012
talk.religion.jewish.conservative	90	2003-2013
talk.religion.jewish.messianic	201	2002-2012
talk.religion.jewish.orthodox	122	2003-2010
talk.religion.jewish.orthodox.chassidic	75	2003-2010
talk.religion.jewish.reconstructionist	33	2003-2010
talk.religion.jewish.reform	78	2003-2011

Table 5: summary data regarding messages in select archives of soc.religion and talk.religion OSNs incorporated into the dataset

These figures indicate how the *amount of participation* within Usenet OSNs was influenced by much earlier decisions about structure and freedoms which I discuss above and in chapter one (p.43), and which took place in the years the UTZoo archive covers. Specifically, we see how the talk hierarchy's more conversational and social behavioural norms supported the creation of increasingly granular and smaller networks. The talk hierarchy did not embody the professional uses common to OSNs in the early Usenet's other hierarchies such as soc and comp. This supports the diversity of talk.religion groups indicated above allowing focused, smaller groups space to connect about their religion, or for others to connect and talk about them. On reviewing my second dataset which contains redacted messages from 1,000 discussion threads, I found some diversity among the subjects discussed within talk.religion though not so much granular focus as may be expected. Regarding religious content, I observed messages proselytising for Christian religion and possibly some prayer. I also observed messages about Christian history and saints and Christian theological discussion. Additionally, I read messages satirising or ridiculing religion, along with discourse expressing scepticism and reason. The latter occurred at an intersection of a minority religion based in Christianity and an atheism discussion group, reflecting the interests participants shared at this intersection. Some other messages were about how to talk with members of the minority group. Therefore, based on this random sample, discussion often reflected aims participants shared, either as part of

their religion (proselytising and prayer) or to discuss religion. Where relevant, this discourse indicated positive affirmation of religious authority, notably the Bible.

By contrast, the soc groups indicate further breadth in the subjects focused on, so far as I could analyse from the archives. Here, I briefly compare and contrast the subjects/foci of messages from three OSNs which are captured in my second dataset, starting with soc.religion.islam. While discussions in this group focus broadly on Islam, they do so from varying perspectives. Some threads focus substantively on historical evidence and events, citing authorities including the Qur'an or Hadith favourably and indicating the writer is a Muslim or is otherwise respectful of these authorities and the religion more broadly. Relatedly, I found some discussion of theological questions regarding the nature of Allah. Other messages discuss morality/behaviour in Islamic life. Still others engage with science, with participants perhaps writing from a position respectful to Islam but less obviously committed to the religion themselves. Such threads accommodate disagreement at times and indicate cooperative discussion. Some situate Islam in wider intellectual traditions of both theology and science, considering Islam's contributions or compatibility. These discussions move the focus from participants having *religious* aims for participation to *discussing* religion or indeed Islam more broadly in (geo)political discussion threads. Other threads take this discussion in more critical directions with varying degrees of civility and constructiveness. Here, interactions again discuss elements of Islamic practice and morality such as child marriage, teaching acceptance of other religions and jihad. Such messages indicate a more critical outsider position, so far as can be inferred from the language used.

The soc.religion.christian (SRC) similarly contains discussion threads indicating contributors are themselves Christian or are sympathetic to Christian religion. These include threads about biblical and subsequent religious figures and which indicate respect for their authority. Messages about the Bible and science at times likewise seek to integrate and avoid/resolve conflicts between these authorities. Other threads proselytise or celebrate religion by posting biblical passages on various themes, or support others in their Christian life and practice by encouraging them and responding to questions, and reflecting on doctrine. Some threads discuss theology or religion without indicating deference to religious authority or indicating their own religiosity, at times considering the boundaries of human knowledge. Some such threads focus on baptism, salvation, and identity and can have a tendency to mark boundaries between those who are perceived by contributors to be Christian and those who are not. Such discussions could impact the degree to which contributors find affinity with each other, which in turn could impact patterns of participation

since some may feel less welcome, though my present analyses do not afford confirmation of this. Other threads are critical of Christian practice such as financial giving and leaders' behaviour. I saw fewer messages here which 'other' Christianity and criticise its practices (or perceptions of Christian practices) compared with the Islamic OSN. I speculate this may be expected given the historical centrality of Christianity in America and Europe and the public and media critical discourse surrounding Islam in the 2000s following the September 11 attacks and others and the rise of New Atheism. I also acknowledge that my second dataset contains a relatively small sample of 500 discussions from the 2000s and 2010s (along with 500 from 1981-1991 considered in chapter four), so I make this contrast tentatively.

Finally, I consider `soc.religion.christian.bible-study`. The discussions I observed focus more deeply and narrowly on Bible discussion than the other groups I reviewed focus on their respective topics. This is unsurprising given the parameters of discussion indicated by the OSN's title and its place in Usenet's hierarchy (within `soc.religion.christian`). It is also a moderated group, providing a more direct gatekeeper role to ensure discussions focus on the group's shared aims. Here, discussions I observed often share a common focus on discussing Christian religion with some degree of respect for biblical authority. Unsurprisingly, some threads discussed certain biblical books or posted passages from the Bible. This can extend to criticising others' interpretations or theological doctrines based on reasoning from the Bible, at times advocating for certain ways to understand texts. Those I reviewed from my 1,000 Thread Dataset tended toward plainer reading and this lends itself to millennial and creationist positions which infer events from the Bible in a way that foregrounds its historical accuracy and prophecy for the future. This can result in some interactions being collegial and detached, and others proselytising ostensibly to seek conversion to Christianity or a particular Christian position on a subject. This diversity of position may lead to disagreement when discussing interpretation or doctrine, such as regarding salvation, history, and future events. Nonetheless the narrow parameters for interactions provided a more clear and consistent focus across the dataset of interactions than I observed elsewhere.

These 'soc' groups saw much greater participation than `talk.religion`, so far as the archives indicate. This may reflect their reputation from Usenet's early years. The early Usenet took soc hierarchy groups more seriously than talk hierarchy groups, so they attracted larger user bases and critical mass of participants which helps sustain groups over time. This benefitted the discoverability of soc groups, with more early Usenet locations providing

access to soc groups than talk groups, with some feeling that carrying talk groups was an unnecessary cost at a time when online data were more expensive to send and receive. Of course the less granular soc groups also had broader remit for discussion, so may have attracted a greater diversity of participants. Taken together, these factors indicate how early decisions around OSN hierarchies' remit, freedoms to create subgroups and the varying distributions of groups across the network, carried forward into the 2000s and shaped later users' experiences.

I also include select soc.culture groups within my dataset, specifically:

- soc.culture.christian
- soc.culture.islam
- soc.culture.jewish
- soc.culture.jewish.moderated

I noted previously (p.152) that these groups afford discussion of a range of ways in which religion influences lives and the world. I focus on soc.culture.jewish.moderated as a case study later in this chapter. This case study illustrates how these groups' interactions about how religion is a force in the world bring contestation as discussions include ways in which religion impinges — or is perceived to impinge — on others' lives. Diverse societies and globalisation result in the butting up of people's values, beliefs, and practices against one another as they are brought into contact. Online interactions accelerate this by connecting spatially dispersed people in near real time. Archives of soc.culture OSNs are therefore particularly fruitful for studying how these interactions can be constructive and the contexts in which incivility occurs.

In addition, Usenet affords the creation of regional OSNs. I don't research those here, instead focusing on the larger, public, and potentially global OSNs that are not confined to a particular region. As religions may be of interest from any country, the OSNs I research can attract participants from around the world. Nonetheless, regional groups focusing on Abrahamic faiths exist, such as uk.religion.christian analysed in Herring's (2010) research (pp.49-50 above). Usenet's decentralised structure allows the creation of local OSNs that are carried across certain sites on the network and not all. This illuminates a challenge of internet archiving, since the hardware and software used to capture content can only record what it can reach. This is another reason for my research to focus on international Abrahamic faith OSNs — their reach is broader and Usenet service providers can capture

them more robustly than they may be able to capture groups whose interactions are conveyed over only a small geographic region within the global internet.

5.2 Computational analyses to illuminate online social lives — characteristics of Abrahamic faith OSNs

I applied my R code to explore high level data about social interactions within and between select OSNs in the 2000s and 2010s. I present the outcomes and my interpretations of them here. In what follows, figures are rounded to two decimal places except for proportions, which are rounded to three places to capture additional nuance. I first analyse the number of participants and their average participation in terms of numbers of messages and length of duration (the time between their first and last posts). I then analyse the proportions of messages (1) from persistent participants, (2) crossposted to multiple OSNs, (3) which indicate incivility, (4) which include keywords indicating shared religious aims being discussed. I also analyse the number of first-time participants, on average, in each discussion across the groups. I focus on the following OSNs from the 2000s and 2010s:

- soc.religion.christian (including the Bible study sub-group)
- soc.religion.christian.bible-study (on its own)
- soc.religion.islam
- soc.culture.jewish.moderated
- talk.religion.christian (including sub-groups)

For contrast, I also present summary data for all the Abrahamic faith OSNs that I incorporated into my dataset (summarised above) taken together, to indicate how certain groups have characteristics that differ from the dataset as a whole.

The following data provide the highest-level overview of the significance of the selected OSNs in terms of how many messages were sent to their participants, how many people participated and the average number of messages per participant:

	All 2000s-2010s messages in Message Dataframe	soc.religion.christian.inc. bible-study	soc.religion.christian.bible-study	soc.religion.islam	soc.culture.jewish.moderated	talk.religion.christian (inc sub-OSNs)
Number of messages	792,301	45,668	19,186	24,518	290,512	6,523
Number of participants	35,892	3,130	1,123	1,676	2,485	949
Mean number of messages per participant	22.07	14.59	17.08	14.63	116.91	6.87
^Standard deviation	301.26	106.56	106.92	87.9	816.12	23.93
Mean length of participation (months)	21.71	6.35	4.32	4.82	4.23	5.86
^Standard deviation	537.01	81.01	37.45	54.18	56.36	49.56
Mean length of thread (messages)	7.63	5.52	7.05	4.93	14.12	2.77
^Standard deviation	28.87	16.43	24.07	9.18	40.01	10.5

Table 6: summary data regarding participants and messages posted across years captured within the 2000s and 2010s

These data show varying levels of participation which reflect the aims of the OSNs included in the metric. For example, I would expect to see less participation in soc.religion.christian.bible-study than soc.religion.islam given the former's narrower focus. Further, the Bible study group attracted fewer participants but, on average, they participated more than the Islam group. This may also reflect the subject of discussion, with the Bible study group attracting a higher level of participation from those who are interested and either

knowledgeable about the subject or wishing to learn. The Islam group can attract people from a broader range of interests but perhaps some of those interests entail less commitment than Bible study — they then participate a little less on average. The extent to which participants contribute to a shared aim may be more limited. The Islam group attracted a slighter longer average duration of participation, though it is a difference of only approximately two weeks. In addition, the Islam group and Bible study groups saw lower standard deviation in participation compared with all the OSNs in my dataset for the 2000s and 2010s, indicating participants' greater variation from the average length of participation. This further reflects how the Bible study group (among others) may have cultivated a stable core of dedicated interlocutors.

The average length of discussion thread also provides insights about interactions within the OSNs. Here, we see short discussions within the talk.religion groups — an average of 2.77 messages per subject. This increases within soc.religion.islam to 4.93 and increases further within soc.religion.christian.bible-study to 7.05. I discuss the standout example of soc.culture.jewish.moderated in the case study below. These figures indicate longer interactions on any one subject within the more narrowly focused Bible study group and this may reflect contributors' shared aims for participation. I must note that if participants changed the subject line of their discussion, then the methodology will detect it as a new discussion. Nonetheless, in many cases this will be accurate, as subject lines may be changed as the topic of discussion moves on, as a person might when responding to an email thread.

This poses a question however: if a narrower focus in the Bible study group was associated with increased participation compared with other groups on average, why do the highly focused talk religion groups (listed in the table above) on average have participants who post less? I argue that this may be because the groups themselves remained small — with a total of 949 participants in the talk.religion groups included in my dataset — so often did not reach a critical mass of participants that would be necessary to sustain the group as people joined, interacted, and left over time. This is visible in the talk.religion subgroup OSNs exhibiting lower average numbers of posts per person than other groups and having fewer posts (as captured in the archives) than other larger, often broader OSNs. It may be that cohesion is benefitted when an OSN has *some* breadth to the focus of its discussions, which reflects my observations in chapter four that soc.culture groups afford broad discussion within the parameters of a certain religious faith. As is seen with

soc.culture.jewish — and its moderated counterpart — this is associated with high levels of participation.

It's also notable that the average length of participation across all religion focused OSNs included in my dataset for the 2000s and 2010s was 21.71 months. This is considerably longer than participation in any one group captured within the table above. This indicates that participation in discussions *in general* did not end with participation in any one group. Usenet software afforded engagement with a number of groups over time, much like Twitter or Reddit users follow many threads of conversation. This observation broadens the focus from what's happening in any one OSN to see participants as contributing to dialogue in a range of online contexts. This leads me to analyse statistics regarding crossposting messages between multiple groups, which helps raise participants' awareness of, and spark contributions to, multiple OSNs over time.

	All 2000s-2010s messages in Message Dataframe	soc.religion.christian.inc.bible-study	soc.religion.christian.bible-study	soc.religion.islam	soc.culture.jewish.moderated	talk.religion.christian (inc sub-OSNs)
Proportion of messages crossposted (0:1)	0.48	0.003	0.003	0.035	0.00228	0.755
Proportion of messages crossposted: threads with first time participants only (0:1)	0.5	0.002	0.002	0.032	0.000379	0.858

Table 7: numbers of crossposted messages, and the proportion in discussion threads that include first time participants

These crossposting figures suggest another factor that affects participation in OSNs. Here, we see that less than one percent of messages sent to the Bible study group are crossposted to other groups. This reflects how the group sees higher than average participation by its contributors and focuses on its subject matter without connecting to other groups frequently. By contrast seventy-five percent of messages sent to talk.religion groups are crossposted. Further, in the case of talk.religion groups (and the moderated Judaism OSN that I discuss below), first time participants were more likely to post messages to threads which were crossposted. This pattern was less distinct on the other OSNs included above and indeed across the dataset on average. Focusing on the talk.religion groups, it is plausible that (1) frequent crossposting to groups that (inevitably) have differing or broader focuses than any one talk.religion group and (2) the low average number of messages posted per participant combine to create a context in which new people typically experience simultaneously a breadth of discussions and infrequent interactions with a limited (or no) stable core of contributors. Together, this helps explain why these narrowly focused groups do not experience persistent participants who grow in number and sustain the group.

This suggests an impact of having new people within a social network: the network may be more likely to persist/grow if new people experience sustained dialogue rather than sporadic messages on diverse topics. This is illuminated by the average length of threads to which new participants contribute.

	All 2000s-2010s messages in Message Dataframe	soc.religion.christian.inc. bible-study	soc.religion.christian.bible-study	soc.religion.islam	soc.culture.jewish.moderated	talk.religion.christian (inc sub-OSNs)
Mean length of thread (messages)	7.63	5.52	7.05	4.93	14.12	2.77
^Standard deviation	28.87	16.43	24.07	9.18	40.01	10.5
Mean length of thread (messages)	28.35	10.81	16.1	8.96	35.12	5.89

(threads w/first-time participants only)						
^Standard deviation	62.22	24.4	37.96	12.85	64.77	17.91
Median length of thread (messages) (threads w/first-time participants only)	10	4	6	4	14	1

Table 8: average numbers of messages in threads from the 2000s and 2010s in the dataset, including figures for threads including first time participants

Here, the trend continues for soc.religion.christian.bible-study exhibiting more sustained contribution than soc.religion.islam and the talk.religion groups. The latter have on average the lowest length of thread, with a mean of 2.77 messages and a median length of just one message. This indicates that many messages go unanswered and, with their high level of crossposting, it is likely that some messages sent to the groups fall outside their aims. This adds further context to my above observation that talk.religion groups see more sporadic participation. The moderated Jewish group sees the longest average lengths of discussion and I analyse this in more detail below. Across all groups, first-time participants are, on average, more likely to be found within longer discussion threads. I identify two factors here. First, longer threads are simply more likely to be visible to new participants as they interact with the discussions they see when encountering the group. If the thread is shorter, it has more chance of dropping off the list of visible messages in the software/app the participant uses. Second, and with more relevance to how OSNs affect social lives, the longer threads may help ‘convert’ people from being passive readers of OSN discussions to becoming active participants in that group for the first time. Longer threads are typically attracting more people interested in the subject. Their responses in turn spark more engagement from others who are also interested and so the length of the thread spikes. The content of these threads, whether novel, contentious, useful, timely, or antagonistic is developed and new people may feel hooked into joining. In addition, if new people feel nervous, joining a larger discussion provides a little less prominence to their contribution as it becomes one among many. Consequently, longer threads may grow the number of participants within an OSN,

though it is reasonable to infer that the number of potential participants will be limited by the subject and behavioural norms within any given group. This explains why a more narrowly focused and largely self-contained Bible study group still attracts relatively few participants overall.

The proportion of messages sent by persistent participants also varies across OSNs. I define persistence as a person who posted messages to any OSNs within the dataset across a two-month period or longer. While average participation length among all users in the dataset was 3.4 months, the threshold of two months nonetheless includes only 20.2% of participants. Indeed, the median length of participation is zero months. While persistent participants account for only a minority, they are a potentially influential group in terms of sustaining an OSN over time, socialising new members into existing behavioural norms, and contributing to sustained discourse in ways that reflect — or do not reflect — an OSN’s aims.

	All 2000s-2010s messages in Message Dataframe	soc.religion.christian inc. bible-study	soc.religion.christian.bible-study	soc.religion.islam	soc.culture.jewish.moderated	talk.religion.christian (inc sub-OSNs)
Proportion of messages from persistent participants (0:1)	0.863	0.881	0.897	0.905	0.983	0.714
^inc shared religious aims keywords	0.879	0.881	0.897	0.908	0.983	0.722
^exc shared religious	0.861	0.881	0.896	0.904	0.983	0.71

aims						
keywords						

Table 9: proportions of messages posted by longer-standing participants and proportions including/excluding words associated with having shared religious aims

In this table, the figures for the proportion of messages that indicate shared religious aims checks whether messages include terms from a list of words I extracted from my dataset either prominently or distinctively (p.113). Here, I used my word embedding (language model) to identify which words were associated — by virtue of co-occurrence — with each of my initial keywords which indicate religion discussion. This provided a shortlist of words, which I pruned to remove those which could easily occur in other contexts beyond religion discussion. Nonetheless, the messages in my dataset are taken from religion-focused OSNs, so it is likely that use of a polysemous word in a religion OSN is likely to be used with its religion-related meaning. I then wrote code to search messages for these keywords and inferred messages were likely to have some minimal or greater proportion of religion discuss where one or more of these words was present. As a sample, the first five of these words in alphabetical order are these words are Allah, Aquinas, Aramaic, baptism, and biblical. There is consistently a very slightly higher proportion of messages sent by persistent participants which include religious keywords than exclude them, across the OSNs (or groups of OSNs) I analysed. Although consistent, I do not conclude that persistent participants are more likely than others to post messages related to religion. The difference in the overall proportion of messages posted by persistent participants varies across the OSNs analysed and is more notable. This affirms the narrative indicted by previous analyses that the talk.religion groups are more transient than others, with 71.4% of messages in those groups (on average) posted by persistent participants, compared with 98.3% in the moderated Judaism group (discussed below), 89.7% in the Bible study group and 90.5% in the Islam group. Those latter OSNs indicate greater cohesion in terms of being comprised to a high degree of people who interact with one another over longer time periods, in contrast with talk.religion.

The following analyses incorporate messages and threads which indicate incivility. My review of my 1,000 Thread Dataset, a subset of 1,000 discussion threads, indicates incivility manifests in these discussions in three broad ways, each with its own diversity. The first is incivility focused on events in the wider world that occurred historically or contemporaneously with the online interaction. Within my dataset, this included antagonistic messages about conflicts between Israel and Palestine and angry language and accusations and derogatory comments regarding prominent political figures in the region

and indeed elsewhere. There was also criticism of western military decisions and spending including anti-terrorism actions. Second, there is incivility focused on the race or religion of certain groups, including comments about destruction of certain people and vulgar stereotypes about cultural and national groups. Antisemitism or criticism of Jewish communities features prominently. This includes swearing and aggressive/violent language, criticism of Jews' morality, conspiracy about Jewish control, and arguments about civility of relations between Jews and others in Israel. Some messages are similarly discriminatory toward Islam, such as flippant or vulgar criticism of Islamic teaching in relation to terror and suicide bombers; aggressive/violent quotations from religious texts and criticisms of Muslims' behaviour. While not all the latter are uncivil, comments around them indicate certain participants hold views strongly. Racist language is also directed at Arabic people and a prominent political leader of colour. Third, I noted incivility in criticism of, and vulgar language directed toward, other participants. This included flaming (trolling), using obscenities and words implying poor mental health. A subset of these messages also suggested certain behavioural norms were expected or were reproduced through discourse. I noted this in terse or critical responses to others who were perceived as lying, bigotry, dishonesty, poor understanding of a subject and flaming in response of others' behaviour. Some of this reflects the behavioural norms of Usenet's earliest years in which a focus on evidence and a rationalised tone (p.155), which perhaps carries over into an abruptness at times. While I do not argue that later participants were socialised into all the behavioural norms of influential contributors decades before them, there is a family resemblance and the early Usenet shaped the social context that later users joined. There is also inevitably some overlap in patterns of human behaviour seen across social network platforms and across time, since the norms of a platform are only one factor that shape interactions among people whose behaviours are informed by their socialised in many other contexts.

Linking this to my preceding paragraphs' discussion of cohesion, I note that cohesion would, of course, be undermined if longer-standing members of social networks actually exhibit *incivility* and undermine the flourishing of their group. While we may infer that is not happening since, if it was, we could reasonably conclude some people would not continue participating and attrition would increase, the following figures are illuminating. They also broaden the perspective to include variations between messages which are crossposted to multiple OSNs and messages which are not.

	All 2000s-2010s messages in Message Dataframe	soc.religion.christian inc. bible-study	soc.religion.christian.bible-study	soc.religion.islam	soc.culture.jewish.moderated	talk.religion.christian (inc sub-OSNs)
Proportion of messages inc shared religious aim keywords which are crossposted	0.388	0.003	0.003	0.033	0.002	0.639
Proportion of messages exc shared religious aim keywords which are crossposted	0.492	0.003	0.003	0.035	0.002	0.776

Table 10: the proportion of messages including and excluding keywords indicating shared religious aims which were crossposted

These figures do not indicate clear distinction between messages *not* indicating a focus on religion discussion (i.e. excluding shared religious aims keywords) and whether they were crossposted, when looking at the soc.religion and soc.culture groups stated above. There is however a distinction within the talk.religion groups. A higher proportion of messages sent to the talk.religion groups which did not indicate shared religious aims were crossposted to multiple groups, compared with the other OSNs analysed above (though not the dataset when reviewed as a whole). Two factors I have already discussed are likely to affect this. First, fewer people connected in talk.religion groups compared with others I analysed and they had shorter discussions on average. Given this, they experienced limited sustained and focused interaction which may be likely to be contained within one OSN that focuses on that subject. That could affect the proportion of off-topic messages. Second, threads within the talk.religion groups may reflect the conversational and discursive tone encouraged in those spaces. They are therefore less likely to be limited to — or less

precisely focused on — discussing religion. If a greater proportion of messages were new and unsuccessful conversation starters, and those posting them understood talk.religion groups had loose boundaries, then we may see a greater proportion of crossposted messages that do not focus on shared aims occurring here than in some other groups. My social network analyses (below) affirm this and add more context. This suggests a broader point about the importance of socialising new people into appropriate behavioural norms to maintain some boundaries around what, and to where, people post messages within an OSN or network of OSNs. The rate at which messages from new people are posted to OSNs can vary.

	All 2000s-2010s messages in Message Dataframe	soc.religion.christian.inc. bible-study	soc.religion.christian.bible-study	soc.religion.islam	soc.culture.jewish.moderated	talk.religion.christian (inc sub-OSNs)
Mean number of first-time participants per thread	0.35	0.34	0.32	0.29	0.11	0.3
Median number of first-time participants per thread	0	0	0	0	0	0

Table 11: average numbers of first-time participants in each discussion thread across the OSNs analysed

These figures show that most discussion threads include no new participants, but the figures vary depending on the group analyses. The soc.culture.jewish.moderated OSN has a lower proportion than the others with 11% including one or more new participants, whereas discussions on talk.religion groups and others are more likely to include people contributing for the first time. I note above how crossposting may be associated with first-time participation. This receives support in the above figures, since the moderated Judaism

group was less crossposted by virtue of its tighter boundaries than the others and, concomitantly, saw fewer new participants on average.

I discuss above that 98% of messages sent to this OSN are from longer-standing participants and that this group has longer average discussions than others at 14 messages per discussion thread on average (see tables 7 and 10). I analyse these figures in more detail below, but they are relevant to my discussion here as they show by contrast how the other groups I analyse may have more a challenging task to socialise new participants. I argue this as new participants appear more frequently in discussions, those discussions are on average shorter so new people are exposed to less dialogue in any given discussion, and a greater proportion of messages are from the new participants themselves. This makes it more challenging to reproduce behavioural norms than in an OSN which experiences longer discussions comprised primarily of established members with new people less frequently. Of course moderation may also benefit socialisation, since the moderator's authority as gatekeeper enables them to reject messages violating behavioural norms and so shape new participants' behaviour should they wish to contribute to the groups' interactions.

Taken together, these computational analyses indicate how OSNs' characteristics — such as the size of the network in terms of people or discourse, and the duration of participation and incivility in messages — can vary. These variations can be understood in terms of the behavioural norms within different groups and patterns of participation. For example, OSNs with less participation within and between message threads also have broader behavioural norms and more crossposting, so are less focused. These characteristics reinforce one another. I now enrich my findings by presenting and interpreting select social network analyses, before drawing this all together to specify how cohesion and incivility manifest across the OSNs I have analysed. I then apply machine learning to identify factors associated with cohesion and incivility and present my applied case study. Together, this addresses my research questions.

5.3 Social network analyses and crossposting

The above section indicates different degrees of crossposting messages between groups. Here, I present select social network analysis graphs. These enrich my preceding quantitative analyses by indicating patterns of crossposting and the subjects/foci of other OSNs that crossposted with religion focused groups.

When analysing the early Usenet (chapter four), I made social network graphs for each OSN I focused on for each quarter then, once that OSN had grown and become established so patterns of crossposting may have settled, each year. The later Usenet is too vast for this approach to be practicable. Instead, I used the insights I presented in the previous section to identify that social network graphs may add insights that are relevant to my research questions where OSNs witnessed high participation and/or indicated uncivil language. This helped me focus on times when social network graphs may add to my understanding of cohesion and incivility, which I then review in the next section. I wrote code to analyse each OSN I am researching (those in the above tables) and identify the 50 days in which (1) the highest number of messages were posted and (2) the highest number of messages *indicating incivility* were posted. I then identified a time period in which those trends were most prominent, typically spanning six or twelve months depending on the number of messages sent to the OSN. I typically selected shorter time periods for higher traffic OSNs as their spikes in participation and messages indicating uncivil language were often concentrated around, say, a six-month period. I found such trends were not as visible for OSNs which had very limited participation. I then made social network graphs using the methodology I set out in chapter three. I found that for OSNs with limited participation one graph could be used to illustrate connections between OSNs that occurred over one month or more, whereas for high traffic OSNs the high number of connections within even one week sometimes rendered a social network graph spanning that week unintelligible as it was too dense. I therefore adjusted the amount of time that each of the following graphs covers accordingly and specify the time period in the figure notes.

Combining the above analyses with my observations from social network graphs, I conclude that highly focused OSNs — in terms of sustained participation and focus on a subject — may have tighter boundaries insofar as they experience limited crossposting. I do not present social network graphs here for `soc.religion.christian.bible-study` or `soc.culture.jewish.moderated` since my above method for identifying potentially interesting periods of their history yielded no messages that had been crossposted between either of those groups and another in the period I examined. A third of one percent, or less, of these OSNs' messages were crossposted. I wrote code to identify those messages in all my data used for this chapter, covering the 2000s-onwards. I found 62 such messages between the Bible Study OSN and others and 661 between the moderated Jewish OSN and others.

The most frequent interaction pattern for soc.religion.christian.bible-study occurred 33 times and included de.alt.test, bit.listserv.test and soc.religion.christian.bible-study. Many of these messages focused on theological or practical matters related to Bible study. 'Test' newsgroups, as indicated, are typically spaces for testing or practising Usenet, so do not focus on any one subject. Such messages may still be propagated around the internet so others may reply. The next most frequent pattern, occurring five times, connected news.announce.newgroups, news.groups.proposals and soc.misc. These messages focus on matters such as proposing groups and seeking help with administrating them. The third most frequent, occurring just three times, was with alt.religion.christian and the regional uk.religion.christian. (I review patterns of interactions between soc.culture.jewish.moderated in my case study below.)

At the other end of the spectrum, a full three quarters of messages posted to talk.religion OSNs were crossposted (as an average of each group). For example, in 2006, interactions between a Jehovah's Witness-focused OSN and others include discussions about religion and science — particularly the book of Genesis, creationism, and atheism — criticism of religious leaders and Jehovah's Witnesses, discussions encompassing other denominations or forms of Christianity and adult spam messages. This signifies how it is difficult for those groups to maintain a focus on shared aims within the group, since discourse may be pulled in many directions. I discussed the typically transient and limited participation that characterises these groups above, along with the looser and conversational behavioural norms embedded in them. These graphs add context by visualising this and showing the diversity of connections. This also helps explain why talk.religion groups had on average 0.3 new participants in each discussion thread — so roughly one new participant for each thread. This is roughly average across the dataset, but far more than the moderated Jewish group's rate of 0.11 new participants despite the latter having far longer discussion threads on average at 14.12 messages compared with talk.religion groups' average of 2.77. Indeed in talk.religion, longer threads were associated with new participants, since the average length of threads involving one or more new participants was 5.89. This paints a picture of talk.religion groups as densely connected with others, and so attracting new participants, but with a trade off in terms of maintaining a core focused group of participants. I argue for the impacts of this in cohesion below.

The soc.religion.islam group presents some contrast. Around 3.4% of messages to this group were also sent to others, around 90% of messages were sent by persistent participants and people participated for average of approximately 4.8 months. Therefore

the group remained much more self-contained than talk.religion groups on average but less than soc.culture.jewish.moderated. When crossposting occurred, the Islam group was embedded in a breadth of connections that at times focused closely on Islam and, at other times, reflect diversity of discussions that could *relate* to Islam. The following social networks indicate this.

