Social Media as Intelligence in Disaster Response: Eyewitness Classification Using Community Detection



Ross Gales

St. Anne's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

This thesis is dedicated to
all those who have been affected
by the disasters studied herein.

Abstract

Disasters cause widespread devastation to both physical infrastructure and the lives of individuals residing in large geographic areas. The disruption caused by disaster events is further compounded by high levels of uncertainty and information scarcity, presenting significant challenges to disaster response organisations and impeding the effectiveness of coordinated response efforts.

The increasing use of digital technologies, such as social media, presents valuable sources of information that are available in real-time from geographically-distributed networks of 'humans as sensors'. The data generated by these technologies can supplement traditional sources of intelligence to build models of situational awareness and inform decision-making, resulting in more effective disaster response operations.

This thesis proposes a method of curating social media data to enhance its usefulness as a source of intelligence for disaster response organisations during crisis events. The research was conducted in four phases:

(i) An ethnographic study developed a conceptual framework of the values and challenges of social media intelligence as perceived by disaster response practitioners. High data volume and low rates of relevance were established as key factors impeding integration with existing intelligence sources.

(ii) Empirical studies of Twitter discourse were conducted during eight disaster events to identify patterns of online behaviour and establish the informative potential of social media data as a rich source of eyewitness reports.

(iii) Geoproximate preferential attachment (homophily) was identified in the structure of Twitter relationship networks. An eyewitness classification model integrated relationship features for data curation. The model was evaluated on temporally-partitioned subgraphs and shown to be effective in real-time environments.

(iv) The classification model was validated in simulated disaster response scenarios conducted with emergency service practitioners. Feedback from participants confirmed the effectiveness of the approach to improving the practical value of social media data as a source of intelligence during disaster response operations.

Contents

List of Figures **ix**

List of Tables **xi**

Glossary **xiii**

1 Introduction **1**
- 1.1 Research Questions . . . 4
- 1.2 Contributions . . . 7
- 1.3 Terminology . . . 8
- 1.4 Method . . . 9
- 1.5 Publications . . . 12
- 1.6 Thesis Structure . . . 13

2 Literature Review **16**
- 2.1 Defining Disasters . . . 17
  - 2.1.1 Terminology . . . 17
  - 2.1.2 Disaster Taxonomies . . . 20
  - 2.1.3 A Working Taxonomy of Event Viability . . . 27
- 2.2 Social Media Data in Disaster Response . . . 28
  - 2.2.1 Situational Awareness and Decision Support . . . 30
  - 2.2.2 Adoption Within Disaster Response Organisations . . . 31
  - 2.2.3 Existing Approaches and Limitations . . . 33
- 2.3 Discussion and Research Gaps . . . 38
- 2.4 Summary . . . 42

3 Qualitative Study of Disaster Response Organisations **45**
- 3.1 Research Design . . . 47
  - 3.1.1 Motivation . . . 47
  - 3.1.2 Methodological Challenges for Disaster Research . . . 49
  - 3.1.3 Qualitative Research Method . . . 51
  - 3.1.4 Participant Selection . . . 56
  - 3.1.5 Ethical Considerations . . . 58

3.1.6 Timeline . . . 58
3.2 Qualitative Analysis . . . 60
3.2.1 Data Summary . . . 60
3.2.2 Thematic Coding . . . 64
3.3 Findings . . . 66
3.3.1 Social Media Intelligence in Disaster Response . . . 69
3.3.2 Challenges to Social Media Intelligence in Disaster Response 78
3.3.3 Supplementary Observations . . . 83
3.4 Discussion . . . 86
3.5 Summary . . . 89

4 Opportunities and Challenges of Social Media Data for Research 91
4.1 Benefits of Social Media Data in Research . . . 92
4.2 Social Media Data Availability for Research . . . 93
4.3 Technical Implications of Using Twitter Data . . . 96
4.3.1 Twitter Data Structure . . . 96
4.3.2 Twitter's Historical Archive and Live Data Streams . . . 98
4.3.3 API Rate Limiting . . . 101
4.3.4 Data Temporality . . . 106
4.4 Collecting Useful Datasets from Twitter . . . 109
4.4.1 Keyword and Hashtag Filtering . . . 110
4.4.2 Key Author Identification . . . 111
4.4.3 Geographic Metadata . . . 113
4.4.4 User Influence and Message Diffusion . . . 115
4.4.5 Social Network Community Detection . . . 117
4.5 Research Ethics and Privacy of Public Data . . . 118
4.5.1 Assumed and Uninformed Consent . . . 119
4.5.2 Author Deletion of Data . . . 121
4.5.3 Publication of Data . . . 122
4.6 Discussion . . . 123
4.7 Summary . . . 130

5 The CrisisData Software 132
5.1 Software Design . . . 134
5.1.1 Requirements Specification . . . 134
5.1.2 Technology Stack . . . 136
5.1.3 Collection Logic . . . 137
5.1.4 Database Structure . . . 139
5.1.5 Interface Design . . . 141
5.1.6 Post-Processing . . . 143

5.2 Data Collection 144
5.2.1 Collection Parameters 145
5.2.2 Reflections on the Data Collection Process 150
5.3 Discussion 156
5.4 Summary 160

6 Quantitative Analysis **161**
6.1 Preliminary Observations 162
6.1.1 Misinformation 162
6.1.2 Geospatial Data 163
6.1.3 Message Content 164
6.1.4 Implications to Method Design 165
6.2 Data Coding 165
6.2.1 Tweet Coding 166
6.2.2 Schema Validation and Inter-Coder Reliability 169
6.2.3 User Coding 171
6.3 Object Analysis 172
6.3.1 Recall and Precision 172
6.3.2 Comparison of Codes and Ground Truth Data 175
6.3.3 Tweet Sources 181
6.4 Discussion 186
6.5 Summary 188

7 Social Network Analysis **189**
7.1 Network Homophily 190
7.2 Hurricane Network Datasets 192
7.3 Visualising Network Structure 193
7.4 Verifying Local Clustering using Modularity 195
7.4.1 Monte Carlo Simulation and Statistical Significance 200
7.5 Identifying Local Communities 201
7.5.1 Community Structure in Networks 202
7.5.2 Community Detection on Hurricane Datasets 204
7.5.3 Evaluating Community Detection Algorithms 209
7.5.4 Characterising and Differentiating Communities 212
7.5.5 Comparison with Early Graph Structure 214
7.6 Discussion 217
7.7 Summary 221

8 Validation of User-Centric Network Based Locality Inference 223
8.1 Digital Tools for Disaster Response Analytics . . . 225
8.1.1 Software Categorisation . . . 226
8.1.2 Limitations of Existing Tools . . . 228
8.2 Prototype Development . . . 229
8.2.1 Requirements Specification . . . 230
8.2.2 Software Design . . . 232
8.3 Evaluation . . . 237
8.3.1 Participant Selection . . . 239
8.3.2 Method . . . 240
8.3.3 Results . . . 243
8.4 Discussion . . . 248
8.5 Summary . . . 254

9 Discussion 257
9.1 Thesis Summary . . . 258
9.2 Research Findings . . . 262
9.3 Research Contributions . . . 266
9.4 Limitations and Challenges . . . 269
9.4.1 Participant Access During Disasters . . . 269
9.4.2 Data Bias and Technology-Induced Privilege of Care . . . 270
9.4.3 Psychosocial Safety and Secondary Trauma . . . 271
9.5 Further Research . . . 272
9.5.1 Preparing for Live Deployment . . . 272
9.5.2 Evaluating Generalisability of Findings . . . 272
9.5.3 Analysing Bidirectional Relationship Data . . . 273
9.5.4 Integrating Approaches to Intelligence . . . 273
9.5.5 Expanding System Scope — the Russo-Ukrainian War . . . 274
9.6 Conclusion . . . 276

Appendices

A Qualitative Research Documents 278
A.1 Interview Questions . . . 278
A.1.1 Exploratory Study . . . 278
A.1.2 Prototype Evaluation . . . 279
A.2 Participant Consent Form . . . 280

B Twitter Event Datasets **282**
B.1 Collection Parameters 282
B.1.1 Mudslides; Sierra Leone 282
B.1.2 Terror Attacks; Barcelona, Spain 282
B.1.3 Hurricane Hato; Hong Kong 282
B.1.4 Hurricane Harvey; Texas, U.S.A. 282
B.1.5 Earthquake; Central Mexico 283
B.1.6 Hurricane Irma; Florida, U.S.A. 283
B.1.7 Wildfires; California, U.S.A. 283
B.1.8 Hurricane Florence; North Carolina, U.S.A. 283
B.1.9 Hurricane Michael; Florida, U.S.A. 284
B.1.10 Hurricane Willa; Sinaloa, Mexico 284

C Twitter Source Metrics **285**
C.1 Twitter Source Relevancy — Hurricane Harvey 285

References **287**

List of Figures

1.1 Research approach 10

2.1 Disaster taxonomy 22

3.1 Adapted GDIA process 54
3.2 Example of the adaptive timeline format 59
3.3 Floor plan of State Control Centre 63
3.4 Intelligence officer desks 64
3.5 Emergency control centre operations floor 64
3.6 Social media integration spectrum 69
3.7 Social media behaviour adoption rates 70
3.8 Social media report interpretation flowchart 71
3.9 Perceived source veracity 81

4.1 Twitter relationship terminology 97
4.2 Twitter object data structure 98
4.3 Twitter authorisation UI 105

5.1 CrisisData technology stack 136
5.2 Application framework 138
5.3 Tweet and user database framework 140
5.4 Data collection in progress. 141
5.5 Geospatial visualisation of data collection parameters and Tweets . 142
5.6 Summary view of detected hashtags, users, and URLs 142
5.7 Geographic bounding boxes used for data collection. 146

6.1 Misinformation — a shark swims on a flooded highway 163
6.2 Tweet codes as proportions of coded data — Hurricane Harvey . . . 168
6.3 Precision and recall equilibrium 175

7.1 Homophily in school friendship networks 191
7.2 Social network structures 193
7.3 Node locality structure — Hurricane Harvey (manually coded). . . 195
7.4 Node locality structure — Hurricane Harvey (geocoded) 198

7.5 Node locality structure — Hurricane Florence (geocoded) 199
7.6 Assortativity of 100 configuration models — Hurricane Harvey . . . 201
7.7 Network of coauthorships in a university department 203
7.8 Distribution of community size by algorithm. 205
7.9 Community network structure — Louvain algorithm 207
7.10 Proportion of local profiles per community — Louvain algorithm . . 208
7.11 Cramér's V (ϕ_c) — node locality and community label. 211
7.12 Normalised runtimes for community detection algorithms. 212
7.13 Growth of graph size measured at four-hour time intervals. 215
7.14 Cramér's V (ϕ_c) — node locality and community label for temporal subgraphs. 215
7.15 Cramér's V (ϕ_c) — node locality and community label for subgraphs. 216
7.16 User centric network based method of binary locality inference. . . . 221

8.1 Example of existing system: TweetDeck 234
8.2 Heatmap figure 235
8.3 Event interface of the prototype. 238

List of Tables

1.1 List of publications 13
1.2 List of presentations 13

2.1 Categorising disasters 21
2.2 Crisis classification matrix 22
2.3 Hazard categories and sub-categories 24
2.4 Working taxonomy of disaster suitability 28
2.5 Research questions 44

3.1 Adapted GDIA application steps 53
3.2 Research participants 62
3.3 Coding matrix with illustrative data excerpts — data value 67
3.4 Coding matrix with illustrative data excerpts — data challenges . . 68

4.1 Twitter API rate limits 102

5.1 Dataset metrics 148

6.1 Tweets by code — Hurricane Harvey 168
6.2 Tweet code contingency matrix 170
6.3 Users by code — Hurricane Harvey 172
6.4 Type I and II errors 173
6.5 Profile location confusion matrix 176
6.6 Profile location as locality predictor 177
6.7 Example profile location strings 178
6.8 Tweet from local region as locality predictor 179
6.9 Comparison of composite predictors 180
6.10 Top 10 Tweet sources 181
6.11 Top 10 Tweet sources — geographic stream 183
6.12 Twitter source taxonomy 185
6.13 Source filtered relevancy measures 186

7.1 Network sizes 192
7.2 Network assortativity coefficient 200

7.3 Results of community detection algorithms. 206
7.4 Cramér's V (ϕ_c) — node locality and community label. 210
7.5 Normalised runtimes for community detection algorithms. 211
7.6 Spearman's rho — community metric variables and community locality. 213

8.1 Feature comparison of analytic software categories. 227
8.2 Summary of key values and challenges of social media data. 231
8.3 Requirement satisfaction matrix for validation software. 237
8.4 Subset of research participants who contributed to evaluation study. 240
8.5 List of tasks performed during observational study of prototype. . . 242
8.6 List of success criteria defined for P_{novel} evaluation. 243
8.7 Evaluation of success criteria for P_{novel}. 248

C.1 Twitter source relevancy. 286

Glossary

Assortativity . The preference for nodes within a network to share edges with similar nodes.

API . . . . . Application Programming Interface.

Geocoding . . Inferring geospatial data from other features (such as text strings).

Geotag . . . . Geospatial data attached to a data object.

Graph . . . . . A data structure in which objects and relationships between them are described.

Homophily . . A sociological concept in which individuals show a tendency to associate with similar others.

LCC . . . . . Largest connected component (of a graph).

Modularity . . A measure of clustering behaviour within a network.

Network . . . See *graph*.

Precision . . . A measure of classification performance — the proportion of retrieved items that are relevant.

Recall . . . . . A measure of classification performance — the proportion of relevant items retrieved.

1 Introduction

Contents

1.1 Research Questions 4
1.2 Contributions 7
1.3 Terminology 8
1.4 Method 9
1.5 Publications 12
1.6 Thesis Structure 13

Disaster events, such as hurricanes, earthquakes, and volcanic eruptions, can affect large regions and populations while introducing high levels of uncertainty, which in turn limit the effectiveness of coordinated emergency response efforts. Information becomes a critical resource that must be gathered and interpreted as efficiently as possible and therefore disaster response efforts are structured around the collection and dissemination of intelligence which informs decision-making processes. These decisions rely on sources being *informative*, *verifiable*, and *timely*, though typically, these three characteristics exist in a state of competition wherein an improvement in one dimension coincides with deterioration of another. Combining information from a range of sources possessing different 'profiles' of these characteristics therefore effectively diversifies the set of data available during disaster response operations.

In recent years, social media has grown from a novel platform for social interaction to, for many people, a primary portal through which its users communicate with the world. These interactions range from intimate messages between friends and family members to global public broadcasts from world leaders. Many of these platforms have also positioned themselves as a source of news content, the feeds of which are curated by the accounts a user chooses to follow. The speed at which information can propagate through these networks makes them an attractive source of information to people affected by disaster events; eyewitness reports created by other users of the platform appear more quickly than those from traditional news outlets, which must first verify their information. Users are able to broadcast their own messages to a wide audience of followers, often enriching these messages with imagery captured by phone cameras. For many platforms, these messages are visible to the public, thus furthering their potential reach.

Social media services therefore become dynamic information-sharing networks during a disaster where first-hand accounts produced within affected areas provide unique perspectives on the impact of the event and the needs of affected population. These publicly-available reports can cover a wide geographic area and are often created within moments of a disaster event occurring. As photo or video footage is increasingly included in public messages, the data observed on social media platforms resemble a geospatially distributed network of sensors.

Social media data therefore present a rich source of timely data which can supplement traditional sources used by disaster response organisations. However, there remain some key challenges preventing a more widespread implementation of these data in disaster information systems. The volume of data created on social media platforms is immense, and the information-seeking behaviour which follows disasters leads to sharp spikes in online activity (Ofli, Imran, et al. 2020). Twitter datasets filtered for event-related keywords in 2017 recorded over 18 million Tweets discussing Hurricane Harvey and 17 million for Hurricane Irma (Littman 2017). Only a small subset of these messages are created by users 'on the ground' at a disaster site, and even fewer contain information useful to response organisations.

Manually inspecting and categorising datasets of this size is clearly beyond the capabilities of a human operator, and thus the free and timely eyewitness reports which exists within these sources of data are often neglected by organisations lacking the capacity to process them.

The keyword filtering used in the collection of these datasets is a broad approach, capturing messages from around the world which are rarely informative in developing a ground-truth understanding. Further automated curation is required to reduce the stream to a volume that is digestible to human operators. For example, if non-local authors are pruned from a stream, the signal (eyewitness) to noise (non-eyewitness) ratio is improved, making the curated stream a more attractive source of data for human analysis. Restricting such feeds by location is difficult: fewer than 2% of messages on Twitter contain geographic coordinate information (Leetaru 2019b), and those that do typically skew towards messages authored by business and public entity accounts. Twitter profiles have an optional location field which can indicate the home location of an author as an open text string, though the feature is unverified and used for diverse purposes.

Emergency responders cannot therefore rely on user-provided location metrics to filter streams of data by removing messages created outside the area of interest. Curating a subset of messages authored by eyewitness users requires a holistic approach to data analysis which evaluates a range of characteristics (Chong and Lim 2017). Automated methods by which media are classified are well-documented in the literature, and most commonly assess the language used in individual messages to infer location (Schempp et al. 2019; A. Kumar and J. P. Singh 2019; Li et al. 2019). However, language-based models are generally constrained by the language in which they have been trained, and thus may not generalise to other events (Ofli, Meier, et al. 2016). Furthermore, model performance entropies as the character of online discourse evolves, deviating from rapidly-outdated training corpora. Finally, a message-centric approach to identification may neglect a proportion of messages from eyewitness authors: for example, where a message contains only photo or video material, or the language used by the author is not appropriately interpreted by the collection

method. Given that eyewitness authors are likely to remain as such throughout the course of an event, adopting an author-centric approach to classification is desirable to ensure that all reports made by a given eyewitness are captured.

This thesis develops a language-agnostic, author-centric method of eyewitness classification. It evaluates metrics that are more robust to culture-driven deviation from training datasets than use of language. For example, the principle of network homophily states that the social relationships of a subject reflect the characteristics of the subject, such as geographic location (McPherson et al. 2001). Based on this principle, the follower/followee networks of Twitter users are shown in this work to provide meaningful predictive properties for location inference and eyewitness classification.

Generalisable models for social media stream curation are informed by perspectives derived from ethnographic studies of disaster response organisations to maintain an alignment between system features and domain requirements. The models developed in this work are implemented in a prototype system which demonstrates the effectiveness of a network-centric approach to filtering data streams for analysis by human operators.

1.1 Research Questions

The purpose of this research is to enhance the utility of social media data to disaster response organisations during disaster events. More specifically:

> How can social media data be made a more useful source of intelligence to disaster response organisations?

This overarching research theme is addressed in the thesis by a series of research questions, each of which builds upon findings from the preceding questions:

> RQ_1 —What opportunities and challenges are presented to the intelligence processes of disaster response organisations by social media data?

Developing automated methods to classify and present data for interpretation by human operators working in disaster response necessitates a deep understanding of the disaster response domain. The constraints under which these organisations operate must inform system design. In exploring this question, the intelligence protocols of response organisations are documented to develop a conceptual framework to understand the ways social media data can augment existing processes and the factors currently impeding the realisation of this value.

These barriers to implementation must be accounted for when considering the potential of social media intelligence and range from the technological to the political. For example, a decree to disregard unreliable social media reports due to the potential for legal repercussions becomes an issue of governance rather than data verification. On the other hand, a lack of manpower available to process social media data can be addressed through automated identification and curation methods. Documenting these challenges is therefore key to directing the focus of research to improve the value of social media intelligence.

> RQ_2 —How can publicly available social media data provide meaningful intelligence to disaster response organisations during disaster events?

The conceptual framework of disaster response intelligence processes is complemented by empirical analyses of online discourse. By documenting patterns of behaviour during disaster events, opportunities are identified for the integration of social media data with existing disaster response information systems. In answering this question, the categories of information published on social media are quantified. The methods of data curation developed in RQ_3 are based on the rates at which information useful to response organisations exists within online discourse.

> RQ_3 —To what extent can graph-structured relationship data inform eyewitness classification for social media data?

Common approaches to social media classification examine message content to establish relevance to the disaster event (Karimi et al. 2013; Zhang and Vucetic 2016; Imran, Mitra, et al. 2016) or infer location (Schempp et al. 2019; A. Kumar and

J. P. Singh 2019; Li et al. 2019). Models which interpret the text of a message are subject to challenges of generalisation as the linguistic structure of online discourse is shaped by the type of event, affected region, or point in time. A user-based approach to classification is more suited to capturing eyewitness reports: given that the author of an eyewitness message is local to the disaster event, there exists a high probability that further messages from the author are also published from within the affected region.

This question explores the significance of an author's social network in predicting their eyewitness status. Social networks often exhibit homophilic (assortative) properties (Newman 2018); users are more connected to other users with whom they share similar traits. The presence of strong geographic assortativity within the graph structure can present as distinct clusters of nodes representing geoproximate communities. The membership of a node to a particular cluster can therefore inform an eyewitness classification model based on the social relationships of the user.

> RQ_4 —How well can a network-based user-centric eyewitness classification approach curate social media data and address the volume constraints of disaster response organisations?

The classification methods developed in this work are evaluated with respect to their ability to address the requirements of the domain of disaster response, as identified in RQ_1. The validity of the methods developed in this research are predicated upon their ability to fulfil the needs of disaster response organisations and the constraints under which they operate. Testing the effectiveness of the findings from RQ_3 therefore requires *in situ* evaluation.

To answer this research question, software is developed in this thesis that implements the models proposed by the findings of the previous study whilst respecting the constraints identified in RQ_1. A prototype system is presented to disaster response practitioners and compared to an existing system during a simulated disaster event. The response data provided by participants is used to evaluate the extent to which the prototype method improves the utility of social media data as intelligence and validate the findings of this research.

1.2 Contributions

The thesis provides the following contributions to the fields of computer science and disaster response research:

C-1 **Empirical studies of online behaviour on Twitter during ten disaster events** contributed valuable perspectives to social media research and informed a conceptual framework of the values presented by these data to disaster response information systems. These analyses extended existing empirical work with results drawn from disaster event data. The novel method of data collection used in this work further augmented the empirical perspectives provided by these results with captured relationship data.

C-2 **A disaster suitability working taxonomy** was synthesised from the literature and contributed a framework within which disaster events were evaluated with respect to how they shaped the character of the resulting online discourse and, in turn, the value of the discourse as a source of intelligence.

C-3 **A conceptual framework of disaster response perspectives of social media data** documented the ways in which social media data presented value to disaster response organisations and the key challenges preventing its further use in the domain. These findings build upon existing studies of intelligence officer needs and introduced perspectives from a broader set of organisations and stakeholders. The ethnographic study of disaster response organisations is constrained by methodological challenges due to the intensity of response operations and therefore remains an underdeveloped area of research, making the framework and reflection on the study's conduct a crucial contribution to the field of disaster response research.

C-4 **Two methods of user locality inference as a means of eyewitness classification** were evaluated in chapter 6: parsing Tweet history for geospatial data and geolocating profile location information. These approaches were augmented with the network-based user-centric location inference method developed in chapter 7 which examined relationship data and community detection methods. These findings comprise a key contribution to computer

science research based on network relationship data and other fields of social media classification research.

C-5 **A novel data collection tool** that enhanced data detected using classic filtering techniques with user relationship data was necessitated by the network-based analyses proposed by this research. The temporally fragile features of network data were captured in real-time using live data streams. The data collection tool has been released as open-source software and, together with the documentation of the design process, presents a practical contribution of this work to computer science research using novel sets of social media data.

C-6 **The development of the disaster intelligence prototype system** used to validate the findings of the research in chapter 8 provided a second minor practical contribution as a reflection on software development processes grounded in the disaster intelligence domain.

1.3 Terminology

The following terminology and distinctions are used throughout this thesis:

- *disaster*, *crisis*, and *emergency* are used interchangeably based on context due to different uses of terminology between organisations and academic discourse. The definition of the terms as used in this work is discussed in section 2.1 and may be loosely conceptualised by the definition provided in Quarantelli (2000):

  > 'Disasters are relatively sudden occasions when, because of perceived threats, the routines of collective social units are seriously disrupted and when unplanned courses of action have to be undertaken to cope with the crisis.'

- *disaster response organisation*, also referred to as *response organisation*, includes any organisation which responds to disasters (as defined above). Within the scope of this research, the term is constrained to disaster response organisations that collect, analyse, or act upon intelligence information (which may include social media data).

- *user* is used within two contexts, in the first instance, it refers to the end user of a software system. In the second, *user* and *author* are used interchangeably to refer to the user account responsible for publishing a social media post. While a user account may be controlled by a number of people or automated by software, rendering the single-person terminology inaccurate, this distinction is unnecessary unless specified within the text.
- *followee* is coined to refer to the subject of a *follower* relationship for online social platforms where user relationships are unidirectional. That is, a *user* account may have *followers* (other user accounts which follow the user) and *followees* (user accounts that the user follows). User accounts may be both a follower and a followee of a given user (i.e. a reciprocal relationship).
- Within the context of this research, the term *social network* presents ambiguity. While the text attempts to provide sufficient context for accurate interpretation, the two meanings are described here for clarity. Hyphenation is added for illustrative purposes:
    1. social-network data — data that have been drawn from online social media platforms.
    2. social network data — network data that document the social relationships of a community.
- A *social media platform* is typically an online web service that facilitates online discourse. The term is analogous to *social network service*, which was avoided to minimise confusion introduced by the conflicting *social network* terminology.

1.4 Method

The primary goal of this research was to improve the value of publicly available social media data to disaster response organisations. Analysis of online public discourse therefore comprised a core aspect of the research approach and was used to develop a conceptual model of social media behaviour from which methods of quantitative curation were developed.

The application of automated methods by which social media data could be manipulated required an understanding of the environmental context within which it was conducted. The complex requirements of disaster response organisations were assessed in section 2.2.1 and demonstrated the importance of aligning system design with informed perspectives drawn from the targeted domain.

Therefore, a mixed-methods approach was adopted by this research (Denzin and Lincoln 2011; Tashakkori and Creswell 2007; Greene 2007; Auerbach and Silverstein 2003) which combined ethnographic studies of disaster response organisations with empirical analyses of social media data. The subsequent development of quantitative approaches to data curation tied together the domain-informed perspectives of social media intelligence with the observed structure and patterns of behaviour within online discourse.

The broad method of the thesis can therefore be considered as a sequence of four related studies implementing distinct submethods. The research approach is illustrated in figure 1.1 and described below.

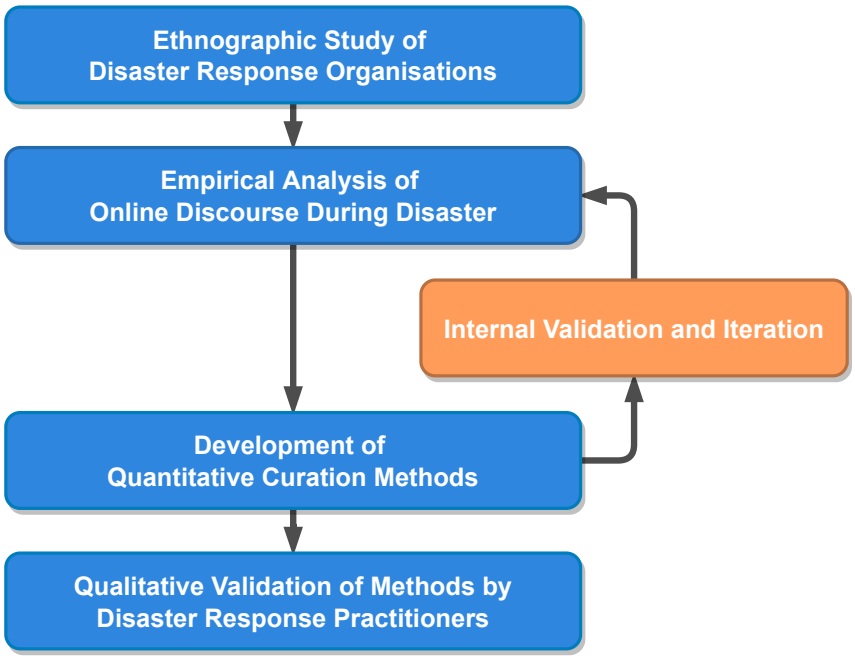


Figure 1.1: Research approach

1. **Ethnographic Study of Disaster Response Organisations:** An ethnographic study of disaster response organisations was conducted to develop a domain-informed understanding of the ways in which social media data enhance intelligence-gathering processes (contribution C-2). In particular, existing implementations of social media intelligence were examined and key factors impeding its integration were identified (contribution C-3).

   Capturing the processes and needs of disaster response practitioners as they work in response to disaster events was conducted as a situated observational study (Ciesielska et al. 2018; Wilson and Sharples 2015; L. Cohen et al. 2002; Dingwall and Miller 1997), however, a number of key limitations were encountered with respect to the disaster response domain. First, response operations are typically conducted within an environment of high stress and workload and therefore the tolerance of practitioners towards an embedded researcher was limited. Second, disaster events are naturally unpredictable and therefore it was recognised that awaiting periods in which *in situ* observation was possible would not have been compatible with the research timeline. For this reason, the observational study was conducted during a period of low alert and supplemented with interview data in which participants were asked to recollect instances of high alert response.

2. **Empirical Analysis of Online Disaster Discourse:** Following the ethnographic study, a substantive framework of online discourse was developed using empirical analyses of online social network data (contribution C-1) captured in real-time during a series of disaster events (contribution C-5). These perspectives contextualised the ethnographic findings and established a domain-informed approach to data curation for disaster response intelligence which was grounded in empirical data.

   Public online discourse published during a series of disaster events was compiled into datasets for analysis. A qualitative study identified broad themes present within the data and characterised online behaviour during disasters. A coding schema was defined to categorise data based on their

value as intelligence. Methods of locality inference were developed to represent ground truth data in the subsequent analyses.

3. **Development of Quantitative Classification Methods:** Network data representing author relationships were examined to facilitate the development of a network-based quantitative approach to author locality inference which matched the patterns of behaviour in online discourse with the organisational requirements of response organisations (contribution C-4). The presence of clustered community structures was established and evaluated in terms of predictive power in user locality classification. A network-based user-centric approach to locality inference was developed and tested on various temporal representations of event data. The algorithm designed in this study was deployed in a prototype tool which implemented a number of supplementary findings based on established organisational needs.
4. **Qualitative Validation of Methods by Disaster Response Practitioners:** The findings of the research were validated by presenting a prototype system to disaster response practitioners in a supervised naturalistic observation study using simulated data (contribution C-6). In this way, the extent to which the system design embedded the domain values derived from the ethnographic study was evaluated and shown to be an effective approach to improving the value of social media data as a source of intelligence to disaster response organisations.

1.5 Publications

The key papers derived from this research and currently in draft are presented in table 1.1 with reference to the chapters from which they were derived. Papers yet to be completed include target journals or conferences for reference, which are contained within brackets. A selection of presentations that were based on the work conducted in this thesis are included in table 1.2.