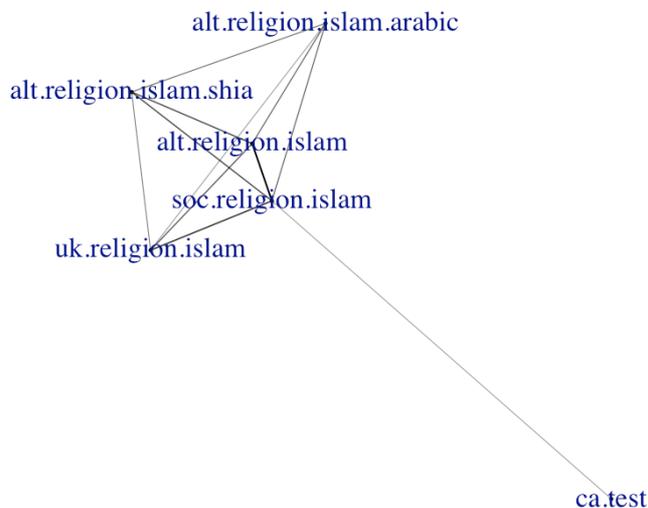


Figure 18: social network graphs showing connections between soc.religion.islam and others from July to December 2005

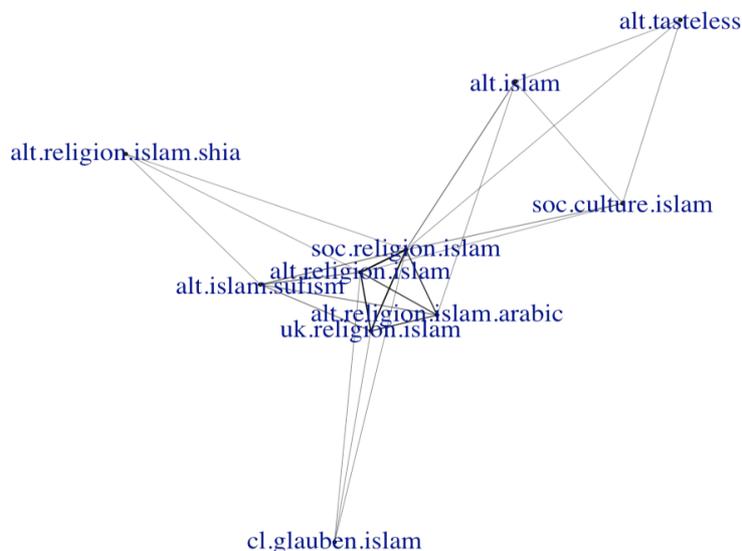


Figure 19: social network graphs showing connections between soc.religion.islam and others from January to June 2006

As above, these graphs reflected time periods in which higher participation discussions took place (in 2005) and incivility was more frequently indicated (in 2006). I reviewed the subjects of discussion in these interactions and found they include discussion of religious figures such as angels, questions regarding the Qur'an, Hadith and Allah, critical discourse around terrorism and Islam, and discussion of Islamic prophets including Jesus. I did not observe

spam when reviewing interactions between the Jehovah's Witness group and others from this period, though more generally spam was sent to the Jehovah's Witness group at times. The former graph includes other OSNs focusing on Islam, so it is reasonable to expect participants across those would be knowledgeable about and/or motivated to contribute to these discussions, thereby driving increased participation. This supports my observation that social networks in which discussions reflect shared aims can exhibit cohesion, seen here in how these discussions are associated with higher participation and are contained within a cluster of relevant subject related OSNs.

The latter figure from 2006 includes a number of groups focusing on politics and, given the 'war on terror', I expect emotions ran high in discussions at these intersections and they could derail the Islam OSN's typical focus on Islamic religion. Therefore incivility may be associated with more diverse patterns of crossposting at intersections that afford contentious topics, but I do not have data to substantiate this clearly, so I must not draw a conclusion from this observation. Instead, I analyse this more substantively in the following section, where I use machine learning to infer associations between different characteristics that OSNs exhibit.

5.4 Machine learning insights

The preceding sections lay my foundation. I first illustrated how social networks flourished in Usenet hierarchies that have different behavioural norms, then analysed select religion-focused OSNs' characteristics, derived by code written and applied to my dataset. These focus on trends in participation and behaviour within each group, indicating how together participants manifest cohesion through their sustained and focused dialogue, and how they may indicate incivility too. This led me to expand my perspective and consider how crossposting messages to multiple social networks appears to reflect those OSNs' behavioural norms and the cohesion and incivility I inferred. With this groundwork done, I then applied association rules mining (ARM) to identify characteristics of OSNs — and people's interactions within them — which are associated with one another (see pp.127-128 and TRA 1.2). ARM's strength is in affording exhaustive analysis of *which* factors co-occur, and the strength of their association, so can provide insights that it is infeasible to obtain with human analysis. However, it analyses categorical data logically and presents numerical outputs. It cannot identify what underlies the associations in terms of human behaviour. Just why may certain associations be present? My preceding analyses of literature relating to Usenet, social networks, and religion (chapter one), insights from some

of Usenet’s earliest interactions which established the hierarchy structure and some behavioural norms that persisted (chapter four), and my above analyses in the present chapter all equip me to make reasonable inferences that explain and enrich my machine learning findings. I can answer my research questions with this computer coding, applied to internet archives, and understood with insights from social network and digital religion studies.

The following machine learning insights relate to the messages in my dataset from the decades of the 2000s-2010s. This is my entire Message Dataframe excluding the 1981-1991 UTZoo archive since that is far too incomplete to afford useful machine learning analyses (see pp.61-63).

5.4.1 First time participation

The machine learning identifies that new people are more likely to be hooked into participating for the first time by responding to a message which was posted across two or groups and was a higher participation discussion thread.

Rule no.	Messages with these factors...	...Are associated with	
1	A reply within a thread High participation subject Participant’s first message	The message being crossposted to multiple groups	
Support	Confidence	Lift	No. of occurrences
0.013	0.810	1.678	10,451

Note — rule numbers reflect the numbering within the machine learning outputs, which are sorted decreasing order by ‘lift’. Support is the proportion of messages in the dataset that include all these characteristics. Confidence is the proportion of messages including the characteristic(s) in the left column *that also* include the association characteristic in the right column. Lift expresses how many times the association occurs in contrast to how frequently it would occur if the left column and right column were independent; for example, if the co-occurrence would be expected 200 times (based on the support statistic) and it occurs 250 times, the lift would be 1.25 (250 divided by 200). (See TA 1.2 for fuller descriptions of these statistics.)

Rule no.	Messages with these factors...	...Are associated with
----------	--------------------------------	------------------------

2	High participation subject Participant's first message	The message being crossposted to multiple groups	
Support	Confidence	Lift	No. of occurrences
0.014	0.762	1.580	10,847

The lift statistics of 1.68 for rule 1 (rounded to two places) indicates that this observation was 68% more likely than would be expected if these characteristics — the message being a reply being a reply, part of a higher participation discussion, and being the person's first message — occurred in crossposted messages by chance. To help express this, 68% indicates that this set of characteristics would be present in crossposted messages roughly sixteen times for every ten times we would expect them to occur if their presence was unrelated to the fact the messages were crossposted to multiple groups. This is a strong tendency when considered that this is an average figure within my entire dataset of messages from the 2000s-2010s. Most trends I will discuss are more subtle, but nonetheless occur frequently and more than would be expected by chance within my large dataset.

A second and closely related insight identifies that participants' first messages in high participant discussion threads are 57% more likely to be posted to two or more groups when that message is not a reply. These messages are, in other words, conversation starters between two or more groups. These insights affirm the value of exposing people to social networks beyond those they may typically read as a means of encouraging their first-time participation. As I discussed above, however, this is a double-edged sword as influxes of new people bring other challenges around maintaining cohesion in a social group. If a network's balance between new and longer-standing members weighs too heavily on the former, there may be insufficient people maintaining the focus of discussion on-topic, socialising new people into the group's behavioural norms, and providing a persistent presence which enables valued (or at least engaging or useful) social relationships to develop over time. This is affirmed by other insights which show the importance of persistent participants.

5.4.2 Incivility and longer-standing participants

Indeed, my manual review of a sample of messages shows longer-standing participants often engaged in higher participation threads and themselves sometimes displayed incivility, while reproducing the free-flowing behavioural norms which enable that alongside

the more civil tone they also supported. My sample of the Giganews data is larger than for UTZoo and working through it suggests themes that emerge.

Antagonistic language appears frequently, with many such messages against the backdrop of discussing religion or events shaped by religion, such as relations between Israel and Palestine. Some messages are antisemitic in the way Jews and/or Israel are discussed, sometimes using block capitalisation (expressive of shouting) and swearing.

George Bush, as US President at the time, is often a locus of antagonistic language. This occurs in at least two ways. Indirectly, Bush is referenced in relation to another subject of antagonism, such as when discussing Bush's conflicts and then using that to draw in antisemitism through vulgar terms used to refer to Jewish people. More directly, the 2003 war in Iraq is a frequent subject of discussion and criticism.

In such threads, Jews often appear as an external group focused upon, rather than as interlocutors participating in the discussion. In a thread on Judaism and Hitler, some on-topic and level-headed responses are posted and the opening post is contested, but antagonistic content likely to be offensive to many abounds.

In UTZoo, discussion I observed regarding Israel and Palestine included criticism of Israel, Islam and racism, and suggests some participants' binary support for one side or another, but there is depth of discussion and detail at times, with occasional wit or humour. Hot button topics of abortion and homosexuality also feature, but I observed less of the more vulgar and offensive content here than in the later Giganews archives.

This suggests that while longer-standing participants are familiar with the social norms of their groups, and at times keep cool heads, in the later years they also reproduce the more caustic tone that gifted Usenet its reputation for, at times, 'flaming' and dysfunction. This may be contrasted with the tone of discussions in UTZoo, which occurred among a narrower demographic of primarily scientists, researchers and tech enthusiasts.

This focus on antagonism in relation to Israel and Judaism also presents a second contrast with the tenor of interactions in the moderated `soc.culture.jewish.moderated` group, which is the focus of a case study below.

Combining this short review with the association presented in section 5.4.1 above, suggests that longer standing members were in the thick of interactions between their core group and others, which were at times heated with the display of antagonistic language toward religious groups or other participants. They could play a particular role in their group's engagement beyond their boundaries, as indicated by the behaviour of newer people who interacted both within and across groups.

5.4.3 Incivility and newer participants

First time participants' messages exhibit antagonistic content at times, such as in content alleging racist discrimination underpinned by Jewish power in power. Here, two responses from other first time participants varied, one tempered and another vulgar.

Some messages from first time participants in discussion of religious groups refer to religious or cultural groups using vulgar stereotypes and slurs, with participants at times appearing antagonised themselves, suggested by a combative tone in opposition to a previous poster's comment. A proportion of messages I observed relate to Arab groups, 'Hindoos' or appear antisemitic. As in longer standing participants' messages, these groups appear to be discussed as 'other', rather than being engaged with as participants in the conversation. This has a refrain similar to that found on social media discussing hot button issues today. While Usenet is not an echo chamber, we may reasonably conclude that people in the groups on which uncivil discussion focus are unlikely to participate. This is affirmed by Chua's (2009) research into the formation of the moderated version of soc.culture.jewish, which created a place more amenable to discussion and community building among Jewish people.

First time participants are also responsible for some spam messages, such as those I observed promoting adult content, a link to a purported online pharmacy and a message of computationally generated text (so it appears, or otherwise vastly off-topic). I observed no evidence of sustained disruption as these messages up a small proportion of those in my sample (around seven messages from among 500 threads). Therefore, from a social network perspective, spam is unlikely to have disrupted engagement in these OSNs.

We cannot know whether participants' first posts reflect their careful observation of an OSN and desire to participate according to its norms (having 'lurked' for some time and discerned how others behave), or are messages in haste having encountered a group for the first time

— particularly where a thread focuses on a topic that animates its readers. Both explanations fit and may occur at different times. Uncivil first-time messages do reflect accepted behavioural norms of antagonistic behaviour on Usenet, but these norms are in part reproduced by longer-standing participants. Some groups disallow this through setting out expectations in FAQs or imposing moderation, but antagonism finds expression in free-flowing Usenet groups, fed by a steady stream of new participants and enabled in part by those who have been engaging for longer.

A tension starts to emerge. Connecting with other groups, perhaps focusing on diverse subjects, draws in new participants but brings challenges to cohesion as those new participants' messages are at times antagonistic in tone and content. New participants are enabled by Usenet's freedoms, the behavioural norms within groups, and are perhaps reacting to their own emotional responses to other posts they have seen, to perpetuate incivility.

5.4.4 Longer-standing participants' contributions to inter-group discussions

The role of our longer standing contributors seems to be, on average, one of participating more than others. In particular they were engaged in replying to higher participation subjects which were crossposted to multiple groups.

Rule no.	Messages with these factors...	...Are associated with	
5	A reply within a thread The message was crossposted to multiple groups The message was posted by a persistent participant	High participation subject	
Support	Confidence	Lift	No. of occurrences
0.307	0.878	1.095	243,345

Rule no.	Messages with these factors...	...Are associated with	
10	A reply within a thread The message was posted by a persistent participant	High participation subject	
Support	Confidence	Lift	No. of occurrences
0.691	0.874	1.091	547,281

Indeed their replies to crossposted messages were 9.5% more likely to be in high participation discussions than would be expected by chance (rule 5). This carries over to their replies not crossposted to other groups, which were 9.1% more likely to occur in high participation subjects than would be expected (rule 10). Discussions which attract more messages are perhaps more engaging due to their subject, argument or due to emotions running high. Core group members are very likely to have an interest in them since they have a stake in the ongoing healthy functioning of the group — since they have invested time and energy into it themselves — and a likely interest in the topics discussed. They therefore help manage tensions between maintaining the group and outreaching to new people, which occurs particularly when messages are posted between two or more OSNs.

5.4.5 Inter-group discussions attracting more participation

Rule no.	Messages with these factors...		...Are associated with
11	A reply within a thread The message was crossposted to multiple groups		High participation subject
Support	Confidence	Lift	No. of occurrences
0.383	0.870	1.086	303,551

Indeed the analyses affirm the value of posting between multiple groups for driving participation. Specifically, replies to crossposted messages are 8.6% more likely to be in higher participation discussions than would be expected if these factors were not associated with each other (rule 11).

5.4.6 In summary — what insights has machine learning delivered alongside the manual review of messages?

What have I learned from using machine learning to mine these association rules? The most pronounced finding is how new participants are drawn into religion-focused discussion when messages to those groups are crossposted with others. The other groups may also be religion-related, but not necessarily as indicated by two non-religion groups in the social network analyses above (figures 19 and 20). This establishes a key catalyst for attracting new participants — reaching out beyond the boundaries of one’s own group. Given the attrition many groups experience — see the average lengths of participation (table 6) — this is a key strategy for growing a group or keeping it replenished with new members.

I found that both first time posts and longer-standing participants' messages can show incivility through antagonistic language in discussion either directly about religious groups or in threads about religion in a less direct context, such as the 2003 war in Iraq. This suggested that new people are enabled to, or socialised into, the option to articulate in a caustic way and depict religious groups or other figures using antagonistic or discriminatory language. This enabled by Usenet's freedoms and the behavioural norms reproduced in groups, which is where more established participants have a role in potentially shaping behaviour.

Indeed, my analyses found an association in which longer-standing contributors are active in messages in longer discussions, including those crossposted to other groups. This indicates a way of managing the tension resulting from reaching out. Here, our more persistent contributors may help to steer discussions, enforce boundaries, discursively include or exclude, and reproduce behavioural norms. As quoted in chapter one, Reid Hoffman said 'every founder of an online community has to shape the culture from day one, because the tone you set is the tone you're gonna keep' (*Masters of Scale* 2018). Our more central members help to reproduce these behavioural norms and have agency — authority — to shape them over time since they make an outsized contribution to discourse. I explore and triangulate these broad findings in my case study below, which focuses on a high-participation moderated discussion group. In this group, fewer messages are crossposted though incivility still occurs. This is therefore a highly relevant case study for analysing (1) the impact of crossposting (since it is more distinctive) and (2) how longer-standing participants reproduce behavioural norms — benefitting cohesion — in this tighter-knit social network.

5.5 Applied case study — cohesion and incivility in a tight-knit OSN focusing on Judaism

The above statistics indicate some distinctiveness about participation within the soc.culture.jewish.moderated (SCJM) discussion group, compared with certain others. This includes the low proportion of messages crossposted, high number of participants and high number of messages *per* participant, and the length of threads. The group is also notable for how it was formed as an OSN distinct from its unmoderated counterpart, in response to uncivil and hateful content. Chua's previous (2009) research illuminated this. Chua explains that by 1987, soc.culture.jewish witnessed antisemitic content, off-topic Middle Eastern political discussion and advocacy for other religions. In response to this, the moderated

counterpart was ultimately created in July 2000. My dataset includes interactions on SCJM from 11 July 2003 to 4 June 2015. These data *suggest* this OSN exhibits cohesion — in terms of high participation and maintaining a focus on shared aims — and needs to manage incivility actively — due to the previous experience of antisemitic content. SCJM therefore represents a case study well-suited to contributing to my research.

This case study has three short sections. First, I confirm the presence and extent of cohesion and incivility indicated by my exploratory quantitative analyses (see tables 7-13). I then illustrate this and add context by analysing crossposting patterns which show the limited connections created with other OSNs. Following this, I use association rules mining machine learning (as above) to identify factors associated with cohesion and incivility.

5.5.1 Overview of exploratory analyses

SCJM provides a suitable case study for analysing factors associated with cohesion and incivility considering (1) the group's previous experiences of antisemitism (Chua 2009) and (2) the ways in which the group contrasts with others. The latter broadens and challenges my understanding of online forums' social dynamics in ways I elucidate here.

The group attract a higher number of participants than others I have analysed, with 2,485 unique contributors to discussion. However, those participants stayed for a mean average of around four months and one week, which is not exceptional. While I cannot infer with certainty, I argue a number of factors may have encouraged participation. First, the engaging subject matter, which attracts a large amount of constructive, inquisitive, and discriminatory content online, reflected in some participants showing higher engagement relative to other groups (pp.203-204). This was the case since the early Internet and, of course, is preceded by antisemitic politics and religious views in other social contexts earlier in the twentieth century. Consequently, many OSN participants had potential exposure to discourse that led them to hold views on — or find out about — Judaism online. Second, the breadth of possible discussions within the wide parameters of Jewish culture may lead to wider discourse and a greater number of messages than, for example, within a Christian OSN which focuses on a smaller and more privatised denomination or sect (see tables 6 and 7). Third, I argue in chapter one that the decisions of early leaders online had ramifications for those joining later and we may observe that here. The roots of SCJM — and specifically its unmoderated counterpart — are found in net.religion.judaism, which was formed in February 1984 (Helland 2007, 961). Therefore, awareness of this group (such as via electronic and print directory listings of Usenet OSNs) may have been greater than some

others and social connections to other groups via crossposting had more opportunity to establish compared with more recently founded OSNs.

Approximately 0.28% of messages to SCJM were crossposted, which is comparatively fewer than in other OSNs I analysed. Indeed, 48% of messages in my entire dataset were crossposted though this masks significant variation. Within this variation, 0.3% of messages to soc.religion.christian (including its Bible study subgroup) and 3.5% of messages to soc.religion.islam groups were crossposted. Moderation may have affected this. Crossposted messages are more likely to vary from the behavioural norms and subjects discussed within SCJM, since these vary across groups, and so may be rejected by moderation. This supports cohesion within the group, since discussions are more easily sustained without the challenges of socialising new people who are used to contributing to other groups (and attracted through crossposting) and the challenge of keeping discussions on topic may be reduced when messages are posted to just one group and need to satisfy just one set of parameters around what is considered suitable for discussion. If it is the case that tighter group boundaries with less crossposting benefits cohesion, we would expect looser boundaries through more crossposting to adversely affect group participation. I observe this in the talk.religion.christian groups that I analyse, which indicated an average of 75.5% of messages crossposted. By contrast with SCJM's mean average of 14.12 messages per thread, the talk.religion.christian groups had an average of only 2.77 messages. The latter also saw far fewer participants overall, with a total of approximately 949 contrasted with SCJM's 2,485.

Indeed lower crossposting could make it challenging to draw in new members. Only 0.038% of messages were crossposted in threads containing new members. This illuminates a challenge. How might a group with tighter boundaries around participation (such as through moderation), so proportionally does not attract as many new participants, be sustained over time when attrition inevitably depletes the size of the group? Indeed this question is relatable to offline religious contexts, such as when religious communities integrate or remain separate from their wider societies in various ways. An answer may lie in the group still being accessible to potential new members — those who have not posted before — since they are able to read messages without needing to contribute themselves. In other words, new people may have been attracted by reading discussions within the group by 'lurking' within it, rather than having their interest piqued when a message was crossposted between SCJM and a group in which they already participate, since that scenario is largely absent. I assert this with my observation that the mean length of threads involving new participants is 35.12 messages, contrasted with 14.12 messages for threads not involving first timers. This contrast of longer threads drawing in new people is seen in other OSNs, though it is

particularly prominent here (see table 8). This suggests that in this case, longer threads may help to draw in new participants since those threads are engaging, contentious or otherwise spark participation. Nonetheless, the impacts of limited crossposting may still be observed in the fewer numbers of new participants per discussion thread compared with other OSNs analysed. In SCJM, there was an average of 0.11 first time participants per thread, which is less than the 0.35 dataset average.

The perception of a tighter social network and boundaries in SCJM compared with other groups is also indicated by the notably prominent role that long-standing contributors have in discussions. In the SCJM data I analysed, those participating for two months or more posted approximately 98.3% of messages. This is higher than in soc.religion.islam, in which the proportion is 90.5%, soc.religion.christian (including the Bible study subgroup) at 88.1% and the talk.religion.christian groups I analysed at 71.4%. SCJM's long heritage affording the development of highly valued longstanding social relationships, the strong communal identity experienced by Jewish contributors connecting across spatial divides online, and low crossposting reducing the incidence of new participants in threads may all contribute to forming a social group in which a subset of contributors cohere through shared aims and persistent participation.

I observed incivility infrequently among SCJM messages in my manual review when compared with other OSNs. A discussion of marriage led to a vulgar sexual comment, I observed *discussion* of antisemitism, rather than the antagonistic antisemitic language exhibited elsewhere. One interaction in which a participant is accused by another of a hateful attitude leads to a coarse reply but is still simply defensive, not vulgar in profane language or discrimination. The tone of discussions is more measured, reflecting the collaborative discourse I observed at times from Usenet in the 1980s (via the UTZoo archive). I witnessed a small amount of spam, with a message extolling against military participation posted, according to one participant, to a large number of OSNs over time. I observed three messages that appear to be computer generated text without discernible topics in relation to SCJM's aims, and an advert for an online pharmacy. These five messages (of 2,033 in my 1,000 Thread Dataset posted to SCJM) would not be sufficient to disrupt engagement in the OSN through 'flooding'.

Taken together, these broad high-level insights paint a picture of an OSN which maintains boundaries by moderation, low crossposting and in which established participants do not often engage antagonistically. While the group has a broad focus, discussions typically remain within those parameters. So while incivility occurs infrequently, this seems not to threaten the persistence and otherwise success of the group in connecting people with

shared identity and/or interests in discussing Judaism. This suggests cohesion and incivility are not a dichotomy but can coexist. I explore this more when concluding the chapter below.

5.5.2 Connections with other OSNs

The limited crossposting makes producing social network graphs less relevant. Since graphs represent slices of time and crossposting was so low, graphs covering shorter time periods are exceptionally sparse and those covering longer time periods are not representative of the group's sparse patterns of crossposting (they could appear to overstate it). Instead, I wrote code to identify the combinations of other groups with which SCJM was connected by crossposting and how frequently these occur. This helps me understand the contexts in which crossposting took place. My analyses indicate the primary reason for crossposting is likely the sharing of (1) updated frequently asked questions documents about Judaism and (2) reading lists on various topics relating to Judaism. In this case, crossposting was frequently to the following groups: soc.culture.jewish.moderated, soc.culture.jewish (i.e. SCJM's unmoderated counterpart), news.answers, and soc.answers. This accounts for around 64.8% of crossposting within my dataset. This is a valuable insight since it explains the relative infrequency of incivility in crossposting by indicating highly intentional crossposting which is aligned with the group's aims. This persists when reviewing crossposting patterns which account for ten or more messages posted between multiple groups including SCJM:

No.	OSNs to which the messages were crossposted
428	soc.culture.jewish, soc.culture.jewish.moderated, news.answers, soc.answers
108	soc.culture.jewish.moderated, sci.crypt
37	soc.culture.jewish, comp.home.automation, alt.usenet.kooks, news.groups, soc.culture.jewish.moderated
18	soc.culture.jewish, soc.culture.jewish.moderated
13	soc.culture.jewish, soc.culture.jewish.parenting, soc.culture.jewish.moderated, news.answers, soc.answers

Table 12: the number of messages crossposted to multiple groups including SCJM within my dataset, excluding patterns occurring fewer than ten times

Notwithstanding some irreverence indicated by the third entry on this table, the alignment of crossposting with SCJM's aims indicates the success of members and moderators in focusing discussions on matters relating to Judaism and other official channels for disseminating updated FAQs and lists, such as news.answers and soc.answers. This leads

to the question of where *did* incivility occur given these crossposting patterns? The answer lies in the statistic that only 0.2% of crossposted messages indicate incivility and, on my analyses, this equates to a single occurrence which did not focus on Judaism and was crossposted at an intersection with a non-religious related group. The methodology likely picked it up due to the subject matter, but it is a distraction to discussions within SCJM. I omit more specific details here to avoid identifying a single message.

Other connections about which, again, I omit specific details to avoid identifying individual messages, nonetheless show that at times content irrelevant to Judaism was posted to the group. However, my analyses indicate that where such connections occurred, they involved very few messages. Indeed, connections that were established for two or fewer messages and appear less relevant — or wholly irrelevant — to SCJM's aims account for around 34 messages within my dataset. This is around 5.1% of all crossposted messages, which themselves are a very small proportion (around 0.23%) of messages posted to SCJM in my dataset.

This subsection shows how crossposting may have attracted new participants if they became aware of SCJM through reading FAQ documents posted to central news.answers and soc.answers groups. However, it also affirms that spontaneous connections instigated by participants were far fewer than I observe in other groups, which reduces the potential breadth of people who may become engaged (i.e. those who are not looking for answers/discussion on topics related to Judaism). This could be by design to maintain the focus of discourse within the group. Indeed providing FAQs can be a means of controlling boundaries around discourse, with people's more basic questions answered by documents which they read or to which they are pointed. This can benefit cohesion within the group if discourse can then focus on more substantive matters. From my review of my 1,000 Thread Dataset — and personal experience — it can be common within OSNs for a new person to experience eye rolling from others who may decline to answer mundane and basic questions about their group's subject yet again. This subsection does not, however, identify what is associated with incivility since, unusually, it scarcely occurred within intergroup discussions. For this, my machine learning insights are required.

5.5.3 Machine learning insights

My analyses identified one *subtle* association which helps illuminate how cohesion manifests and is managed in SCJM. This is that longer-standing participants reply to high participation subjects slightly more frequently than would be expected:

Rule no.	Messages with these factors...	...Are associated with	
20	High participation subject Sent by a longstanding participant	The message being a reply within a thread	
Support	Confidence	Lift	No. of occurrences
0.853	0.976	1.043	247,730

This association is modest at a 4.3% increased likelihood than would be expected by chance. Related to the first insight, this suggests core group members have a role in maintaining/progressing longer threads, which are potentially those more impactful in terms of the size and breadth of readership and engagement. On the one hand, this benefits cohesion since they are likely to reproduce the behavioural norms that have persisted in the group and improve the proportion of messages which receive replies. On the other hand, since there is a modest association between replies in longer threads indicating incivility, it indicates the potential difficulty of avoiding incivility in longer discussion threads. I draw on Chua's (2009) research to suggest that the subject matter — Judaism — may be attracting antagonism which, even with moderation applied, makes it difficult to avoid contentious interactions. My manual analysis of a subset of discussion threads reflects this.

5.6 Conclusion

Here, I summarise how my above analyses answer my two research questions:

1. In what ways, and to what extents, do cohesion and incivility manifest in a substantial sample of prominent religion-focused historical Usenet OSNs?
2. Which factors are associated with cohesion and incivility?

In doing so, my empirical findings and conclusions represent an original contribution to knowledge. I focus here on narrative, not representing statistics, since online social lives — while quantifiable — are perhaps best explained via qualitative description of trends rooted in quantitative data and qualitative analyses, which together reflect the stories of many interactions contained in my dataset.

I have found many facets of Usenet render it an effective site for analysing how cohesion and incivility can manifest online, including: its structure, in which OSNs with differing social norms can be established dependent on the hierarchy; its freedoms in regards to the flexibility it affords to connect with other OSNs; forms of gatekeeping such as moderation which can be enacted; and Usenet being broadly accessible via interfaces from Google and others.

I found that the variations in exploratory statistics — including the numbers of participants, average lengths of discussion threads and duration of participation, and the frequency with which new people joined — varied in ways which are associated with the behavioural norms of different hierarchies. The ‘talk’ groups experienced looser and less stable social networks than the ‘soc.religion’ and ‘soc.culture’ groups I analysed. This affirms the ongoing impacts of early decisions made to structure Usenet in certain ways, which I explored in chapter one.

This relates to my finding that crossposting messages between different groups can influence cohesion. I observed particularly high levels of crossposting, on average, in talk.religion groups compared with others. These crossposted patterns were associated with a lower level, on average, of ‘on topic’ messages which reflected shared religious aims. I nonetheless also found that crossposting could be associated with drawing in new participants. This makes sense intuitively but presents a tension in which crossposting may bring in new members — essential for any group which experiences attrition in participation over time — while also diluting focus. How can participants socialise new members while maintaining their focus?

My analyses regarding incivility are tempered by the imprecision of natural language processing, though my triangulation through manual review mitigates this. Indeed my analyses support the association between crossposting and incivility in groups which experience higher levels of crossposting. This threatens cohesion since these messages can create noise among the on topic civil discussions between existing participants. However, the effective socialisation of new people, where possible, can help the group to persist. I observed this to be a challenge in talk.religion groups at times, since there was little participation overall so little socialisation and consequently little ongoing participation. People didn’t stick around, even though many joined, proportionate to the size of the OSNs.

I found that longstanding members may be important for resolving this tension, where groups cultivated such a set of contributors. They may be in the thick of interactions, replying to discussion threads that have higher levels of participation than average and thereby reproduce the behavioural norms of the group which, to some degree inevitably, they embody themselves as the core participants who are often responsible for a high proportion of messages posted to the group. I found these longer-standing participants had prominent roles in the SCJM case study and in the soc.religion.islam and soc.religion.christian archives that I analysed.

Indeed, SCJM proves a valuable case study for illustrating how cohesion can be supported through *limiting* crossposting and focusing discussions on the topics at hand. This includes posting FAQs addressing common questions/setting parameters for discussions, establishing behavioural norms which can be valuable for maintaining focus. The group nonetheless experienced incivility and attracted new participants. Perhaps both occurred due to interest and contention around the subject discussed. In any case, core group members were responsible for a notably high proportion of messages so can socialise and manage interactions.

Therefore, I observe that cohesion and incivility are not a dichotomy. They can coexist, though I found groups with more cohesion in terms of focus in discussions, continued participation by committed people and behavioural norms that encourage this behaviour — such as being within a part of Usenet's hierarchy which encouraged this or managing it through moderation — have more tools available to manage incivility. The topic/subject on which the group focuses is also a factor, with more contentious subjects such as Judaism — as indicated by hostility evidenced in extant literature and elsewhere online and in wider society — being associated with incivility, notwithstanding their potentially tight-knit network of people who connect and discuss this subject. Taken together, these observations indicate that the following factors may be influenced to support cohesion and mitigate incivility, to the degree that they are practical and desired:

- Behavioural norms of the platform/protocol within which the OSN is established — such as the 'soc' or 'talk' hierarchies in the present research
- Platform affordances, in terms of its structures and freedoms
- Subjects/topics on which the OSN focuses
- Number of established, committed participants and the proportion of messages they contribute

- Frequency with which new people are brought into the group through outreach or openness
- Moderation, computational or manual, of messages to include/exclude certain topics, language and/or senders
- New contributors continuing to participate so a (relatively) stable core of participants develops and can socialise new members
- Responding to messages to develop longer discussion threads which, while risking incivility, may also attract new members and increase social network connections between existing members

The presence and strength of these factors — and interactions between them — varies within OSNs. In the next chapter, I conceptualise ways in which they inter-relate. I then use this as an interface to identify how my research in this dissertation enhances academic understanding of the concepts of community and authority in religion-focused social networks online.