Paper	Journal/Conference	Chapters
Detecting Eyewitness Twitter Users During Hurricane Florence	ISCRAM 2023, Omaha	6, 7, 8
CrisisData: Twitter Relationship Data Collection for Crisis Response	ACM SIGKDD 2023, Long Beach	5
Crisis Response Organisations and Attitudes to Social Media Intelligence	(ISCRAM 2024)	3
Methods for Capturing Live and Large-Scale Social Media Data	(SIGKDD 2024, WWW 2024)	4
#HurricaneHarvey: Community Sensemaking During Hurricane Events	(CSCW)	6

Table 1.1: List of publications

Presentation	Location
Network Analysis for Social Media Data	University of Oxford, 2022
Location Inference Using Network Data	Deepmind, London, 2021
Understanding Twitter Data for Sociologists	University of Oxford, 2020
Network Analysis for Information Discovery	PSC-E Conference, Bled, 2018
Bushfire Tracking with Social Media Streams	University of Melbourne, 2017
Social Media in Disaster Response	TU Graz, 2018

Table 1.2: List of presentations

1.6 Thesis Structure

Chapter 2 — Literature Review: This chapter examines the characteristics distinguishing disaster events and demonstrates how these influence the approach taken by this research. A working taxonomy is developed to constrain the scope of this research to a subset of event types according to their suitability. A review of disaster research literature documents existing cases in which social media data are used as intelligence and notes the established challenges facing its further use. Gaps in the literature are identified and formulated as the research questions examined by this research.

Chapter 3 — Qualitative Study of Disaster Response Organisations: An qualitative study is conducted to evaluate the extent to which disaster response organisations use social media intelligence. Existing intelligence processes are documented and the viability of implementing social media intelligence is assessed. The findings of this work are distilled into a conceptual framework of social media in disaster response from which subsequent studies are informed.

Chapter 4 — Opportunities and Challenges of Social Media Data for Research: The landscape of social media research is explored with regard to data availability and ethical implications. For reasons of data access, Twitter is selected as the focus of this work. The technical challenges and limitations of Twitter as a source of research data are documented and provide context for the design of a data collection protocol.

Chapter 5 — The CrisisData Software: Custom software is developed to capture live Twitter streams enhanced by user relationship data. The network-driven analysis proposed by this research requires relationship data captured in real time due to the impermanence of network edges in social networks (Vieweg, Palen, et al. 2008) and therefore existing capture methods or archived datasets could not be used. Novel datasets for a series of eight disaster events are recorded and examined.

Chapter 6 — Quantitative Analysis: Two disaster event datasets are are analysed. Messages are manually coded and the presence of message classes that align with the domain requirements is established. User locality is coded as a binary value and used to evaluate the performance of methods of location inference which provide approximations of ground truth data used to evaluate the network-based inference approach in the following chapter.

Chapter 7 — Social Network Analysis: The graph structure of the event data is examined and assortative patterns of mixing between users are established. Strong community structure is shown to exist within each dataset and correlate with the 'eyewitness' status of a user. A network-based user-centric method of locality inference is developed and tested on temporal representations of event data.

Chapter 8 — Validation of User-Centric Network Based Locality Inference: The extent to which domain-specific challenges have been addressed by this research is measured through an *in situ* evaluation by disaster response practitioners. A prototype system implementing the findings of the previous chapters is populated with a synthesised data stream simulating a disaster response event and compared to existing tools.

Chapter 9 — Discussion: A reflection on the research is presented. Key outcomes are drawn out and examined, and directions for future research are established.

2
Literature Review

Contents

2.1 Defining Disasters 17
2.1.1 Terminology 17
2.1.2 Disaster Taxonomies 20
2.1.3 A Working Taxonomy of Event Viability 27
2.2 Social Media Data in Disaster Response 28
2.2.1 Situational Awareness and Decision Support 30
2.2.2 Adoption Within Disaster Response Organisations . . . 31
2.2.3 Existing Approaches and Limitations 33
2.3 Discussion and Research Gaps 38
2.4 Summary 42

This review of the literature sets out the theoretical and methodological frameworks used in this research. The first part of this chapter examines the scope of the term 'disaster' as used in existing literature, and discusses key defining characteristics with respect to their influence on online public discourse and consequently, the potential for these data to provide meaningful intelligence to disaster response organisations. A *disaster suitability working taxonomy* is constructed with which candidate events were evaluated in chapter 5 as subjects of the research (contribution C-2).

The second part of the chapter documents empirical studies of the role of social media data as intelligence in disaster response organisations and explores the opportunities and challenges presented by these data. While the value of social media

data as intelligence was widely acknowledged by disaster response practitioners, integration of these data with existing systems was limited. Perspectives provided by disaster response practitioners inform the construction of a framework from which existing approaches to social media data curation for disaster response are evaluated and later enhanced by a qualitative study conducted in chapter 3. Factors contributing to the misalignment between the needs of response organisations, the design of existing systems, and the characteristics of online discourse are identified and shape the formulation of the core research questions. Gaps in existing work are discussed and shape the formulation of four research questions examined in this thesis.

2.1 Defining Disasters

The term 'disaster event' which is used throughout this thesis encapsulates a broad range of scenarios, each with its own unique characteristics. The set of dimensions across which each event is made distinct from another had a significant impact on the methodologies adopted by this research and the extent to which its findings were generalisable to other types of event. For example, the long, drawn-out nature of a bushfire event is associated with patterns of behaviour and social action that have little in common with that of a sudden terrorist attack.

For this reason, the characteristics differentiating the events for which disaster response organisations are responsible were explored with respect to their influence on patterns of public discourse. This analysis positioned event types within a working taxonomy which was used in chapter 5 to evaluate and select the events used in this research to develop intelligence systems.

2.1.1 Terminology

While often used synonymously, the terms 'disaster', 'crisis', and 'emergency' have distinct meanings within academic discourse, and are associated with separate fields of research. While disaster research encapsulates all phases of a disaster life cycle (response, recovery, preparedness and mitigation) (Mileti et al. 1975), crisis research

typically focuses on limiting the effects of an emerging or escalating incident during the period in which the disaster may only exist as a potential threat (Boin et al. 2018), and emergency research is generally associated with the medical domain. For the purposes of this research, the distinction was unimportant; the scope of the research was predicated upon the outcomes of this literature review and an initial exploratory analysis of the domain. Furthermore, the terms were used interchangeably by the subjects of this research and therefore appear as such within this work.

The set of events to which the term 'disaster' may refer is not as intuitive as one may expect. Instinctually, the reader may first imagine large-scale *natural disasters* such as volcanic eruptions, earthquakes and bushfires. While these events, or *hazards*, may certainly fall within the scope of the term, they must be evaluated with respect to their impact on local human (or even animal) populations. Minor earthquakes are commonplace in cities such as Tokyo, which are built upon tectonic plate boundaries. In Australia, bushfires are a natural summertime phenomenon — yet they only infrequently cause disaster-level events. Therefore, the broad categorisation of an event does not provide sufficient information for classification as a disaster: contextual features must be included.

White et al. (1978) defines a disaster as an extreme event that arises when a *hazard agent* intersects with a *human-use system*. This perspective views hazards as naturally occurring events which only become disasters once they affect a human population. Quarantelli (2005) argues that the hazard-centric approach views the disaster as an epiphenomenon rather than a central focus of the definition, though it is equally true that 'a disaster is but a moment or materialization of [important] underlying conditions' (Birkmann et al. 2014). Within the context of this research, the range of events for which disaster response organisations are responsible exceeded those caused only by natural phenomena and therefore a framework that included anthropogenic events was sought.

While governments develop 'mandated definitions' of disaster for the purpose of distributing funds and other resources (Britton 2005; Buckle 2005), for many disaster response organisations, defining a strict disaster taxonomy introduces unnecessary

scoping constraints which are unable to capture novel types of event. The United Nations Office for Disaster Risk Reduction (UNDRR) defines a disaster broadly by the measure to which it impacts society:[1]

> '[A disaster is] a serious disruption of the functioning of a community or a society at any scale due to hazardous events interacting with conditions of exposure, vulnerability and capacity, leading to one or more of the following: human, material, economic and environmental losses and impacts.'

Within academic research, a definitional consensus does not exist, nor is there an expectation that this will be achieved (Alexander 2005; Quarantelli 1987). Disaster research comprises an array of academic disciplines, and as such, the scope of the term and the level of granularity considered informative is contingent upon the goals of the research (Perry 2018).

Early definitions of the term 'disaster', which arose within research following World War II, identified the disruption to normal patterns of behaviour as a fundamental element. Killian (1954) proposes that disasters disrupt social order and produce physical destruction and death, causing a departure 'from the pattern of norm expectations'. Wallace (1956) characterises disasters as 'extreme situations' that involve the threat of 'an interruption of normally effective procedures for reducing certain tensions, together with a dramatic increase in tensions'. H. E. Moore (1958) agrees that new patterns of behaviour are a defining feature of disasters, however the author believed that 'the loss of life is an essential element'. The definition in Charles Edward Fritz (1961) establishes the place of the social in disasters, emphasising that 'essential functions of the society [are] prevented'. These qualities were seen to limit disasters to rapid-onset events and were later expanded to include a broader set of events (Quarantelli 1984). More recently, disasters have come to be defined more firmly by their social elements rather than the physical agents (Barton 1989; Quarantelli 2000). Russell R Dynes (1998) defines disasters as situations in which norms fail, causing communities to engage in extraordinary efforts 'to protect and benefit some social resource'. In the field of humanitarian

[1] https://www.undrr.org/terminology/disaster (accessed 2022-09-07)

logistics, Van Wassenhove (2006) uses the definition of 'a disruption that physically affects a system as a whole and threatens its priorities and goals'.

The discussion surrounding the definition of the term 'disaster' is nuanced and ongoing (see Rodríguez et al. (2018)). As the goal of this research was not to strictly define disaster terminology, nor to add to the literature in this area, further exploration of this area of debate was not necessary. The consensus definition put forward by Quarantelli (2000) was broad enough to capture the essence of the term for the purposes of this work:

> 'Disasters are relatively sudden occasions when, because of perceived threats, the routines of collective social units are seriously disrupted and when unplanned courses of action have to be undertaken to cope with the crisis.'

The relevance of a given type of disaster event to this research was predicated primarily upon the remit of the disaster response organisations; only the subset of disasters to which these organisations were tasked with responding were valid candidates for examination. Additionally, events which were not considered disasters, but were relevant to response organisations, were included in the analysis (for example, organised protests). Therefore, the definitions discussed above were adhered to only loosely, and used primarily to develop an initial taxonomy of disaster types from which key differentiating features were explored.

2.1.2 Disaster Taxonomies

As shown in the preceding section, disasters comprise a range of events which each have unique characteristics. Many of these defining features have a measurable influence on the patterns of behaviour observed during the event. As identifying these patterns, within the context of social media communication, was a key outcome of this research, disaster taxonomies were examined to identify these features and evaluate their implications on method design. This subsection outlines existing taxonomical approaches and how they related to this work, and establishes a *working taxonomy* which was used in this research to inform the selection of subject events.

Van Wassenhove (2006) proposes a disaster classification matrix based on provenance (natural vs. human-induced) and speed of onset (see table 2.1). Olteanu, Vieweg, et al. (2015) draws from existing work in the sociology of disaster research (Perry and Quarantelli 2005) to define three dimensions with which to classify disaster events — hazard type (natural vs. human-induced, sub-categorised by phenomenon), temporal development (instantaneous vs. progressive), and geographic impact (focalised vs diffuse).

Taxonomies that target specific subdomains adopt more nuanced differentiations: Burnett (1998) examines the defining features of organisational crises, developing a matrix to rank their severity using four key 'inhibiting characteristics', including 'degree of control' and 'number of response options available' (table 2.2). This section explores the three dimensions most commonly used in the field — hazard type, temporality, and geographic spread — and discusses other dimensions which differentiated events in terms of their candidacy for use as a source of social media intelligence.

	Natural	**Anthropogenic**
Sudden-onset	Earthquake	Terrorist Attack
	Hurricane	Coup d'état
	Tornado	Chemical Leak
Slow-onset	Famine	Political Crisis
	Drought	Refugee Crisis
	Poverty	

Table 2.1: Categorising disasters (Van Wassenhove 2006)

Threat Level:	Time Pressure: / Degree of Control: / Response Options:	Intense		Minimal	
		Low	High	Low	High
Low	Many	Level 2	Level 1	Level 1	Level 0
	Few	Level 3	Level 2	Level 2	Level 1
High	Many	Level 3	Level 2	Level 2	Level 1
	Few	Level 4	Level 3	Level 3	Level 2

Table 2.2: Crisis classification matrix (Burnett 1998)

Hazard Type

The most intuitive dimension by which disaster events are classified is the underlying physical agent: the sociological characteristics of a hurricane making landfall are very different from a terror attack. In Guha-Sapir et al. (2015), The Centre for Research on the Epidemiology of Disasters (CRED) presents a taxonomy of six natural disaster sub-groups (figure 2.1), to which must be added anthropogenic events such as technological disasters (e.g. rail crashes) (Tatham et al. 2013) and human conflict (e.g. wars, protests and terror attacks) (Eshghi and Larson 2008).

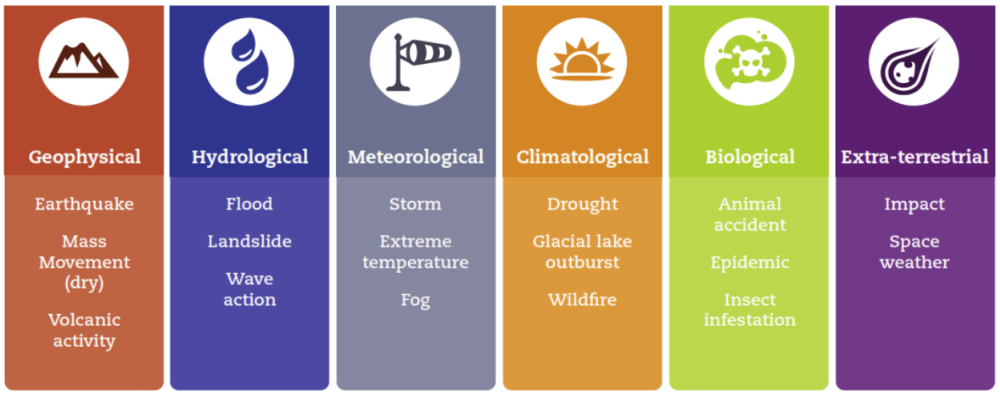


Figure 2.1: Natural disaster subgroup classification (Guha-Sapir et al. 2015)

Olteanu, Vieweg, et al. (2015) defines two primary categories which delineate these sub-types — natural and human-induced (table 2.3). Research indicates

that there exist fundamental differences between disasters arising from conflict and consensus situations (Peek and Sutton 2003; Quarantelli 1993; Quarantelli 2005; Waugh 2007). This binary distinction is common, though there is some nuance in the categorisation; Palen (2014) differentiates between events caused by *endogenous* and *exogenous agents*. Exogenous agents are outside of human control and cannot be apprehended; these events include natural hazards such as earthquakes or hurricanes. Endogenous agents are a product of human society; for example, criminal enterprises, oil spills, or pandemics. The distinction is significant in terms of the effect on discourse during and after the event, though is becoming less clear as the role of human activity as a cause of natural hazards becomes increasingly apparent.

Within online discourse concerning a disaster event, the community looks for the 'most salient problems to solve' (Palen 2014). In the case of endogenous agents, the social behaviour may be directed at identifying and apprehending individuals, assigning blame, and seeking justice. Exogenous agents present a more diffuse set of problems to solve: as they cannot be controlled or apprehended, discourse focuses on recovery and mitigation and therefore the social structures of the crowd are much different (Palen and A. Hughes 2018). During a disaster, online participation focuses on problems common to participants. When problems are less commonly shared, the discourse fractures and organises into smaller sub-groups. These sub-groups exhibit distinct behavioural patterns and therefore can degrade the fit of qualitative and statistical models when compared to more unified discourse. The endogeneity or exogeneity of the disaster event can predict this effect (Palen and Anderson 2016) and was therefore a significant consideration in event selection for this research.

Category	Sub-category	Examples
Natural	Meteorological	tornado, hurricane
	Hydrological	flood, landslide
	Geophysical	earthquake, volcano
	Climatological	wildfire, heat/cold wave
	Biological	epidemic, infestation
Human-Induced	Intentional	shooting, bombing
	Accidental	derailment, building collapse

Table 2.3: Hazard categories and sub-categories (Olteanu, Vieweg, et al. 2015)

Temporal Development

The duration of a disaster and the suddenness with which it appears are strong influencing factors in patterns both of information-seeking and sharing behaviour. A sudden, unexpected event such as an earthquake creates an immediate period of uncertainty and chaos, during which communication focuses on developing an understanding of the situation. In contrast, a slow-developing event, or one which has been forecast, instils a comparatively lower (though rarely trivial) degree of uncertainty. and therefore, public discourse contains more instances of coordination, recovery, and offers of aid.

The UN distinguishes between *sudden-onset* and *slow-onset* disasters. Sudden-onset disasters are typically discrete events that occur quickly with little to no warning. Slow-onset disasters evolve gradually from incremental changes occurring over many years or from an increased frequency or intensity of recurring events (UNFCCC 2012). Here, the breadth of the terminology is expansive and includes ocean acidification, sea level rise, and glacial retreat. These protracted phenomena present their own set of challenges which fall outside the scope of this research, and therefore a more fine-grained distinction was developed. Adams (1970) classifies events temporally as either *instantaneous* or *progressive*. Instantaneous events occur without warning and do not allow pre-disaster mobilisation of workers

whereas progressive events are preceded by a warning period. Parsons (1996) adds a third category to differentiate between emerging and sustained disasters:

1. Immediate: Disasters for which little or no warning is provided (e.g. earthquakes, terror attacks, volcanic eruptions)
2. Emerging: Disasters that are slower to develop, allowing for advance preparation or prevention (e.g. hurricanes, floods, protests)
3. Sustained: Disasters that may last for weeks, months or even years (e.g. bushfires, pandemics, civil unrest)

Within the context of characterising disasters, a single event may progress through each of these stages: an immediate or emerging crisis, such as an earthquake or hurricane respectively, may develop into a sustained event lasting weeks as the affected region attempts to recover or is subject to ancillary effects such as flooding. An ongoing bushfire season is characterised as a sustained event but may develop into an immediate or emerging crisis following a sudden change in conditions that cause the fires to threaten a populated area. Therefore, public behaviour during a disaster event should not be considered temporally invariant and may exhibit emerging or fading behavioural patterns as the event moves through these phases (Tapia and K. Moore 2014).

Geographic Spread

The size of the geographic area which experiences the impact of a disaster affects both the potential for harm and the response of the population. A widespread event that displaces a densely populated city presents a set of challenges that are distinct from an event of a smaller geographic scale. Adams (1970) discusses this concept from the perspective of Red Cross disaster relief, categorising events as *diffused* or *focalised*. Diffused events affect a larger area than focalised events, and often displace more people than can be housed by friends or relatives. The ongoing provision of mass shelter and the personnel to manage their operation is therefore an example of a problem unique to this class of event.

The character of online discourse is moulded by the set of challenges that it seeks to address, as well as the extent to which participants have been impacted by the event (Palen and Anderson 2016). As both of these factors are influenced by the geographic spread of the event, the distinction between diffused and focalised is important when modelling online behaviour. Olteanu, Vieweg, et al. (2015) observe that the proportion of Tweets they classified as containing 'caution and advice' ('information about warnings issued or lifted, guidance and tips') was higher in events that were geographically diffuse.

Culture and Technology

In addition to the key distinguishing features discussed above, a number of supplementary factors were considered for their potential to shape online discourse and therefore influence the viability of deriving meaningful intelligence from social media platforms. These characteristics included, for example, the cultural and demographical characteristics of the affected population (Xiao et al. 2015), and their access to communication technologies (compare, for example, Sakaki et al. (2010), A. Hughes and Palen (2010), Al-Saggaf and Simmons (2015), Murthy and Longwell (2013), and Stieglitz, Bunker, et al. (2018)).

An early dataset collected from Twitter during the 2017 mudslides in Sierra Leone (discussed in chapter 5) contained little informative data from the affected population, who had neither the telecommunication infrastructure nor the inclination to contribute to online discourse. In contrast, high levels of online public participation during other disaster events have provided sufficient levels of data from which event detection has been successfully conducted (Tanev et al. 2017).

The focus of this research was therefore placed upon events during which members of the affected population had access to, and the means to share, information useful to responders. Initial background research suggested that the following dimensions were significant in measuring the potential value of social media discourse as a source of disaster intelligence, and suggest further comparative research:

- Availability of social-media-enabled technology

- Demographic stratification of social media participants
- Cultural attitude towards social media
- Literacy of affected population
- Unity of language within a population
- Level of telecommunication infrastructure
- Resilience of telecommunication infrastructure to local disasters
- Presence of government censorship
- Obscurative presence of other dominant topics on social media

2.1.3 A Working Taxonomy of Event Viability

The dimensions drawn from the literature described above were synthesised into a *working taxonomy* (presented in table 2.4) designed to provide a preliminary framework with which disaster events could be evaluated with respect to their viability as subjects of the quantitative approaches proposed by this research. The *temporal development* category was split into the subcategories *duration* and *speed of onset*. *Geographic spread* was split into subcategories *affected population* and *magnitude*.

The extent to which each characteristic impacted the viability of an event as a source of social media intelligence was evaluated based on existing work and exploratory observations of online discourse conducted in early stages of the study. Further research into these effects is encouraged to provide important insights supporting the generalisation of the methods developed within this work to other classes of disaster event.

Dimension	Viability		Importance
	Low	High	
Hazard Type	Endogenous Agent	Exogenous Agents	Medium
Duration	Short-Term	Long-Term	Medium
Speed of Onset	Sudden-Onset	Slow-Onset	Low
Affected Population	Low-Population	High-Population	High
Magnitude	Focalised	Diffused	High
Culture and Technology	Low-Tech	High-Tech	High

Table 2.4: Working Taxonomy of disaster suitability for the purpose of quantitative intelligence classification methods

Table 2.4 distilled the key characteristics of disaster events identified in the literature which shape online behaviour into a working taxonomy that contextualised the evaluation of the events considered as subjects for this research. The event weighting provided by the taxonomy ensured that events were selected for which sufficient rates of informative eyewitness reports could be captured by the data collection processes conducted in chapter 5. These perspectives formed a conceptual framework with which empirical studies of social media use by disaster response organisations were examined and discussed in the following section.

2.2 Social Media Data in Disaster Response

Participation in online discourse on social media platforms continues to grow around the world as an outcome of the rise in availability of communication technology and infrastructure, and resulting growth in technological literacy and cultural acceptance. As more people turn to digital forms of communication, the data that are generated on social media platforms during disaster events increase in volume, diversity, and consequently, value as a source of intelligence to disaster response organisations (Zubiaga, Aker, et al. 2018). Real-time data generated on these platforms can contribute to the development of *situational awareness*

and lead to better outcomes in disaster response operations (Madey et al. 2007; Karami et al. 2020).

The phenomenon of *attention convergence* onto the physical sites of disasters has been well documented in sociological studies (Russell Rowe Dynes 1970; Charles E Fritz and Mathewson 1957; Kendra and Wachtendorf 2003) and is paralleled in online behaviour (A. Hughes, Palen, et al. 2008; Palen 2008; Palen, Vieweg, et al. 2007). The population affected by the event seeks to form ad-hoc information networks to share, seek, and broker information (Palen and Sophia B Liu 2007; Palen, Vieweg, et al. 2007; Qu et al. 2009) to mitigate the exposure of themselves and others to dangerous situations (Castillo 2016).

Studies of online discourse during disasters are diverse and have been conducted with respect to all stages of the disaster life cycle (Reuter and M.-A. A. Kaufhold 2018; Reuter, A. Hughes, et al. 2018). These include the detection of sudden-onset events (such as earthquakes) (Avvenuti et al. 2014; Sakaki et al. 2012; Earle et al. 2012; Fallis 2013; Robinson et al. 2013; Crooks et al. 2013; A. T. Chatfield et al. 2013; A. Chatfield and Brajawidagda 2012; Sakaki et al. 2010) or slow-onset events (such as viral outbreaks) (Alessa, Faezipour, et al. 2019; Wakamiya, Kawai, et al. 2018; Thapen et al. 2015; Chon et al. 2015; Santos and Matos 2014; Lamb et al. 2013; Aramaki et al. 2011), collecting situational intelligence to inform response operations (Dashti et al. 2014), tracking population movement (S. Y. Han et al. 2019), detecting disinformation (A. Gupta, Lamba, et al. 2013), and assessing damage for reconstruction (Cresci et al. 2015).

Analyses of public behaviour on social media platforms have become a significant component of disaster research and include studies of rumour generation (Oh et al. 2010), message classification and content analysis (Q. Huang and Xiao 2015b; Kongthon et al. 2014; Sreenivasan et al. 2011; Sinnappan et al. 2010), information dissemination patterns (Takahashi et al. 2015), geolocated message detection (de Albuquerque et al. 2015; Herfort et al. 2014), public behaviour and participant role emergence (Shaw et al. 2013; Miyabe et al. 2012; Vieweg, A. Hughes, et al. 2010; Palen, Starbird, et al. 2010), public sentiment over time (Neppalli et al.

2017; Doan et al. 2011), correlation between author location and URL sharing (Murthy and Longwell 2013), and communication from authoritative bodies (A. Hughes, St. Denis, et al. 2014; Procter, Vis, et al. 2013; Procter, Crump, et al. 2013).

2.2.1 Situational Awareness and Decision Support

A timely and effective response to a disaster event can play a significant role in reducing deaths, injuries, and secondary disasters, and minimise the period of social instability and economic loss (Shklovski et al. 2010). Disasters are inherently disruptive events, during which responding organisations must confront severe uncertainty in their decision-making processes. Decisions made under more informed conditions are naturally correlated with better outcomes and therefore a critical objective in response operations is the reduction of uncertainty and development of *situational awareness* (Q. Huang and Xiao 2015a; Vieweg 2012; Verma et al. 2011).

Situational awareness is defined in Sarter and Woods (1991) as 'all knowledge that is accessible and can be integrated into a coherent picture, when required, to assess and cope with a situation'. In the context of disaster response, the development of situational awareness is primarily concerned with obtaining and interpreting reliable, accurate, and timely information describing conditions of the affected environment (Karami et al. 2020). As situational awareness is increasingly developed through the accumulation of useful information, critical decision-making processes become better informed and therefore lead to more effective response operations.

The focus of early research in developing situational awareness has, however, been criticised as being unnecessarily constrictive. The aim of developing a general model of knowledge leads to large amounts of data which can overload responders seeking information relevant to the decisions they currently face (Zade et al. 2018; Ofli, Imran, et al. 2020). Furthermore, models informed by social media data that fall outside the scope of situational awareness, such as forecasting and behavioural modelling, can provide meaningful support to decision-making processes. Approaches based on *decision support* shift the focus of system design from the development of a general

model of knowledge to 'mission specific' tools which best serve the diverse needs of response organisations (Imran, Castillo, F. Diaz, et al. 2015).

Eyewitness Reports

Throughout this thesis, the class of data which contributes to the development of the situational awareness and decision support of disaster response organisations is referred to as 'intelligence', and constrained in later chapters to 'eyewitness reports' based on the qualitative study of response organisations (chapter 3). Eyewitness reports are widely acknowledged to be highly valuable to disaster response organisations (Zahra, Imran, and Frank O. Ostermann 2020) and the focus on this class of data is aligned with the findings of similar work (Zahra, Imran, and Frank O. Ostermann 2020; Zahra, Imran, Frank O Ostermann, et al. 2018; Tanev et al. 2017; Truelove et al. 2015).

2.2.2 Adoption Within Disaster Response Organisations

The extent to which social media data is implemented in disaster response processes varies significantly between organisations. A 2009 study investigated the use of social media by the Los Angeles Fire Department, which included both community engagement and intelligence monitoring (Latonero and Shklovski 2011). This early case was unusual for the time, and the authors suggest that the advanced adoption could be attributed to a social media *evangelist* within the department. More recent studies have documented broader recognition of the value of social media data (Zade et al. 2018; Power and Kibell 2017; Marbouti and Maurer 2016). Social media monitoring has alerted response organisations to events more quickly than traditional intelligence channels (Mason and Power 2015), supported the correction of rumours and misinformation (Palen 2014; A. Hughes and Palen 2012), coordinated volunteer efforts (Elbanna et al. 2019), and augmented information from traditional sources (Vieweg, Castillo, et al. 2014). The growing acceptance of the role of social media by authorities is demonstrated in the response to Hurricane Harvey in 2017, during which access to the emergency phone line (911) was disrupted leading to an

update in policy by the United States Coast Guard to allow the launch of rescue operations based on individual Twitter and Facebook posts (Stosz 2017).

A study of public information officers in Colarado, U.S.A. noted that despite the professed interest from the participants in using social media data, a lack of permission, institutional support, and training limited its use (A. Hughes and Palen 2012). More recent surveys conducted in the U.S. and Europe have found that only around 50% of participants report using social media in their work (Plotnick et al. 2015; Reuter, Ludwig, M.-A. Kaufhold, and Spielhofer 2016). Observational studies analysing the participation of response organisations with social media discourse during disaster events have found inadequate levels of community engagement (Roshan et al. 2016; A. Hughes, St. Denis, et al. 2014; Ehnis and Bunker 2012), demonstrating an unfamiliarity of many organisations with the medium.

The disparity between levels of social media engagement within the disaster response community reveals the infancy of the discipline. Several persistent challenges constrain the more widespread integration of these sources of data into formal response efforts. Most significantly, the insurmountable volume of data generated by social media platforms has been cited as a key barrier to integration (Palen and A. Hughes 2018; Stieglitz, Mirbabaie, et al. 2018; Marbouti and Maurer 2016; Palen and Anderson 2016; Castillo 2016; Vieweg, Castillo, et al. 2014; Reuter, Marx, et al. 2012), further exacerbated by limitations on personnel availability during periods of high intensity (Stieglitz, Mirbabaie, et al. 2018; Potter 2016; Tapia, Bajpai, et al. 2011).

Other challenges commonly reported include data veracity and issues of misinformation (Stieglitz, Mirbabaie, et al. 2018; Giasemidis et al. 2016; Nurse et al. 2015; Hiltz, Kushma, et al. 2014; Reuter, Ludwig, M.-A. Kaufhold, and Pipek 2015; A. Hughes and Palen 2012), lack of adequate contextual data (Martínez-Rojas, Pardo-Ferreira, and Rubio-Romero 2018; Laylavi et al. 2017; Martínez-Rojas, Pardo-Ferreira, López-Arquillos, et al. 2019; Spielhofer et al. 2016), the requirement of receiving reports in near real time (Vieweg, Castillo, et al. 2014; Mason and

Power 2015; Ofli, Meier, et al. 2016; Francalanci and Pernici 2018), and inadequate integration with existing information systems (Hiltz, A. Hughes, et al. 2020).

2.2.3 Existing Approaches and Limitations

Implementing eyewitness classification tools that conform to the temporal requirements of the response organisations is an effective approach to addressing many of the key challenges discussed in the previous section. Primarily, a classification approach curates the raw data derived from social media platforms subject to organisational constraints such that a richer and more informative stream of data is presented to the operator.