Chapter six — conceptual conclusions and next steps

Chapters four and five include findings which address my research questions. First, I explained the results of my analyses factually and then interpreted them in light of my concepts of cohesion and incivility. At the end of chapter five, I synthesised my individual findings to specify factors which my research indicates can influence cohesion and incivility. In this final shorter chapter, I draw conceptual conclusions by working through elements of my conceptual framework, specifying how each informs and shapes my findings and relates to the concepts of community and authority well-established in digital religion literature. In doing so, I situate my work in the broader study of digital religion and seek to indicate its scholarly merit. I then outline how I seek to develop my research in future and what is required to enable its future impact through collaboration and expansion to consider other research sites and social processes beyond cohesion and incivility.

The development of my conceptual framework and research design were framed by theoretical concept of religious-social shaping of technology. This and the (non-religious) framing of the social shaping of technology (pp.67-68), emphasise how technologies' users influence the acceptance, rejection and modification of technology to meet their ends. My conceptual framework includes ways in which early and later Usenet participants had agency to shape the technology, with my particular focus being on its shaping/adaptation to accommodate religion.

The elements of my conceptual framework are as follows:

Within the *digital substrate* (pp.28-30), which structures and facilitates social network interactions:

- Decentralised network
- Packet switching technology
- Gateway machines and gatekeepers

Within *Usenet* (pp.34-38), the affordances enabled by the digital substrate and Usenet's technologies:

- Asynchronous communication
- Anonymous/pseudonymous communication

Relatedly, I identified that values/behavioural norms can persist on Usenet over time as new people are socialised into online social networks (OSNs) and *Usenet's structure* influences behaviour, such as through OSNs' titles (e.g. religion or politics groups) and the hierarchy encouraging different tenors of interaction in, for example, soc hierarchy groups compared with those in alt or talk hierarchies.

The *social processes* that I analyse (see pp.59-61 and pp.120-126), which manifest in participants' interactions and are enabled by the digital substrate and Usenet, are:

Cohesion:

- Discussions indicating shared (religious) aims
- Cooperative discussions, notwithstanding disagreement
- A high signal to noise ratio within discussion threads

Incivility:

- Antagonistic discourse
- Flooding an OSN with messages, undermining social structure

I now work through each of these concepts, illustrating which of my metrics and findings relate to them, affirming how my research using this conceptual framework cumulatively answers my research questions (as specified in chapters four and five) and informs academic understanding of community and authority.

6.1 Digital substrate

In chapter one, I explore how decentralised networks afford many simultaneous near-synchronous connections between spatially dispersed people. This facilitates the crossposting between Usenet OSNs that I explore in chapters four and five which, in turn, connects people with diverse interests, beliefs and reasons for participating. Further, crossposting encourages new connections with people participating in discussions between the OSN with which they typically engage and others which are new to them. My analyses found that the hierarchical structure of organising Usenet OSNs influenced patterns of crossposting and engagement with new participants. For example, the talk hierarchy groups analysed in chapter five typically experienced greater frequency of crossposting and new participation than groups which focused on narrow topics more seriously, with less focus on chat and light conversation.

This relates to community as decentralised networks connecting people in this manner need to manage how they accommodate disagreement. I found that participants often did so in focusing on their subjects while disagreeing without becoming antagonistic (pp.190-192). Therefore this element of community — and indicator of cohesion — is present at times, though the subject on which the OSN focuses can influence this. In a specific example, Chua (2009, discussed on p.51 above) noted antisemitic content in Judaism discussions which, by its nature, is not accommodating disagreement.

Cumulatively, this illuminates how authority can manifest where a person has an influential position in a network. Usenet's decentralised network means that, broadly speaking, anyone can send a message to multiple (inter)national groups of their choice and those messages are likely to propagate around the world. But not all messages are received equally. In addition to formal means of gatekeeping (discussed below), some people's messages receive more replies than others'. This can indicate that some people have an informal authority, whether because they are a longer-standing participant so have built rapport with others, or are known to multiple social networks and experience engagement when they post and thereby form a bridge between them. This encourages a distinctively *social network* perspective to analysing authority in online social networks that focus on religion. This form of authority is fluid, since the person successes influences others to engage with them, but this is contingent on others continuing to find them engaging and worthy of their time.

Usenet's decentralised network also affords freedoms to form and propagate OSNs on myriad subjects and engage others in discussion with few limitations. These freedoms benefit the formation of community since they enable the development of persistent (online) relationships between participants who may first encounter one another through connections enabled by Usenet's flexibility. The persistence of these connections between people is indicated by the duration of their participation in groups. Here, I found that average lengths of participation varied across groups and that approximately 20% people joined and engaged groups for more than two months, which was much longer than others on average. This also varied between groups, with soc.religion.christian (when including its Bible study subgroup) attracting longer participation than others I analysed at approximately 6.35 months.

Packet switching technology (see p.29) is the second of three elements of the digital substrate I analysed which enables rapid new connections on Usenet. This technology routes data across decentralised networks efficiently, supporting rapid one-to-many communications within and between social networks. Along with affording valued social connections, such as those indicated in the metrics I discuss above, this also affords cheap propagation of spam messages. Here, my dataset provides insights regarding spikes in answered messages which can indicate this. This has the potential to undermine cohesion and community by making it difficult for people to form and sustain connections amidst floods of messages which may be off-topic, commercial or offensive in content, though I found such messages were typically insufficiently concentrated within any given time period to be significantly disruptive to the OSNs that I analysed (p.240). Further, although imperfect, my analyses of the proportion of messages indicating a focus on religion showed that this varied across OSNs. Here, Usenet's structure, freedoms and the packet switching technology combine to allow discourse beyond the bounds of OSNs' intended focuses.

This brings to the fore my third element of the digital substrate, which is the role of gatekeepers. Here, the digital substrate allows certain people to prevent access to a social network or influence their participation within it. Access could be prevented in the 1980s, albeit decreasingly through the decade as the adoption of Internet protocol increased, by administrators who managed Usenet sites (locations) that had central backbone locations (pp.32-34). They could choose which OSNs' messages were forwarded to sites downstream and which were not. More widely, gatekeeping manifests through (1) moderation which can impose manual review of messages before publication, and (2) through publishing FAQs or other documents for prospective participants which state the group's purpose and behavioural norms (though these terms themselves may not be used). I found these two latter means of gatekeeping occurred in my case study of a Judaism-focused group in chapter five (pp.224-230). My metric indicating the proportion of messages focused on religion-discussion, although imprecise, is useful for indicating some impacts of such gatekeeping. Here, my findings indicate that moderation and clearly defined aims may increase the proportion of messages targeted more closely to discussing religion. I found that around 98.3% of long-standing participants' messages to the moderated Judaism discussion group I analysed in chapter five indicate discussion of religion, reflecting the group's aims, whereas this was 72.2% on average across the talk.religion groups I analysed, which do not exhibit this gatekeeping. Similarly, in the Bible study group that I analysed, which also has tightly defined aims, 89.7% of long-standing participants messages indicate a focus on religion. This relates to community in online social networks

since those more focused groups' discussions cultivate interactions oriented toward shared aims of discussing one or more of the Abrahamic religions; this supports the formation of social relationships online. However, the persistence of these groups is not visible in mean average lengths of participation, which I found to be approximately 4.23 months in the moderated Judaism discussion OSN and 4.32 in the Bible study group compared with, for example, 5.86 in the looser talk.religion groups. The length of participation may not equate to qualities of relationship established between participant. It also indicates a form of religious authority since those in control appear to have effectively oriented discussion among long-standing participants towards their aims, through their moderation and/or otherwise setting direction.

6.2 Usenet's affordances

In chapter one, I identify how asynchronous communication (enabled by the network of the digital substrate, pp.28-30) and pseudonymous/anonymous participation help enable and shape interactions on Usenet (pp.36-38). They do so in ways which relate to my findings and community and authority. Starting with the former, we can measure the reach of interactions through my summary statistics including the number of participants in each OSN and the number of replies received to each thread. The latter is particular to asynchronous communication since participants can reply from their local time zone around the world whenever they next check for messages which would not have been as frequently as many people check their social network messages today, given more widespread use of smartphones. This enabled the formation of persistent social relationships indicating an element of community and provides an indicator of a 'soft' form of authority since we can observe the proportion of users' messages which are higher participation discussions, suggesting those people have some influence over the direction of discussion. To this end, of the approximately 683,415 messages sent by long-standing contributors (exc. my UTZoo data as it is too fragmentary), I found that approximately 82.3% were posted to higher participation discussions. By contrast, approximately 66.7% of messages sent by contributors with shorter tenures (two months or less) were to higher participation discussion threads. Across the entire dataset, 14.97% of discussion threads were higher participation (one or more messages longer than average), approximating the Pareto principle in which approximately 20% of causes are responsible for around 80% of effects. Some threads shape discussions and social connections disproportionately.

This complicates authority a little as these contributors may not be using their agency and influence positively. People may spark others' engagement with them by kindling their frustration or a negative response which feeds into their replies. This is authority in a sense, since it is a display of power and may be influential over others' behaviour.

The affordance to interact anonymously or pseudonymously relates to three of my metrics and multiple findings. First, the number of participants or messages. People may reasonably find it difficult to build rapport if they do not 'see' others fully due to their anonymous or pseudonymous participation. Consider a social networking profile on Twitter/X or Instagram which includes the person's name and photo or likeness, compared with one that has neither. The former affords more opportunity to build rapport through a first impression. On the other hand, discussing contentious topics in religion may benefit from pseudonymity/anonymity (pp.36-38) and we would expect to see this manifest in Usenet OSNs which focus on more thorny or sensitive matters. That, in turn, could help feelings of community to form. This could arguably be inferred by contrasting the metrics for numbers of participants, their participation levels, and duration of participation in the Bible study group I analysed in chapter five compared with the less-effectively managed (or unmanaged) talk.religion groups. I found that within the Bible study group, each of the 1,123 contributors I identified posted 17.08 messages across 4.3 months on average, whereas in the talk.religion groups the 949 contributors posted only 6.87 messages across 5.86 months on average. This indicates that Bible study participants were able to discuss potentially challenging subjects while continuing participation. Usenet's freedoms permit a spectrum of choice from anonymity to self-disclosure and this may be a factor here since anonymity/pseudonymity, if desired, may encourage an openness and safety to engage with critical discussion in ways that may not be possible in a participant's offline social context.

We could also expect anonymous/pseudonymous participation to spark uncivil discourse due to the protection it offers and since it may lead to the development of limited rapport. I cannot isolate the impact of this, since by design my dataset does not capture whether or not a contributor is using their (ostensibly) real name. Nonetheless, contrasting the proportions of messages indicating incivility across OSNs which are more and less conducive with building rapport through sustained focused on defined topics is illuminating. Approximately 0.78% of messages sent to our Bible study group indicated incivility contrasted with 1.44% (roughly double) among talk.religion groups on average. This seems to support, very tentatively, my expectation that sustained focus on defined topics (such as

Bible study) builds rapport and may reduce instances of antagonistic messages between strangers who perhaps do not even know one another's names. However, this figure is 2.11% of messages sent to soc.religion.islam. This complicates the picture, since discussion subject may be a factor given particular tensions online around Islam post-September 11. This affirms how it's difficult to identify the impact of anonymous/pseudonymous participation on incivility since other relevant factors occur concurrently.

Lastly, if pseudonymous/anonymous participation disinhibits behaviour that breaks behavioural norms, and affords the propagation of spam, we would expect to see spikes in the numbers of messages unanswered. This could be because messages are off-topic, overtly spam, or otherwise defying of behavioural norms in ways which leads them to be ignored. While highly concentrated spam messages would be easy to spot on the OSNs I analysed — and I did not observe this — there is some nuance around messages which are simply more careless or antagonistic to their target group and (perhaps consequently) do not receive replies. I observed this when reviewing data for days on which larger numbers of messages did not receive replies. A further complication, affirming the need for this manual review, is that some messages are simply informational, such as a quotation of the day, and so were unlikely to solicit replies in any case.

Together, these concepts and findings relate to community and authority. Regarding community, the stability of core group members' participation may be affected by the rapport and substance of interactions which are arguably influenced by flexibility to interact anonymously or pseudonymously. This may also relate to how effectively disagreement is accommodated, since my analyses (pp.219-221, pp.222-223) suggest the importance of core members for maintaining the group over time. Therefore factors could interact — anonymity/pseudonymity, stability of the core group, and accommodating disagreement — to influence the emergence of community. Relatedly, this affirms the authority of core group members to influence the group's discourse over time. Their agency can keep a group functioning over time which, in turn, could help temper any adverse behaviour from those emboldened by anonymous/pseudonymous participation.

6.3 The persistence of shared aims and/or behavioural norms

I identify that the persistence of shared (religious) aims and concomitant behavioural norms within an OSN can indicate cohesion. Two of my metrics inform this: the proportion of

messages indicating shared aims and persistence of participation. Further, I summarise above the (potentially) cohesive role of having a core group of persistent participants and OSNs with shared *religion*-focused aims may facilitate cohesion motivated by faith and evangelism.

Like all such measures, my metric and findings regarding the proportion of messages indicating religion-discussion based on featuring certain keywords distinctively or prominently is imperfect but suggests overall trends. I found trends exist with more narrowly focused and more controlled groups indicating a tighter focus on religion discussion when focusing on longer-standing group members, who are those likely to be socialised into groups' behavioural norms including the focuses of discussion. We can contrast the talk.religion groups, in which approximately 72.2% of those group members' messages included such keywords with 98.3% within the moderated Judaism discussion group I analysed and 90.1% within the Islam discussion group. These latter groups are therefore more likely to host discussion focused on subjects which tap into the potential for faith to build connections incorporating elements of community. Whether participants felt this is beyond my gaze, but this is supported by both the longer than average lengths of discussion and higher proportion of messages sent by longer-standing members in the Judaism and Islam groups compared with the talk.religion groups as a whole (tables 7 and 10), since these metrics suggest those participants sustained discussions with one another. Elements of authority also relate to having a (relatively) stable core of contributors. In particular, we can expect to see these individuals gain some influence over the behavioural norms and topics of conversation due to their prominent roles and likely centrality in their social networks. I explore this a little above (pp.84-85) and consider it an expression of authority invested in the core group since Usenet participants are, to some varying degrees, socialised into behavioural norms.

6.4 Cohesion

6.4.1 Shared (religious) aims

Three of my analytic metrics identify ways in which an OSN's shared aims may be challenged: the presence of uncivil language, crossposting with other groups and the frequency with which people new to religion discussion in these groups join in. Regarding language, I identify that this could undermine constructive behavioural norms into which people are socialised since people may repeat the behaviour they witness within a group.

The value of socialisation is reflected in frustrations found in difficulty doing so when many new participants joined Usenet in the 1990s (pp.42-43). If enough people are disruptive, an OSN may be unable to cultivate discussion on the topic on which the group focuses. This was seen in the unmoderated version of soc.culture.jewish. The corollary is that well-focused discussion groups with low levels of incivility, such as the Bible Study group I analysed in chapter five, have a smoother path to achieving their aims.

A similar effect holds regarding crossposting. Where this occurs very frequently, and particularly between multiple disparate OSNs focusing on different subjects, the resulting diversity of discussion topics, lack of shared intentions for participating and non-existent/limited rapport between contributors, may make it difficult to maintain focus on a common interest and develop discussions that serve that. Crossposting also draws in new participants (pp.218-219) who contribute to this scenario, so the frequency with which new people join can affect the level of challenge that core group members experience regarding keeping the group on track.

Cumulatively, this relates to a factor in building community online (and an indicator of cohesion in my definition), which is how OSNs accommodate disagreement while maintaining a focus on engaging together about the focus of their group. To the extent this is successful, contributors may hear opposing views from one another while continuing to engage, as I saw in the case study of a moderated Judaism discussion group in chapter five. Here, the groundwork was laid by the group having clear expectations around behaviour — publicised in FAQ documents — and moderation which enabled gatekeeping to ensure adherence to these parameters while also enabling argument. Without these parameters, and given the potentially contentious nature of religion discussion, uncivil interactions could lead to attrition among the core group of participants. This attrition occurred among talk groups which often did not build a core stable group of people interacting, so far as I identified from the archives. Absence of these parameters can also undermine the authority a group invests in religious texts. I observed this in critical discussions about the Bible. Although some were constructive and focused on hermeneutics, others (such as found in chapter four's analysis of 1980s OSNs) were concerned with creationism. Some participants in these discussions certainly appear not to have shared others' views on the plain truth of Genesis accounts and this led to highly critical views on creationist arguments. This can make for lively interactions but indeed undermines the use of the Bible as an authority in, for example, these discussions that intersect science concerned with cosmology or Earth's history.

6.4.2 Cooperative discussions

Two of my metrics relate to cooperative discussions, in which participants are supportive of one another and progress discussions in line with the topic of the thread and the OSN's broad aims: people continue to participate longer than average for Usenet OSNs and a higher proportion of messages reflect the discussion of religion. The presence of these would indicate a context in which people desire to sustain engagement with others in the group and this engagement is aligned with discussing religion, which could include practising religion such as through prayer or evangelisation. I found evidence of this within the groups I analysed which have a tighter focus on aspects of religion or religious groups, including within soc.religion.christian and its Bible study sub-group. Within these OSNs, people contributed for 6.35 months on average and 25.07% of messages indicate discussion on shared religious aims. This contrasts with 5.86 months on average and 20.36% of messages in talk.religion groups on average.

Where present, this helps indicate community in two ways. The first is the potential for shared faith to bring people together around common beliefs or practices which are expressed in their discourse. This relates to authority since the religious tradition which provides that shared foundation may, in doing so, provide participants with a means of engaging cooperatively with each other, with each person recognising the authority of the text or leader being discussed, or the value of prayer as a practice.

The second way in which (sustained) cooperative discussions indicates community is, relatedly, the development of valued social relations. These may flow from shared faith or, in a self-selecting way, because people who have sustained their engagement with one another for a period of months (or years) have likely developed social relationships that hold some value, else they would not continue participating. I recognised this in the longer than average participation in soc.religion.christian (as noted above), though this figure is less within soc.culture.jewish.moderated at 4.23 months. This is, *prima facie*, surprising. It is explained by my observation that of 2,485 unique users in the moderated group, 492 stayed two months or more and 155 for six months or more. This core group are likely to have valued the relationships formed in some ways that they continued to participate far longer than others. Based on the methodology, 10.9% of messages indicate shared *religious* aims, but this likely reflects the broader discussions about myriad aspects of Judaism and Jewish life, which could be engaging for many.

6.4.3 High signal to noise ratio

In chapter one, I explain how cohesion may be indicated where a high proportion of messages are relevant to the OSN in which it is being posted. Where this is not the case, it may be difficult to sustain focused dialogue, build rapport and encourage sustained participation. I use two metrics to infer a high signal to noise ratio: the proportion of messages indicating religion discussion and spikes in unanswered messages. I include the latter since high numbers of unanswered messages may indicate they are irrelevant to the OSN's focus, are spam, or are otherwise found disengaging so do not support cohesive dialogue. Chapter one indicates how Usenet's platform structure and freedoms permit the dissemination of such messages, since without active gatekeeping a user can send any message to any group that is accessible from their network location (see p.30, pp.32-33). These messages propagate quickly and can undermine discussion in established OSNs, as Chua (2009) found regarding the unmoderated soc.culture.jewish group. My findings here indicate that some groups are more effective than others in sustaining a high signal to noise ratio. For example, the creation of soc.culture.jewish.moderated addresses the problem Chua identified. I found that of the messages sent to the group which related to its broad focus on Judaism and its intersections with other related topics, approximately 98% were from longer-standing participants. This figure was a little lower in the other groups I analysed, save for the talk religion for which the figure was notably lower at approximately 72% on average. Here, community is seen in how social relationships have the opportunity to form and be sustained, free from being undermined by bad actors disrupting the group or frequent off-topic messages diluting the group's focus. This is particularly important since the Usenet groups I researched are public, so participants and observers have access to the full stack of threads and messages, with those more and less relevant together in a list (similar to visiting a discussion forum with threads presented sequentially, rather than content pushed algorithmically as on Twitter/X's For You tab). This benefits the authority of core group members (pp.238-239) as their social network connections to others, and the prominence of their own messages, are maintained when not crowded out by noise.

6.5 Incivility

6.5.1 Antagonistic discourse

Most clearly, my concise summaries of how incivility manifests in the OSNs I researched are relevant to understanding some of the tenor and relative extents of antagonistic language and spam. For example, I note infrequent presence of antagonism in the moderated Judaism OSN I researched compared with other groups that had free-flowing behavioural norms and in which participation was open to all (see pp.181-185, 190-192, 219-222 and 224-227). However, my crossposting metrics and insights also figure as this crossposting often brings disparate groups into contact resulting in a change of tone and subject of discussion within a group. This has potential to undermine cohesive interactions between established participants in the manner I describe in the previous paragraph. A few factors can temper this. First, moderation may nip it in the bud since behavioural norms can be enforced. I found limited incivility occurred in the moderated Judaism group I analysed in chapter five, though the potential for control remains present and indeed this was a key motivation for forming the group. This group experienced little crossposting in contrast with other OSNs (table 7) suggesting moderators succeeded in managing likely attempts to crosspost messages between disparate OSNs, as other groups experienced and which can lead to conflict.

Second, groups which have a stable core network of contributors — at least relatively stable, since I find groups experienced attrition and new joiners over time — have a resource which can sit at the intersection of their group and those to which messages are crossposted and influence interactions to reflect their own behaviours and their group's chosen topic. This may be by *not* responding at times where messages are egregious in subject or content, or can involve reproducing behavioural norms that accept uncivil participation. I explored this further in a case study of people connecting online to criticise and protest Scientology (Kinnear 2023) and my analyses in chapter five indicate this can also occur in other groups.

Third, the subject on which a group focuses — and the concomitant behavioural expectations placed upon contributors — can influence antagonistic discourse. Simply, certain subjects are likely to attract more uncivil messages than others. I observed this to varying degrees across the religion OSNs, with the notable recurrence of antisemitic language and caustic interactions regarding Israel and Palestine, and other contentious topics such as homosexuality and abortion. Relatedly, I found limited incivility in interactions about religion and science in chapter four (pp.190-192). Here, the subject was treated at times with some distance, in the manner that Usenet's early demographic of researchers and academics were doubtless used to treating objects of intellectual interest.

Together, these factors can influence community as they may create strain by enabling disagreement and offensive content which needs to be managed. This, in turn, could affect the stability of longstanding contributors' participation. This is because they bear a proportion of this strain since they may (1) participate more than others, (2) have experience of managing such content — such as I found when messages were crossposted between potentially disparate groups, drawing in new people (cf. pp.222-223) — and (3) they bear some responsibility for socialising new people into their online community by affirming and reproducing the behaviour established within the group. The success of accommodating this disagreement and sustaining a cohesive core group of contributors indicates elements of community are present.

6.5.2 Flooding undermining social structure

Lastly, my conceptual framework and review of extant research on Usenet identifies that flooding a group with unwanted messages, which typically receive no replies, can undermine an OSN's social structure. This is because OSNs benefit from regular reproduction of social connections between new and existing contributors by the posting of messages to others in the network and receiving replies. This increases engagement — people respond to interesting discussion threads — helps reproduce social norms, and builds rapport, sometimes based around shared religious identities signalled within messages. This is disrupted when spurious connections are formed by those spamming groups within irrelevant messages, such as those wildly off-topic or promoting adult content. This is enabled by Usenet's platform structure, which affords dissemination of content across distributed networks reaching OSNs around the world. Usenet's freedoms also come into play, since people can post anonymously and the content posted is unregulated. Indeed, so far as Usenet is decentralised with no central point of control, content *cannot* be regulated. In practice, individual OSNs manage this by gatekeeping through moderation or, for example, keyword-based message blocking. My analyses found that adult content permeated the Usenet OSNs that I analysed where such gatekeeping was not in place. It was, therefore, present within the unmoderated Judaism group and (at least largely) absent from the moderated Judaism one. I found that such content was present within soc hierarchy groups as well as those in the talk hierarchy. This partly undermines the intentions of those early key figures in Usenet's history who created the hierarchies with different types of discourse in mind. To some extent the distinction holds, as my analyses contrasting soc and talk groups finds (pp.154-165), but it suggests the decisions at some sites to not provide access to talk groups — in order to focus on

more serious and substantive interactions — was not entirely successful since the OSNs to which they provided access were still receiving undesired content. Nonetheless, I found little evidence of a high volume of messages within a short period of time, which would truly indicate disruptive *flooding*, as was narrated in Peckham's (1998) paper regarding alt.religion.scientology. I ran analyses which identify the days on which the highest number of unanswered messages were sent to the OSNs that I researched. This allowed me to review messages on these dates, considering their subject lines and the analytical descriptors (distinctive and prominent keywords, and whether these indicate religion discussion). Given this, I found that posting adult content, adverts, and irrelevant messages, while indicating the presence of spam, did not indicate disruptive flooding (cf. p. 227 and p.240, where I note such findings). This is conducive with cultivating elements of community online. It also indicates that gatekeepers' authority can successfully protect online groups, though without this the frequency of spam messages appeared insufficient to undermine the potential for cohesive social relations.

In summary, this section has summarised how Usenet's structure and freedoms, and consequent affordances, simultaneously enable cohesion which sustains online social networks and the incivility which can undermine them. These forces combine in varying ways — and with different outcomes — in different social networks. This indicates that cohesion and incivility are influenced by many people's actions from those who established Usenet's decentralised structure; early actors who formed the hierarchies and (to a more limited extent) behavioural norms which outlived their tenure on Usenet; those in charge of services that provide access to Usenet and can block access to certain groups or shape users' experiences through the services they provide to enable participation; individual people who formed OSNs within Usenet, set out behavioural norms through their actions or in introductory/FAQ documents; and, crucially, the actions of the many people — at times tens of thousands of people — who participated in a group through posting messages, creating new connections between groups by crossposting, deciding whether to continue or leave, and reproducing accepted social norms themselves or undermining an OSN by posting off-topic content or being offensive.

6.6 Major findings in relation to community and authority

Here, I step back from the conceptual framework to summarise major findings and relate them directly to concepts of community and authority in the sociology of religion, as discussed in sections 1.4 and 1.5. This is in addition to my comments in preceding subsections of this chapter which relate my findings regarding individual elements of my

conceptual framework to community and authority. This represents a contribution to academic understanding of community and authority as it manifests in Usenet, which is a less-studied and historically important site of digital religion, and which is in limited ways relatable to more current sites such as Reddit.

The major findings of the research in relation to community and authority relate to how OSNs' characteristics vary in ways that affects the presence of elements of community and how authority is exercised. The findings are presented here as themes, drawn from empirical analyses in chapters four and five.

6.6.1 Regarding community

I found that historical developments on Usenet, in which certain groups were setup earlier in Usenet's history, can benefit their prominence. Here, the 'soc' hierarchy groups benefitted from wider distribution in Usenet's earlier years, in contrast to 'talk' hierarchy groups, to which users had more limited access. This correlates with greater participation in 'soc' groups into the 2000s, which had become well-established and, on the datasets I observed, had higher participation than 'talk' hierarchy groups. Further, smaller groups can struggle to acquire the critical mass necessary to sustain participation and growth amidst attrition. So effective development of community may be influenced by historical lineage on the network and having sufficient members.

But other factors affect development of community. The breadth of parameters established for discussion (i.e. the topics discussed) can be impactful – wider parameters can encourage participation, as perhaps seen in 'soc.culture' groups which encourage broad discussion of a topic at hand, such as Judaism. Parameters that are too narrow may not attract critical mass to sustain community development, but too wide – such as where groups experience high levels of crossposting to disparate groups – may limit interactions as people observe fewer discussions focused onto their interests and thereby reduce the duration of participation. This may inhibit community growth.

Crossposting, however, also brings benefits. It develops the embeddedness of an OSN in the wider Usenet network, affording prominence and exposure to the OSN among those who do not join, and may thereby encourage new participants. This can benefit community development where sufficient focus is held on the OSN's aims and topics.

The bringing of existing 'offline' religious identity and community online affords shared aims and focus to an online group, helping create longer-standing engagement and the possibility of relationship development through affinity and homophily. I observed this, for example, in Jewish and Christian group discussions, though not within the talk.religion hierarchy which has looser participation.

Another factor of participants' wider 'offline' cultural milieu that affects community is how behavioural norms may encourage civil participation and thereby sustain relationships and community. The more rationalised tone of discussion in religion and science discussions in the 1980s on UTZoo is an example here. I refer to this as 'measured disagreement', since those opposed to one another often displayed disagreement without resorting to incivility that I observed more in the 2000s and onwards. This tone may help enable the sustaining of relationships and thereby (professional) community, while closely-held opinions on sensitive topics differ.

Lastly, an aspect of Usenet groups illuminates an interesting tension regarding community development. While groups' openness, through crossposting and the exposure this brings, and not imposing moderation, can benefit the visibility of an OSN on Usenet and attract new members, it may also widen the range of topics and diversity of participants in a way that reduces the effectiveness of community through people building relationships with affinity, shared aims, and longstanding participation. I find tentatively that longer-standing participants may have a particular role, operating as they do within social norms to reproduce them for new members, and engaging with higher participation threads and new participants' messages. I would wish to complete more research regarding this tension and its impacts on online community development.

The interplay of Usenet's structures and freedoms, affordances and how these enable cohesion and incivility, discussed here and in the preceding subsection, affects whether *elements* of community can be present in Usenet's OSNs. At times they certainly can and the methodology identifies the following (pp.59-61 and pp.73-77):

- The development and persistence of relationships between participants as represented through their discourse
- The stability of core participants' membership
- The identification of shared group aims, represented in cooperative discussions and ongoing participation
- The accommodation of disagreement

- The apparent presence of shared faith, indicated by group practice such as prayer, shared language, and topics of discussion

However, on reflection following completing this research, I feel cautious about using the term community going forward. First, only certain *elements* of community are present, insofar as it is a diverse umbrella term which manifests in different ways for different people (pp.72-73), but also because community appears to be a term used predominantly in a positive way. Anecdotally, I have not noted from literature any uses of ‘community’ to describe a group negatively. This matters, since cohesion and incivility (in their various forms) coexist in diverse ways on Usenet OSNs. A contributor may form longstanding social connections through agreeable discussions about religion, but in doing so also be exposed to connections with people antagonistic to their religious beliefs and practice, or other aspects of their experience such as their LGBT+ or ethnic identity, along with exposure to adult content and spam messages. The use of ‘community’ does not, I argue, capture this coexisting positive and negative experience which, my data indicate, would be fairly common in many OSNs. The presence of authority is more clear cut insofar as: early actors had authority to shape later participants’ experiences; administrators on the early Usenet – when access was at fewer more central locations such as universities or government research institutions – could enable or withhold access to certain Usenet OSNs or hierarchies; moderators can act as gatekeepers to permit or block content from being posted; and religious texts and leaders were at times cited as authorities such as in the Bible study discussion group.

6.6.2 Regarding authority

In part building on section 6.5, I identify the following major findings that relate to authority.

Behavioural norms affect engagement and their reproduction through socialisation, moderation or adherence to FAQ documents (which set out how participants engage within an OSN) may be seen as an act of authority by longer-standing or otherwise influential members, such as moderators. This is a different nuance on authority than I typically found covered in academic scholarship.

Relatedly, and as with community, early participants’ actions can affect those who join later. In one example, gatekeepers blocking access to talk.religion – in doing so exercising authority over digital access – likely contributed to those OSNs’ less-established nature compared with ‘soc’ groups for people later joined in larger numbers and found ‘soc’

groups to be well-established and highly-populated, since those groups had exposure and were well-connected in Usenet's network. Here then, authority over technology shapes access to online religious spaces and discourse to some degree.