Crowdsourcing and Digital Volunteerism

Methods of data curation are a common focus of research on disaster response intelligence augmentation using social media data. Early approaches aimed to support the efforts of members of the public as digital volunteers (Starbird and Palen 2011). The *crowdsourcing* approach has been effective in mobilising a geographically distributed workforce and included tools for translating messages (Sophia B. Liu 2014), created crowdmaps (Meier and Brodock 2008; Morrow et al. 2011), and supporting digital volunteerism (Reuter, Ludwig, M.-A. Kaufhold, and Pipek 2015; Ludwig et al. 2015; Rogstadius et al. 2013), however, tools built for this purpose are limited by the willingness of the volunteers to perform the tasks therein and therefore do not satisfy the reliability constraints of response organisations. Furthermore, a liability concern is introduced where digital volunteers are not covered under Good Samaritan laws (in the U.S.A.) because the volunteers seek situations in which to assist (Palen and A. Hughes 2018).

Quantitative Approaches to Data Collection

Several tools have been developed by researchers to automatically collect and analyse social media data, implementing various filtering methods by which the large volumes of data are curated to be more useful to disaster response organisations. All datasets drawn from social media platforms are subject to an initial filtering

effect defined by the parameters with which the data are collected. In effect, this is a case of intentional *sampling bias*, intended to return a dataset relevant to the given application.

Data collection is most commonly conducted using keyword-based queries (Bevensee et al. 2020; Ashktorab et al. 2014; Bruns and Liang 2012; Abel et al. 2012; Yin et al. 2012; Marcus et al. 2011). This approach is intuitive and aligns with the concept of tagging (e.g. *hashtags* on Twitter), where selected terms are included by message authors to denote the topics of their message and support discoverability.

A number of tools that collect eyewitness reports from Twitter implement geographic filtering methods which return Tweets authored within an operator-defined region, where geospatial data are optionally attached to the message (Tsou et al. 2017; Burnap et al. 2015; Rogstadius et al. 2013; S. Kumar et al. 2011). Given that eyewitness reports must naturally be authored within the affected area, filtering messages by location eliminates material published by users around the world who may be discussing the event or whose messages were otherwise ensnared by the keyword parameters used for data capture.

Geospatially augmented messages are highly valuable to response organisations, as the exact locations from which messages are published provide useful context to the reports (Martínez-Rojas, Pardo-Ferreira, and Rubio-Romero 2018). However, only a small proportion (0-2%) of Twitter data include geographic coordinates (Leetaru 2019b) and therefore a reliance on coordinate based detection methods may fail to capture a significant proportion of eligible material.

Finally, identifying and monitoring key users, or 'persons of interest' (Mason and Power 2015), ensures that all messages published from their accounts are captured regardless of whether they include matching keywords or geographic data (Borra and Rieder 2014; Caragea et al. 2011). Naturally, this approach relies on the initial detection and classification of users as notable, for which various heuristics are used.

The volume of data collected during disaster events as a product of these methods exceeds the capacity of disaster response organisations to interpret manually and therefore motivates the development of data curation systems to identify

eyewitness reports. While existing tools provide automated filtering methods designed for analysing and exploring Twitter data, many fail to support the requirements of disaster response practitioners (A. L. Hughes and Shah 2016) and focus only on interpreting the textual content (Hiltz, A. Hughes, et al. 2020; Reuter, A. Hughes, et al. 2018).

Text-based classification models rely heavily on pre-existing coded datasets which are highly contextual to specific events and therefore do not generalise well to other events. Crowdsourced volunteering has been integrated to perform continuous labelling of live data which is used to retrain a classification model throughout an event (Ofli, Meier, et al. 2016; Imran, Castillo, Lucas, et al. 2014), though this requires ongoing motivated participation from volunteers and therefore does not present a reliable long-term solution. Furthermore, text-based models are vulnerable to changes in public behaviour, may perform poorly in areas where multiple languages are spoken, and do not generalise well to other regions where discourse behaviour deviates from the training data. For these reasons, approaches that examine supplementary features of social media data present attractive alternatives for the development of more robust classification methods.

Images and Video

The most common formats of social media messages typically include or combine text, image, and video components, the distribution of which varies by platform and user behaviour. Images from social media sources have been identified as highly useful by disaster response organisations (Hiltz, A. Hughes, et al. 2020; Power and Kibell 2017; Dashti et al. 2014) and are more resilient to the key challenges identified above of datum volume and veracity. Visual media require less time for a human operator to identify as relevant than text, allowing them to process a data stream more efficiently. Propagating misinformation through image and video material is more difficult due to the technical requirements, and easier for an operator to verify compared to text-based reports (Mason and Power 2015).

Images derived from social media sources have been integrated into existing disaster response studies to support situational awareness (Fernandez-Marquez et al. 2017; Rosser et al. 2017; Fohringer et al. 2015), and typically rely on geotagged media, though hybrid approaches have been explored (X. Huang et al. 2019). Text-based methods of classification are naturally ineffective on messages which contain photo or video material with little or no accompanying text. Given the demonstrated value of this class of data, approaches to classification that are robust to variations in message medium are useful.

Geoinference and Locality Classification

The textual or visual data which constitute the messages observed on social media platforms are augmented with additional informative *metadata*. The metadatum material attached to a message object varies by platform and may include both features defined by the author and those over which the author has no direct control. Where the challenges of text based classification are grounded in variant author behaviour, metadata provide alternative predictive features with which classification may be performed.

Geospatially augmented messages include geographic coordinates or other geospatial features describing the location from which they were published. The value of geospatial data in disaster response operations is well established (Hiltz, A. Hughes, et al. 2020; Laylavi et al. 2017; Martínez-Rojas, Pardo-Ferreira, López-Arquillos, et al. 2019; Spielhofer et al. 2016). Spatial representations of geographically augmented data are highly desired by disaster response practitioners (Francalanci and Pernici 2018), and geographic proximity to an event is a feature highly correlated with the informative value of a message (Palen and A. Hughes 2018), given the established goal of capturing eyewitness reports (section 2.2.1). Therefore, where available, the geographic location from which a message is published is a desirable trait by which to filter (Vieweg, Castillo, et al. 2014).

Posting messages publicly which include geographic data is highly revealing and exposes the author to significant privacy-related risks. For this reason, geographic

data are rarely included in online messages: fewer than 2% of Tweets contain coordinate metadata (Leetaru 2019b; Morstatter et al. 2013; Laylavi et al. 2016). *Geoinference* (or *geolocation*) is a process by which messages are spatially augmented through location inference informed by supplementary features. Content-based approaches extract geographic context from message text (Schempp et al. 2019; A. Kumar and J. P. Singh 2019; Li et al. 2019; Ren et al. 2012; Wakamiya, Lee, et al. 2011; Z. Cheng et al. 2010), though these approaches are vulnerable to the same risks as content-based filtering discussed in the previous section. Crowdsourced methods have been effective for content containing suitably informative data such as photos (Tong et al. 2020; H. Hu et al. 2016), however volunteer-based approaches are not suitable for sustained deployment.

Network homophily is a concept in the study of social networks[2] which states that nodes within a network demonstrate preferential attachment to other nodes with which they share similar characteristics (Newman 2018). Within the context of online social behaviour, homophily manifests as clustering within friend networks or interaction graphs. The preferential attachment behaviour is induced by latent traits, though the significance of geographic proximity is well established in the study of offline social networks (McPherson et al. 2001).

Existing research has leveraged the phenomenon of geoproximate homophily to conduct geoinference of social media users based on graphs representing reciprocal user 'mention' interactions (indicating a bidirectional conversation) (Rahimi et al. 2018; Rahimi et al. 2015; Do et al. 2017; Miura et al. 2017; Jurgens 2021; Compton et al. 2014) and follower/followee relationships (Rodrigues et al. 2016; Kong et al. 2014; Davis Jr. et al. 2011). User mention graphs have become the dominant focus of analyses due to the increasing difficulty with which follower/followee network data may be collected as a result of platform-instituted limitations (Jurgens 2021).

While these studies have achieved impressive results, the target outcome of user *location* represents a granularity in excess of what is required for disaster response classification. Where user accounts local to the affected area are sought, a

[2]Social network analysis is the process of investigating social structures through the use of networks and graph theory, and not to be confused with the study of social network web platforms.

binary *locality classification* is sufficient. That is, determining the granular location of *non-local* user accounts is an unnecessary level of detail that constrains the performance of classification methods. The binary classification approach therefore presents a novel avenue of research in the field of geoinference for the purposes of eyewitness identification. A secondary benefit to a binary approach is that lower amounts of sensitive user location data are inferred as they are not necessary for the application and introduce ethical concerns.

2.3 Discussion and Research Gaps

The studies discussed in this chapter demonstrate the enormous potential of social media data to provide useful insights to disaster response practitioners which better inform their decision-making processes during disaster events. An effective implementation of social-media-informed intelligence systems that supports the needs of decision-makers can improve response outcomes and reduce the loss of life, economic damage, and period of social disruption.

Subject Event Selection

The types of event for which intelligence derived from social media data is informative are varied, however, there are a number of key features which define the classes of disaster most suitable for the approaches explored in this research. Fundamentally, a data-driven process based on discourse within social media platforms relies upon high levels of public engagement, through which the public shares perspectives not otherwise available to disaster response organisations.

Disasters that cause prolonged periods of disruption over large geographic regions create conditions in which it is difficult or impossible for response organisations to develop a comprehensive model of the conditions on the ground. A population of distributed and connected users of social media within the affected area publish valuable and timely perspectives which can supplement traditional sources of intelligence. By contrast, during comparatively brief and geographically focused

events (such as some terror attacks), the capacity for the public to share eyewitness perspectives considered informative to response operators is limited.

These considerations suggested that the focus of this research be placed on *long-term*, *geographically diffuse* disaster events in regions with *high populations*. Familiar events which satisfy these criteria include earthquakes, hurricanes, and bushfires. Secondary considerations were based on traits that influenced behavioural patterns within online discourse such as the speed of disaster onset, technological attitudes of the affected population, and whether the event was anthropogenic or caused by external forces. By selecting events that closely aligned with these characteristics, the potential for public discourse to provide valuable intelligence to disaster response organisations was best realised.

Data Bias

There remains an ethical concern with respect to the participation bias on social media platforms. While data collected from these sources may be seen as representative of the voice of affected people, there are significant critiques of the ethical limitations and potential harms of its use during disaster response (Madianou 2019). Where these data are used to inform and activate response operations, there is a risk that vulnerable demographics of people not equably represented are provided a lower quality of care. Causes of unequal representation may include lower levels of participation in online discourse, for example, due to lower technological literacy or cultural adoption of social media platforms (Martí et al. 2019; Xiao et al. 2015), or a sampling bias in the collection software (e.g. where other languages are used) (Rains and Brunner 2015).

Furthermore, the interface constraints of each platform and the algorithms which define how messages are spread introduce epistemological limitations which determine what kinds of information are published online (Crawford and Finn 2015), and data collection methods which are based on matching a set of defined keywords naturally introduce data bias (Murzintcev and C. Cheng 2017). These

issues must be considered when adapting data as disaster intelligence such that vulnerable groups are not disadvantaged.

Understanding Organisational Requirements

The range of events for which disaster response organisations are responsible is highly variant in nature and introduces considerable challenges to the design of decision support systems for disaster response. Furthermore, the challenges faced by decision-makers evolve as disaster events progress through the phases of the disaster life cycle, requiring adaptive intelligence processes. As the scope of research in disaster response intelligence broadens to examine the role of social media data in decision support systems, it is imperative that perspectives from response organisations are embedded in design processes to ensure that system features are aligned with operator needs.

As the landscape of social media discourse evolves due to changes in technology, platform availability, and public behaviour, so too does its role in augmenting decision support systems. While a number of studies exist which have documented system requirements for a selection of response organisations, there exists a continual motivation to expand upon these findings and incorporate more recent technological developments. Therefore, the first research question was formulated as:

> RQ_1 —What opportunities and challenges are presented to the intelligence processes of disaster response organisations by social media data?

Online Discourse and Data Access

Understanding how social media data may augment existing disaster response intelligence processes required an examination of patterns of online behaviour during disaster events to determine to extent to which valuable information was generated by online users and the features by which valuable information could be characterised for quantitative detection processes. Additionally, the availability of social media data, as provided by the source platforms, governs the roles that they can fulfil and must therefore be considered in conjunction with the information within the data.

The responsibilities of response organisations encapsulate a broad range of events, the unique characteristics of which bear significant implications to the role of public discourse (see table 2.4). Social media messages have been characterised in numerous studies adopting various perspectives (Rudra et al. 2018; Finch et al. 2016; Vieweg, Castillo, et al. 2014; Lachlan et al. 2014), however there remains considerable impetus to conduct further research in this area. Patterns of online public behaviour undergo continual evolution due to changes in social norms, updates made to the underlying infrastructure, and migration to alternative platforms. For example, in 2017 Twitter expanded its iconic character limit from 140 to 280, and is currently trialling an edit feature.[3] Changes to the environment within which discourse takes place influence the patterns of discourse and therefore ongoing contributions to the analyses in this field of research is important.

Furthermore, during the ethnographic study conducted to address RQ_1, a number of unique requirements were derived which shaped the analysis of social media communication with respect to the needs of the organisations examined, formalised as:

> RQ_2 —How can publicly available social media data provide meaningful intelligence to disaster response organisations during disaster events?

User-Based Geoinference Using Network Data

Eyewitness reports (whether text, image, or video) have been identified in a number of studies as one of the most valuable classes of message published on social media platforms(Imran, Ofli, et al. 2020) and were the primary focus of the methods developed in this research. As social media platforms increasingly obfuscate identifying geographic information from publicly available data, developing methods of inferring the location from which a message was published grows as an active area of research.

A precise message-centric approach to geoinference was considered unnecessarily complex for the identification of eyewitness reports: once an author was classified as non-local, they could be discarded from the dataset and therefore a more granular

[3] https://twitter.com/TwitterComms/status/1511456430024364037 (accessed 2022-05-22)

determination was not required. The focus on a binary local/non-local classification was motivated by requirements derived from response organisations and introduced the opportunity to pursue novel methods of analysis.

The concept of network homophily states, in terms of social media data, that users are more likely to exhibit relationships with other users with which they share common attributes, such as geographic proximity. Strong homophilic behaviour leads to node clustering within a graph and can be measured to inform classification methods. Analysing the extent to which homophilic behaviour caused by geographic proximity existed within data derived from social media disaster discourse was selected as a focus of this work. Network data is not typically examined by social media classification methods, due primarily to the difficulty with which it is gathered. The motivation for this approach was therefore based on assessing the value of the network approach to user classification, formulated as the research question:

> RQ_3 —To what extent can graph-structured relationship data inform eyewitness classification for social media data?

Eyewitness Classification for Data Curation

Finally, an evaluation of the degree to which the eyewitness classification system addressed the requirements informed by domain perspectives was conducted through an *in situ* deployment of prototype software. Disaster response practitioners were observed interacting with the system and provided feedback which supplemented the findings from RQ_1 and validated the outcomes of RQ_3, thus ensuring that the methods developed in this research aligned with the needs of the organisations from which the requirements were drawn.

> RQ_4 —How well can a network-based user-centric eyewitness classification approach curate social media data and address the volume constraints of disaster response organisations?

2.4 Summary

This chapter has discussed the scope of the term ‘disaster’ as it is defined in the literature. The set of events encapsulated by the term was then constrained

to represent only those that were relevant to disaster response organisations. A taxonomy of disaster events was developed to distinguish candidate events across selected dimensions and evaluate their suitability as subjects of this research.

A review of the literature on social media data in disaster response organisations identified the key areas in which online discourse augments existing intelligence processes and the primary challenges limiting its more widespread implementation. Messages containing eyewitness reports from users in affected areas were considered highly useful by responders, however, the large volume of data generated on social media platforms, and therefore the effort required to detect this class of message, was an obstructive limitation.

Existing approaches to data classification and curation were examined and found to be heavily reliant on text content and therefore vulnerable to changes in patterns of behaviour, and unable to adequately detect useful photo and video material. The limitations of these methods exemplified the importance of including domain perspectives in system design such that organisational needs are adequately addressed. Geoinference approaches to message classification were shown to be effective at eyewitness detection, though this remains a relatively young and emergent field of research. Significantly, the requirement of location classification for the purpose of eyewitness detection was reformulated as a binary *local* and *non-local* classification problem, presenting an opportunity for novel research.

The research questions developed in this chapter are repeated in table 2.5 with the chapters in which they are explored.

	Research Question	Chapter
RQ_1	What opportunities and challenges are presented to the intelligence processes of disaster response organisations by social media data?	3
RQ_2	How can publicly available social media data provide meaningful intelligence to disaster response organisations during disaster events?	4, 5, 6
RQ_3	To what extent can graph-structured relationship data inform eyewitness classification for social media data?	7
RQ_4	How well can a network-based user-centric eyewitness classification approach curate social media data and address the volume constraints of disaster response organisations?	8

Table 2.5: Research questions

3

Qualitative Study of Disaster Response Organisations

Contents

3.1 Research Design 47
3.1.1 Motivation 47
3.1.2 Methodological Challenges for Disaster Research 49
3.1.3 Qualitative Research Method 51
3.1.4 Participant Selection 56
3.1.5 Ethical Considerations 58
3.1.6 Timeline 58
3.2 Qualitative Analysis 60
3.2.1 Data Summary 60
3.2.2 Thematic Coding 64
3.3 Findings 66
3.3.1 Social Media Intelligence in Disaster Response 69
3.3.2 Challenges to Social Media Intelligence in Disaster Response 78
3.3.3 Supplementary Observations 83
3.4 Discussion 86
3.5 Summary 89

This chapter describes the design, conduct, and results of a qualitative study documenting disaster response organisations. The study develops an understanding of how social media data are currently used as intelligence by participant organisations and illustrates the research challenges and requirements. A conceptual

framework of the role of social media data in disaster response intelligence is developed and shapes the direction of later work in this thesis (contribution C-3). A qualitative research method was chosen to elicit perspectives and requirements from disaster response organisations in terms of the role that social media data represent as a source of intelligence.

The system development conducted in chapter 5 draws on a set of software requirements informed by the framework developed in this chapter to align software features with domain perspectives. The contributions of this chapter are therefore sociotechnical in nature, furthering the field of disaster information system research by presenting findings grounded in the perspectives of social media intelligence operators. The research question addressed by the study presented in this chapter was formulated as:

> RQ_1 —What opportunities and challenges are presented to the intelligence processes of disaster response organisations by social media data?

Specifically, the purpose of the study was threefold:

1. To observe existing implementations of social media data as intelligence.
2. To identify areas in which social media intelligence may enhance the performance of response organisations during disaster events.
3. To document constraints limiting the use of social media intelligence in disaster response.

This ethnographic study comprised interviews with disaster response intelligence practitioners from a range of domains and a substantive observational study in an emergency control centre. The substantive outcome of this study, which shaped the research conducted in later chapters, was the formulation of a set of five key values in which social media data were seen to provide informative potential to disaster response practitioners, and four challenge areas.

3.1 Research Design

This study implemented qualitative research methods to observe and document the operations of disaster response organisations during a disaster event with respect to the use of incoming intelligence data. The research was interested primarily in the organisations' use of social media data during a disaster, or lack thereof. Of significant importance were the perceived advantages to the organisations which drive the use of these data, and barriers which prevent its greater adoption. The research was also interested in the wider role of information collected during an event in order to model the process by which incoming data are received, verified, and acted upon within an organisation. Understanding this process using inductive techniques allowed for the development of tools that more accurately addressed the needs of disaster response organisations based on these findings. Furthermore, by establishing a conceptual framework from which system requirements could be defined using domain perspectives, tools developed through this process may be validated against real-world measures rather than metrics selected by the researcher. The findings of this chapter informed the outcomes of chapter 7, which developed a focused quantitative approach for data curation grounded in the needs of participant organisations.

3.1.1 Motivation

To improve the utility of social media data to disaster response organisations, it is important to understand the challenges faced by these teams as they respond to emerging crises. Social media data provide a rapid, cheap, unverified and diluted source of information that may supplement existing sources once processed appropriately. There is much discussion within the information systems literature as to how these data can provide meaningful information to readers during disaster events (A. Hughes, Palen, et al. 2008; Lotan et al. 2011; Olteanu, Vieweg, et al. 2015; Vieweg, A. Hughes, et al. 2010), however the ways in which the data may best be processed to meet the needs of response organisations are less commonly examined.

While it is apparent that identification, sorting, and verification are key issues in the use of these data, the extent to which these matter, and therefore the relative impact of addressing these issues, is unclear. The needs of disaster response organisations during disaster events may be influenced by a number of factors beyond the scope of deductive reasoning, and therefore difficult for an outsider to understand. For example, to an outside observer, it is unknown whether these organisations lack the knowledge to use existing social media tools, are unable to allocate the resources to monitor the traffic, or do not consider unverified sources valuable. Designing systems that neglect to understand or consider these factors may result in a failure to provide value and therefore tools that have limited use in a disaster event.

Much of the quantitative literature written on social media data in disasters fails to consider the perspectives of response organisations. These projects implement and evaluate designs to capture, filter and verify the data based on self-determined metrics (see, for example, A. Gupta, Kumaraguru, et al. (2014)). Therefore, the true utility of the data may not be adequately realised by disaster response operators in practice. While many of these methods return promising results, without validation methods that include the users for which the systems are intended, it is difficult to realistically measure the practical relevance of these approaches to disaster response practitioners. The qualitative approach of the research presented in this thesis addressed this limitation and developed an understanding of the practices and perspectives of disaster response organisations, providing context to the development of new social media systems suited to the needs of these organisations.

This study implemented an adapted requirements engineering approach to document user needs and identify obstacles with respect to the use of social media data as a source of intelligence. The findings of the research presented in this chapter informed the quantitative approaches used in the subsequent chapters of this thesis. By integrating a qualitative approach to requirements gathering with the design of quantitative analysis tools, the classification of useful data in social networks was grounded in findings derived from domain-validated data. This mixed-method approach more effectively captured the practices and perspectives of

disaster response organisations and therefore more accurately addressed the needs of information officers through the informed design of novel methods and systems.

3.1.2 Methodological Challenges for Disaster Research

From a social science perspective, disaster research presents a unique set of conditions that set it apart from other fields. Stallings (2007) identifies the key elements that make disaster research unique:

- Timing: Disaster events typically occur with little warning and evolve rapidly. Researchers therefore have limited time to prepare for fieldwork. The rapid evolution of disaster events increases the *perishability* of data: the quality of the collected data is degraded based on the delay between the initial event occurrence and moment of capture.
- Access: Researchers may have limited access to individuals, organisations, or areas affected by a disaster. Within the context of this work, response organisations are under considerable workloads during disasters and therefore not likely to prioritise responding to requests from an external researcher. In some cases, the simple act of being observed or the presence of an additional person in a control room may affect the ability of the unit to function during a life-threatening situation and therefore the tolerance towards external researchers may be limited.
- Generalisability: The unique factors defining disaster events limit the degree to which findings can be abstracted to other scenarios. Each type of disaster affects a population in different ways, and disasters of the same type are rarely alike in other respects.
- Interdisciplinarity: Donner and W. Diaz (2018) adds a fourth element, noting that the link between the environmental origins of a disaster and its effects on society necessitates collaboration between researchers. Considered from a requirements engineering perspective, this quality also describes the interoperability that exists within a disaster response team, where multiple organisations collaborate to coordinate a response (Hiltz, P. Diaz, et al. 2011).

To address the first two of these elements, researchers either adapt empirical methods to capture perishable data before the social behaviour unique to the event returns to normal (Palen, Vieweg, et al. 2007; Quarantelli 1997; Palinkas, Downs, et al. 1993; Vieweg, Palen, et al. 2008) or evoke the conditions of past experiences through, for example, simulations and scenarios (McLennan et al. 2006; Helsloot 2005; Wood and Büscher 2012; Bharosa et al. 2010; Yang et al. 2015; Militello et al. 2007).

Quick Response Research

The *quick response research* (QRR) methodology (also referred to as *rapid assessment procedures*) attempts to capture transient data surrounding disaster events. It adapts existing disaster study frameworks to rapidly deploy researchers to disaster sites and collect perishable data as close to an event occurrence as possible (Stallings 2003; Quarantelli 1997; Palinkas, Downs, et al. 1993; Palen and Sophia B Liu 2007). The goal of this approach is to capture first-hand observations and timely interview data to provide depth to the understanding of an event and the community context, often to precede a more formal research approach (Palinkas, Prussing, et al. 2004; Scrimshaw and Hurtado 1987).

Simulations and Scenarios

Scenarios and simulations offer an alternative to QRR by recreating the features of a disaster event within a controlled environment and observing the subject behaviour as they interact with the environment. Immersive live simulations are often used by disaster response organisations for training purposes, providing researchers with planned and accessible events from which to record data. More abstracted simulations can target specific teams or aspects of behaviour and are generally less resource-intensive than live simulations. These can range from virtual-reality environments to board games or simple textual representations of an event chain (Bailly and Adam 2017; van Ruijven 2011; Ley et al. 2012; Dugdale et al. 2010; Yang et al. 2015; McLennan et al. 2006).

Proxy Events

The study of proxy events represents a compromise between the authenticity of QRR and the convenience of planned simulations. By observing events that share similar characteristics to those of interest, the researcher may reduce the severity of the challenges described above (Tapia, Giacobe, et al. 2015; Boersma et al. 2009; The Institute of Medicine 2004). For example, large congregations of people at sporting events, concerts or protests may lead to tension and instances of emergency which, while not generally of the severity of a typical disaster, are often monitored by the same organisations and therefore may serve as appropriate proxies for disaster communication behaviour.

Adapting Approaches to Disaster Ethnography

These research approaches address key challenges of disaster response ethnography related primarily to the difficulty in performing direct observation of behaviour during live disaster events. In terms of this work, quick response research was dismissed due to the geographically distributed nature of research participants such that the required travel was not possible. Simulations and scenarios were used in the validation study in chapter 8 to simulate a disaster response activity within the software environment. For the study in this chapter, observation work was conducted during periods of high alert (though not active disaster response). Activity during these periods modelled behaviour of a full response activation and minor events effectively acted as proxies from which insights were derived.

3.1.3 Qualitative Research Method

This stage of the research examined the behaviour and motivations of disaster response personnel within the constraints defined by the research question. The explorative analysis took an inductive approach that allowed for the investigation and discovery of new analytic ideas which may not otherwise be noticed in a deductive approach. The research design drew from qualitative sociological methods which provided a framework from which a robust approach was developed.

This research adopted an ethnographic approach to data collection in which immersion and engagement facilitated the generation of *descriptive representations* of participating organisations (Moran 2002; Sokolowski 2000; Cooney 2010). Methods were chosen to best suit the inquiry posed by the research question (Holloway and Todres 2003) and were comprised primarily of semi-structured interviews and observations with the aim of developing a conceptual framework of the relationships between disaster response organisations and social media data. Analyses were conducted using the *grounded theory methodology*, which has become the gold standard approach to inductive theory-building, by which a theory is derived from data that is then illustrated by characteristic examples of data (Glaser and A. L. Strauss 1967; A. L. Strauss 1987; Charmaz 2014). Grounded theory helps the researcher move from a description of events to an understanding of why the events are happening (J. Corbin and A. L. Strauss 2014), thus better meeting the aim of the study. While initially proposed by Glaser and A. L. Strauss (1967), the two authors have since developed competing schools of thought most clearly delineated by their approach to analysis (Glaser 1992; J. Corbin and A. L. Strauss 2014). The more open Glaserian analytic framework was favoured over the formulaic Straussian method to better accommodate an informal approach to research structure.

This study sought to document domain perspectives on social media intelligence systems and identify viable opportunities for intervention. Therefore an inductive, or 'bottom-up' approach was taken in which the researcher does not test a pre-established hypothesis, but rather develops 'data-led' hypotheses throughout the course of the study as a product of the study itself. Glaser and A. L. Strauss (1967) describes the process of *constant comparison*, a method that is concerned with generating and suggesting (rather than testing) hypotheses and characteristics related to the subject, and is compatible with incomplete (that is, growing) datasets. This method integrated well with the *theoretical sampling* approach (discussed in section 3.1.4); questions that emerged from ongoing analyses guided the collection of further data. Where access to participants was limited, the adaptability of the

constant comparison approach ensured that each new iteration of data remained relevant to the evolving research question.

The research framework was based upon the concept of *goal-directed information analysis* (GDIA) proposed in Yang et al. (2015) which in turn adapts *goal-directed task analysis* (GDTA). Goal-oriented analysis has been proposed in requirements engineering literature as a method by which to capture domain properties and expectations about the environment in addition to the software specifications (Lapouchnian 2005). It is therefore a useful approach for cases where many domain-specific characteristics exist, such as disaster management. Goal-based approaches offer an intuitive way to elicit requirements from participants who do not typically think of their role within the context of software processes (Ali et al. 2010). Goals are defined as something which the stakeholders hope to achieve in the future (Rolland et al. 1998). In the context of this work, the goal is explicitly predefined by the research question: to *extract useful intelligence from social media data*. The research method therefore combines the task analysis of GDTA with the scenario building of GDIA and is presented as an adaptation of the GDIA process in table 3.1. Figure 3.1 shows the sequence of steps and the validation loop between steps 4 and 5. The path between steps 4 and 2 represents the iterative process occurring between data collection events as described above.