On more familiar ground, OSN moderation, as an expression of authority, can influence the connectedness of that group to others. In my case study of `soc.culture.jewish.moderated`, I found far lower levels of crossposting of messages between that group and others, compared with other groups in my research. This provided tighter boundaries, enabling closer management of participation, the subjects and tenor of discourse, and maintaining focus on the group's aims to connect and discuss regarding Jewish culture.

Further, expectations may be imposed on participants within religious OSNs, where participants are from that religion themselves. This is an example of the authority of religious tradition or texts, reflecting how, in the theory of religious social shaping of technology, there is co-shaping of digital spaces through technologies' affordances and the influence of religion.

Less clear cut but, I argue, still a form of authority, is longstanding participants' influence over others by shaping experience through replying, or not replying, to others' messages and (at times) choosing to reply on-topic and keeping their cool when doing so. I note that this has particular (potential) influence when engaging with new people hooked in by messages crossposted between the groups I analysed and other groups, potentially focusing on wildly different topics. They can shape discourse in ways which encourage others to engage along those same lines, such as by cultivating long discussion threads on interesting subjects (to which others may then also reply), ignoring irrelevant or otherwise undesirable messages (which then reduces their visibility to others since threads remain short), and socialising others into the language, humour or other ways of behaving which may then be reproduced. This indeed is how longstanding contributors cultivate cohesion and elements of community (see the previous paragraph).

6.7 Significance of findings for future research

I argue that my present research indicates the usefulness of historical OSN interactions for analysing social processes that occur online. If that is the case, then I must indicate potential broader applications to affirm the relevance of this approach for further research. I identify three ways in which I could broaden this research: analysing other social processes,

research other social networking platforms, and analysing subjects other than the religions on which I focus presently.

6.7.1 Using the methodology to analyse various social processes

While I focus on cohesion and incivility, I could adapt the methodology to focus on other social processes. I have not written the necessary code but have specified how I would identify the following social processes in archives of online social network messages:

Social processes that bring individuals or clusters of participants together:

- Cooperation between participants in terms of discussion on shared interests
- Accommodation of two or more overlapping or opposed social networks within a platform (in terms of people or topic)
- Assimilation of two or more social networks together
- Accumulation of participants or engagement over time

Social processes that may fracture clusters of participants:

- Competition for engagement
- Contravening behavioural norms (flooding a group with messages, off-topic messages etc.)
- Differentiation between clusters of people/social networks developing over time
- Disintegration of social networks and decreased engagement
- Isolation of individuals or clusters of people within social networks

Therefore, the methodology could be used to analyse the consequences of social network interactions for (online) social lives that are broader than I undertake in the PhD.

6.7.2 Researching other online discussion group platforms alongside Usenet and Google Groups

My work focuses on archives of messages sent across Usenet and Google Groups, but I can adapt it to analyse archives from other prominent historical platforms such as Yahoo! Groups, which was similarly popular, or users engaging with one another via comments on other forums or websites. The challenge of platformisation in which owners of current social media platforms restrict or prohibit researchers' access to data (pp.99-101) increases the challenge of research on current dominant platforms and increases the relevance of seeking

out ethically curated historical archives from different platforms, especially since I have observed ways in which behaviour on historical platforms is relatable to how people interact on social media today. To incorporate other platforms' archives, I would need to adapt the code that extracts messages so it can compile datasets from archives which have different formats.

6.7.3 Researching other subjects

My research analyses interactions focused on religion, but the methodology could be used to analyse the substance and social impacts of interactions in other subjects. Since, my conceptual framework is drawn from studies of Usenet and digital religion, it may be that the methodology is best suited to studying social contexts which function similarly, such as new religious movements and millennial groups. For example, I could research historical interactions regarding political or religious groups which are of interest to government agencies, contextualising those groups' current activities. Or I could research online discussions about climate science, identifying what may help draw people's interest and persuade/nudge them toward more environmentally friendly behaviour. Additionally, I can research religion at the intersection of other topics such as LGBT+ lives and religion (as for my forthcoming work with *Explaining Atheism*, a John Templeton Foundation funded project for which I am researching how historical online social networks catalyse engagement with atheism). Although historical online social network platforms have been superseded, I argue that my research shows how vast troves of data illustrate how people interact and how their online social lives and discourse are, in turn, influenced by interacting online.

6.7.4 Collaborating with others

While I will undertake potential future research case studies myself, it could be very fruitful to reduce barriers to others so they can analyse message archives to research social processes without the need for specialist computer programming. This could unlock collaboration between myself and other researchers at a greater scale.

With this in mind, I will seek to broaden my impact by incorporating my computer code and datasets of historical online interactions regarding religion (including data from c.1981-2015) into a user-friendly online interface. I am researching how I can use R Shiny to create an application which would afford this while maintaining strong ethical safeguards in line with the approach I take here. This will enable other researchers to collaborate with me by

accessing sociological insights from historical social network interactions relating to shared research interests. The interface would provide access to a range of historical online social networks that focus on mainstream and minority religions, and groups that are functionally similar to religions such as millennial or militia groups. The visibility of minority religions online may be impactful, since research indicates that exposure to such religions via social media may influence people's awareness of those groups and willingness to engage with them (Masood and Skoric 2023). This could, consequently, influence social relations between the minority group and others. I will also seek to broaden analyses to include the broader range of social processes specified.

6.8 Conclusion

My empirical findings in chapters four and five, following my original analyses of data, underpin my above conceptual discussions here. This empirical and conceptual work represents an original contribution to knowledge in the field of digital religion studies, interfacing with extant research on how community and authority manifest online. In addition, my methodological approach represents a contribution to knowledge of what is possible at an intersection of digital religion studies, internet archiving and computer coding. I summarise this in the dissertation conclusion which follows, along with my contributions to developing an ethical framework for such work. I hope to broaden the impact of my work in future by reducing the barriers to accessing ethically sensitive analyses of historical OSNs. I am particularly interested in illuminating the experience of marginalised groups, since my research has shown how those connecting online to discuss religion — who may themselves be a minority group in their wider society — can also experience marginalisation online at intersections of, for example, LGBT+ or ethnic identities. By better understanding people's experiences of online social networks, we can make recommendations to improve cohesion and reduce harms. I have found that many actors (summarised above) can have a role to play here, and although Usenet has faded from prominence among other social media platforms, it has some overlapping affordances with other formats of online discussion forum and services such as Reddit and Discord, so its data remain a test laboratory for understanding some ways in which people connect and behave online. The data therefore remain useful for not only understanding the online past but identifying ways in which online interactions could be shaped for the better today.

Reflections in conclusion

The conclusions of chapters four and five (pp.192-194, 230-233) indicate how my empirical findings, and interpretations of them, answer my research questions. Chapter six draws higher-level conceptual conclusions, showing how the breadth of my conceptual framework (developed from chapter one onwards) informs my analyses of the development of Usenet online social networks and the engagement and social processes I observe within them. Both summarise the original contribution that my empirical findings make to scholarship. Here, I discuss my engagement with computational methodology and ethical considerations. While the methods used and ethical safeguards employed are already well-established in academic research, my research entailed applying them in a novel context – large historical archives of Usenet OSN interactions.

7.1 Deploying computational methodology and ethical safeguards in my research context

As summarised above, the computational methodology I deployed to address my research aims is subject neutral and can be used to research myriad historical OSNs and, with some additional code, a range of social processes. I am immensely grateful for Dame-Griff's (2019) paper which, while using Python (rather than R) and meeting different research aims so not delivering the full range of insights afforded by my approach, was an encouraging and informative discussion of how a researcher can develop a computational methodology in this space. Notwithstanding its limitations (pp.136-140), the salience of the methodology deployed (presented in detail in chapters two and three) is affirmed by engagement I have received from other scholars after my conference presentations, particularly at the American Academy of Religion annual meeting and at the International Society of Media, Religion, and Culture doctoral colloquium, both in 2023. It is further affirmed by my work with *Explaining Atheism*, funded by the John Templeton Foundation at Queen's University Belfast. This research entails applying elements of the methodology to historical online interactions regarding atheism online, often at the intersection of religion focused OSNs and others, such as those focusing on LGBT+ lives. In this research, using elements of the methodology, I identify what draws people into engaging with atheism focused OSNs and add some empirical data to existing academic understandings of the importance of early social media for such engagement which could help catalyse the loss of religious belief and identity, and connect and resource those doubting their faith. This work has already been presented as an academic poster at Queen's University Belfast and as a conference

paper at the American Academy of Religion annual meeting. I am also presenting an invited seminar and conference in 2024 and further publications are forthcoming. These are tangible outcomes beyond the PhD, made possible by my learning and deploying the computational methodology.

The ethical safeguards I deploy required consideration that the research sits at an interdisciplinary nexus that affords the use and analysis of data which are not amenable to ethical frameworks for undertaking ethnography of *current* online interactions for which permissions and consent would be sought from community gatekeepers or individual research subjects (i.e. replicating, to some degree, ethical protocols well-established for human subject research in offline contexts). Nonetheless, ethical safeguards were required beyond the considerations necessary for cleansed and pre-published quantitative datasets, such as the European Values Study or other similar data. My approach is embedded in my methodological chapters two and three (and discussed particularly in sec. 3.6, pp.140-146), but here I summarise key aspects, affirming that other established ethical safeguards could be deployed. My guiding principle was to embed privacy by design and ensure proportionate ethical safeguards for each artefact that my research generates: my computer code, Message Dataframe, 1,000 Thread Dataset, quantitative analyses, the dissertation, presentations, and publications. Here, privacy meant pseudonymisation/anonymisation and data minimisation, retaining only the data necessary at each stage for fulfilling the research objectives (see pp.87-88) and aggregating where suitable and beneficial. The code I wrote in R is key here, since in generating my Message Dataframe it replaces usernames with randomly generated alphanumeric strings and replaces messages with analytic descriptors, specifically prominent and frequently used terms, and a categorical indicator of whether the message relates to religion discussion. For my 1,000 Thread Dataset, I computationally redacted strings of text that fit the format of full names, URLs, and email addresses. All these steps required me to learn and adapt coding techniques to suit my context and were necessary since manual approaches were infeasible due to the size of the datasets. I ensured my natural language processing was completed satisfactorily and then excluded granular personally identifiable data from my datasets. In the dissertation, presentations, and publications, I go further by reporting on trends and aggregating data when presenting quantitative insights in figures such as the duration of contributors' participation within a specified OSN and social network graphs. To strike this balance, I learned R functionality, reflected critically on approaches taken in other research which uses historical archives of online data, considered emerging legislation and GDPR in particular, learned data encryption techniques and best practice within macOS and Microsoft environments, and discussed my intended approach with my

university's convened ethical review board meeting before proceeding. This ethical approach emerges from this engagement between technology, data sources, research aims, and the sensitive nature of the subjects I research. It is relatable to the research which I intend to undertake in future and I promote it in papers when presenting my research findings.

7.2 Metareflexivity - coding to analyse digital artefacts

This conclusion now takes a step back to reflect on computational methods for studying digital artefacts and, consequently, the strength of learning to code in order to sit at the intersection that I occupy. This leads me to consider how I identified a temporary disciplinary location in dialogue with coding and internet history literature. The position of my work in the academy informs — and provides critical perspectives on — my use of coding to research social interactions online. I also reflect on how the project developed over time, leveraging and responding to developments in computational methods and internet history as a discipline during my candidature. I close by exploring how I seek to cultivate future impact through stepping beyond this space to others. The purpose of these observations is to articulate my development as a researcher in an environment which can discretise academic specialisms but which, if a researcher bridges between them, can afford innovative and practically useful scholarship.

Throughout my research, I was alert that the interactions I analysed were shaped by network infrastructure and technologies which I term a digital substrate (pp.28-30), and that the archives I used were shaped by individual and commercial archivists' actions. In addition, the methodology illuminated certain facets of data and social processes. This meant that the decisions for my digital methodology were shaping my engagement with OSN participants' digital interactions which were, by the time they reached me, shaped by digital archiving practices. I had to be reflexive about how these facets of my work interplayed and this required me to understand how each of them influenced my empirical data and what was possible with it, thereby shaping research design — and specifically coding — to capture what I required for my research questions with sufficient accuracy.

I was able to do this with confidence as I learned the coding necessary for this work. Humanities and social science scholars may collaborate with computer scientists to undertake computational humanities or social research. However, by developing enough capability in both of these domains, to the extent required for this research, I was able to ensure that my religion-focused research questions, qualitative analyses, and

computational methodology combine to form one cohesive methodology in which I fully understand all components and how they fit together. I argue that this is a strength of learning to code and consequently situating myself at an intersection of disciplines. This required a particular investment in learning the coding and this results in methodological chapters two and three.

7.3 Interdisciplinary work in fast-moving fields

In occupying a disciplinary location within digital religion studies while in dialogue with coding and internet history literature, I was exposed to the rapid development of these fields during my PhD candidature from 2018-2024. I observed the advance of computational methodologies in two related ways. The first is that greater processing power, particularly across distributed systems connecting resource from multiple computers, and innovations in how data can be fed into programs with more data in parallel, led to larger language models being produced than models such as BERT (pp.267-269), with which I experimented, and the style of word embedding I used (pp.112-113) to support the development of analytic descriptors of messages in my dataset. These models are also accessible, most notably at present in ChatGPT. However, these are computationally expensive and lack perspicuity. I found my approach to be sufficient, proportionate, ethically defensible (a challenge with large language models which send data offsite for processing), computationally lighter, and more environmentally friendly as building large language models entails many gigabytes or terabytes of data transfer alongside use of server farms, widespread use of which is responsible for significant carbon emissions.

Internet history also evolved with, for example, new outreach from the Web archives team at the British Library, the ongoing development of the Web archive analysis toolkit developed by Archives Unleashed, and Dame Griff's (2019) computational analysis of transgender Usenet OSNs. Taken together, these developments provided me with the challenge of a moving foundation on which to assemble the PhD, though they also provided a strength to my work in two ways.

First, the development of new computational innovations gave me new natural language processing approaches to consider for my research. This afforded me careful consideration of what is right for my project from the panoply available. This strengthens my work, rather than defaulting for the approach which is in vogue at the time. Indeed this can happen, as I observed waves of research first using Latent Dirichlet Allocation topic

modelling (see Technical Appendix One) and then word embeddings (pp.112-113). Instead, my approach to analysing messages in the manner required for my research synthesises elements of different approaches to meet my needs (pp.108-113).

Second, advances in the field of Internet archiving enabled me to appreciate the distinctiveness of my research while positioning it within a broader field which is primarily, though not exclusively (cf. Milligan 2013; Paloque-Bergès 2019; Dame-Griff 2019), focused on archiving and analysing webpages rather than OSN interactions. From that research, I learned about the analyses people are undertaking such as link analysis, in which forms social networks based on webpages which link to each other (cf. Meyer *et al.* 2017) or contrast historical webpages to observe the portrayal of, for example, news stories over time. This encourages me to play to the strengths of the methodology by leaning into the research projects that leverage its distinctiveness, such as my work with Explaining Atheism.

Lastly, I close by reflecting on the position of this dissertation within digital religion studies. This is a fast-moving field and presently heading into to a fifth wave of research, which is distinguished by studies of existential questions of living in the digital (Campbell 2023). It seems to me, from papers I have heard at specialist conference panels reflecting the latest research in this field at the time of writing, that my PhD research case studies and work on Explaining Atheism, fit within digital religion studies by virtue of their empirical focuses on (non)religion and how I orient my PhD research to help understand how community and authority — key concepts in digital religion studies — manifest in online religion-focused OSNs. Nonetheless, my interest in religion is for how it is a factor in the formation and evolution of online networks and shapes people's interactions within them. This broader landscape, within which religion fits — as do, for example, my other interests of science and LGBT+ lives — is the field in which I seek to research in future. In doing this, I hope to use my literacy about religion and these other fundamental aspects of human discourse and identity, and which form social groups, worldviews and lead to progress and harms in society, combined with the potential of computational analyses, to engage with other fields of research in collaboration with other scholars. For example, at the time of writing I am tentatively scoping a project at an intersection of religion, sexuality and accessing sex education online, collaborating with psychology professionals. I hope that such scholarship will broaden the reach and appeal of my research agenda. If so, it will have been sparked by the present research, the intellectual and professional development it has afforded me as a researcher, and the toolkit that I have started developing through my

candidature alongside a wider community of coders, internet archivists and scholars of (non)religion. For this, and for those scholars, I am immensely grateful.

Technical Appendices

TA 1.1 Common facets and challenges of ML NLP, with reference to a dominant approach

Discussing various algorithms in some in detail justifies my choices and indicates where my approach fits into the methodological landscape used in computational NLP research, as is done in other doctoral dissertations that focus on computational methods for mining data from text (cf. Cheng 2017; Zhou 2017). The detailed discussion assures the reader, I hope, and avoids the error of seeing algorithmic processes as autonomous ‘thinking’ agents that bestow insights in the manner that a human makes sense of the world (Krasmann 2020, 4).

I discovered algorithmic approaches that may be useful for extracting analytic identifiers through engaging with literature in preparation for my research. This was not exhaustive, due to the size of the field, but through my reading I reached a saturation in which I was no longer discovering approaches that were fundamentally different from those I had encountered before, among those suitable for the size and aims of my project. The scope for my reading included academic books on data mining and machine learning, digital humanities and internet histories scholarship, and computer science articles. I also learned through the courses I completed and the formal documentation that accompanies R. I then experimented with configuring and applying a number of these approaches to contrast their suitability, since their success is relative (Kim *et al.* 2018). Nonetheless, my understanding of how the algorithms operate technically was crucial in assessing their suitability and I explain the relevant facets of select algorithms in this section to argue for my choices.

I refer frequently to Latent Dirichlet Allocation (LDA) in this subsection. It has been a popular algorithmic tool in recent years for NLP and a base for related tools, but it has certain assumptions and limitations. Referring to LDA allows me to introduce various important facets and challenges that are common in NLP, while justifying not using LDA myself. With this in mind, I then introduce select other approaches to analysing texts and my chosen methodology, affirming the strengths of my approach through contrast. I show how my work is aligned with, and differs from, computational approaches used in other research and how it addresses common NLP tasks and challenges.

LDA is a form of topic modelling, which identifies subjects of discussion in a dataset of documents that focus on a range of different matters. It assumes there exists a number of

'topics', which means something different from the common use of the term. Here, topics are sets of words from the corpus of words used in the documents across the dataset, and which are expected to co-occur within documents and not be distributed evenly across all documents. For example, a topic about artificial intelligence may include the words algorithm, artificial, intelligence, ethics, neural *et cetera*. LDA may be considered a form of factor analysis (Taris 2000; Bainbridge 2020, 254-260), which is to say it seeks to identify *latent* factors that have structured the dataset or informed its content, in this case the topics. It has been used successfully to analyse online discussion of religion, as in Elwert, Tabti and Pfahler's (2020) research of Evangelical and Salfist discussion forums which showed changes in the presence of topics over time once the researchers had developed models of 100 and 200 topics and opted for the latter. Sanders and Ferré's (2020) analyses of responses about Answers in Genesis' Ark Encounter on Reddit used a model with 13 topics. Specifying the number of topics that are present is a challenge that I discuss later and address through my approach, since the topics are considered to exist separately from the dataset that they structure. In practice, this is not how a person reading texts would summarise topics; they would likely infer them as a property that emerges *from* the dataset.

NLP very often requires pre-processing to tokenise the text, resulting in documents comprised of individual words that the algorithm(s) consider separately, rather than sentences or paragraphs. Certain words that contain limited or no semantic content within the resulting 'bag of words', such as conjunctions 'and', 'but' or 'if', are then removed algorithmically based on a pre-defined list. These are termed *stop words*. I removed such words algorithmically, using open-source lists of names, in order to exclude these personal details from my analyses. To be clear, words such as conjunctions may of course be important to the structure of a sentence, but when considered in isolation they are less informative. The words are often then stemmed to a root, so words similar in content are represented with the same word. This can either be done based on the word structure, in which an algorithm removes common suffixes as in R's 'Snowball' stemming function, or based on a dictionary defined or imported by the researcher. For example, chair, chairs, chairing, and chaired may all be stemmed to 'chair'. In summary, *tokenising* documents into (here) words, *removing stop words* and *stemming* comprise three pre-processing steps that are common to many NLP algorithms including LDA. I use all three steps in the methodology.

Focusing specifically on LDA for two paragraphs, I now summarise its technical process in a non-mathematical manner, drawing on Lesmeister (2019) and Blei (2012), official R

documentation and sources that apply LDA (including those cited above). This illuminates its strengths and limitations, which I considered when assembling the methodology. An LDA algorithm is applied iteratively to predict the words contained within the topics that, it is presupposed, underly and give rise to the documents. The algorithm first assigns a random topic number (from total number of topics specified) to each word in each document in the corpus so words that occur frequently result in a document being affiliated more with that topic. This is intuitive since words representative of a document's topic are likely to occur more frequently in that document than in documents unrelated to that topic. The name Dirichlet comes from the term Dirichlet distribution, which refers to this prior (starting) distribution of topics across the corpus of words and documents. This weights topics to documents at the first pass and results in two calculations: first, a summary of how many times each word was assigned to each topic, for example the word chair may have been assigned to topic two seven times and topic four twice (in practice, there may be hundreds of topics); second, a summary is calculated listing how many times each topic appears in each document, based on the assignment of words (within documents) to topics.

The following process is then undertaken for each word in each document. First, the topic number assigned to the word is removed and the two above summaries are recalculated accordingly, reducing the count of (1) how many times the topic was assigned to that word and (2) the number of times the topic appears in that document. The word is then reassigned a topic number based on a calculation of how many times the word was assigned to that topic across the corpus multiplied by how much each topic is represented in the present document, based on other assignments of topics to words in that document. The word is reassigned to the topic that delivers the largest output from those calculations and this is repeated for each word in each document in the corpus. This process is undertaken iteratively and, consequently, words are affirmed probabilistically as part of a topic. It is probabilistic as words remain assigned to multiple topics, though may converge around one topic where a document has been identified as having a high presence of a topic and the word is often identified with that topic (as may happen over multiple passes). Therefore the output indicates the overall probability of each word being present in a latent set of underlying topics. Researchers typically then take several of the words, often twenty, that are most probably indicative of each topic and analyse them to identify the subject to which they relate. Since LDA also calculates the inferred proportion of each document that is allocated to each topic, the researcher can then identify topics present within each document. Of course, topics here differ from an intuitive human understanding of a topic. I expect a human would rarely conceive of a topic as a set of words distributed over a corpus.

The researchers who developed LDA recognise this, stating that they ‘refer to the latent multinomial variables in the LDA model as topics, so as to exploit text-oriented intuitions, but we make no epistemological claims regarding these latent variables beyond their utility in representing probability distributions on sets of words’ (Blei, Ng and Jordan 2003, 996).

Returning to a challenge I mentioned, how do researchers undertaking topic modelling identify a suitable number of topics? A starting principle is that a higher number is more likely to capture the granularity of language used (Xu *et al.* 2019, 567-570) though this also has potential to add significant noise as there may not be enough diversity within the corpus of documents for each topic to have distinctiveness. Topics could nonetheless be grouped together by the researcher post hoc. Alternatively, researchers may specify a number of topics used commonly in other research (Brookes and McEnery 2018, 7) but since the diversity and number of documents varies extensively between corpora, this is not a reliable approach to yielding optimal results. Preferable to this is applying the algorithm iteratively with different numbers of topics and reviewing the semantic coherence of the resulting topics, such as by considering how they reflect the co-occurrence of words within documents (Fonseca *et al.* 2019). Similarly, Genovese (2018) modelled topics from Pope Benedict XVI’s Twitter messages and computed the complexity score which indicates how well the set of topics will represent the actual distribution of words in a portion of the corpus that was excluded when the LDA model was calculated.

Short texts can also pose a problem for topic modelling, including LDA, since they may not contain words that become representative of topics. Researchers have developed responses to this challenge, such as combining texts by the same author or inferring topics from longer texts or external sources such as WordNet and then applying the resulting model of topics to the corpus including shorter texts (Yi, Jiang and Wu 2020). WordNet is a database that clusters semantically similar words (Princeton University 2010); employing tools such as WordNet can help bridge a gap between short texts that may refer to a topic also present in longer texts in the same corpus but which use semantically related words. Relatedly, researchers may append hypernyms to a document (Xu *et al.* 2019) to help link the document to others by using words that refer more generally to a topic and are consequently more likely to be present in other texts discussing that topic.

In addition to the challenge of short texts this brief summary indicates limitations of LDA, which provide starting points for considering other approaches to NLP and which, consequently, I must consider in light of my research aims and datasets. First is that LDA

is stochastic, since it has a randomised Dirichlet distribution and predicts topics probabilistically. This means the topics inferred will, by default, vary to some degree each time the algorithm is run using the same data. This makes it difficult to (1) know if the topics are optimal so whether the results should be re-run and (2) see clearly why LDA arrived at its resulting topic set, since it would arrive at others when using different Dirichlet distributions. This relates to another limitation, namely the challenge of knowing when to stop running the LDA process — how many iterations/passes of (re)assigning topics to words and documents are optimal?

In summary, I have introduced various facets of NLP that are common to many approaches that researchers take. These include:

- Pre-processing
 - tokenising
 - removing stop words
 - stemming words
- Perceiving documents as ‘bags of words’
- Assigning words to or one more topics/groups of words

And I have introduced the following challenges:

- Extracting topics from short texts
- Identifying if stochastic processes have delivered outputs that reflect the corpus’ underlying structure optimally
- Knowing at what stage to stop applying the algorithm and use the outputs to infer that structure (such as subjects of discussion)
- Specifying the optimal number of topics before commencing computation

I refer to these facets and the above challenges as I now discuss other approaches. This enables critical evaluation and justification of my chosen methodology for NLP.

It is said that big data science projects require researchers to be part artist and part scientist (Simon 2013, 16-17). Grappling with the above challenges requires knowledge of the domain being analysed, understanding of how one can test the accuracy of the algorithms employed and, in light of this, a sense of how well the language, numbers and visualisations that one forms from the results fit the underlying social reality captured in the datasets. This

shows the reliance of the researcher on algorithms for their creations. I sought to explore algorithmic approaches that, while they have other strengths and weaknesses, do not incorporate the limitations I summarised above and grant fuller insight into, and control over, the creation of analytic descriptors that I can add to my dataset. These descriptors include attributes of the messages' language and subjects, helping me achieve my research aims. This reflects Nanni's observation, while discussing LDA, that 'researchers should not employ a computational tool because it is widely adopted in the community, on the opposite they should critically question it, especially because it is so widely used' (Nanni in Brügger and Milligan 2019, 7).

TA 1.1.1 Clustering documents

Clustering documents based on the words that co-occur between them is an alternative means of identifying themes/subjects of discussion within a corpus of texts. One established approach to this, which I mention here given its popularity, would be to plot documents on a graph based on their semantic similarity and then partition the graph using a k-means algorithm (Lesmeister 2019, 197). K-means would cluster the documents by first assigning a random document as the mean, or centre, of each cluster. The remaining documents or, more formally, observations, would be assigned to the cluster mean to which it is closest. A two-step process is then applied iteratively. First, the mean would then be recalculated and located at the centroid of the new cluster, minimising the distance between observations assigned to that cluster. Second, the observations are then reassigned to their closest cluster based on the relocated mean. This process is repeated until the cluster to which observations are assigned no longer changes. There are however two principal limitations here. The first is that results may vary when the k-means algorithm is applied multiple times to the same dataset as each time the algorithm is run, the mean centres of each cluster are each assigned to an observation at random. Second, the researcher needs to specify the number of clusters ('k') prior to applying k-means, similar to how the researcher must specify the number of topics before running LDA. In addition, given my aims of identifying analytic descriptors, this approach requires substantial post-classification analysis to identify whether the partitioning of the documents into groups really reflects distinctions in their themes/subjects, as imparted by their use of overlapping terms, and then labelling each partition accordingly while considering the possibility of boundary cases which don't neatly fit the cluster to which they have been assigned. Consequently, having evaluated its potential and limitations, I chose not to use an approach which clusters documents via k-means or a similar algorithm.

TA 1.1.2 Producing word embeddings

Word embeddings are digital objects that represent words in a corpus numerically by plotting the words within a high dimensional space so that words' spatial proximity within the space (termed a vector space) represents the closeness of some facet of their relationship to each other. This is often their semantic relationship so that, say, the terms religion and spirituality would appear closer to one another within the space than the words religion and cricket.

Word embeddings are trained by analysing a large corpus of text, often to identify word co-occurrence within and across texts, for example by iteratively improving the weightings of words, which indicate the words' locations in the vector space, through training that seeks to predict words that have been masked in the training data. State-of-the-art embeddings can capture different facets of a word's meaning, potentially enabling them to distinguish between different meanings of a word depending on its surrounding context.

When exploring methodological options, I tested generating a word embedding in which semantic similarity between the *subject lines* of messages is represented by those subject lines having spatial proximity when they are plotted in high dimensional vector space. I undertook this test using a small sample of 100 dummy message subjects that are not part of my PhD dataset. I plotted the embedding of the subject lines using Google's uncased base BERT model, made accessible by the AI company Hugging Face and via the 'text' package of software in R. BERT is a neural network 'deep learning' model and is an acronym for Bidirectional Encoder Representations from Transformers. The uncased base version of BERT does not distinguish between upper and lower case in text. The acronym refers to how BERT was trained on a large corpus of text from BookCorpus (11,038 texts) and English Wikipedia (Hugging Face, no date) to develop a model of the English language using sentences in which the location of each word in the sentence was specified, rather than considering sentences as bags of words (Devlin *et al.* 2018). This reflects BERT's early function for translating texts (Vaswani *et al.* 2018) as the location of words in sentences may differ based on the languages used, so the locations must be specified as part of the input. The term transformers refers to a means of processing the input in which all words in an input sentence are processed at the same time, enabling the *attention* part of the process to weight the focus on different words which contribute to the context of the sentence. This way of processing, rather than taking single words as inputs, accelerates the rate at which

training texts can be ‘fed into’ the model, thereby increasing the number of texts which can be used for training and enhancing the model’s representation of the English language compared with previous state of the art neural network approaches.

BERT uses these position embeddings along with attention and self-attention which in part see the model learn, through masking 15% of input words and seeking to predict them, which *other* words in the sentence require attention to predict the missing word. This affords contextual understanding that disambiguates polysemous words. For example in the sentences ‘the train was late’ and ‘I will train my colleague’ the words surrounding train afford disambiguation to which BERT can be *attentive*. BERT was also trained by concatenating sets of two sentences and seeking to predict whether the two sentences actually followed one another in the original text (Devlin *et al.* 2018). This further enhances BERT’s representation of words’ semantic relationships based on their usage in English language.

Using BERT via R to train my word embedding enabled me to generate an embedding of unique message subject lines, in which each subject line is represented by 1,536 numbers. For example, a test of this methodology produced -0.518, 0.231, -0.124 and 0.193 as the first four numbers. Taken together, the numbers representing each unique subject can be used as coordinates to plot their location (i.e. embed them) in a hypothetical high dimensional vector space. The locations of pairs of subjects can then be contrasted to identify those subjects’ semantic proximity. These locations are informed by BERT’s pre-training on the corpuses of English text summarised above. Consequently, when the subject lines from my dataset are provided as input the model, the coordinates of subjects in my dataset can have proximity if they are semantically similar but do not contain the same words since their locations are generated based the much larger corpuses used in BERT’s pre-training. This is an advance from LDA, which relies on word co-occurrence within the corpus being analysed, and enabled my word embedding to identify semantic proximity in short texts which do not share any specific words.

BERT has been used in research publications for NLP tasks related to my aims, such as Egger and Yu’s (2022) paper on BERTopic which is an approach that uses BERT to infer topics from a corpus of texts, as may be sought using LDA. However, for my research project I identified some shortcomings to this approach. First, it is computationally expensive and time consuming to generate the embedding for each subject line and contrast them to identify those which are similar. Indeed my choice to use *subject* lines reflects constraints

in this approach — it is not suitable for entire messages and requires substantial computational resource for even short texts. Second, this ‘cost’ has environmental burdens when relying on large energy-consuming server farms and I argue that researchers have a responsibility for the environmental ethics of their methodology. Other approaches, as I shall demonstrate in this dissertation, are suitable for my tasks at hand. At the time of writing, large language models (such as ChatGPT) are performing well in a range of natural language processing tasks such as question answering and text summarisation. However, researchers must select appropriate and *proportionate* methodologies, especially given the environmental impacts of large computer server farms on the internet. Such server farms are typically used for some of the mathematical heavy lifting required. Third, and relatedly, using external servers brings ethical concerns about processing research data on computers under the control of a third party beyond the researcher’s gaze, including how the data may otherwise be used by that third party, who would have access to it, and how it would be stored and deleted. My ethical approach keeps tight control over the pseudonymity/anonymity of my data and its secure storage, so I considered the use of external servers incompatible with my aims.

Fourth, using subject lines *exclusively* would assume that the subject accurately reflects the substance of discussion. In fact, I observed some messages within my dataset that do not reflect the tone or substance of the subject line, either because the subject line is inaccurate or because the subjects/themes changed over the course of that discussion thread. Consequently, after experimenting with large transformers-based models, given these limitations, I chose not to use this technique though I did use a more economical (but still sufficient) algorithm to generate a word embedding for part of my approach. I will now summarise the approach I use, along with my wider methodology for analysing messages computationally.

TA 1.2 Association Rules Mining

To summarise association rules mining, I require a dummy example of an association, so I imagine the association between a message including discussion of a minority religious group (‘x’) *and* indications of incivility (‘y’):

Support: the proportion of messages in the dataset that indicate both discussion of a minority religious group *and* incivility, expressed as the number of such messages divided by the total number of messages in the dataset. This indicates the extent to which an

association (i.e. a rule) occurs within the dataset. A rule with high support may be interesting, but equally may merely be obvious and of little explanatory use. If discussions of a certain minority religious group consistently indicate incivility, this suggests a potentially interesting research focus but is simply a starting point and researchers may then be more interested in instances where the former is *not* associated with the latter, to understand why.

Confidence: the proportion of messages in the dataset that indicate both discussion of a minority religious group *and* incivility, expressed as the number of such messages divided by the total number of messages that indicate discussion of a minority religious group. This indicates the likelihood of incivility occurring when the minority religious group is being discussed and provides a good measure for the usefulness of a rule.

Lift: identifies how many more times the association, in this case 'x' discussion of a minority religious group and 'y' indicators of incivility, occurs than would be expected if x and y were statistically independent in the dataset. To obtain the lift statistic, the confidence figure (see above) for the association between x and y is divided by the result of multiplying the proportion of the messages indicating x with the proportion of messages indicating y. While an association that has low 'lift' but high support may still be relevant to my research interests, the lift statistic helps identify those rules where two things appear related, since they occur more frequently than would otherwise be expected. This potentially indicates, where considered across my large datasets, social phenomena that might be analysed to understand *why* they are occurring together — if such outcomes are of concern, say incivility around discussions of certain minority groups, recommendations might be possible to help mitigate their negative effects. This measure also helps avoid Bruce's (2018, 204-207) concern that a correlation between quantitative data provides a starting point, but that it must be explainable in terms of how people actually behave. With the lift statistic, we can evidence that the social phenomena being analysed are strongly associated beyond what would be expected by chance, so we have a credible insight into social behaviour for which we can then seek more understanding. If the association was not credible, it would be difficult for researchers to explain it plausibly.

While my research generates traditional association rules, other approaches may be used. Fan *et al.* (2015) identify graph-pattern association rules (GPARs) for use in social media marketing. For example, let us take a group of three friends in which two are members of a certain OSN. There is some probability that the third person will in future join that OSN due

to the social network property of homophily, in which similar people find affinity and form social networks (birds of a feather flock together). GPARs could model such a scenario using social network graphs, but given the strength of insights derivable from applying traditional association rules to my datasets, and that my ontology of data and epistemology of what I can 'see' through my data are perfectly explorable through traditional ARM, I took that approach.

Nonetheless, within ARM a broad range of algorithms may be applied to identify potentially interesting association rules. I must therefore briefly explain and justify my choice. Ghafari and Tjortjis (2019) survey algorithmic approaches to mining association rules. In addition to the Apriori algorithm (Agrawal and Srikant 1994), they classify approaches as algorithms based on Apriori, heuristic approaches and other algorithms that are not based on Apriori. I chose to use the Apriori algorithm. Its strength is that it is an exhaustive approach to mining itemsets within parameters that the researcher specifies, but this means it is also computationally intensive. A range of heuristic approaches seek to exploit duplication or evaluation of otherwise unnecessary itemsets, and some start with some initial limited distribution of itemsets to be analysed, before iterating to analyse more (Ghafari and Tjortjis 2019). These are particularly useful when very large datasets are being analysed though my work is bounded to my selected research sites, closed to new data for the purpose of my current project and, relative to the world of big data machine learning, is not exceptionally large. Consequently the exhaustive nature of Apriori helped ensure I answer my second research question robustly and the computational power required was still manageable. One feature of Apriori that required particularly careful consideration is that the researcher must specify the minimum levels of confidence and support (explained above) that rules must have. This is to help identify signal among the noise generated by the huge numbers of association rules that Apriori delivers. The algorithm, implemented in R, starts by extracting one itemset — a combination of elements that co-occur in the dataset. In my case this could be between indicators of certain analytic descriptors indicating a subject of discussion and the length of a discussion thread. If this itemset has the minimum specified level of support then additional elements are added to the itemset for so long as the minimum level of support is achieved, identifying association rules that occur among elements (Lin *et al.* 2019, 700-701). Consequently, defining the parameters directly influences the number of resulting rules. I iteratively specified parameters for minimum levels of lift to deliver a number that I had the resources to review while minimising the risk that I would miss rules that provide valuable insight regarding the social processes in my datasets. I have confidence that I am unlikely to have missed significant associations

since ARM algorithms typically extract multiple rules that convey the same information to the researcher (Gahar *et al.* 2018, 228). While this introduces redundancy, it mitigates against missing valuable insights.

TA 2.1 Anonymised extract of 1,000 Thread Dataset

Samples are included here to indicate the format of the 1,000 Thread Dataset and the content of labels I applied to messages within it. The messages are not reproduced here in accordance with the ethical parameters of the research; the message subjects provide a broad indicator of content.

Message subject	My note
Peace Requires Justice for Refugees from Palestine	Shared aims (intersection), geopolitics, Israel, Palestine, accommodating disagreement
Peace Requires Justice for Refugees from Palestine	Shared aims (intersection), geopolitics, Israel, Palestine, accommodating disagreement
Peace Requires Justice for Refugees from Palestine	Shared aims (intersection), geopolitics, Israel, Palestine, accommodating disagreement
The Black Stone Idol [Repost]	Shared aims (discuss religion), Islam, Quran, reasoning
911, What is the truth?	Conspiracy
911, What is the truth?	Conspiracy

Faith as Narcissism	Shared aims (discuss religion), reasoning, knowledge boundaries
Faith as Narcissism	Shared aims (discuss religion), reasoning, knowledge boundaries, accommodating disagreement
Faith as Narcissism	Shared aims (discuss religion), reasoning, knowledge boundaries, accommodating disagreement
Catskill Mountain Jews to Participate	Shared aims (intersection), Israel, Palestine, arguing
Catskill Mountain Jews to Participate	Shared aims (intersection), Israel, Palestine, arguing
Tawheed or Trinity ala Craig and God's love	Shared aims (discuss religion), philosophy, accommodating disagreement, reasoning, Trinity

Who are the rebbes?	Shared aims (discuss religion), cooperative discussion, accommodating disagreement
Help, How could God have....	Shared aims (religious), creationism, science, reasoning, cooperative discussion
Help, How could God have....	Shared aims (religious), creationism, science, reasoning, cooperative discussion
Help, How could God have....	Shared aims (religious), creationism, reasoning, critical
Help, How could God have....	Shared aims (religious), creationism, science, reasoning, cooperative discussion
Teaching non-Jews about Judaism	Shared aims (discuss religion), cooperative discussion, Judaism, Christianity, apologetics, theology, law, behavioural norms

Theology	Shared aims (discuss religion), theology, science, cooperative discussion
Theology	Shared aims (discuss religion), theology, science, cooperative discussion
Don't Ever Be Late!	Joke
Theology	Shared aims (discuss religion), theology, science, cooperative discussion
Banks	Irrelevant
Closed communities, blogs and sexual abuse	Shared aims (not religious), morality

TA 2.2 Sample analysed to evaluate the religion discussion classifier

I analysed a random sample of 100 discussion threads (one third from UTZoo 1981-1993 and two thirds from Giganews 2003-2015) which in total comprise 359 messages, to identify the subjects captured by the computer code as indicating religion discussion in accordance with the keywords I specified appearing prominently or distinctively, which relate to Abrahamic faiths (see pp.108-113).

Messages with the following subjects were captured as indicating religion discussion (sorted alphabetically):

"Old" Testament
(!) "MERRY CHRISTMAS" IN 33 LANGUAGES
Ancient Hebrew Poetry
exclusive salvation
Festivals of Lights, etc
For [REDACTED] and [REDACTED]
Is Judaism backward?
Lawyers!
Loose Cannon [REDACTED] totally humiliated !!!!!
LORD JESUS MISSIONS.
Muslim Spain
Our Reality Consistent with Impersonal or No God only
Palestinian Violations of International Law
question about grammer in sefer shirim in tanach
Re Eruvin
Re: Why Do We Christians Worship On
Resurrection of Jesus: 100% Proven Historical Fact
The Coming of Christ
The physical evidence
The Pope Excommunicated Communists, but not Nazis Re: Pope Pius XII Opposition to the Nazis
THE SECOND HALF OF THE EIGHTEENTH CENTURY
Was: God will save all Israel
Why God the father sent His only Son on earth
Women in Islam and Jewelry not exempt from Zakat

This sample indicates that the discussions identified by the classifier as relating to religion are indeed related to it. False positives occur – ‘Lawyers!’ and ‘Loose Cannon [REDACTED] totally humiliated !!!!!’ – but the classifier also generates some false negatives, as seen in the following.

Messages with the following subjects were captured as *not* indicating religion with the specific keywords in mind:

"God's General" unmuzzled!

"Misha: A Memoir of the Holocaust Years." Part IV - More \$\$\$\$
\$156 million for terror victim's family.

About this mind...

AD&D PBM GAME: MORE INFO, PROPAGANDA, AND ADVERTISEMENT.

Another One Bites The Dust

Answer to [REDACTED] on the Problem of the Gas chambers

Anti-Jews should be ashamed of themselves

Back on Topic: Trading Spouses: Jewish Mom Vs. Christian Mom
beyond honorifics

Bitburg - my 2 cents worth

Canonical Collection of LB Jokes

Cathbad and 3 wise men

Checking your mind at the door

cmsg newgroup net.religion

Compresseur micro tracteur

Enduro trying-to-ride Report

Entusiasmo e disillusione

FATAL HERITAGE; I need a hint, PLEASE!

Festivals of Lights, etc

For [REDACTED]

FORGERY Virginia Bill Requiring Doctors Lie To Their Patients

FORGERY Re: RUSSIA TO ARM SAUDI ARABIA

Greedy jews shot themselves in the foot

Hamas could be run from Finland would Jews try to assassinate me then IS THIS A
CONFESSION?

headlove versus heartlove

IAC/Vets For Peace/ANSWER/Wespac coordination?

INCREASE IN BUSINESS

Interesting Paragraph (Historical Inaccuracy)

Interpretations (cont)

Is "sen~or" contemptuous?

ISLAM WAS BEFORE DINOSAURS

Islamic Contribution To Science

Islamic VERMIN detonate 4 bombs in London

Israel: Former intel chief slams Netanyahu on Iran

JAP joke (not obscene)

Jewgle Torah. Parshat Chayei Sarah 5771. Isaac Meets Rebecca

[REDACTED]

[REDACTED] could not be a pop

kit information requested

[REDACTED] on dhimmitude

List of Active Newsgroups (Last changed: 16 May 1987)

Loose Cannon ([REDACTED]) totally humiliated !!!!!
Malaysia's PM: Islam inconsistent with democracy and human rights
Man avoids jail for accidental deadly fire
Maths question - totally correct actually
Mossad staged and taped "celebration"
Muslim gang attacks young mother on French train, cutting her hair, drawing swastikas on her
Muslim Spain
NEC V20 chip
New e-book
Offensive to Ancient Egyptians
Our Reality Consistent with Impersonal or No God only
Palestinian Violations of International Law
Parallel compilation (and linking) bibliography?
Project Ben-Yehuda - the classics of Hebrew literature
question about grammar in sefer shirim in tanach
QUESTION FASHION
[REDACTED]
Racism on MTV...isn't anyone concerned?
Re Eruvin
Re: Fund raising techniques
Re: When is the Sabbath?--or--the TOTAL prayer experience: CONSTANTLY!
Re: Why Do We Christians Worship On
Resurrection of Jesus: 100% Proven Historical Fact
Russell Crowe Comes to Aid of Jewish School
Shabat Shalom to all !!
Software thieves (was Re: Software theives)
SUN Modula-2 Linker problem (Why does it loop?).
Sweden rallies around Malmo Jews Ah a blame the victims politician gets comeuppance?
Telecom Italia mi ha rovinato il WE
Terrorism: definition: re to Tanenbaum
The 6-Million-Heresy that will Dissolve Jewry
The Coming of Christ
The physical evidence
The Pope Excommunicated Communists, but not Nazis Re: Pope Pius XII Opposition to the Nazis
THE PROMISE OF WORLD PEACE (Part 2 of 5)
The Second Coming of Christ will go something like this.
THESE ARE JEWS, THE SUFFERING SUCCOTASH
Tipos de sacerdotes
U.S. Jew charges CIA with anti-Semitism
Unable to upload windows updates
Was there a Jesus? Of course there was a Jesus ñ many! But only one is God.
Was: God will save all Israel
Where the Deer and the Cantaloupe Play

Why God the father sent His only Son on earth
Why Gogu is a HYPOCRITE !
Why use an LA instruction?

Here, false negatives are apparent in some threads, see 'Islamic Contribution To Science', 'Re: When is the Sabbath? [...]', 'Re: Why Do We Christians Worship On Was there a Jesus [...]', and 'Why God the father sent His only Son to earth'. Notwithstanding this, the classifier has mostly correctly identified threads that do not focus closely on religion discussion. And nuance is added regarding putative false negatives when considering there is some overlap between the subjects in the two lists. Each of these subjects appears in both lists (for different messages):

Festivals of Lights, etc
For [REDACTED] and [REDACTED]
Loose Cannon ([REDACTED]) totally humiliated !!!!!
Muslim Spain
Our Reality Consistent with Impersonal or No God only
Palestinian Violations of International Law
question about gramer in sefer shirim in tanach
Re [REDACTED]
Re: Why Do We Christians Worship On
Resurrection of Jesus: 100% Proven Historical Fact
The Coming of Christ
The physical evidence
The Pope Excommunicated Communists, but not Nazis Re: Pope Pius XII Opposition to the Nazis
Was: God will save all Israel
Why God the father sent His only Son on earth

This is suggestive of discussions that evolve in their subject over time, either initially or later meeting the criteria for inclusion, but not at other times. Alternatively a message reply may not include keywords included in the opening post. This would be expected given the diverse length and topics of Usenet posts.

I refer to the above findings in chapter three (p.134). Given this sample, and since I iterated the methodology during development when identifying where greater accuracy was needed, I have cautious confidence that it provides a tentative indication of religion discussion. In chapters four and five, wherever the classifier present potential insights regarding religion discussion being associated with other factors of messages (such as crossposting or first

time participation), I always use this algorithmic suggestion as a starting point for closer manual review, which I present in my qualitative discussion of OSN interactions.

TA 3 Extracts of the Message Dataframe

This appendix presents an extract of the Message Dataframe, indicating the format and content of the much larger versions in the research datasets.

Note: an extract of the term document matrix (TDM) is not provided. The TDM is a very sparse matrix including one axis for every message and other for every 'stemmed' word that occurs frequently in the dataset. The cells at the intersection of the message and word are populated to indicate the occurrence of that word in the dataset. This enables the following TF-IDF and TF analyses, and potentially enables other analyses, but its very large and sparse nature render it impractical to export from R and reproduce an extract here. The following five pages present metadata and the results of statistical analyses for 100 messages in the Message Dataframe. These data are:

- Date: the date on which the message was posted
- OSN: the online social network(s) to which it was posted
- Subject: the subject line
- Reply: whether the message was a reply (True or N/A)
- Userhash: a unique alphanumeric mix referring to the participant who posted it
- MessageID: a unique alphanumeric mix representing the message
- ADuncivillan: the result of sentiment analysis seeking infer incivility (not relied upon in my analyses)
- TFIDF: the start of the results of Term Frequency – Inverse Document Frequency analysis to identify words distinctive to the message
- TF: the start of the results of Term Frequency analysis to identify the most frequently occurring words (note: the original Message Dataframe includes the full results for TF-IDF and TF analyses, and these analyses are partly redacted in the following table)
- HighPartSubject: whether the message was part of a higher than average length discussion thread
- First: whether the message was the first posted by the participant within the dataset
- InterOSN: whether the message was crossposted
- PersistentPart: whether the message was posted by a longer-standing participant
- RelSharedAimWord: whether the message included one or more religion keywords either frequently or distinctively

Date	OSN	Subject	Reply	Userhash	MessageID	Aduncivillian	TFIDF	TF	HighPart Subject	First	InterOSN	Persistent Part	RelShared AimWord
03/06/1983	net.religion	What is God doing Today? - (nf)	TRUE	a11d560c	vaytODpm	FALSE	c(agnost = "1.2952756", = "0.9726809", mo [...])	c(agnost = "2", benefit = "1", care = "1", caus = [...])	FALSE	FALSE	FALSE	TRUE	FALSE
03/06/1983	net.religion	Views on Religions	TRUE	a2deef6f	LBKkA2uB	FALSE	c(cathol = "0.1503133", priest = "0.1446119", met [...])	c(cathol = "4", church = "3", live = "3", certain [...])	FALSE	TRUE	FALSE	TRUE	FALSE
03/06/1983	net.religion	Baptism question - (nf)	TRUE	32269eea	8zaTY5pP	FALSE	c(= "1.0370677", tc = "1.037067 [...])	c(doctrin = "2", agre = "1", bibl = "1", care = "1 [...])	FALSE	TRUE	FALSE	FALSE	TRUE
03/06/1983	net.flame	racism and hi-tech	TRUE	9f5bd963	6przirQG	FALSE	c(white = "0.10559591", pattern = "0.10511421", di [...])	c(white = "6", black = "5", discrimin = "4", male [...])	FALSE	TRUE	FALSE	FALSE	FALSE
03/06/1983	net.religion	reply to [NAME]	TRUE	4ab46708	WCShgC	FALSE	c(= "0.09280192", tim = "0.07094784", unbe [...])	c(believ = "5", bibl = "5", understand = "5", god [...])	FALSE	FALSE	FALSE	TRUE	FALSE
03/06/1983	net.religion	Views on Religions	TRUE	4ab46708	CCI8vecP	FALSE	c(religion = "0.3093384", hipocr = "0.1598961", ca [...])	c(religion = "11", peopl = "5", cathol = "3", ma [...])	FALSE	FALSE	FALSE	TRUE	FALSE
04/06/1983	net.religion	Views on Religions	TRUE	d1a11ad3	Yjh3HEh4	FALSE	c(conspicu = "0.2694140", rmembr = "0.2694140", c [...])	c(church = "4", marriag = "3", requir = "3", valid [...])	FALSE	FALSE	FALSE	TRUE	FALSE
04/06/1983	net.religion	Views on Religions - (nf)	TRUE	5a97f812	bYW7cT6G	FALSE	c(religion = "0.2562197", lber = "0.18210 [...])	c(religion = "8", one = "3", popular = "3", cathol [...])	FALSE	TRUE	FALSE	FALSE	FALSE
04/06/1983	net.flame	Fighting fire with fire	NA	aafefb7	d8onCNOI	FALSE	c(= "0.29663980", flame = "0.27101580", net [...])	c(flame = "27", item = "20", netgene r = "12", post [...])	FALSE	TRUE	FALSE	FALSE	FALSE
04/06/1983	net.religion	reply to [NAME]	TRUE	cc074e09	JZKBpoaz	FALSE	c(= "0.2628039", = "0.2628039" [...])	c(one = "3", bibl = "2", each = "2", interpret = " [...])	FALSE	FALSE	FALSE	FALSE	FALSE
04/06/1983	net.religion	Resurrection and the Burden of Proof	TRUE	cc074e09	33kRctNt	FALSE	c(christ = "0.23422784", gottal = "0.10215964", sj [...])	c(christ = "9", know = "4", death = "3", hope = "3 [...])	FALSE	FALSE	FALSE	FALSE	TRUE
04/06/1983	net.religion	Views on Religions	TRUE	cc074e09	hrv6vthL	TRUE	c(= "0.4270564", jk = "0.4270564" [...])	c(intent = "2", religion = "2", beheld = "1", clea [...])	FALSE	FALSE	FALSE	FALSE	FALSE
05/06/1983	net.religion	Views on Religions - (nf)	TRUE	ad17bcd4	YRKTHWzB	FALSE	c(flamer = "0.4572388", = "0.4390807", writ [...])	c(one = "2", anywher = "1", belief = "1", bibl = " [...])	FALSE	FALSE	FALSE	TRUE	FALSE
05/06/1983	net.religion	Orphaned Response - (nf)	TRUE	e1ae0dfb	ToMqsZpD	FALSE	c(= "1.9667219", = "1.8667 [...])	c(entir = "1", general = "1", = "1", jes [...])	TRUE	FALSE	FALSE	TRUE	FALSE
05/06/1983	net.religion	'net.lobotomy revisited'	NA	da2a4936	EmWgRTPD	FALSE	c(belief = "0.2088032", mam = "0.2081837", defend [...])	c(belief = "5", defend = "4", one = "4", believ = [...])	FALSE	FALSE	FALSE	FALSE	FALSE
05/06/1983	net.religion	"Ark Confusion - (nf)"	TRUE	d523a49a	YDun5sKC	FALSE	c(moon = "0.14685509", flood = "0.12105236", speck [...])	c(god = "9", moon = "9", flood = "7", believ = "5" [...])	FALSE	TRUE	FALSE	FALSE	FALSE
05/06/1983	net.philosophy	absolute value systems	TRUE	e29c6e27	AwODVMkl	FALSE	c(absolut = "0.31224512", = "0.12306237", [...])	c(absolut = "34", law = "14", moral = "11", system [...])	FALSE	TRUE	FALSE	TRUE	FALSE
05/06/1983	net.religion	reply to [NAME]	TRUE	e29c6e27	4ksbnthS	FALSE	c(scriptur = "0.10372268", cor = "0.09394391", tim [...])	c(christian = "8", reason = "8", god = "7", script [...])	FALSE	FALSE	FALSE	TRUE	TRUE
05/06/1983	net.religion	What is God doing Today?	TRUE	98728834	BatVeQHP	FALSE	c(zen = "1.7769324", netreligion = "1.4495188", co [...])	c(articl = "1", cours = "1", god = "1", netreligio [...])	FALSE	TRUE	FALSE	TRUE	FALSE

Date	OSN	Subject	Reply	Userhash	MessageID	ADuncivilian	TFIDF	TF	HighPart Subject	First	InterOSN	Persistent Part	RelShared AimWord
05/06/1983	net.flame	Racist and hi-tech	NA	024f3e34	MJm4Jbl2	FALSE	c(flame = "0.8172738", naaaaa = "0.6086067", [...])	c(flame = "7", status = "3", symbol = "3", anyth = [...])	FALSE	FALSE	FALSE	FALSE	FALSE
06/06/1983	net.religion	Abortion arguments	NA	9fb75c	43MfPpR	FALSE	c(proabort = "0.4441991", antiabort = "0.4149959", [...])	c(argument = "4", antiabort = "2", proabort = "2", [...])	FALSE	TRUE	FALSE	FALSE	FALSE
06/06/1983	net.religion	Is the jury biased?	TRUE	9ab9992b	5iAtPw6B	FALSE	c(0.4273646", resu [...])	c(challeng = "2", christian = "2", faith = "2", ma [...])	FALSE	TRUE	FALSE	TRUE	TRUE
06/06/1983	net.religion	Resurrection proof et al	TRUE	1ffc9dd	jHKAo025	FALSE	c(= "0.3539855", = "0.3401473" [...])	c(argument = "2", legal = "2", logic = "2", point [...])	FALSE	FALSE	FALSE	FALSE	FALSE
06/06/1983	net.groups.con trol,net.news.g roup	to many groups	NA	9f4a2bfb	Bb9acyNn	FALSE	c(netmicro = "0.11248321", netunix wizard = "0.1101 [...])	c(group = "6", made = "3", netmicro = "3", netunix [...])	FALSE	TRUE	TRUE	FALSE	FALSE
06/06/1983	net.religion	Christ the Anti Christ	NA	9b1ee471	KsyPJboh	FALSE	c(adversari = "0.16908839", angel = "0.11785569", [...])	c(adversari = "4", angel = "4", corrupt = "4", cou [...])	FALSE	TRUE	FALSE	TRUE	FALSE
06/06/1983	net.religion	Views on Religions - (nf)	TRUE	4ab46708	TnlcevB	FALSE	c(religion = "0.14412357", ligion = "0.14021905", [...])	c(religion = "9", thing = "4", catholic = "3", make [...])	FALSE	FALSE	FALSE	TRUE	FALSE
07/06/1983	net.religion	Religions based on "holy books"	TRUE	ad17bcd4	wu0dOwE7	FALSE	c(descrip = "0.4123656", unc = "0.3101888", relati [...])	c(consid = "2", relationship = "2", sense = "2", an [...])	FALSE	FALSE	FALSE	TRUE	FALSE
07/06/1983	net.religion	Reply to [NAME]	NA	e3cc490f	J7d5i144	FALSE	c(worldview = "0.12374808", = "0.08716253", ex [...])	c(god = "6", mean = "6", belief = "5", find = "5", [...])	FALSE	TRUE	FALSE	TRUE	FALSE
07/06/1983	net.religion	Religions based on "holy books"	NA	852e7745	f43RjWUo	FALSE	c(book = "0.3756321", holi = "0.2725029", valid = [...])	c(book = "14", holi = "8", follow = "5", particu [...])	FALSE	TRUE	FALSE	TRUE	FALSE
07/06/1983	net.religion	Views on Religions - (nf)	TRUE	d1a11ad3	brYNPvaV	FALSE	c(unfrock = "0.3548173", assur = "0.2834439", prot [...])	c(assur = "2", catholic = "2", church = "2", hold = [...])	FALSE	FALSE	FALSE	TRUE	FALSE
07/06/1983	net.religion	Ridiculing Genesis	NA	385a9612	zBpgrGH8	FALSE	c(creation = "0.13511872", genesis = "0.12499026", [...])	c(creation = "4", think = "4", ancient = "3", big [...])	FALSE	TRUE	FALSE	TRUE	FALSE
08/06/1983	net.religion	Discussion of Religion	TRUE	629f384e	Ndch5kV6	FALSE	c(christian = "0.14625268", = "0.10497806" [...])	c(christian = "9", exist = "4", source = "4", debat [...])	FALSE	FALSE	FALSE	FALSE	TRUE
08/06/1983	net.religion	What if...	TRUE	4a6138a2	304tRqZg	FALSE	c(master = "0.22119758", self = "0.21213314", free [...])	c(free = "7", master = "5", self = "5", christian [...])	FALSE	FALSE	FALSE	TRUE	TRUE
08/06/1983	net.religion	A portrait of Messiah	NA	4e269e99	n4sQu4uw	FALSE	c(isaiah = "0.4029251", micah = "0.2778745", bethl [...])	c(isaiah = "2", aspect = "1", before = "1", bethleh [...])	FALSE	FALSE	FALSE	TRUE	TRUE
08/06/1983	net.religion	Orphaned Response - (nf)	TRUE	1e0e738c	FmxW6hF3	FALSE	c(pj) = "0.3767137", enquiri = "0.2195941", tidbit [...])	c(author = "2", debat = "2", know = "2", littl = " [...])	TRUE	TRUE	FALSE	FALSE	FALSE
08/06/1983	net.religion	Christ the Anti Christ	TRUE	e29c6e27	YVwdr8C7	FALSE	c(adversari = "0.09154467", angel = "0.07643873", [...])	c(god = "8", law = "7", angel = "6", good = "6", a [...])	FALSE	FALSE	FALSE	TRUE	TRUE
09/06/1983	net.religion	'What is God doing today?'	NA	1fd01b87	ERK2MIPY	FALSE	c(cbn = "0.1563674", pat = "0.1332242", ebi = "0.1 [...])	c(christian = "11", god = "8", pat = "7", well = [...])	FALSE	TRUE	FALSE	FALSE	TRUE
09/06/1983	net.religion	Retransmission of Mail: Reply to [NAME]	NA	e3cc490f	aHefrR8	FALSE	c(worldview = "0.12190109", = "0.08586159", ex [...])	c(god = "6", mean = "6", belief = "5", find = "5", [...])	FALSE	FALSE	FALSE	TRUE	FALSE