Step	Title	Phase
1	Context discovery	Study preparation
2	Establish scenarios and select participants	
3	Identify physical tasks	Data collection
4	Identify subgoals	
5	Identify data value and impediments — online	
6	Identify data value and impediments — offline	Validation

Table 3.1: Adapted GDIA application steps

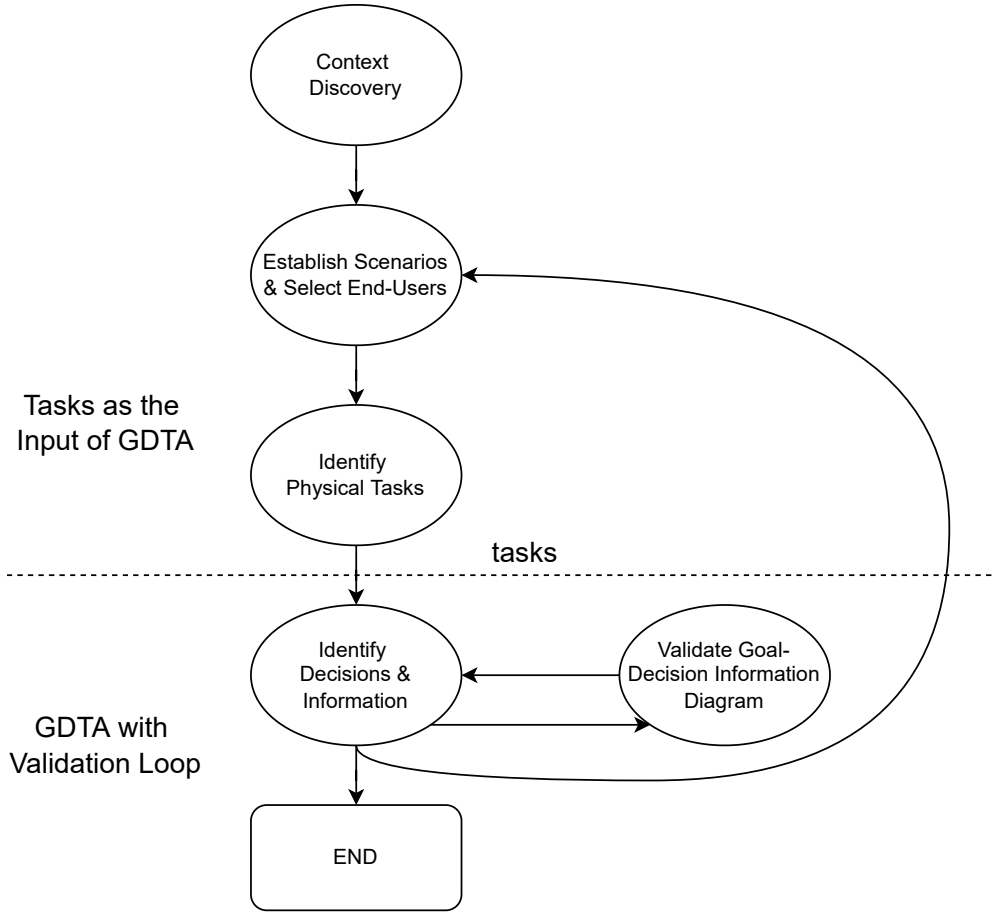


Figure 3.1: Adapted GDIA process

Step 1 — Context Discovery

Context discovery is the process by which the researcher develops an initial understanding of the domain under investigation. The discovery is facilitated through early interviews and discussions with practitioners, leading to the development of an interview strategy and informing participant selection decisions.

Step 2 — Establish Scenarios and Select Participants

Scenarios address the access challenges facing ethnographic studies of disaster response organisations (see section 3.1.2). A scenario is a sequence of events and conditions designed to simulate a realistic situation within which the participant is asked to operate. Pre-established scenarios provide structure to the interview process, allowing the researcher to focus the discussion within a framework comfortable to

the participant. Scenarios can challenge participants' assumptions and draw out *tacit knowledge* (Polanyi 1958), thus providing an informative representation of a participant's decision-making process to the researcher. For this work, the set of scenarios predefined by the researcher was regularly enriched with more specific scenarios formulated or recalled by participants during their interview, ensuring scenarios remained relevant to the participant's field of expertise.

Participant selection was ongoing throughout the process and followed a *snowball sampling* approach described in section 3.1.4.

Step 3 — Identify Physical Tasks

This step represents the first line of questioning directed towards the interview participants. The scenario framework is presented to the participant and they are asked to describe the tasks they would seek to accomplish at each phase of the event. Tasks are represented by clearly-defined actions a participant may take to achieve a broader goal.

Scenarios are often used to discover and define goals in requirements engineering (Rolland et al. 1998). The goal-oriented approach contrasts with the qualitative work of this research, which investigates the processes (or tasks) directed towards achieving a goal predefined by the research question. A task-oriented approach captures links between the decisions a participant is faced with making and the overall goals of the organisation. Furthermore, the tasks identified by each participant may vary within an organisation based on individual factors such as seniority or experience, whereas the goals are typically shared by the team. Task identification is therefore able to capture the individual perspectives of participants with respect to a broader goal and is better-suited to disaster response research, where practitioners operate within a highly structured environment and therefore may not have a goal-level perspective of their actions (Yang et al. 2015).

Step 4 — Identify Subgoals

In the context of this work, the distinction between a goal and a task may be broadly thought of as the difference between answers to the questions: 'Why do you collect intelligence data' (goal) and 'How do you collect intelligence data' (task). The tasks identified in step three inform the identification of the subgoals that exist under the predefined goal. During a semi-structured interview, the interviewer may probe the participant to explore why certain tasks are performed to encourage introspection and subgoal discovery. Thus, identifying tasks (step 3) and their associated subgoals (step 4) is an explorative and iterative process.

Step 5 — Identify Data Value and Impediments — Online Analysis

The objective of this study was to identify key areas in which social media data could be used as intelligence, and the primary factors preventing or limiting such use. Analysis of the collected data was performed in two stages. First during an interview, the researcher identified these features and presented them to the participant for discussion and validation. This choice of *online analysis* was motivated by the limitations of participant access: as many of the participants were too busy to offer follow-up interviews, an online validation process was required.

Step 6 — Identify Data Value and Impediments — Offline Analysis

The second stage of analysis was conducted offline — that is, once each round of data collection had ended. As recommended by the constant comparison method, this analysis informed the selection of participants and the structure of subsequent interviews. A more formal analysis was conducted once the entire data collection period was completed and is discussed in section 3.3.

3.1.4 Participant Selection

A *theoretical sampling* approach governed the process of participant selection. Theoretical sampling was guided by the emerging theory; new samples were selected to make comparisons with existing data. The method was directed towards

generating a conceptual theory of social media intelligence in disaster response organisations rather than providing a descriptive account (Breckenridge and Jones 2009). This flexible approach encouraged the pursuit of leads within the data as they arose and progressively iterated data collection methods to best integrate the emerging theory (Glaser and A. L. Strauss 1967). Furthermore, the adaptive theoretical approach mitigated the challenge of inconsistent access to participants caused by the unpredictable nature of their workloads.

Disaster response organisations are typically bureaucratic and under-resourced and therefore it was challenging to elicit the commitment required from disaster response practitioners to participate in an external study. A pragmatic approach to sampling examined a number of participant organisations spanning multiple sectors. These included non-profit aid organisations, state-run disaster management departments, and police forces. Later rounds of research were directed by the emerging theory and sought to clarify or investigate key themes, as encouraged by Glaser and A. L. Strauss (1967).

The initial round of selection was enabled through introductions facilitated by members of the University community and the personal network of the researcher. Making direct contact with relevant personnel in this way proved more productive than later rounds of solicitation made through public channels. After each interview, participants were asked whether they could introduce to the researcher colleagues from other organisations for recruitment into the study. The selection method thus resembled a *snowball sampling* approach (Goodman 1961).

The risk of sampling bias introduced by the snowball approach was mitigated by the inclusion of constraints in the selection process such that participant organisations spanned a range of domains. A continuous process of lead generation through alternative means was conducted in parallel to further diversify the breadth of participant backgrounds.

While participating in an interview was a relatively minor commitment for an organisation, requiring the time of only a small number of personnel, the prospect of observational work was more disruptive and demanded more resources of the

participant organisations, therefore eliciting participation for observational studies was more difficult. During the interview process, the possibility of extending the study to an on-site observation (where possible) was discussed and a willing organisation was selected as the subject for this stage of the study.

The chosen organisation operated as a state-level emergency control centre and integrated eight response agencies including search and rescue, police, and fire services. Their intelligence division was therefore responsible for a range of hazard categories and benefitted from the combined resources of its constituents. A number of social media intelligence protocols were already in place and provided a useful standard upon which novel approaches were assessed.

3.1.5 Ethical Considerations

Due to the sensitive nature of the personal data collected during this study in the form of participant responses, care was taken to ensure compliance with ethical guidelines. Participants were made aware of the ways in which their data would be used and given the right to withdraw consent at any time. Participants and their organisations were anonymised and represented only in broad illustrative terms.

The study met the appropriate ethical standards of the Central University Research Ethics Committee (CUREC) and was granted approval under application numbers `R49170/RE001` and `R49170/RE002`. Research data were handled according to the University's data protection policy[1] and stored in a manner that was compliant with legal obligations.

The participant consent form which was presented to each interview participant is provided in appendix A.2.

3.1.6 Timeline

The interviews for this study were conducted over a period of 24 months. First, an initial round of interviews provided context for the development of further studies. Later interviews with different organisations offered alternative perspectives and

[1] https://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-data-supporting-research-outputs/ (accessed 2022-10-06)

were used to pursue questions that had emerged during the study. This unstructured approach was taken for two reasons: First, pursuing leads to reach the relevant candidate within an organisation could take time in the order of weeks or months; responding to an external study recruitment email was understandably considered of low priority to organisations managing time-critical operations.

Second, including interviews throughout the course of the research provided an opportunity to explore emergent themes with domain experts during later phases of analysis (see the discussion of the constant comparison approach above). For this reason, the interview schedules were designed in a responsive manner and were largely independent of one another. The study method was therefore designed to progress based on a small initial interview dataset which was supplemented as more participants were recruited concurrently into later studies, thus minimising the effect of delays to the research without excluding useful data points.

The observational work with the control centre closely followed an initial interview with a participant working in the centre. The observation periods included a number of on-site interviews which were conducted in the same manner as those done remotely.

Follow-up interviews were conducted with participants where possible. The scheduling of these was similarly constrained by the demands of the domain and therefore also adopted the responsive approach described above. Figure 3.2 exemplifies the opportunistic approach taken to the qualitative work of this study.

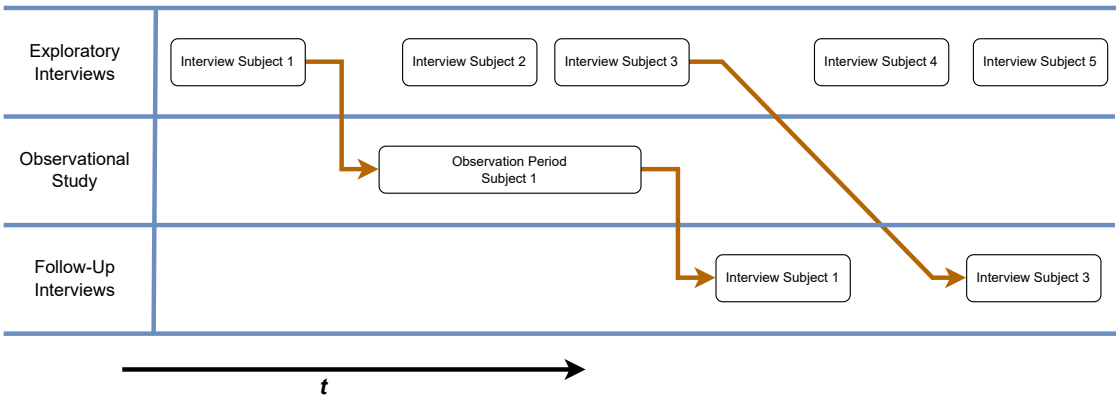


Figure 3.2: Example of the adaptive timeline format

Interviews constituted the majority of the data collection process and were conducted primarily as single points of contact with each participant. In certain cases, follow-up interviews were conducted with participants to further explore emerging themes, thereby exploiting an advantage of the constant comparison approach. The structure of each interview was predicated upon the outcomes of those that preceded it. Formally, the interviews adopted a semi-structured approach which provided the interviewer the flexibility to more deeply explore the answers given by each participant, and shape the direction of the interview based on the participant's relationship with their organisation's intelligence procedures. The initial structure of the interview is provided in appendix A.1.1.

The observational work took a similar approach, adapting to the limitations of participant availability which would change by the hour based on ongoing events. Observations provided organisational context to the findings that emerged from the interview data and were conducted using an inductive approach, in which the researcher operates from outside of the observed environment. Document analysis supplemented observations in defining key processes and organisational structure.

3.2 Qualitative Analysis

This section presents the results of the data collection process and describes how iterative data analysis was conducted to derive a domain-informed understanding of disaster response organisations and their use of social media as intelligence. The analysis addresses RQ_1 and contributes to a conceptual framework from which a set of requirements were defined that inform the software development presented in chapter 8.

3.2.1 Data Summary

In total, sixteen participants from eight organisations were interviewed throughout of study. Each interview explored themes based upon the findings of ongoing comparative analysis as described in the preceding section. The role of each participant (translated into British terminology) and their associated domain is presented as

a summary in table 3.2. Participants interviewed during the observational work are also included here. Names have been redacted for publication as discussed in section 3.1.5 and domains have been consolidated into illustrative categories.

In most cases, interviews were conducted remotely due to the diverse locations of the participant organisations. Where possible, a small number of interviews were held in person. In both instances, interviews would run for approximately one hour. Follow-up interviews were conducted with eight participants at various points in the collection period.

The observational work comprised three site visits over the course of two weeks during a period of high alert. Each visit lasted for a day. While the control centre operates on a twenty-four-hour basis, much of the staff, including the intelligence officers, maintain 'regular' hours unless a state of emergency is declared. The study was therefore constrained to these hours.

Domain	Participant	Position
Police	P1	Chief Superintendent
	P2	Superintendent
	P3	Superintendent
Fire Department	P4	Chief
	P5	Team Manager
	P6	Team Manager
Emergency Response	P7	Control Centre Manager
	P8	Control Centre Administrator
	P9	Intelligence Officer
	P10	Deputy Head
	P11	Coordinator
	P12	Specialist
	P13	Specialist
	P14	Deputy Head
Humanitarian Organisation	P15	CEO
	P16	Manager

Table 3.2: Research participants

A floor plan of the control centre's main floor is provided in figure 3.3. The desks of the intelligence analysts whose responsibilities included monitoring social media feeds are highlighted in orange. The blue desks represent other intelligence officers with whom the analysts most commonly interacted. The green desks were occupied by the public outreach team whose responsibilities included managing official social media accounts. While the focus of these staff was on messages sent directly to the organisation (as opposed to intelligence extracted from social media data), there was some information exchange between the roles.

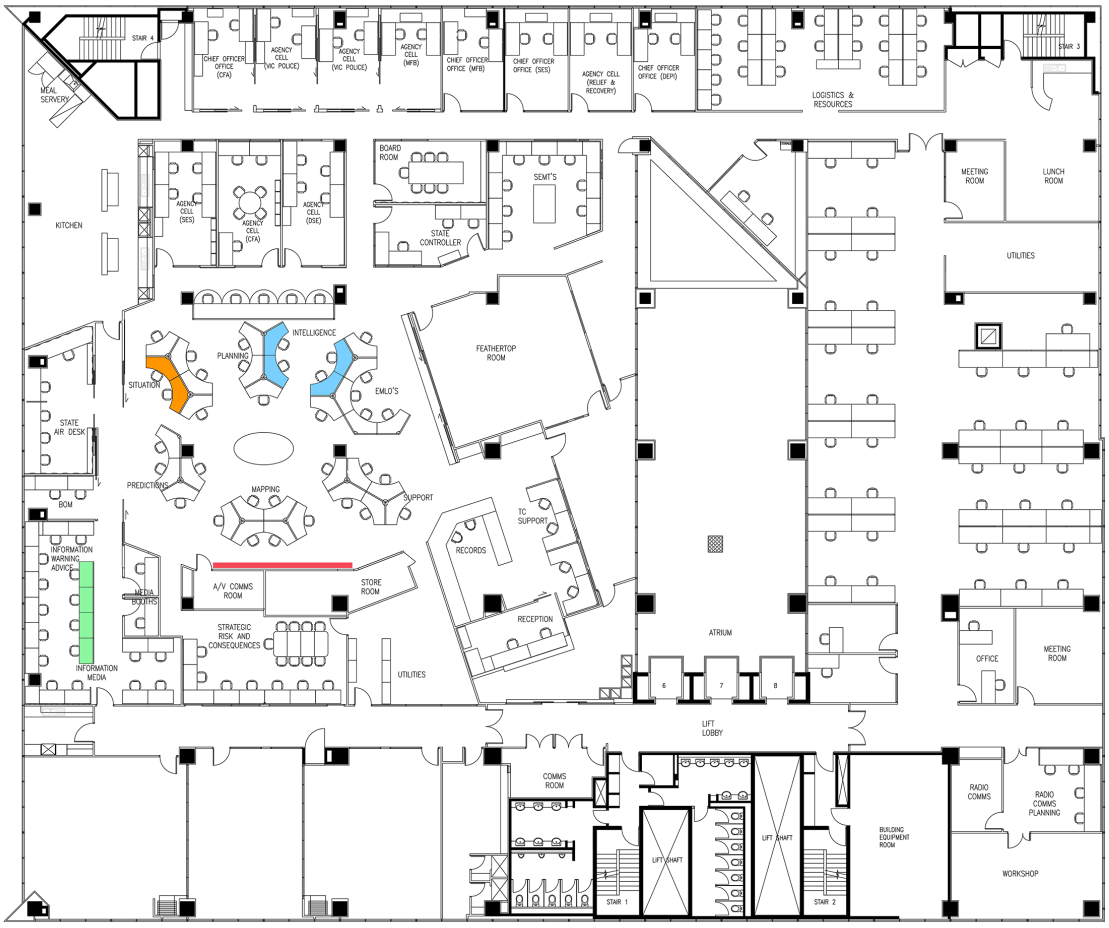


Figure 3.3: Floor plan of State Control Centre. Orange desks represent the position of intelligence officers whose responsibilities included monitoring social media feeds. Blue desks represent other intelligence roles. Green desks were occupied by public engagement staff. The red line denotes the large shared screens and therefore the focal point of the main room.