Date	OSN	Subject	Reply	Userhash	MessageID	ADuncivilian	TFIDF	TF	HighPart Subject	First	InterOSN	Persistent Part	RelShared AimWord
09/06/1983	net.religion	What do YOU think?	NA	f698f6c4	5HMU8gOb	FALSE	c(= "0.1425161", = "0.1425161", m [...]	c(question = "6", individu = "4", feel = "3", inte [...]	FALSE	TRUE	FALSE	FALSE	FALSE
09/06/1983	net.religion	A burden both ways	NA	13e28e7c	wjpxC2th	FALSE	c(resurrect = "0.10238369", bias = "0.08428751", t [...]	c(evid = "10", resurrect = "10", bias = "9", one [...]	FALSE	FALSE	FALSE	TRUE	FALSE
09/06/1983	net.religion	Not about the Resurrection	NA	13e28e7c	87JGYzWn	FALSE	c(massa = "0.2004427", slave = "0.1982533", cower [...]	c(slave = "4", love = "3", articl = "2", cower = " [...]	FALSE	FALSE	FALSE	TRUE	FALSE
09/06/1983	net.religion	Views on Religions - (nf)	TRUE	8248d820	ESgjpIDV	FALSE	c(church = "0.20055235", noncathol = "0.13373842", [...]	c(church = "8", cathol = "4", offici = "3", roman [...]	FALSE	TRUE	FALSE	TRUE	TRUE
09/06/1983	net.religion	Orphaned Responce - (nf)	TRUE	4ab46708	Vcjb2v9	FALSE	c(harmless = "0.2346865", ayetollah = "0.1787929", [...]	c(cours = "3", harmless = "3", religi on = "3", des [...]	TRUE	FALSE	FALSE	TRUE	FALSE
09/06/1983	net.religion	Resurrection proof et al	TRUE	4ab46708	eqj3P9Qx	FALSE	c(= "0.1370547", = "0.1204534", p [...]	c(point = "5", understand = "4", argu ment = "3", i [...]	FALSE	FALSE	FALSE	TRUE	FALSE
09/06/1983	net.religion	Views on Religions - (nf)	TRUE	4ab46708	UIHSXnKO	FALSE	c(cathol = "0.2340757", church = "0.2263869", ctah [...]	c(church = "7", cathol = "6", protest = "3", belie [...]	FALSE	FALSE	FALSE	TRUE	FALSE
10/06/1983	net.religion	A burden both ways	TRUE	ad17bcd4	f3ocqjUT	FALSE	c(heresay = "0.2910525", resurrect = "0.2589928", [...]	c(event = "6", resurrect = "5", descr ib = "4", bas [...]	FALSE	FALSE	FALSE	TRUE	FALSE
10/06/1983	net.jokes	Brain Jokes	NA	41dc05f1	5G8FOPJE	FALSE	c(salesman = "0.7435594", brain = "0.5908603", dol [...]	c(brain = "3", cost = "2", custom = "2", dollar = [...]	FALSE	TRUE	FALSE	FALSE	FALSE
10/06/1983	net.jokes	Jewish & Leaper	NA	4a76b661	xKBI0zXX	FALSE	c(leaper = "1.2467529", marshmello w = "0.6345113", [...]	c(call = "2", leaper = "2", blender = "1", car = " [...]	FALSE	TRUE	FALSE	FALSE	FALSE
10/06/1983	net.religion	Born again	NA	fd15379b	TPvrNGHe	FALSE	c(born = "0.5701552", spirit = "0.2782971", flesh [...]	c(born = "9", again = "6", spirit = "4", christ = [...]	FALSE	TRUE	FALSE	TRUE	TRUE
10/06/1983	net.religion	"Ark Confusio n"	TRUE	341f9923	Vh6gKtIW	FALSE	c(bibl = "0.05764709", = "0.05746787", s [...]	c(thing = "8", bibl = "6", articl = "5", believ = [...]	FALSE	TRUE	FALSE	FALSE	FALSE
10/06/1983	net.religion	Paradise vs. Resurrection	NA	385a9612	LwE61e7S	TRUE	c(= "0.2787941", resurrect = "0.2689541", [...]	c(die = "2", heaven = "2", resurrect = "2", suppor [...]	FALSE	FALSE	FALSE	TRUE	FALSE
10/06/1983	net.philosophy	More on absolute value systems.	NA	e29c6e27	94DatbJF	FALSE	c(valu = "0.07747378", absolut = "0.05963047", god [...]	c(valu = "22", god = "21", absolut = "18", think = [...]	FALSE	FALSE	FALSE	TRUE	TRUE
10/06/1983	net.religion	A burden both ways	TRUE	4ab46708	b6iXeC3K	FALSE	c(nirvana = "0.2559308", i = "0.2218313", buddh [...]	c(friend = "4", nirvana = "4", see = "4", wit = "4 [...]	FALSE	FALSE	FALSE	TRUE	FALSE
10/06/1983	net.religion	What if...	TRUE	4ab46708	bcAnOvCP	FALSE	c(master = "0.13518544", self = "0.0803807", free [...]	c(free = "7", master = "6", christian = "5", self [...]	FALSE	FALSE	FALSE	TRUE	TRUE
11/06/1983	net.religion	exclusive salvation	TRUE	d1a11ad3	0MGMCbnA	FALSE	c(alien = "0.2317097", salvat = "0.1591931", [...]	c(alien = "4", christ = "4", well = "4", god = "3 [...]	FALSE	FALSE	FALSE	TRUE	TRUE
11/06/1983	net.sf-lovers	various random good books	NA	0b0fb597	AOnpw9k7	FALSE	c(book = "0.10749910", trilog = "0.05786207", kill [...]	c(book = "18", out = "10", sort = "7", earth = " [...]	FALSE	TRUE	FALSE	TRUE	FALSE
11/06/1983	net.religion	Born again	TRUE	4a6138a2	oZl0bxgj	FALSE	c(ruti = "0.5694675", hutch = "0.4831565", view = [...]	c(christian = "3", view = "3", anyon = "1", church [...]	FALSE	FALSE	FALSE	TRUE	TRUE
11/06/1983	net.religion	A burden both ways	TRUE	341f9923	hubGU9Xl	FALSE	c(= "0.1827388", = "0.1300707", apostl [...]	c(believ = "7", account = "4", apostl = "3", [...]	FALSE	FALSE	FALSE	FALSE	FALSE

Date	OSN	Subject	Reply	Userhash	MessageID	ADuncivilian	TFIDF	TF	HighPart Subject	First	InterOSN	Persistent Part	RelShared AimWord
11/06/1983	net.religion	Rational descriptions of miracles? Huh?	NA	5592d0e3	FZQOhtaG	FALSE	c(transmut = "0.2497349", carbon = "0.1845528", br [...])	c(bread = "5", organ = "5", way = "5", carbon = "4 [...])	FALSE	FALSE	FALSE	TRUE	FALSE
11/06/1983	net.religion	Jewish Baptism	NA	021a4376	rZpH2pn4	FALSE	c(immers = "0.4140627", mikvah = "0.3280448", ritu [...])	c(immers = "4", ritual = "4", baptism = "3", befor [...])	FALSE	FALSE	FALSE	TRUE	TRUE
11/06/1983	net.singles	How did your parents meet?	TRUE	021a4376	IN811Ow	FALSE	c(= "0.7281226", = "0.6152600" [...])	c(andi = "1", appar = "1", ball = "1", bell = "1", [...])	FALSE	FALSE	FALSE	TRUE	FALSE
12/06/1983	net.religion	Question on immersion in Judaism	TRUE	c919d628	cyinMhDe	FALSE	c(immers = "0.6847960", mikveh = "0.5144825", twel [...])	c(immers = "6", mikveh = "4", befor = "3", themsel [...])	FALSE	TRUE	FALSE	TRUE	TRUE
12/06/1983	net.religion	what do you think gang?	NA	8248d820	132FvcRa	FALSE	c(unpleasnt = "0.1855398", religion = "0.1631588", [...])	c(religion = "5", think = "5", peopl = "4", church [...])	FALSE	FALSE	FALSE	TRUE	TRUE
12/06/1983	net.religion	Orphaned Response - (nf)	TRUE	341f9923	yJTrmvFC	FALSE	c(baptism = "1.1182879", = "0.5955880", pra [...])	c(baptism = "3", jew = "2", practic = "2", around [...])	TRUE	FALSE	FALSE	FALSE	TRUE
12/06/1983	net.sf-lovers	Locus Summary and Review of new Varley Book	NA	35dbc2e9	HiWnW7Bw	FALSE	c(= "0.16724345", book = "0.07990505", dcie [...])	c(book = "6", veri = "6", end = "4", more = "4", t [...])	FALSE	TRUE	FALSE	TRUE	FALSE
13/06/1983	net.religion	Reasons for Belief	NA	514d4a65	hUj2bGMd	FALSE	c(belief = "0.19966810", scientif = "0.18353462", [...])	c(belief = "6", believ = "4", evid = "4", question [...])	FALSE	TRUE	FALSE	FALSE	FALSE
13/06/1983	net.religion	Father of Jesus	NA	a715597b	7wGen989	FALSE	c(mari = "0.8474339", husband = "0.7817585", close [...])	c(husband = "2", mari = "2", closet = "1", first = [...])	FALSE	TRUE	FALSE	FALSE	FALSE
13/06/1983	net.religion	A burden both ways	TRUE	3cf3c5c	4hFGOsQY	FALSE	c(resurrect = "0.4661871", = "0.3512645", sp [...])	c(christian = "2", histor = "2", resurrect = "2", [...])	FALSE	TRUE	FALSE	TRUE	TRUE
13/06/1983	net.religion	Paradise vs. Resurrection	TRUE	3cf3c5c	bizYXNMT	FALSE	c(afterlif = "0.2968348", death = "0.2590571", chr [...])	c(death = "4", one = "4", christ = "3", experi = [...])	FALSE	FALSE	FALSE	TRUE	TRUE
13/06/1983	net.philosophy	absolute value systems	TRUE	51de952a	d3bgliB4	FALSE	c(valu = "0.2905267", = "0.2728970", bibl = [...])	c(bibl = "3", system = "3", valu = "3", ask = "2", [...])	FALSE	TRUE	FALSE	TRUE	FALSE
13/06/1983	net.religion	Who killed Christ?	NA	13159dee	ggQUMLyM	FALSE	c(l = "0.2869095", arrang = "0.1739964", christe [...])	c(arrang = "3", parti = "3", show = "3", trial = [...])	FALSE	TRUE	FALSE	TRUE	TRUE
13/06/1983	net.religion	Not about the Resurrection - (nf)	TRUE	c1d089a4	eOTi8Gx9	FALSE	c(valid = "0.2472399", judgement = "0.2390315", rq [...])	c(valid = "4", judg = "3", judgement = "3", worri [...])	FALSE	TRUE	FALSE	FALSE	FALSE
13/06/1983	net.religion	Re: Ark Confusion - (nf)	TRUE	c1d089a4	A82CNk3y	FALSE	c(= "1.4048014", = "1.3 [...])	c(claim = "2", god = "2", know = "2", 'doesn't' = [...])	FALSE	FALSE	FALSE	FALSE	FALSE
13/06/1983	net.religion	Paradise vs. Resurrection	TRUE	f78d191a	vKxYmse	FALSE	c(paradis = "0.2246288", christ = "0.1687356", hea [...])	c(christ = "10", heaven = "8", parad is = "8", go [...])	FALSE	FALSE	FALSE	TRUE	TRUE
13/06/1983	net.religion	"Re: "Ark Confusion" - (nf)"	NA	8429471b	ugRBDnpT	FALSE	c(god = "0.09290484", question = "0.06799753", ani [...])	c(god = "18", such = "14", question = "13", ask = [...])	FALSE	TRUE	FALSE	FALSE	FALSE
13/06/1983	net.religion	What are Mormons	NA	4aebd1a1	8svfm3rX	FALSE	c(mormon = "1.2417951", chicotekl abslynnf = "1.03 [...])	c(mormon = "3", believ = "1", cbosg = "1", chicote [...])	FALSE	TRUE	FALSE	TRUE	TRUE
13/06/1983	net.philosophy	absolute value systems	TRUE	66f08b9d	cFe9sYNp	FALSE	c(valu = "0.2362005", surviv = "0.2223859", risk = [...])	c(valu = "6", surviv = "5", anim = "4", risk = "4" [...])	FALSE	TRUE	FALSE	FALSE	FALSE

Date	OSN	Subject	Reply	Userhash	MessageID	ADuncivilian	TFIDF	TF	HighPart Subject	First	InterOSN	Persistent Part	RelShared AimWord
13/06/1983	net.religion	'net.lobotomy revisited'	TRUE	a11d560c	Fv9Oa3ot	FALSE	c(belief = "0.2469325", believ = "0.2142131", mere [...])	c(believ = "6", belief = "4", mean = "3", peopl = [...])	FALSE	FALSE	FALSE	TRUE	FALSE
13/06/1983	net.religion	I Hate Religion	NA	6b8a2943	HY72QkSi	FALSE	c(nicodemius = "0.21526496", syste m = "0.15848739" [...])	c(god = "11", religion = "11", syste m = "10", life [...])	FALSE	FALSE	FALSE	TRUE	FALSE
13/06/1983	net.religion	I Hate Religion	NA	12911099	KqJtm19s	FALSE	c(nicodemius = "0.21526496", syste m = "0.15848739" [...])	c(god = "11", religion = "11", syste m = "10", life [...])	FALSE	TRUE	FALSE	FALSE	FALSE
13/06/1983	net.suicide	More thoughts	TRUE	3d19cba2	iZwglw6w	FALSE	c(suicid = "0.27482278", shintoistbu ddhist = "0.12 [...])	c(suicid = "8", peopl = "6", right = "5", chang = [...])	FALSE	TRUE	FALSE	TRUE	FALSE
14/06/1983	net.religion	Discussion of Religion	NA	629f384e	bcGEQZK	FALSE	c(christian = "0.14547886", = "0.10442262" [...])	c(christian = "9", exist = "4", sourc = "4", debat [...])	FALSE	FALSE	FALSE	FALSE	TRUE
14/06/1983	net.flame,net.p olitics	A Flame at Affirmative Action	TRUE	9ab9992b	k21iDJIB	FALSE	c(discrimin = "0.27643479", "0.15697039", st [...])	c(discrimin = "12", job = "5", peopl = "5", hire [...])	FALSE	FALSE	TRUE	TRUE	FALSE
14/06/1983	net.religion	The Importance of the Resurrection - (nf)	NA	aa23d5a3	fMeyQbyR	FALSE	c(sin = "0.1683949", perfect = "0.1355486", resur [...])	c(sin = "5", jesus = "4", perfect = "4", anim = "3 [...])	FALSE	TRUE	FALSE	FALSE	TRUE
14/06/1983	net.religion	MAN and APE S	NA	727b66d	JTiOZso	FALSE	c(= "1.1301410", ap e = "0.5418334 [...])	c(ape = "1", believ = "1", between = "1", dont = " [...])	FALSE	TRUE	FALSE	FALSE	FALSE
14/06/1983	net.religion	"Re: Re: "Ark Confusion" - (nf)"	TRUE	4a6138a2	SWO1NVNd	FALSE	c(druidic = "0.18351005", witch = "0.10707575", co [...])	c(christian = "4", contradict = "3", g od = "3", po [...])	FALSE	FALSE	FALSE	TRUE	TRUE
14/06/1983	net.religion	Born Again	TRUE	4a6138a2	oFYJbtw	FALSE	c(nicodemus = "0.23724689", spirit = "0.15732875", [...])	c(born = "13", spirit = "12", one = "11", god = "1 [...])	FALSE	FALSE	FALSE	TRUE	FALSE
14/06/1983	net.religion	Orphaned Response - (nf)	TRUE	4a6138a2	KIF0kzQ0	FALSE	c(oversimplist = "0.4800674", asissi = "0.2935406" [...])	c(religion = "3", give = "2", oversim plist = "2", [...])	TRUE	FALSE	FALSE	TRUE	FALSE
14/06/1983	net.flame	Smoking... (Slow Motion Suicide)	TRUE	1ffc9dd	CKnCDWbT	FALSE	c(= "0.33005526", submiss = "0.18276230", [...])	c(= "7", those = "4", discuss = "3", submiss = [...])	FALSE	FALSE	FALSE	FALSE	FALSE
14/06/1983	net.religion	Father of Jesus	TRUE	34119923	sa7CVvGi	FALSE	c(= "0.7357263", proofread = "0.6790057", g [...])	c(articl = "1", bad = "1", care = "1", cours = "1" [...])	FALSE	FALSE	FALSE	FALSE	FALSE
14/06/1983	net.religion	"Re: Re: "Ark Confusion" - (nf)" - (nf)"	TRUE	a11d560c	zIGjyxD	FALSE	c(rainbow = "0.5633702", symbol = "0.4235991", noa [...])	c(rainbow = "4", symbol = "4", befor = "3", mean = [...])	FALSE	FALSE	FALSE	TRUE	FALSE
14/06/1983	net.religion	Orphaned Response - (nf)	TRUE	a11d560c	9fci9mFr	FALSE	c(ayahtollah = "0.4573772", selfdelu s = "0.2799503 [...])	c(peopl = "5", believ = "4", 'don't' = "4", kill = [...])	TRUE	FALSE	FALSE	TRUE	FALSE
14/06/1983	net.religion	What do YOU think?	TRUE	a11d560c	6wAv9Txc	FALSE	c(death = "0.2845494", = "0.1845244", beacaus [...])	c(death = "7", life = "4", accept = "3", experi = [...])	FALSE	FALSE	FALSE	TRUE	FALSE
14/06/1983	net.religion	A burden both ways	TRUE	e29c6e27	SrVAOdbO	FALSE	c(nirvana = "0.27873314", buddist = "0.20268247", [...])	c(nirvana = "9", see = "7", buddist = "6", wit = " [...])	FALSE	FALSE	FALSE	TRUE	FALSE
14/06/1983	net.religion	Some questions on religion	NA	1ea90ee2	XT1vx0SN	FALSE	c(religion = "0.4117816", nonbeliev = "0.2126586", [...])	c(religion = "10", anoth = "3", better = "3", at [...])	FALSE	TRUE	FALSE	FALSE	FALSE
14/06/1983	net.religion	exclusive salvation	TRUE	4ab46708	PWYzZmZM D	FALSE	c(alien = "0.17966797", salvat = "0.13166773", chr [...])	c(god = "6", alien = "5", christ = "4", 'don't' = [...])	FALSE	FALSE	FALSE	TRUE	TRUE

Date	OSN	Subject	Reply	Userhash	MessageID	ADuncivillan	TFIDF	TF	HighPart Subject	First	InterOSN	Persistent Part	RelShared AimWord
14/06/1983	net.misc	Rock Music Considered Harmful to Unborn (nf)	TRUE	ad66065f	IzTe50b	FALSE	c(= "0.4603947", = [...]	c(discoveri = "2", = "2", religion = "2", [...]	FALSE	TRUE	FALSE	TRUE	FALSE
15/06/1983	net.news.grou p	List of Active Newsgroups	NA	82abb9c6	j2BnB1E7	FALSE	c(subgroup = "0.65823720", digest = "0.09575017", [...]	c(subgroup = "50", etc = "13", comput = "12", disc [...]	TRUE	TRUE	FALSE	TRUE	FALSE
15/06/1983	net.religion	Discussion of Religion	TRUE	ad17bcd4	IDiQqEOG	FALSE	c(expert = "0.29081569", protomark = "0.11377444", [...]	c(expert = "7", christian = "4", document = "3", m [...]	FALSE	FALSE	FALSE	TRUE	TRUE
15/06/1983	net.religion	Reasons for Belief	TRUE	ad17bcd4	3geJ75BQ	FALSE	c(belief = "0.3042561", how = "0.1810532", unc = " [...]	c(belief = "5", one = "3", anoth = "2", believ = " [...]	FALSE	FALSE	FALSE	TRUE	FALSE

TA 4 Extracting data from the archives and organising them into datasets

I used the following steps to extract and organise datasets from the large text archives that I had obtained/derived. This resulted in data frames which include one row per message and one column for each of the following, reflecting my above decisions about what features of the data to include:

- A row number
- The date on which the message was sent/received
- List of newsgroups (OSNs) to which the message was posted, with each separated by a comma if more than one
- The subject line of the message
- Whether the message was a reply to an existing thread
- Alphanumeric string of characters representing
 - Each message
 - The message sender
- Any analytic descriptors identified from the message indicating subjects of discussion
- Whether the message
 - was the first posted by that user, so far as could be inferred
 - posted to more than one newsgroup/OSN
 - posted by a longer-standing (persistent) participant

A screenshot of a dataset's column headers, displayed in a dark-themed interface. The headers are arranged in two rows. The first row contains: 'date', 'OSN', 'subject', 'reply', 'userhash', 'messageID', and 'ADuncivillan'. The second row contains: 'TFIDF', 'TF', 'HighPartSubject', 'first', 'interOSN', 'PersistentPart', and 'RelSharedAimWord'. Each header has a small downward-pointing arrow to its right, indicating it is a dropdown menu.

date	OSN	subject	reply	userhash	messageID	ADuncivillan
TFIDF	TF	HighPartSubject	first	interOSN	PersistentPart	RelSharedAimWord

Figure 20: a screenshot of dataset column headings indicating the dataset structure

Extracting these features was challenging to varying degrees, particularly given the heterogeneous structure and content of the UTZoo dataset. I now summarise my approach, with the analytic descriptors requiring a separate section which follows. I applied these steps to a temporary object I had created in R which was simply a list all messages from the archives with one message per row, rather than trying to extract data directly from the very large archives with which I was working. Having the dataset delineated between messages in this manner enabled me to extract features from individual messages rather than potentially mixing up properties of different messages. This required me to identify where each message ended and began within the text of the archives. For this, I identified delimiters, which are numbers, letters, or other characters (potentially including line breaks) that indicate where one message (including its metadata such as sender and date) ends

and the next one starts. I identified these delimiters, which varied between and sometimes within archives, and then specified them in my R code as the cut off for each message.

Regular expression, or 'regex', was required for much of my work to extract, transform, and load message data into my datasets. Regex is code written to identify certain portions of text based on patterns of alphabetical and numeric characters, punctuation, other symbols, and other properties such as line breaks (as if the return key had been pressed). Regex is employed in many programming languages including R. The regex code specifies the pattern and is embedded in one or more lines of code that instruct actions to be taken when that pattern is encountered. Here is a simple example of a regex (emboldened) embedded in a line of my R code:

```
fpdf$forums <- str_extract(fpdf$Full_Post, "Newsgroups: (.*)\n")
```

Focusing on the regex, this code extracts text within each message that starts with exactly 'Newsgroups:' followed by one space, then followed by any number of any characters, which is represented by '(*)' until the end of that line in the message, instructed by '\n'. The backslash ensures the regex looks for a new line, since the backslash is an 'escape' character meaning the computer interprets the character after the backslash as representing a property of the text that has been ascribed to that character in regex rules, rather than the representing the character as read plainly. In this case, '\n' leads regex to search for a new line, rather than the character 'n'.

Extracting the sender's name and substituting it with a random alphanumeric string entailed two steps. First, I used regex to extract the sender's name from the header information of each message. In Giganews messages and later UTZoo messages this was relatively straightforward as the header includes a 'From:' line, as in email. Earlier UTZoo messages did not contain this and conflate the route that the message took through the network to reach the recipient with the sender's username. I inferred this was the case by referring to appendices in Hauben and Hauben (1997, 191-194, 198) which reproduce Usenet messages in which the network path is appended by a username which matches the name of the sender as stated in the message. The same is the case in messages I spot-checked, where the user indicates their name. This represented a typical pattern in a format of 'path!path!path!username' in which 'path' is a step on the network that the message had travelled (with each step in the path differentiated by an exclamation mark). I applied regex

to extract the username in accordance with this sample pattern, capturing the characters after the last element of the network path:

```
utzool$user <- str_extract (utzool$full_post, "![[[:alpha:]]-]{1,}\\n")
```

This was imperfect since I do not know if users in different locations may have used the same username, though this does not impact my work as I do not seek to infer patterns of participation or network density from UTZoo as the archive is too fragmentary. I nonetheless include the username from UTZoo messages for consistency, since I *do* utilise the representation of the username derived from later Giganews archive messages. For Giganews and later UTZoo messages, I used regex to extract the sender's username/email using the 'From:' line of the header:

```
gnrel$user <- str_extract (gnrel$full_post, "From: (.*)\\n")
```

I then used R code to identify each unique username/email and associate each one with a random alphanumeric string of eight characters, in the style of 'j8fd9kv5', and wrote code to confirm that (however unlikely) the same string had not been allocated to more than one username/email. This enabled me to exclude original usernames/email addresses from my datasets while still being able to discern high-level trends/patterns in participation across time since all messages from one username/email address shared the same string.

I then extracted the names of the newsgroup(s) (OSNs) to which each message was posted. Early UTZoo messages were typically posted to one group but crossposting to multiple groups increased over time due to the number of groups available, the growth of the network and the number of participants. I was generally able to extract these by using regex to capture the names of groups in the header 'Newsgroups:' line (see regex above). However, the sparse and heterogeneous headers of very early UTZoo messages between 1981 and 1982, which used the more basic 'A News' Usenet software (Hauben and Hauben 1997, 190), lacked this line which made it more challenging to extract the newsgroup names. These messages were posted to the 'net.' and '.fa' Usenet hierarchies, with the specific group name appended to those 'top level' hierarchies such as net.religion when it was created later in 1983. '.fa' newsgroups included digests of messages from equivalent groups on the ARPANET such as 'fa.sf-lovers' for science fiction fans, though if a Usenet

participant posted to a '.fa' newsgroup (as in the '.fa' messages in my dataset) it would only be seen by other Usenet participants by default (Hauben and Hauben 1997, 191).

Due to the heterogenous formats as summarised above, it was challenging to use R code to extract dates from messages. First, I used regex to remove extraneous information, such as the day of the week and punctuation from the text that conveyed the date. I then used regex to extract the year, month, and day into temporary columns. This required careful review of how date formats in messages evolved over time as the software used to send and receive messages developed, with different patterns requiring different regex. For later UTZoo messages, I developed 'nested' regex patterns with the computer code searching for each of multiple patterns in turn the data were extracted successfully.

Once I had extracted years, months, and dates, I recombined them in one column in day-month-year order. I then used the date function in R to convert the resulting string of numbers into a date format that R could work with.

As noted above, some messages record dates and times for when the message was posted and when it was received. In these cases, I included the posted date as that most closely represents when the participant contributed to the discussion; the received date may vary depending on one's time zone and location in the network.

Extracting the subject line of the message was straightforward for later UTZoo and Giganews archive messages since the header includes a subject line. I applied regex to extract the subject without the preface 'Subject:'. I also removed the 'RE:' from messages indicating that they were replies helping me to group together messages that form one discussion by improving the consistency of their subject lines. Early UTZoo messages do not have the text 'Subject:' prefacing the subject in the header, but these were few and the subjects still occupied a distinct line of the header, so I was able to extract them. This approach has limitations since participants may change the subject line, though the analytic descriptors identified in the messages before and after the change would bear similarity if the conversation remained substantively similar, including if previous messages were incorporated into the reply. It is also possible that conversations could use a subject line that has been used before. I checked for this after assembling my datasets by running code that compared the dates of messages in each of the one hundred longest threads of conversation and checked for suspiciously long discussion threads which could in fact be multiple distinct conversations. I found one, with a single word subject. Nonetheless, I conclude the other threads were one long discussion. These often appeared to focus on

the sort of contentious and engaging subjects (often at intersections of religion and another topic) that I explore in this research. Therefore, the problem of threads appearing long due to them sharing the same subject appears to be minimal and not affect the broad trends and patterns that my analyses focus on. Lastly, I wrote a line of code to add one sequential number (1, 2, 3...) to each row, giving each message a unique reference number.

These steps delivered an organised dataset, extracting from the archives all the features I required for each message, organised into one row per message and one column per feature, except for my analytic descriptors of message content and before I appended results of statistical analyses, summarised in the bullet points at the start of this subsection. Those analyses required bespoke R code and the analytic descriptors require careful consideration and configuration of NLP algorithms to infer topics of discussion. I have appended a sample of this dataset as Technical Appendix three (TA 3).

Bibliography

Abdulla, R. A. (2007) 'Islam, Jihad, and Terrorism in Post-9/11 Arabic Discussion Boards', *Journal of Computer-Mediated Communication*, 12(3), pp. 1063-1081.

Acker, A. and Donovan, J. (2019) 'Data craft: a theory/methods package for critical internet studies', *Information, Communication & Society*. DOI: 10.1080/1369118X.2019.1645194

Across the Board (2014) BBC Radio 4, 31 October, 12:04.

Al-Garadi, M. A., Varathan, K. D., Ravana, S. D., Ahmed, E., Mujtaba, G., Khan, M. U. S. and Khan, S. U. (2018) 'Analysis of Online Social Network Connections for Identification of Influential Users: Survey and Open Research Issues', *ACM Computing Surveys*, 55(1), pp. 1-37. DOI: 10.1145/3155897

Ali, I., Asif, M., Hamid, I., Sarwar, M. U., Khan, F. A. and Ghadi, Y. (2022) 'A word embedding technique for sentiment analysis of social media to understand the relationship between Islamophobic incidents and media portrayal of Muslim communities', *Peer J Computer Science*, 8(e838). DOI: 10.7717/peerj-cs.838

Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B. and Kochut, K. (2017) 'Text Summarization Techniques: A Brief Survey', accessed via arXiv. DOI: 10.48550/arXiv.1707.02268

Allmer, T., Fuchs, C., Kreilinger, V. and Sevignani, S. (2014) 'Social Networking Sites in the Surveillance Society: Critical Perspectives and Empirical Findings' in Jansson, A. and Christensen, M. (eds.) *Media, Surveillance and Identity: Social Perspectives*, pp. 49-70. New York: Peter Lang.

Altopia (2021) accessed at 'Usenet', *Wikipedia*. Available at: <https://en.wikipedia.org/wiki/Usenet> (Accessed: 30 July 2021).

Attali, D., Almende, B. V. (2022) *timevis: Create Interactive Timeline Visualizations in R*. R package version 2.1.0, <https://CRAN.R-project.org/package=timevis>.

Bainbridge, W. S. (2020) *The Social Structure of Online Communities*. Cambridge: Cambridge University Press.

Bainotti, L., Caliandro, A. and Gandini, A. (2020) 'From archive cultures to ephemeral content, and back: Studying Instagram Stories with digital methods', *New Media & Society*. DOI: 10.1177/1461444820960071

Balazka, D., Houtman, D. and Lepri, B. (2021) 'How can big data shape the field of non-religion studies? And why does it matter?', *Patterns*, 2(6). DOI: 10.1016/j.patter.2021.100263

Baker, J. D. (1995) *Christian Cyberspace Companion*. Grand Rapids, Michigan: Baker Books.

Bannerman, S. (2018) 'Relational privacy and the networked governance of the self', *Information, Communication & Society*. DOI: 10.1080/1369118X.2018.1478982

- Baran, P. (1964) 'On Distributed Communications Networks', *IEEE Transactions on Communications Systems*, 12(1), pp. 1-9.
- Barker, E. (2005) 'Crossing the boundary: new challenges to religious authority and control as a consequence of access to the Internet' in Højsgaard, M. T. and Warburg, M. (eds.) *Religion and Cyberspace*. Abingdon, Oxon; New York: Routledge.
- Batten, D., Catchpoole, D., Sarfati, J. and Wieland, C. (2014) *The Creation Answers Book*. Powder Springs, Georgia: Creation Book Publishers.
- Bay, M. (2018) 'Conversation with a pioneer: Leonard Kleinrock on the early days of networking, the ARPANET...and winning in Las Vegas', *Internet Histories*, 2(1-2), pp. 140-152. DOI: 10.1080/24701475.2018.1446239
- Bay, M. (2019a) 'Conversation with a pioneer: Larry Roberts on how he led the design and construction of the ARPANET', *Internet Histories*, 3(1), pp. 68-80. DOI: 10.1080/24701475.2018.1544727
- Bay, M. (2019b) 'Hot potatoes and postmen: how packet switching became ARPANET's greatest legacy', *Internet Histories*, 3(1), pp. 15-30. DOI: 10.1080/24701475.2018.1544726
- Baym, N. K. (1995) 'The Emergence of Community in Computer-Mediated Communication' in Jones, S. G. (ed.) *CyberSociety Computer-Mediated Communication and Community*. Thousand Oaks, CA; London; New Delhi: SAGE Publications, Inc.
- Bednar, K., Spiekermann, S. and Langheinrich, M. (2019) 'Engineering Privacy by Design: Are engineers ready to live up to the challenge?', *The Information Society*, 35(3), pp. 122-142. DOI: 10.1080/01972243.2019.1583296
- Beer, D. and Burrows, R. (2013) 'Popular Culture, Digital Archives and the New Social Life of Data', *Theory, Culture & Society*, 30(4), pp. 47-71. DOI: 10.1177/0263276413476542
- Bennett, J. M. (2012) *Aquinas on the Web? Doing Theology in an Internet Age*. London; New York: T&T Clark International.
- Bermejo, F. (2007) *The Internet Audience Constitution and Measurement*. New York: Peter Lang Publishing, Inc.
- Beuving, J. J. (2019) 'Ethnography's future in the big data era', *Information, Communication & Society*, 23(11), pp. 1625-1639. DOI: 10.1080/1369118X.2019.1602664
- Blanchard, A. L. and Markus, M. L. (2004) 'The Experienced "Sense" of a Virtual Community: Characteristics and Processes', *ACM SIGMIS Database*, 35(1), pp. 64-79. DOI: 10.1145/968464.968470
- Blei, D. M. (2012) 'Probabilistic Topic Models', *Communications of the ACM*, 55(4), pp. 77-84. DOI: 10.1145/2133806.2133826
- Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3, pp. 993-1022.
- Blondheim, M. and Rosenberg, H. (2017) 'Media Theology: New Communication Technologies as religious constructs, metaphors, and experiences', *New Media & Society*, 19(1), pp. 43-51. DOI: 10.1177/1461444816649915

- Bouchet-Valat, M. (2023) *SnowballC: Snowball Stemmers Based on the C 'libstemmer' UTF-8 Library*. R package version 0.7.1, <https://CRAN.R-project.org/package=SnowballC>.
- Bourdieu, P. (1992) 'The Practice of Reflexive Sociology (The Paris Workshop)' in Bourdieu, P. and Wacquant, L. J. D. *An Invitation to Reflexive Sociology*. Chicago; London: The University of Chicago Press.
- Boy, J. D., Uitermark, J. and Wiersma, L. (2018) 'Trending #hijabfashion: Using Big Data to Study Religion at the Online–Urban Interface', *Nordic Journal of Religion and Society*, 31(1), pp. 22-40. DOI: 10.18261/issn.1890-7008-01-02
- Brookes, G. and McEnery, T. (2018) 'The utility of topic modelling for discourse studies: A critical evaluation', *Discourse Studies*, 21(1), pp. 3-21.
- Burrows, R. and Savage, M. (2014) 'After the crisis? Big Data and the methodological challenges of empirical sociology', *Big Data & Society*, April-June 2014, pp. 1-6.
- Bramer, M. (2007) *Principles of Data Mining*. London: Springer.
- Brasher, B. E. (2001) *Give Me That Online Religion*. San Francisco: Jossey-Bass Inc.
- Broström, G. (2012) *Event History Analysis with R*. Boca Raton, Florida: CRC Press.
- Bruce, S. (2011) *Secularization: In Defence of an Unfashionable Theory*. Oxford: Oxford University Press.
- Bruce, S. (2018) *Researching Religion*. Oxford: Oxford University Press.
- Brügger, N. and Milligan, I. (2019) 'Internet histories and computational methods: a "round-doc" discussion', *Internet Histories*. DOI: 10.1080/24701475.2019.1639352.
- Brumfiel, G. (2008) 'Creationists launch 'science' journal', *Nature*, 451, pp. 382-383. Available at: <http://www.nature.com/news/2008/080123/full/451382b.html> (Accessed: 8 March 2023).
- Cachia, R., Compañó, R. and Da Costa, O. (2007) 'Grasping the potential of online social networks for foresight', *Technological Forecasting and Social Change*, 74, pp. 1179-1203.
- Campbell, H. A. (2001) *An Investigation of the Nature of the Church Through an Analysis of Selected Email-Based Christian Online Communities*, PhD dissertation, University of Edinburgh.
- Campbell, H. A. (2005) *Exploring Religious Community Online We are One in the Network*. New York: Peter Lang Publishing, Inc.
- Campbell, H. A. (2007) 'Who's Got the Power? Religious Authority and the Internet', *Journal of Computer-Mediated Communication*, 12, pp. 1043-1062. DOI: 10.1111/j.1083-6101.2007.00362.x
- Campbell, H. A. (2010a) *When Religion Meets New Media*. Abingdon, Oxon; New York: Routledge.

- Campbell, H. A. (2010b) 'Religious Authority and the Blogosphere', *Journal of Computer-Mediated Communication*, 15(2), pp. 251-276. DOI: 10.1111/j.1083-6101.2010.01519.x
- Campbell, H. A. (2012) 'Religion and the Internet: A microcosm for studying Internet trends and implications', *New Media & Society*, 15(5), pp. 680-694. DOI: 10.1177/1461444812462848
- Campbell, H. A. (2013) 'Community', in Campbell, H. (ed.) *Digital Religion: Understanding Religious Practice in New Media Worlds*. Abingdon, Oxon; New York: Routledge.
- Campbell, H. A. 'New Books in Digital Religion Studies and Asking the Hard Questions', *Annual Meeting, American Academy of Religion*, San Antonio, Texas, 18-21 November 2023. Henry B. Gonzalez Convention Center.
- Campbell, H. A. and Golan, O. (2011) 'Creating digital enclaves: Negotiation of the internet among bounded religious communities', *Media, Culture & Society*, 33(5), pp. 709-724. DOI: 10.1177/0163443711404464
- Campbell, H. A. and Vitullo, A. (2016) 'Assessing changes in the study of religious communities in digital religion studies', *Church, Communication and Culture*, 1(1), pp. 73-89.
- Campbell, H. A. and Altenhofen, B. (2016) 'Methodological Challenges, Innovations and Growing Pains in Digital Religion Research' in Cheruvallil-Contractor, S. and Shakkour, S. (eds.) *Digital Methodologies in the Sociology of Religion*. London; New York: Bloomsbury Academic
- Can, U. and Alatas, B (2019) 'A new direction in social network analysis: Online social network analysis problems and applications', *Physica A: Statistical Mechanics and its Applications*, 535. DOI: 10.1016/j.physa.2019.122372
- Cheng, W. (2017) *Statistical Data Mining for Sina Weibo, a Chinese Micro-blog: Sentiment Modelling and Randomness Reduction for Topic Modelling*, PhD dissertation, London School of Economics.
- Cheong, P. H. (2013) 'Authority' in Campbell, H. (ed.) *Digital Religion: Understanding Religious Practice in New Media Worlds*. Abingdon, Oxon: New York.
- Cheong, P. H. (2014) 'Tweet the Message? Religious Authority and Social Media Innovation', *Journal of Religion, Media and Digital Culture*, 3(3), pp. 1-19.
- Cheong, P. H. (2017) 'The vitality of new media and religion: Communicative perspectives, practices, and changing authority in spiritual organization', *New Media & Society*, 19(1), pp. 25-33. DOI: 10.1177/1461444816649913
- Chiovaro M., Windsor L.C., Windsor A. and Paxton A. (2021) 'Online social cohesion reflects real-world group action in Syria during the Arab Spring', *PLoS ONE*, 16(7). DOI: 10.1371/journal.pone.0254087
- Chua, C. E. H. (2009) 'Why Do Virtual Communities Regulate Speech?', *Communication Monographs*, 76(2), pp. 234-261. DOI: 10.1080/03637750902828420

- Churchill, R. H. (2009) *To Shake Their Guns in the Tyrant's Face: Libertarian Political Violence and the Origins of the Militia Movement*. Michigan: The University of Michigan Press.
- Coco, A. (2008) 'Pagans Online and Offline: Locating Community in Postmodern Times', *Sociological Spectrum*, 28(5), pp. 510-530.
- Coleman, E. G. (2013) *Coding Freedom The Ethics and Aesthetics of Hacking*. New Jersey: Princeton University Press.
- Cooper, A-P., Kolog, E. A. and Sutinen, E. (2020) 'Exploring the Use of Machine Learning to Automate the Qualitative Coding of Church-related Tweets', *Fieldwork in Religion*, 14(2), pp. 140-159.
- Corvo, E. and De Caro, W. (2019) 'Social network and grief: a text mining analysis', *European Journal of Public Health*, 29 (supplement 4), pp. 334-335.
- Costa, E. (2018) 'Affordances-in-practice: An ethnographic critique of social media logic and context collapse', *New Media & Society*, 20(10), pp. 3641-3656.
- Cox, A. M. (2008) 'An exploration of concepts of community through a case study of UK university web production', *Journal of Information Science*, 34(3), pp. 327-345.
- Crooks, R. and Currie, M. (2021) 'Numbers will not save us: Agonistic data practices', *The Information Society*, 37(4). DOI: 10.1080/01972243.2021.1920081
- Csardi, G., Nepusz, T. (2006). 'The igraph software package for complex network research' *InterJournal, Complex Systems*, 1695, <https://igraph.org>.
- Dame-Griff, A. (2019) "Herding the 'Performing Elephants:' Using Computational Methods to Study Usenet", *Internet Histories*, 3(3-4), pp. 223-244. DOI: 10.1080/24701475.2019.1652456
- Davidson, B. I., Jones, S. L., Joinson, A. N. and Hinds, J. (2019) 'The evolution of online ideological communities', *PLoS ONE*, 14(5). DOI: 10.1371/journal.pone.0216932
- Davis, J. L. and Love, T. P. (2019) 'Generalizing from social media data: a formal theory approach', *Information, Communication & Society*, 22(5), pp. 637-647.
- Davison, A. (2021) 'Machine Learning and Theological Traditions of Analogy', *Modern Theology*, 37(2), pp. 254-274. DOI: 10.1111/moth.12682
- Dawson, L. L. (2000) 'Researching Religion in Cyberspace: Issues and Strategies' in Hadden, J. K. and Cowan, D. E. (eds.) *Religion on the Internet: Research Prospects and Promises*. Bingley: Emerald Group Publishing Ltd.
- Dawson, L. L. (2004) 'Religion and the Quest for Virtual Community' in Dawson, L. L. and Cowan, D. E. (eds.) *Religion Online: Finding Faith on the Internet*. New York; London: Routledge.
- Dawson, L. L. (2005) 'The mediation of religious experience in cyberspace' in Højsgaard, M. T. and Warburg, M. (eds.) *Religion and Cyberspace*. Abingdon, Oxon; New York: Routledge.

- Dawson, L. L. and Cowan, D. E. (2004) 'Introduction' in Dawson, L. L. and Cowan, D. E. (eds.) *Religion Online: Finding Faith on the Internet*. New York; London: Routledge.
- Dawson, L. L. and Hennebry, J. (1999) 'New religions and the internet: Recruiting in a new public space', *Journal of Contemporary Religion*, 14(1), pp. 17-39. DOI: 10.1080/13537909908580850
- Dervan, P. A. (2015) *Social Networking Theory and the Rise of Digital Marketing in the Light of Big Data*, PhD dissertation, University of Southampton.
- van Deursen, A. and van Dijk, J. (2019) 'The first-level digital divide shifts from inequalities in physical access to inequalities in material access', *New Media & Society*, 21(2), pp. 354-375. DOI: 10.1177/1461444818797082
- Devlin, J., Chang, M-W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', access via arXiv: 1810.04805 [cs.CL]
- Dobber, T., Trilling, D., Helberger, N. and de Vreese, C. (2018) 'Spiraling downward: The reciprocal relation between attitude toward political behavioral targeting and privacy concerns', *New Media & Society*, 21(6), pp. 1212-1231. DOI: 10.1177/1461444818813372
- Donath, J. S. (1999) 'Identity and deception in the virtual community' in Smith, M. A. and Kollock, P. (eds.) *Communities in Cyberspace*. London; New York: Routledge.
- Dougherty, M., Meyer, E.T., Madsen, C., van den Heuvel, C., Thomas, A. and Wyatt, S. (2010) *Researcher Engagement with Web Archives: State of the Art*. London: JISC.
- Downing, J. and Dron, R. (2019) 'Tweeting Grenfell: Discourse and networks in critical constructions of British Muslim social boundaries on social media', *New Media & Society*, 22(3), pp. 449-469. DOI: 10.1177/1461444819864572
- Driscoll, K. and Paloque-Bergès, C. (2017) 'Searching for missing "net histories"', *Internet Histories*, 1(1-2), pp. 47-59. DOI: 10.1080/24701475.2017.1307541
- Dumbleton, S. P. H. (2016) *Social Privacy: Perceptions of Veillance, Relationships, and Space with Online Social Networking Services*, PhD dissertation, University of Bedfordshire.
- Eagle, N. and Greene, K. (2014) *Reality Mining*. Cambridge, Massachusetts; London: The MIT Press.
- Ecklund, E. H., Park J. Z. and Veliz, P. T. (2008) 'Secularization and Religious Change among Elite Scientists', *Social Forces*, 86(4), pp. 1805-1839.
- Ecklund, E. H. and Scheitle, C. P. (2018) *Religion vs. Science What Religious People Really Think*. New York: Oxford University Press.
- Egger, R. and Yu, J. (2022) 'A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts', *Frontiers in Sociology*, vol. 7(886498). DOI: 10.3389/fsoc.2022.886498

- El-Kassas, W. S., Salamaa, C. R., Rafea, A. A. and Mohamed, H. K. (2021) 'Automatic Text Summarization: A Comprehensive Survey', *Expert Systems with Applications*, 165(113679). DOI: 10.1016/j.eswa.2020.113679
- Elwert, F., Tabti, S. and Pfahler, L. (2020) 'Me, myself and the other. Interreligious and intrareligious relations in neo-conservative online forums', *Religion*. DOI: 10.1080/0048721X.2020.1754603
- Ess, C., Kawabata, A. and Kurosaki, H. (2007) 'Cross-Cultural Perspectives on Religion and Computer-Mediated Communication', *Journal of Computer-Mediated Communication*, 12, pp. 939-955. DOI: 10.1111/j.1083-6101.2007.00357.x
- Etzioni, A. (2004) 'On Virtual, Democratic Communities' in Feenberg, A. and Barney, D. (eds.) *Community in the Digital Age Philosophy and Practice*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.
- Everton, S. F. (2018) *Networks and Religion Ties that Bind, Loose, Build-up, and Tear Down*. Cambridge: Cambridge University Press.
- Facebook (2021) Introductory text for Facebook Groups. Available at: <https://duckduckgo.com/?q=facebook+groups> (Accessed: 21 January 2021).
- Faimau, G. and Behrens, C. (2016) 'Facebooking Religion and the Technologization of the Religious Discourse: A Case Study of a Botswana-based Prophetic Church', *Online - Heidelberg Journal of Religions on the Internet*, 11, pp. 66-92.
- Fan, W., Wang, X., Wu, Y. and Xu, J. (2015) 'Association Rules with Graph Patterns', *Proceedings of the VLDB Endowment*, 8(12), pp. 1502-1513. DOI: 10.14778/2824032.2824048
- Fayard, A-L. and DeSanctis, G. (2005) 'Evolution of an Online Forum for Knowledge Management Professionals: a Language Game Analysis', *Journal of Computer-Mediated Communication*, 10(4). DOI: 10.1111/j.1083-6101.2005.tb00265.x
- Feenberg, A. and Bakardjieva, M. (2004) 'Consumers or Citizens? The Online Community Debate' in Feenberg, A. and Barney, D. (eds.) *Community in the Digital Age Philosophy and Practice*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.
- Feinerer, I., Hornik, K. (2024) *tm: Text Mining Package*. R package version 0.7-13, <https://CRAN.R-project.org/package=tm>.
- Feinerer, I., Hornik, K., Meyer, D. (2008) 'Text Mining Infrastructure in R.' *Journal of Statistical Software*, 25(5). DOI:10.18637/jss.v025.i05
- Feldt, L. and Høgel, C. (2018) 'Reframing Authority - The Role of Media and Materiality' in Feldt, L. and Høgel, C. (eds.) *Reframing Authority: The Role of Media and Materiality*. Sheffield; Bristol, CT: Equinox Publishing Ltd.
- Fidler, B. (2019) 'The Evolution of Internet Routing: Technical Roots of the Network Society', *Internet Histories*, 3(3-4), pp. 364-387. DOI: 10.1080/24701475.2019.1661583
- Fincham, A. and Burton, N. (2020) 'Religion and social network analysis: the discipline of early modern quakers', *Journal of Management History*, 27(3), pp. 339-358. DOI 10.1108/JMH-02-2020-0011

- Finke, R. and Stark, R. (1992) *The Churching of America 1776-1990 Winners and Losers in Our Religious Economy*. New Brunswick; New Jersey: Rutgers University Press.
- Fisher, D. (2003) 'Studying Social Information Spaces' in Lueg, C. and Fisher, D. (eds.) *From Usenet to CoWebs Interacting with Social Information Spaces*. London: Springer.
- Fisher, D. and Lueg, C. (2003) 'Appendix: Studying Online Newsgroups' in Lueg, C. and Fisher, D. (eds.) *From Usenet to CoWebs Interacting with Social Information Spaces*. London: Springer.
- Fisher, D., Smith, M. and Welser, H. T. (2006) 'You Are Who You Talk To: Detecting Roles in Usenet Newsgroups' in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, 3. Waikoloa, Big Island, Hawaii: IEEE.
- Foley, J. P. (2002) *The Church and Internet*. Available at: https://www.vatican.va/roman_curia/pontifical_councils/pccs/documents/rc_pc_pccs_doc_20020228_church-internet_en.html (Accessed: 1 August 2021).
- Foltz, F. and Foltz, F. (2003) 'Religion on the Internet: Community and Virtual Existence', *Bulletin of Science, Technology & Society*, 23(4), pp. 321-330. DOI: 10.1177/0270467603256085
- Fonseca, A. F., Bandyopadhyay, S., Louçã, J. and Manjaly, J. (2019) 'Caste in the News: A Computational Analysis of Indian Newspapers', *Social Media + Society*, 5(4), pp. 1-7.
- Fox, M. F., Johnson, D. G. and Rosser, S. V. 'Some Unintended Consequences of New Reproductive and Information Technologies on the Experience of Pregnancy Loss' in Fox, M. F., Johnson, D. G. and Rosser, S. V. (eds.) *Women, Gender and Technology*. Urbana; Chicago: University of Illinois Press.
- Fristrup, J. A. (1994) *USENET: Netnews for Everyone*. New Jersey: Prentice Hall PTR.
- Gagolewski, M. (2022) 'stringi: Fast and portable character string processing in R.', *Journal of Statistical Software*, 103(2). DOI:10.18637/jss.v103.i02
- Gahar, R. M., Arfaoui, O., Hidri, M. S. and Hadj-Alouane, N. B. (2018) 'An Ontology-driven MapReduce Framework for Association Rules Mining in Massive Data', *Procedia Computer Science*, 126, pp. 224-233.
- Gelfgren, S. (2011) 'Virtual Churches, Participatory Culture, and Secularization', *Journal of Technology, Theology, & Religion*, 2(1).
- Gelfgren, S. (2014) 'Virtual Christian Places Between Innovation and Tradition', *Online - Heidelberg Journal of Religions on the Internet*, 6, pp. 42-65.
- Genovese, F. (2018) 'Politics @Pontifex: International Crises and Political Patterns of Papal Tweets', *PS: Political Science and Politics*, 52(1), pp. 7-13.
- Ghafari, S. M. and Tjortjis, C. (2019) 'A survey on association rules mining using heuristics', *WIREs Data Mining and Knowledge Discovery*, 9(e1307). DOI: 10.1002/widm.1307

- Gilbert, S., Vitak, J. and Shilton, K. (2021) 'Measuring Americans' Comfort With Research Uses of Their Social Media Data', *Social Media + Society*, 7(3). DOI: 10.1177/20563051211033824
- Golder, S. A. and Donath, J. (2004) 'Social Roles in Electronic Communities', *Association of Internet Research (AoIR) conference Internet Research 5.0*, Brighton, England, 19-22 September 2004. University of Sussex.
- Graham, E. (2019) 'Boundary maintenance and the origins of trolling', *New Media & Society*, 21(9), pp. 1-19. DOI: 10.1177/1461444819837561
- Granovetter, M. S. (1973) 'The Strength of Weak Ties', *American Journal of Sociology*, 78(6), pp. 1360-1380.
- Grossman, W. M. (1995) 'alt.scientology.war', *Wired*. Available at: <https://www.wired.com/1995/12/alt-scientology-war/> (Accessed: 1 August 2021).
- Grossman, W. M. (1997) *net.wars*. New York; London: New York University Press.
- Hafner, K. and Lyon, M. (1996) *Where Wizards Stay Up Late The Origins of the Internet*. New York; Touchstone.
- Hahsler, M. (2017) 'arulesViz: Interactive Visualization of Association Rules with R', *R Journal*, 9(2). DOI: 10.32614/RJ-2017-047
- Hahsler, M. (2024) *arulesViz: Visualizing Association Rules and Frequent Itemsets*. R package version 1.5.3, <https://CRAN.R-project.org/package=arulesViz>.
- Hahsler, M., Buchta, C., Gruen, B., Hornik, K. (2023) *arules: Mining Association Rules and Frequent Itemsets*. R package version 1.7-7, <https://CRAN.R-project.org/package=arules>.
- Hahsler, M., Chelluboina, S., Hornik, K., Buchta, C. (2011) 'The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Datasets', *Journal of Machine Learning Research*, 12, pp. 1977-1981.
- Hahsler, M., Gruen, B., Hornik, K. (2005) 'arules - A Computational Environment for Mining Association Rules and Frequent Item Sets', *Journal of Statistical Software*, 14(15). DOI:10.18637/jss.v014.i15
- Haigh, T., Russell, A. L. and Dutton, W. H. (2015). 'Histories of the Internet: Introducing a Special Issue of *Information & Culture*', *Information & Culture*, 50(2), pp. 143-159. DOI: 10.7560/IC50201
- Ham, K. (ed.) 2013 *The New Answers Book 4*. Green Forest, Arkansas: Master Books.
- Hanegraaff, W. J., 2001. 'Prospects for the Globalization of New Age: Spiritual Imperialism Versus Cultural Diversity' in Rothstein, M. (ed.) *New Age Religion and Globalization*. Aarhus; Oxford; Oakville, Connecticut: Aarhus University Press.
- Hauben, M. and Hauben, R. (1997) *Netizens On the History and Impact of Usenet and the Internet*. Los Alamitos, California: IEEE Computer Society Press.
- Heard, J. (2008) *Re-evangelising Britain? An ethnographic analysis and theological evaluation of the Alpha course*, PhD dissertation, King's College London.

- Helland, C. (2005) 'Online Religion as Lived Religion Methodological Issues in the Study of Religious Participation on the Internet', *Online - Heidelberg Journal of Religions on the Internet*, 1(1).
- Helland, C. (2007) 'Diaspora on the Electronic Frontier: Developing Virtual Connections with Sacred Homelands', *Journal of Computer-Mediated Communication*, 12, pp. 956-976.
- Henderson, L. (2012) *USENET and the Future of Anonymity*. Great Britain: printed by Amazon.
- Herring, D. (2005) 'Virtual as contextual: a Net news theology' in Højsgaard, M. T. and Warburg, M. (eds.) *Religion and Cyberspace*. Abingdon, Oxon; New York: Routledge.
- Herring, D. (2010) 'Perichoresis and Praxis in Usenet' in Leaning, M. and Pretzsch, B. *Visions of the Human in Science Fiction and Cyberpunk*. Oxford: Inter-Disciplinary Press.
- Herring, S. C. (2004) 'Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior' in Barab, S. A., Kling, R. and Gray, J. H. (eds.) *Designing for Virtual Communities in the Service of Learning*. Cambridge: Cambridge University Press.
- Hipps, S. (2009) *Flickering Pixels How Technology Shapes Your Faith*. Grand Rapids, Michigan: Zondervan.
- Hocquet, A. and Wieber, F. (2018) 'Mailing list archives as useful primary sources for historians: looking for flame wars', *Internet Histories*, 2(1-2), pp. 38-54. DOI: 10.1080/24701475.2018.1456741
- Horsfall, S. (1999-2000) 'Use of the Internet in the Unification Movement', *Journal of Unification Studies*, 3, pp. 101-116.
- Howard, J. R. and Streck, J. M. (1999) *Apostles of Rock The Splintered World of Contemporary Christian Music*. Lexington, Kentucky: The University Press of Kentucky.
- Howard, R. G. (1997) 'Apocalypse in Your In-Box: End-Times Communication on the Internet', *Western Folklore*, 56(3-4), pp. 295-315.
- Howard, R. G. (2011) *Digital Jesus The Making of a New Christian Fundamentalist Community on the Internet*. New York; London: New York University Press.
- Hugging Face (no date) *BERT base model (uncased)*. Available at <https://huggingface.co/google-bert/bert-base-uncased> (Accessed: 31 March 2024).
- Hutchby, I. (2001) 'Technology, Texts and Affordances', *Sociology*, 35(2), pp. 441-456.
- Hutchings, T. (2010) *Creating Church online: an ethnographic study of five Internet-based Christian communities*, PhD dissertation, Durham University.
- Hutchings, T. (2013) 'Considering religious community through online churches', in Campbell, H. (ed.) *Digital Religion: Understanding Religious Practice in New Media Worlds*. Abingdon, Oxon: New York.
- Hutchings, T. (2015) 'Real Virtual Community', *Word & World*, 35(2), pp. 151-161.

- Jacobsen, B. N. (2020) 'Algorithms and the narration of past selves', *Information, Communication & Society*. DOI: 10.1080/1369118X.2020.1834603
- Jockers, M. L. (2015) *Syuzhet: Extract Sentiment and Plot Arcs from Text*, <https://github.com/mjockers/syuzhet>.
- Johnen, M., Jungblut, M. and Ziegele, M. (2018) 'The digital outcry: What incites participation behavior in an online firestorm?', *New Media & Society*, 20(9), pp 3140-3160. DOI: 10.1177/1461444817741883
- Jones, S. G. (1995) 'Understanding Community in the Information Age' in Jones, S. G. (ed.) *CyberSociety Computer-Mediated Communication and Community*. Thousand Oaks, CA; London; New Delhi: SAGE Publications, Inc.
- Kaggle (2017) *20 Newsgroups A collection of ~18,000 newsgroup documents from 20 different newsgroups*. Available at: <https://www.kaggle.com/crawford/20-newsgroups> (Accessed: 24 March 2022).
- Kesvani, H. (2019) *Follow me, Akhi The Online World of British Muslims*. London: C. Hurst & Co. (Publishers) Ltd.
- Kim, H., Jang, S. M., Kim, S. and Wan, A. (2018) 'Evaluating Sampling Methods for Content Analysis of Twitter Data', *Social Media + Society*, 4(2). DOI: 10.1177/2056305118772836
- Kimmons, R., McGuire, K., Stauffer, M., Jones, J. E., Gregson, M. and Austin, M. (2017) 'Religious Identity, Expression, and Civility in Social Media: Results of Data Mining Latter-Day Saint Twitter Accounts', *Journal for the Scientific Study of Religion*, 56(3), pp. 637-657.
- Kinnear, T. 'Computer coding to research social impacts of discussing contentious subjects online', *Doctoral Colloquium Pre-Conference of the International Society for Media, Religion, and Culture*, Bochum, Germany, 1 August 2023. Ruhr Universität Bochum.
- Kinney, J. (1995) 'Net Worth? Religion, cyberspace and the future', *Futures*, 27(7), pp. 763-776.
- Kolaczyk, E. D. and Csárdi, G. (2014) *Statistical Analysis of Network Data with R*. New York; Heidelberg; Dordrecht; London: Springer.
- Kolkman, D. (2020) 'The (in)credibility of algorithmic models to non-experts', *Information, Communication & Society*, 25(1), pp. 93-109. DOI: 10.1080/1369118X.2020.1761860
- Kołodziejska, M. (2018) *Online Catholic Communities Community, Authority, and Religious Individualization*. Abingdon, Oxfordshire; New York: Routledge.
- Kołodziejska, M. and Neumaier, A. (2017) 'Between individualisation and tradition: transforming religious authority on German and Polish Christian online discussion forums', *Religion*, 47(2), pp. 228-255.
- Kong, L. (2001) 'Religion and technology: refiguring place, space, identity and community', *Area*, 33(4), pp. 404-413.
- Krämer, N.C., Sauer, V. and Ellison, N. (2021) 'The Strength of Weak Ties Revisited: Further Evidence of the Role of Strong Ties in the Provision of Online Social Support', *Social Media + Society*, 7(2). DOI: 10.1177/20563051211024958

- Krasmann, S. (2020) 'The logic of the surface: on the epistemology of algorithms in times of big data', *Information, Communication & Society*, 23(14), pp. 2096-2109. DOI: 10.1080/1369118X.2020.1726986
- Krüger, O. (2005) 'Discovering the Invisible Internet Methodological Aspects of Searching Religion on the Internet', *Online - Heidelberg Journal of Religions on the Internet*, 1(1).
- Lancichinetti, A., Radicchi, F., Ramasco, J. J. and Fortunato, S. (2011) 'Finding Statistically Significant Communities in Networks', *PLoS ONE*, 6(4), pp. 1-18.
- Lane, J. E. (2021) *Understanding Religion Through Artificial Intelligence*. London; New York; Dublin: Bloomsbury Academic.
- Lane, J. and Shults, F. L. (2020) 'The Computational Science of Religion', *Journal for the Cognitive Science of Religion*, 6(1-2). DOI: 10.1558/jcsr.38669.
- Larsen, E., Rainie, L., Fox, S., Horrigan, J. Lenhart, A., Spooner, T., Lewis, O. and Carter, C. (2000) 'Wired churches, wired temples: Taking congregations and missions into cyberspace', *Pew Internet & American Life Project: Online life report*. Washington.
- Larson, E. J. (1985) *Trial and Error The American Controversy Over Creation and Evolution*. New York; Oxford: Oxford University Press.
- Larson, E. J. (1997) *Summer for the Gods The Scopes Trial and America's Continuing Debate Over Science and Religion*. Cambridge, MA; London: Harvard University Press.
- Lee, C. (2019) 'How are 'immigrant workers' represented in Korean news reporting?—A text mining approach to critical discourse analysis', *Digital Scholarship in the Humanities*, 34(1), pp. 82-99. DOI: 10.1093/llc/fqy017
- Lesmeister, C. (2019) *Mastering Machine Learning with R*. 3rd edn. Birmingham: Packt Publishing Ltd.
- Lewis, J. R. (2014) *Sects & Stats Overturning the Conventional Wisdom about Cult Members*. Sheffield; Bristol, Connecticut: Equinox Publishing Ltd.
- Lewis, S. C., Zamith, R. and Hermida, A. (2013) 'Content analysis in an era of big data: a hybrid approach to computational and manual methods', *Journal of Broadcasting & Electronic Media*, 57(1), pp. 34-52. DOI: 10.1080/08838151.2012.761702
- Lickliger, J. C. R. and Taylor, R. W. (1968) 'The Computer as a Communication Device', *Science and Technology*, April 1968.
- Lin, H-K., Hsieh, C-H., Wei, N-C. and Peng, Y-C. (2019) 'Association rules mining in R for product performance management in industry 4.0', *Procedia CIRP*, 83, pp. 699-704. DOI: 10.1016/j.procir.2019.04.099.
- Linderman, A. and Lövheim, M. (2003) 'Internet, Religion and the Attribution of Social Trust' in Mitchell, J. and Marriage, S. (eds) *Mediating religion: Conversations in media, culture and religion*. London; New York: T&T Clark.
- Lochhead, D. (1997) *Shifting Realities Information Technology and the Church*. Geneva: WCC Publications.

- Lőrincz, L., Koltai, J., Győr, A.F. and Takács, K. (2019) 'Collapse of an online social network: Burning social capital to create it?', *Social Networks*, 57, pp. 43-53. DOI: 0.1016/j.socnet.2018.11.004.
- Loutfi, D., Andersson, N., Law, S., Salsberg, J., Haggerty, J., Kgakole, L and Cockcroft, A. (2019) 'Can social network analysis help to include marginalised young women in structural support programmes in Botswana? A mixed methods study', *International Journal for Equity in Health*, 18. DOI: 10.1186/s12939-019-0911-8
- Lövheim, M. (2004) 'Young People, Religious Identity, and the Internet' in Dawson, L. L. and Cowan, D. E. (eds.) *Religion Online: Finding Faith on the Internet*. New York; London: Routledge.
- Lövheim, M. (2013) 'Identity', in Campbell, H. (ed) *Digital Religion: Understanding Religious Practice in New Media Worlds*. Abingdon, Oxon; New York: Routledge.
- Lövheim, M. and Campbell, H. A. (2017) 'Considering critical methods and theoretical lenses in digital religion studies', *New Media & Society*, 19(1), pp. 5-14.
- Lövheim, M. and Hjarvard, S. (2019) 'The Mediatized Conditions of Contemporary Religion: Critical Status and Future Directions', *Journal of Religion, Media and Digital Culture*, 8(2), pp. 206-225.
- Lövheim, M. and Linderman, A. G. (2005) 'Constructing religious identity on the Internet' in Højsgaard, M. T. and Warburg, M. (eds.) *Religion and Cyberspace*. Abingdon, Oxon; New York: Routledge, pp. 121-137.
- Lucas, DEwcbA., Tuszynski, J., Bengtsson, H., Urbanek, S., Frasca, M., Lewis, B., Stokely, M., Muehleisen, H., Murdoch, D., Hester, J., Wu, W., Kou, Q., Onkelinx, T., Lang, M., Simko, V., Hornik, K., Neal, R., Bell, K., de Queljoe, M., Suruceanu, I., Denney, B., Schumacher, D., Chang, W., Attali, D., Chirico, M. (2024) *digest: Create Compact Hash Digests of R Objects*. R package version 0.6.35, <https://CRAN.R-project.org/package=digest>.
- Luke, D. A. (2015) *A User's Guide to Network Analysis in R*. Cham; Heidelberg; New York; Dordrecht; London: Springer.
- Lundby, K. (2011) 'Patterns of Belonging in online/offline interfaces of religion', *Information, Communication & Society*, 14(8), pp. 1219-1235. DOI: 10.1080/1369118X.2011.594077
- Lyman, P. and Kahle, B. (1998) 'Archiving Digital Cultural Artifacts Organizing an Agenda for Action', *D-Lib Magazine*, July-August 1998. Available at: <http://www.dlib.org/dlib/july98/07lyman.html> (Accessed: 19 March 2022).
- Lyu, Z. (2023) 'Cross-cutting interaction, inter-party hostility, and partisan identity: Analysis of offensive speech in social media', *New Media & Society*. DOI: 10.1177/14614448231180654
- Mackenzie, A. (2018) 'From API to AI: platforms and their opacities', *Information, Communication & Society*, 22(13), pp. 1989-2006. DOI: 10.1080/1369118X.2018.1476569
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2023) *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.6.

- Mancosu, M. and Vegetti, F. (2020) 'What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data', *Social Media + Society*, 6(3). DOI: 10.1177/2056305120940703
- Markham, A. N., Tiidenberg, K. and Herman A. (2018) 'Ethics as Methods: Doing Ethics in the Era of Big Data Research—Introduction', *Social Media + Society*, 4(3), pp. 1-9. DOI: 10.1177/2056305118784502
- Marsden, G. M. (1994) *The Soul of the American University From Protestant Establishment to Established Nonbelief*. New York; Oxford: Oxford University Press.
- Masood, M. and Skoric, M. M. (2023) 'From Silicon to Himalayan Valley(s): Tales of Serendipitous Exposure and Interfaith Contact on Social Media'. *Social Media + Society*, 9(2). DOI: 10.1177/20563051231167335
- Masters of Scale (2018) *Build a More Human Internet* [Podcast]. Available at: <https://mastersofscale.com/caterina-fake-build-a-more-human-internet/> (Accessed: 31 July 2021).
- Masur, P. K., Bazarova, N. N. and DiFranzo, D. (2023) 'The Impact of What Others Do, Approve Of, and Expect You to Do: An In-Depth Analysis of Social Norms and Self-Disclosure on Social Media', *Social Media + Society*, 9(1). DOI: 10.1177/20563051231156401
- McClure, P. K. (2016) 'Faith and Facebook in a Pluralistic Age: The Effects of Social Networking Sites on the Religious Beliefs of Emerging Adults', *Sociological Perspectives*, 59(4), pp. 818-834.
- McQuillan, D. (2018) 'People's Councils for Ethical Machine Learning', *Social Media + Society*, 4(2), pp. 1-10. DOI: 10.1177/2056305118768303
- Meyer, E. T., Yasseri, T., Hale, S. A., Cowls, J., Schroeder, R. and Margetts, H. (2017) 'Analysing the UK web domain and exploring 15 years of UK universities on the web', in Brügger, N. and Schroeder, R. (eds.) *The Web as History*. London: UCL Press.
- Mellor, A. S. (2017) *Monitoring and Modelling of Social Networks*, PhD dissertation, University of Leeds.
- Miller, T. W. (2015) *Web and Network Data Science Modeling Techniques in Predictive Analytics*. New Jersey: Pearson Education Inc.
- Miller, T., Paloque-Bergès, C. Dame-Griff, A. (2022) 'Remembering Netizens: an interview with Ronda Hauben, co-author of Netizens: on the history and impact of Usenet and the Internet (1997)', *Internet Histories*, 7(1), pp. 76-98. DOI: 10.1080/24701475.2022.2123120
- Milligan, I. (2013) 'Exploring the USENET Archive: Early Thoughts', *Ian Milligan Digital History, Web Archives, and Contemporary History*, 6 March 2013. Available at: <https://ianmilli.wordpress.com/2013/03/06/exploring-the-usenet-archive-early-thoughts/> (Accessed: 19 March 2022).
- Milligan, I. (2016) 'Lost in the Infinite Archive: the Promise and Pitfalls of Web Archives', *International Journal of Humanities and Arts Computing*, 10(1), pp. 78-94.

- Misra, G. (2017) *Recommending Access Control Decisions To Social Media Users*, PhD dissertation, Lancaster University.
- Moreno, J. L. and Jennings, H. H. (1938) 'Statistics of Social Configurations', *Sociometry*, 1(3-4), pp. 342-374.
- Morse, T., and Birnhack, M. (2022). 'The continuity principle of digital remains', *New Media & Society*. DOI:10.1177/14614448221133535
- Mullan, M. (2015) 'Constructing an Identity Online: Logging-On as "Catholic"', *Journal of Religion, Media and Digital Culture*, 4(1), pp. 96-125. DOI: 10.1163/21659214-90000102
- Nagar, I., Hoter, E. and Hasler, B. S. (2021) 'Intergroup Attitudes and Interpersonal Relationships in Online Contact between Groups in Conflict', *Journal of Global Information Technology Management*, 24(3), pp. 208-223. DOI: 10.1080/1097198X.2021.1953318
- Nakissa, A. (2021) 'Cognitive and quantitative approaches to Islamic studies: Integrating psychological, socioeconomic, and digital-cultural statistics', *Religion Compass*, 15(12). DOI: 10.1111/rec3.12424
- Narbona, J. (2016) 'Digital leadership, Twitter and Pope Francis', *Church, Communication and Culture*, 1(1), pp. 90-109. DOI: 10.1080/23753234.2016.1181307
- Nash, D. (2002) 'Religious Sensibilities in the Age of the Internet: Freethought Culture and the Historical Context of Communication Media' in Hoover, S. M. and Clark, L. S. (eds.) *Practicing Religion in the Age of the Media*. New York: Columbia University Press.
- Naughton, J. (2015) *A Brief History of the Future The Origins of the Internet*. London: Weidenfeld & Nicolson.
- Neuhaus, F. and Webmoor, T. (2012) 'Agile Ethics for Massified Research and Visualization', *Information, Communication & Society*, 15(1), pp. 43-65. DOI: 10.1080/1369118X.2011.616519
- Neumaier, A. (2019) 'Christian Online Communities Insights from Qualitative and Quantitative Data', *Online - Heidelberg Journal of Religions on the Internet*, 14, pp. 20-40.
- Neumaier, A. (2020) 'The big friendly counter-space? Interreligious encounter within social media', *Religion*, 50(3), pp. 392-413. DOI: 10.1080/0048721X.2020.1754605
- Nissenbaum, H. (2011) 'A Contextual Approach to Privacy Online', *Daedalus*, 140(4), pp. 32-48. DOI: 10.1162/DAED_a_00113
- Novotná, M., Macková, A. and Rossini, P. (2023) 'Incivility and Intolerance in COVID-19 Discussions on Facebook', *Social Media + Society*, 9(4). DOI: 10.1177/20563051231207848
- Numbers, R. L. (2006) *The Creationists From Scientific Creationism to Intelligent Design*. Cambridge, MA; London: Harvard University Press.
- Nyland, R. and Near, C. 'Jesus is My Friend: Religiosity as a Mediating Factor in Internet Social Networking Use', *AEJMC Midwinter Conference*, Reno, Nevada, 23-24 February 2007. Brigham Young University.

- Obar, J. A. and Oeldorf-Hirsch, A. (2018) 'The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services', *Information, Communication & Society*, 23(1), pp. 128-147. DOI: 10.1080/1369118X.2018.1486870
- O'Leary, S. D. (2005) 'Utopian and dystopian possibilities of networked religion in the new millennium' in Højsgaard, M. T. and Warburg, M. (eds.) *Religion and Cyberspace*. Abingdon, Oxon; New York: Routledge.
- O'Leary, S. D. And Brasher, B. E. (1996) 'The Unknown God of the Internet' in Ess, C. (ed.) *Philosophical perspectives on computer-mediated communication*. New York: State University of New York Press.
- O'Reilly, T. (2005) 'What Is Web 2.0 Design Patterns and Business Models for the Next Generation of Software'. Available at: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1> (Accessed: 31 July 2021).
- Olson, R. S. and Neal, Z. P. (2015) 'Navigating the massive world of reddit: using backbone networks to map user interests in social media', *PeerJ Computer Science*, 1(e4). DOI: 10.7717/peerj-cs.4
- Ortiz, S.M. (2020) 'Trolling as a Collective Form of Harassment: An Inductive Study of How Online Users Understand Trolling', *Social Media + Society*, 6(2). DOI: 10.1177/2056305120928512
- Özkula, S.M., Reilly, P.J. and Hayes, J. (2022) 'Easy data, same old platforms? A systematic review of digital activism methodologies', *Information, Communication & Society*, 26(7), pp. 1470-1489. DOI: 10.1080/1369118X.2021.2013918
- Paloque-Bergès, C. (2015) 'Multi-layered archives of electronic conversations: how the past of the Internet becomes critically present again on Usenet Web archives', *RESAW 2015 conference*, Aarhus, Denmark, 8-10 June 2015. Aarhus University.
- Paloque-Bergès, C. (2017) 'Usenet as a web archive: Multi-layered archives of computer-mediated-communication' in Brügger, N. (ed.) *Web 25: Histories from the First 25 Years of the World Wide Web*. New York: Peter Lang Publishing, Inc.
- Paloque-Bergès, C. (2019) 'La Mémoire Culturelle D'Internet: Le Folklore de Usenet', *Le Temps des Médias*, 18(1), pp. 111-123. DOI: 10.3917/tm.018.0111
- Paloque-Bergès, C. (2021) 'The real "poor man's Arpanet"? A conversation about Unix networks with Kilnam Chon, godfather of the Asian Internet', *Internet Histories*, 5(3-4), pp. 376-388. DOI: 10.1080/24701475.2021.1982167
- Paloque-Bergès, C., Peuvrelle, V. and Schafer, V. (2019) 'Behind the scenes: an interview with Pierre Beyssac', *Internet Histories*, 2(3-4), pp. 315-328. DOI: 10.1080/24701475.2018.1431496
- Paloque-Bergès, C. and Schafer, V. (2019) 'Arpanet (1969–2019)', *Internet Histories*, 3(1), pp. 1-14. DOI: 10.1080/24701475.2018.1560921
- Panek, E., Hollenbach, C., Yang, J. and Rhodes, T. (2018) 'The Effects of Group Size and Time on the Formation of Online Communities: Evidence From Reddit', *Social Media + Society*, 4(4). DOI: 10.1177/2056305118815908

- Pangrazio, L and Selwyn, N. (2018) “‘It’s Not Like It’s Life or Death or Whatever’: Young People’s Understandings of Social Media Data’, *Social Media + Society*, 4(3). DOI: 10.1177/2056305118787808
- Peckham, M. (1998) ‘New Dimensions of Social Movement/Counter-movement Interaction: The Case of Scientology and Its Internet Critics’, *The Canadian Journal of Sociology*, 23(4), pp. 317-347.
- Petit, A. (2022) “‘Do female anime fans exist?’ The impact of women-exclusionary discourses on rec.arts.anime’, *Internet Histories*, 6(4), pp. 352-368. DOI: 10.1080/24701475.2022.2109265
- Pfaffenberger, B. (1996) “‘If I Want It, It’s OK’’: Usenet and the (Outer) Limits of Free Speech’, *The Information Society*, 12(4), pp. 365-386. DOI: 10.1080/019722496129350
- Pfaffenberger, B. (2003) “‘A Standing Wave in the Web of Our Communications’’: Usenet and the Socio-Technical Construction of Cyberspace Values’ in Lueg, C. and Fisher, D. (eds.) *From Usenet to CoWebs Interacting with Social Information Spaces*. London: Springer.
- Pfeifer D. and Leidner J. L. (2019) ‘Topic Grouper: An Agglomerative Clustering Approach to Topic Modeling’ in Azzopardi L., Stein B., Fuhr N., Mayr P., Hauff C. and Hiemstra D. (eds.) *Advances in Information Retrieval. ECIR 2019. Lecture Notes in Computer Science*, 11437. Springer; Cham. DOI: 10.1007/978-3-030-15712-8_38
- Phillips, P. M. (2019) *The Bible, Social Media and Digital Culture*. Abingdon, Oxon; New York: Routledge.
- Phillips, P. M., Schiefelbein-Guerrero, K. and Kurlberg, J. (2019) ‘Defining Digital Theology: Digital Humanities, Digital Religion and the Particular Work of the CODEC Research Centre and Network’, *Open Theology*, 5, pp.29-43. DOI: 10.1515/oph-2019-0003
- Phillips, P. M. (2023) ‘Digital Theology and a Potential Theological Approach to a Metaphysics of Information’, *Zygon Journal of Religion and Science*, 58(3), pp. 770-788. DOI: 10.1111/zygo.12883
- Polonski, V. W. (2017) *Commitment Issues: Toward an Understanding of Young People’s Social Media Choices in the Multi-Platform Era*, DPhil dissertation, University of Oxford.
- Pomson, A. (2008) ‘Look who’s talking: Emergent evidence for discriminating between differences in listserv participation’, *Education and Information Technologies*, 13, pp. 147-163.
- Pope Francis (2019) “‘We are members of one another” (Eph 4,25). From social network communities to the human community. Available at: http://www.vatican.va/content/francesco/en/messages/communications/documents/papa-francesco_20190124_messaggio-comunicazioni-sociali.html (Accessed: 24 January 2021).
- Pope John Paul II (2002) *Internet: A New Forum for Proclaiming the Gospel*. Available at: https://www.vatican.va/content/john-paul-ii/en/messages/communications/documents/hf_jp-ii_mes_20020122_world-communications-day.html (Accessed: 30 July 2021).

- Porter, M. A., Onnela, J-P. and Mucha, P. J. (2009) 'Communities in Networks', *Notices of the AMS*, 56(9), pp. 1082-1097.
- Possamai-Inesedy, A. and Nixon, A. (2017) 'A place to stand: Digital sociology and the Archimedean effect', *Journal of Sociology*, 53(4), pp. 865-884.
- Preece, J. and Maloney-Krichmar, D. (2005) 'Online Communities: Design, Theory, and Practice', *Journal of Computer-Mediated Communication*, 10(4). DOI: 10.1111/j.1083-6101.2005.tb00264.x
- Princeton University (2010) *About WordNet*. Available at: <https://wordnet.princeton.edu> (Accessed: 20 March 2022).
- Proferes, N., Jones, N. Gilbert, S., Fiesler, C. and Zimmer, M. (2021) 'Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics', *Social Media + Society*, 7(2). DOI: 10.1177/20563051211019004
- Quarterman, J. S. and Hoskins, J. C. (1986) 'Notable Computer Networks', *Communications of the ACM*, 29(10), pp. 932-971.
- R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, <https://www.R-project.org>.
- Ramo, J. C. and Burke, G. (1996) 'Finding God on the Web', *Time*, 148(27), 16 December 1996.
- Rau, G. (2012) *Mapping the Origins Debate Six Models of the Beginning of Everything*. Downers Grove, Illinois; Nottingham: Inter-Varsity Press.
- Recuero, R., Zago, G. and Soares, F. (2019) 'Using Social Network Analysis and Social Capital to Identify User Roles on Polarized Political Conversations on Twitter', *Social Media + Society*, 5(2). DOI: 10.1177/2056305119848745
- Refaee, E. A. A. (2016) *Sentiment Analysis for Micro-Blogging Platforms in Arabic*, PhD dissertation, Heriot-Watt University.
- Rega, R., Marchetti, R. and Stanziano, A. (2023) 'Incivility in Online Discussion: An Examination of Impolite and Intolerant Comments', *Social Media + Society*, 9(2). DOI: 10.1177/20563051231180638
- Reid, M. and Gray, C. (2007) 'Part 1. Past and Present', *Searcher*, 15(7), pp. 33-43.
- Rheingold, H. (2000) *The Virtual Community Homesteading on the Electronic Frontier*. Cambridge, Massachusetts; London: The MIT Press.
- Richterich, A. (2018) 'Tracing controversies in hacker communities: ethical considerations for internet research', *Information, Communication & Society*, 23(1), pp. 76-93. DOI: 10.1080/1369118X.2018.1486867
- Robinson, L. (2005) 'Debating the Events of September 11th: Discursive and Interactional Dynamics in Three Online Fora', *Journal of Computer-Mediated Communication*, 10(4). DOI: 10.1111/j.1083-6101.2005.tb00267.x

- Roth, S. and Luczak-Roesch, M. (2018) 'Deconstructing the data life-cycle in digital humanitarianism', *Information, Communication & Society*, 23(4), pp. 555-571. DOI: 10.1080/1369118X.2018.1521457
- Ruse, M. (2005) *The Evolution-Creation Struggle*. Cambridge, Massachusetts; London: Harvard University Press.
- Salganik, M. J. (2018) *Bit by Bit Social Research in the Digital Age*. Princeton, New Jersey; Woodstock, Oxfordshire: Princeton University Press.
- Sanders, W. S. and Ferré, J. P. (2020) 'Reader Responses to Religion News: Discussions about Ark Encounter on Reddit', *Journal of Religion, Media and Digital Culture*, 9(1), pp. 107-130.
- Schafer, M. H. (2018) 'Who Talks Religion and What Are the Consequences for Social Ties? Unpacking a Sensitive Discussion Topic in Close Networks', *Sociology of Religion: A Quarterly Review*, 79(4), pp. 395-424. DOI: 10.1093/socrel/srx069
- Schmidt, B. (2015) 'The Usenet Archive', *Ben's Bookworm Blog*, 8 May 2015. Available at: <http://bookworm.benschmidt.org/posts/2015-05-08-Usenet.html> (Accessed: 19 March 2022).
- Schneider, N. (2021) 'Admins, mods, and benevolent dictators for life: The implicit feudalism of online communities', *New Media & Society*, 24(9), pp. 1965-1985. DOI: 10.1177/1461444820986553
- Schneider, N. (2022) 'Admins, mods, and benevolent dictators for life: The implicit feudalism of online communities', *New Media & Society*, 24(9), pp. 1965-1985. DOI: 10.1177/1461444820986553
- Schroeder, R. (2019) 'Big data and cumulation in the social sciences', *Information, Communication & Society*, 23(11), pp. 1593-1607. DOI: 10.1080/1369118X.2019.1594334
- Schroeder, R. and Brügger, N. (2017) 'Introduction: The Web as History', in Brügger, N. and Schroeder, R. (eds.) *The Web as History*. London: UCL Press.
- Schroeder, R., Heather, N. and Lee, R. M. (1998) 'The Sacred and the Virtual: Religion in Multi-User Virtual Reality', *Journal of Computer-Mediated Communication*, 4(2). DOI: 10.1111/j.1083-6101.1998.tb00092.x
- Scott, E. C. (2004) *Evolution vs. Creationism An Introduction*. Berkeley; Los Angeles; London: University of California Press.
- Selivanov, D., Bickel, M., Wang, Q. (2023) *text2vec: Modern Text Mining Framework for R*. R package version 0.6.4, <https://CRAN.R-project.org/package=text2vec>.
- Sharon, S. (2021) 'User-generated nationalism: interactions with religion, race, and partisanship in everyday talk online', *Information, Communication & Society*, 24(13), pp. 1854-1869. DOI: 10.1080/1369118X.2020.1748088
- Si, H., Zhou, J., Chen, Z., Wan, J., Xiong, N. N., Zhang, W. and Vasilakos, A. V. (2019) 'Association Rules Mining Among Interests and Applications for Users on Social Networks', *IEEE Access*, 7, pp. 116014-116026. DOI: 10.1109/ACCESS.2019.2925819

- Silge, J. and Robinson, D. (2017) *Text Mining with R A Tidy Approach*. Sebastopol, California: O'Reilly Media Inc.
- Simon, P. (2013) *Too Big to Ignore*. Hoboken, New Jersey; John Wiley & Sons, Inc.
- Šimunjak, M. and Caliandro, A. (2019) 'Twiplomacy in the age of Donald Trump: Is the diplomatic code changing?', *The Information Society*, 35(1), pp. 13-25. DOI: 10.1080/01972243.2018.1542646
- Smith, M. A. (1999) 'Invisible crowds in cyberspace: Mapping the social structure of Usenet' in Smith, M. A. and Kollock, P. (eds.) *Communities in Cyberspace*. London; New York: Routledge.
- Smith, M. A. (2003) 'Measures and Maps of Usenet', in Lueg, C. and Fisher, D. (eds.) *From Usenet to CoWebs Interacting with Social Information Spaces*. London: Springer.
- Squirrell, T. (2019) 'Platform dialectics: The relationships between volunteer moderators and end users on reddit', *New Media & Society*, 21(9), pp. 1910-1927. DOI: 10.1177/1461444819834317
- Stadtfeld, C., Takács, K. and Vörös, A. (2019) 'The Emergence and Stability of Groups in Social Networks', *Social Networks*, 60, pp. 129-145. DOI: 10.1016/j.socnet.2019.10.008
- Stark, R. and Finke, R. (2000) *Acts of Faith Explaining the Human Side of Religion*. Berkeley and Los Angeles, California; London: University of California Press.
- St Mary's University (2020) *Doctoral Research Student Handbook*. <https://simmspace.stmarys.ac.uk/prog-admin/research-students/Documents/PGR-Student-Documents/Regs-and-CoP/St%20Mary%27s%20PGR%20Handbook%20for%20students%20registered%20with%20Liverpool%20Hope%202020-21%20FINAL.pdf> (Accessed: 7 February 2021).
- Stein, R. E., Corcoran, K. E., Kowalski, B. M. and Colyer, C. J. (2020) 'Congregational Cohesion, Retention, and the Consequences of Size Reduction: A Longitudinal Network Analysis of an Old Order Amish Church', *Sociology of Religion*, 81(2), pp. 206-229.
- Stockinger, A., Schäfer, S. and Lecheler, S. (2023) 'Navigating the gray areas of content moderation: Professional moderators' perspectives on uncivil user comments and the role of (AI-based) technological tools', *New Media & Society*. DOI: 10.1177/14614448231190901
- Stoycheff, E., Liu, J., Xu, K. and Wibowo, K. (2018) 'Privacy and the Panopticon: Online mass surveillance's deterrence and chilling effects', *New Media & Society*, 21(3), pp. 602-619. DOI: 10.1177/1461444818801317
- Sun, Q. and Shen, C., 2021 'Who would respond to A troll? A social network analysis of reactions to trolls in online communities', *Computers in Human Behavior*, 121(106786). DOI: 10.1016/j.chb.2021.106786
- Svensson, J. (2020) 'Computing Swedish Salafism: An Example of a Digital Humanities Approach to Collecting, Organizing, and Analyzing Data on Web-based Salafi Missionary Activity in Sweden', *Journal of Religion in Europe*, 13, pp 1-22. DOI:10.1163/18748929-20201472

- Tan, M. J. P. (2020) 'Community in the Digital Body of Christ' in Campbell, H. A. (ed.) *The Distanced Church Reflections on Doing Church Online*. Digital Religion Publications (Accessed at: <https://oaktrust.library.tamu.edu/handle/1969.1/187891>).
- Taris, T. W. (2000) *A Primer in Longitudinal Analysis*. London; Thousand Oaks, California; New Delhi: SAGE Publications.
- Thomas, D. (2018) 'Quantitative Text Mining Analysis of the Testimonies for the Church: Implications for Non-Native English Readers', *Journal of Research on Christian Education*, 27(2), pp. 145-156. DOI: 10.1080/10656219.2018.1502104
- Thomas, J., Al Shehhi, A. and Grey, I. (2019) 'The sacred and the profane: social media and temporal patterns of religiosity in the United Arab Emirates', *Journal of Contemporary Religion*, 34(3), pp. 489-508. DOI: 10.1080/13537903.2019.1658937
- Trifiro, B., Paik, S., Fang, Z. and Zhang, L. (2021) 'Politics and Politeness: Analysis of Incivility on Twitter During the 2020 Democratic Presidential Primary', *Social Media + Society*, 7(3). DOI: 10.1177/205630512111036939
- Tromble, R. (2021) 'Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age', *Social Media + Society*, 7(1). DOI: 10.1177/2056305121988929
- Tulin, M., Pollet, T. V. and Lehmann-Willenbrock, N., 2018 'Perceived group cohesion versus actual social structure: A study T using social network analysis of egocentric Facebook networks', *Social Science Research*, 74, pp. 161-175. DOI: 10.1016/j.ssresearch.2018.04.004
- Turner, T. C., Smith, M. A. Fisher, Y. and Welser, H. T. (2005) 'Picturing Usenet: Mapping Computer-Mediated Collective Action', *Journal of Computer-Mediated Communication*, 10(4). DOI: 10.1111/j.1083-6101.2005.tb00270.x
- Urban, H. (2011) *The Church of Scientology A History of a New Religion*. New Jersey; Woodstock, Oxfordshire: Princeton University Press.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2018) *Attention Is All You Need*. Available at: <http://nlp.seas.harvard.edu/2018/04/03/attention.html> (Accessed: 31 March 2023).
- Villani, C. (2015) *Birth of a Theorem A Mathematical Adventure*. London: The Bodley Head.
- Vitullo, A. (2019) 'Multisite Churches Creating Community from the Offline to the Online', *Online - Heidelberg Journal of Religions on the Internet*, 14, pp. 41-60.
- Vogt, B. (2011) *The Church and New Media*. Huntington, Indiana: Our Sunday Visitor, Inc.
- Wang, W. and Bullfrog, J. (2004) *Steal This File Sharing Book: What They Won't Tell You About File Sharing*. San Francisco: No Starch Press.
- Wang, L., Yang, A. and Thorson, K. (2021) 'Serial participants of social media climate discussion as a community of practice: a longitudinal network analysis', *Information, Communication & Society*, 24(7), pp. 941-959. DOI: 10.1080/1369118X.2019.1668457

- Ward, Jr., J. H. (1963) 'Hierarchical Grouping to Optimize an Objective Function', *Journal of the American Statistical Association*, 58(301), pp. 236-244.
- Weber, M. (2017) 'A common language', *Internet Histories*, 1(1-2), pp. 26-38. DOI: 10.1080/24701475.2017.1317118
- Weber, M., Viehmann, C., Ziegele, M. and Schemer, C. (2020) 'Online Hate Does Not Stay Online – How Implicit and Explicit Attitudes Mediate the Effect of Civil Negativity and Hate in User Comments on Prosocial Behavior', *Computers in Human Behavior*, 104. DOI: 10.1016/j.chb.2019.106192.
- Weeber, S. C. and Rodeheaver, D. G. (2003) 'Militias at the Millennium: A Test of Smelser's Theory of Collective Behavior', *The Sociological Quarterly*, 44(2), pp. 181-204. DOI: 10.1111/j.1533-8525.2003.tb00554.x
- Wellman, B. and Gulia, M. (1999) 'Virtual communities as communities: Net surfers don't ride alone' in Smith, M. A. and Kollock, P. (eds.) *Communities in Cyberspace*. London; New York: Routledge.
- Wertheim, M. (1999) *The Pearly Gates of Cyberspace A History of Space from Dante to the Internet*. London: Virago Press.
- Wessinger, C. (2000) *How the Millennium Comes Violently From Jonestown to Heaven's Gate*. New York: Seven Bridges Press.
- Whittaker, S., Terveen, L., Hill, W. and Cherny, L. (2003) 'The Dynamics of Mass Interaction' in Lueg, C. and Fisher, D. (eds.) *From Usenet to CoWebs Interacting with Social Information Spaces*. London: Springer.
- Wickham, H. (2023) *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.6.1, <https://CRAN.R-project.org/package=stringr>.
- Wickham, H., François, R., Henry, L., Müller K., Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H., Hester, J., Bryan, J. (2024) *readr: Read Rectangular Text Data*. R package version 2.1.5 <https://CRAN.R-project.org/package=readr>.
- de Wildt, L. and Aupers, S. (2019) 'Pop theology: forum discussions on religion in videogames', *Information, Communication & Society*, 23(10), 1444-1462. DOI: 10.1080/1369118X.2019.1577476
- Williams, J. R. (2019) 'The use of online social networking sites to nurture and cultivate bonding social capital: A systematic review of the literature from 1997 to 2018', *New Media & Society*, 21(11-12), pp. 2710-2729.
- Winters, J. (2017) 'Breaking in to the mainstream: demonstrating the value of internet (and web) histories', *Internet Histories*, 1(1-2), pp. 173-179.
- Wu, C., Wang, G., Zhu, J., Lertvittayakumjorn, P., Hu, S., Tan, C., Mi, H., Yadan, X. and Xiao, J. (2019) 'Exploratory Analysis for Big Social Data Using Deep Network', *IEEE Access*, 7, pp. 21446-21453. DOI: 10.1109/ACCESS.2019.2898238

- Xia, W. and Doss, H. (2020) 'Scalable Hyperparameter Selection for Latent Dirichlet Allocation', *Journal of Computational and Graphical Statistics*, 29(4), pp. 875-895. DOI: 10.1080/10618600.2020.1741378
- Xu, Z., Harzallah, M., Guillet, F. And Ichise, R. (2019) 'Modular Ontology Learning with Topic Modelling over Core Ontology', *Procedia Computer Science*, 159, pp. 562-571. DOI: 10.1016/j.procs.2019.09.211
- Yi, F., Jiang, B. and Wu, J. (2020) 'Topic Modeling for Short Texts via Word Embedding and Document Correlation', *IEEE Access*, 8, pp. 30692-30705. DOI: 10.1109/ACCESS.2020.2973207
- Zeller, B. E. (2014) *Heaven's Gate America's UFO Religion*. New York; London: New York University Press.
- Zhao, Y. (2013) *R and Data Mining Examples and Case Studies*. San Diego, California; Waltham, Massachusetts; London; Amsterdam: Elsevier Inc.
- Zhou, Y. (2017) *Understanding the Topics and Opinions from Social Media Content*, PhD dissertation, University of Warwick.
- Ziembowicz, K., Roszczyńska-Kurasińska, M., Rychwalska, A. and Nowak, A. (2021) 'Predicting conflict-prone disputes using the structure of turn-taking: the case of Wikipedia', *Information, Communication & Society* 25(13), pp. 1987-2005. DOI: 10.1080/1369118X.2021.1924224
- Zimmer, M. (2018) 'Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity', *Social Media + Society*, 4(2). DOI: 10.1177/2056305118768300
- Zuboff, S. (2019) 'Surveillance Capitalism and the Challenge of Collective Action', *New Labour Forum*, 28(1), pp. 10-29. DOI: 10.1177/1095796018819461

Appendix – evidence of ethical approval

Note, the project title was later amended with no changes to the substance of the research.



St Mary's
University
Twickenham
London

2 July 2019

SMEC_2018-19_047

Timothy Kinnear (Theology): 'What were the lasting impacts of prominent Christian social media groups on Usenet from 1983-2014?'

Dear Tim

University Ethics Sub-Committee

Thank you for re-submitting your ethics application for consideration.

I can confirm that all required amendments have been made and that you therefore have ethical approval to undertake your research.

Yours sincerely

A handwritten signature in purple ink, appearing to read 'Jamie North'.

Dr Jamie North
Chair, Ethics Sub-Committee

Cc Prof Stephen Bullivant