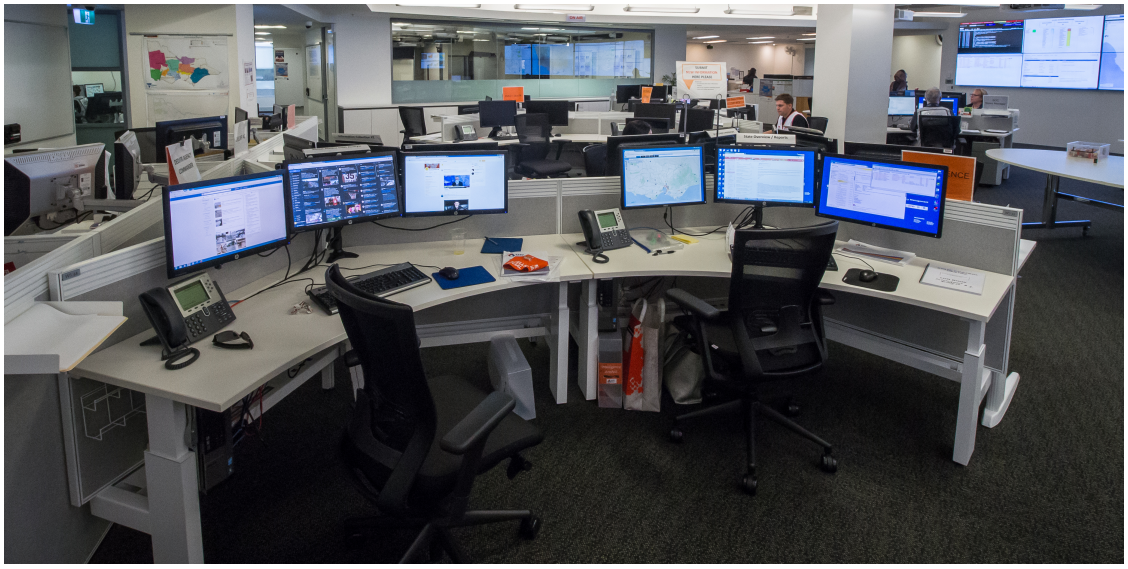


Figure 3.4: A view of the intelligence officer desks. The leftmost screen displays a Facebook feed. The second screen from the left monitors Twitter feeds (via the TweetDeck interface).

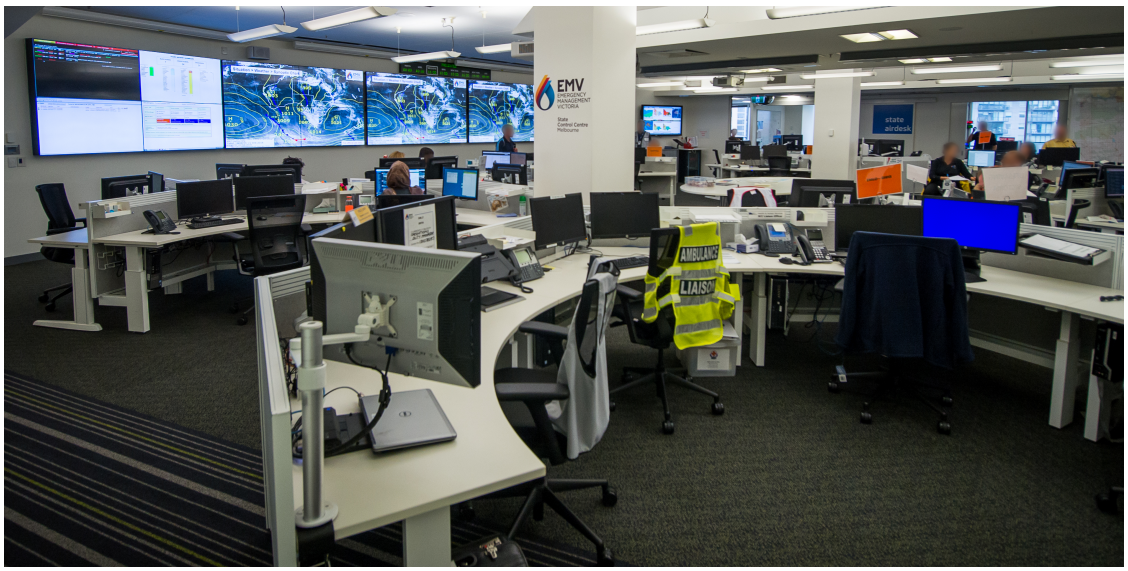


Figure 3.5: The operations floor of the emergency control centre demonstrating the common main screens and open layout.

3.2.2 Thematic Coding

The three phases of inquiry described in A. L. Strauss (1987) are *induction*, *deduction* and *verification*. All three processes continued throughout the duration of the

research project. Inductive analysis based on experiential data and early interviews led to the discovery of provisional hypotheses. Throughout the work, induction examined data for emerging categories and relationships. Deductive phases drew implications from hypotheses for the purpose of verification. Verification of hypotheses through partial qualification or negation was conducted both by examining existing data and through guided discussion with participants in subsequent interviews.

Analyses were conducted informally during early stages of the data collection process based upon the constant comparison method described in Boeije (2002). The limitations of the participants' schedules necessitated an extended period of data collection, encouraging a comparative approach throughout which data are constantly examined and research objectives redefined. Interaction between the data collection and the data analysis phases of research is called for by Hammersley and Atkinson (2007). As such, while presented sequentially in this chapter, the stages of data collection and data analysis were inextricably linked; each iteration of the process informed the next. A comprehensive formal analysis followed the period of embedded observational work, by which point a *theoretical saturation* had been reached.[2]

The coding approach followed practices described by J. M. Corbin and A. Strauss (1990), wherein three stages of coding are prescribed. First, an *open coding* approach was used to explore theoretical possibilities and identify interesting phenomena. The *coding paradigm* proposed in A. L. Strauss (1987) provided a framework by which categories were initially documented. The paradigm encourages the researcher to code data for the following phenomena: *conditions*, *interaction among actors*, *strategies and tactics*, and *consequences*. These distinctions closely resemble the principles of software engineering documentation and allowed this strategy to integrate the resulting categories smoothly into the format required for software development.

[2]Theoretical saturation is the point at which further analysis no longer leads to the emergence of new properties, dimensions or relationships (J. Corbin and A. L. Strauss 2014). The researcher observes similar instances with each new datum and therefore can be empirically confident that their categories are saturated (Glaser and A. L. Strauss 1967). At this point, the theory is sufficiently developed in terms of density and variation and data collection can cease.

Axial coding built a texture of relationships and interactions between each code, reestablishing the structure in the data that had been fractured by open coding (Charmaz 2014). As the structure emerged, *selective coding* was introduced to more systematically code for categories relevant to the research question and move towards conceptual integration. A review of the data using these overarching categories empirically grounded the findings, from which a set of requirements was synthesised to inform the next stage of the research. The refined categories are shown in tables 3.3 and 3.4 alongside illustrative excerpts of data. The categories are labelled as relating either to the *value* or *challenge* of social media data intelligence. Categories that were adequately saturated but not refined into overarching themes provided useful insights and context to the research.

3.3 Findings

This section describes the findings of the qualitative analysis. Thematic codes derived from the data are formulated as a set of features and challenges which inform the prototype design presented in chapter 8. These are presented in descending order of significance with a rationale supported by excerpts from the data and a discussion of their implications to the research. The data excerpts were drawn from interviews with participants and field notes made by the researcher during the observational study. Examples were chosen to illustrate wider patterns of behaviours within the dataset. Due to constraints of space, some data are omitted and referenced by their participant's label. Participant labels follow the format Px and their sector affiliations are provided in table 3.2.

Supplementary findings are then presented, comprising insights drawn from the data which were not captured by a theme. These findings provided useful contextual information which informed the design of the software prototype presented in chapter 8.

ID	Thematic Category	Comment	Example Datum
1-V	Detect Emerging Events	Detecting emerging events using online conversation was a common theme. This most often referred to detecting sub-events within an active event (for example, agitation within a protest or the coordination of ad-hoc civilian responses) rather than initial detection of an event.	*'Events in which we're interested are often first discussed or even organised on public platforms, so we'd like to notice that as it happens.'*
2-V	Situational Awareness	Developing and maintaining a sense of situational awareness or ground truth knowledge is vital during disaster response. The perspectives of those posting on social media was seen as a valuable potential resource, supplementing existing sources of data.	*'Each person becomes a sensor for us, provides us updates of the conditions on the ground. We can't trust it in the same way but it helps paint a picture.'*
3-V	Rumour Correction	Rumours causing confusion and harm spread quickly and may be the product of bad actors or innocent mistakes. Intercepting rumours and providing accurate information from a position of authority is effective at limiting their growth.	*'If we're able to put a voice of authority in the chain early, a lot of the misinformation stops spreading.'*
4-V	Identify Urgent Needs	Messages alerting responders to urgent needs could be directly addressed to organisations or observed incidentally.	*'We've had cases where people ask for help on Twitter before calling dispatchers. It's hard to know whether it's a matter of network coverage and wait times or it's just more habitual, instinctual to use these apps first.'*
5-V	Public Engagement	Using social media platforms to engage with the public is an established practice. Pushing messages out on these platforms was a natural extension of traditional protocols, while using them to field questions or receive reports from the public was less common.	*'People will ask us questions on our [Facebook] page. If we don't respond quickly, there's a risk they then use less reliable sources.'*

Table 3.3: Coding matrix with illustrative data excerpts — data value

ID	Thematic Category	Comment	Example Datum
6-C	Datum Volume and Noise	The insurmountable volume of data generated on social media and the lack of tools with which it could be sorted was seen as a fundamental issue. The data was considered to have an untenable *signal-to-noise ratio.*	*'There's plenty of things we could use on there but there's just no easy way to find them. We don't have the resources to sort through the junk.'*
7-C	Datum Veracity	Data from social media were considered untrustworthy. While some concern existed over *vandalism* and deliberate misinformation, the greater issue was the unreliability of reports made by untrained observers and the issues of incomplete data.	*'If we go out there and the data was wrong, that's a wasted resource. We can't trust what we're seeing there.'*
8-C	Organisational	In some cases, where an appetite may have existed, organisational protocols prevented increased use of social media intelligence. These ranged from restrictions based in law, to work cultures resistant to adopting new procedures.	*'Even if we saw [a message calling for help], we wouldn't be able to use it. We have rules that set out what can trigger a response.'*
9-C	Integration	Facilitating communication between groups is an ongoing challenge of disaster information system literature. An additional tool to observe social media presents another input which may not integrate well enough with existing software and organisational protocols to justify its use.	*'There are so many platforms out there that it becomes a job just to send information between them.'*

Table 3.4: Coding matrix with illustrative data excerpts — data challenges

3.3.1 Social Media Intelligence in Disaster Response

Detailed analysis of the data led to the development of three categories of activity defining the use of social media platforms by disaster response organisations: *broadcasting*, *receiving* and *observing*. Where the adoption of social media as an informational tool varied between organisations, these classes of behaviour demarcated an *integration spectrum*. That is, each behaviour represented a step further along the path of social media integration, demanding more resources and acceptance than those that preceded it: later stages were not observed in absence of the earlier. This is illustrated in figure 3.6: as social media policies mature, an organisation progresses rightward, thus implementing and maturing each behaviour in sequence. A summary of behaviour prevalence within the participant organisations is presented in figure 3.7. The findings presented here were developed iteratively and presented to participants throughout the process to validate the results, thus taking advantage of the constant comparison method and sporadic interview timeline imposed by organisational constraints.

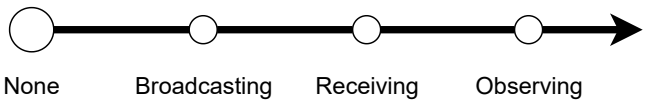


Figure 3.6: Social media integration spectrum

Broadcasting is the most natural extension of existing intelligence processes and was observed universally within participant organisations. In its most simple form, this describes the act of publishing updates to the public using one-to-many platforms such as Facebook pages or official Twitter accounts. *Receiving* messages from the public builds upon the audience established on these platforms: the public direct queries and updates to an organisation through these accounts, which are typically handled by the same public engagement team responsible for broadcasting messages. Finally, *observing* describes actively monitoring social media communication for useful information. This is distinct from *receiving* behaviour in that the target messages are not those directed specifically towards the organisation.

Rather, the messages comprise, for example, undirected broadcasts from other users or conversations between users which are visible to the public. In this way, consumption of the information by the response organisation is *incidental*.

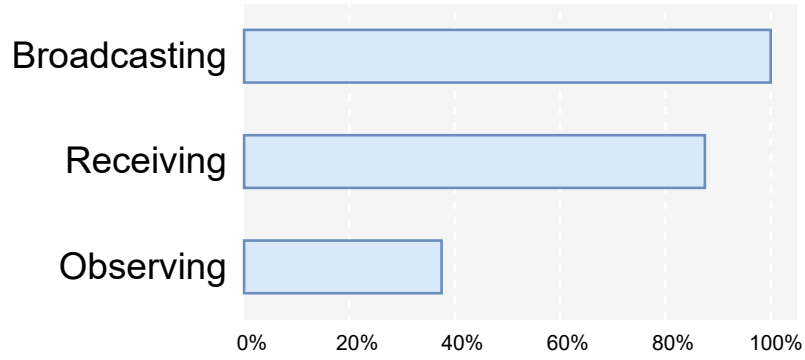


Figure 3.7: Social media behaviour adoption rates

Incidental information may be discovered within any publicly accessible social media content. This represents a vast pool of media with a unique set of characteristics. Integrating these data with existing processes is therefore a complicated process, requiring significant attention and expertise. Of the eight organisations studied, only three had established observation protocols, each operating at a different level of maturity. Participants from the remaining five organisations were able to clearly identify opportunities wherein the use of incidental information could improve their ability to respond.

A flowchart diagram modelling the generalised process of interpreting incidental information from social media sources was developed based on the responses from the three organisations with mature processes and is presented in figure 3.8. The diagram demonstrates three levels of data qualification — first, the relevance of the data is judged with respect to the current needs of the organisation; second, the veracity of relevant data is evaluated using features drawn from the social media source, such as an inspection of the author account; finally, the report is formally verified using information from sources other than social media in a process of *method triangulation*, which is discussed below.

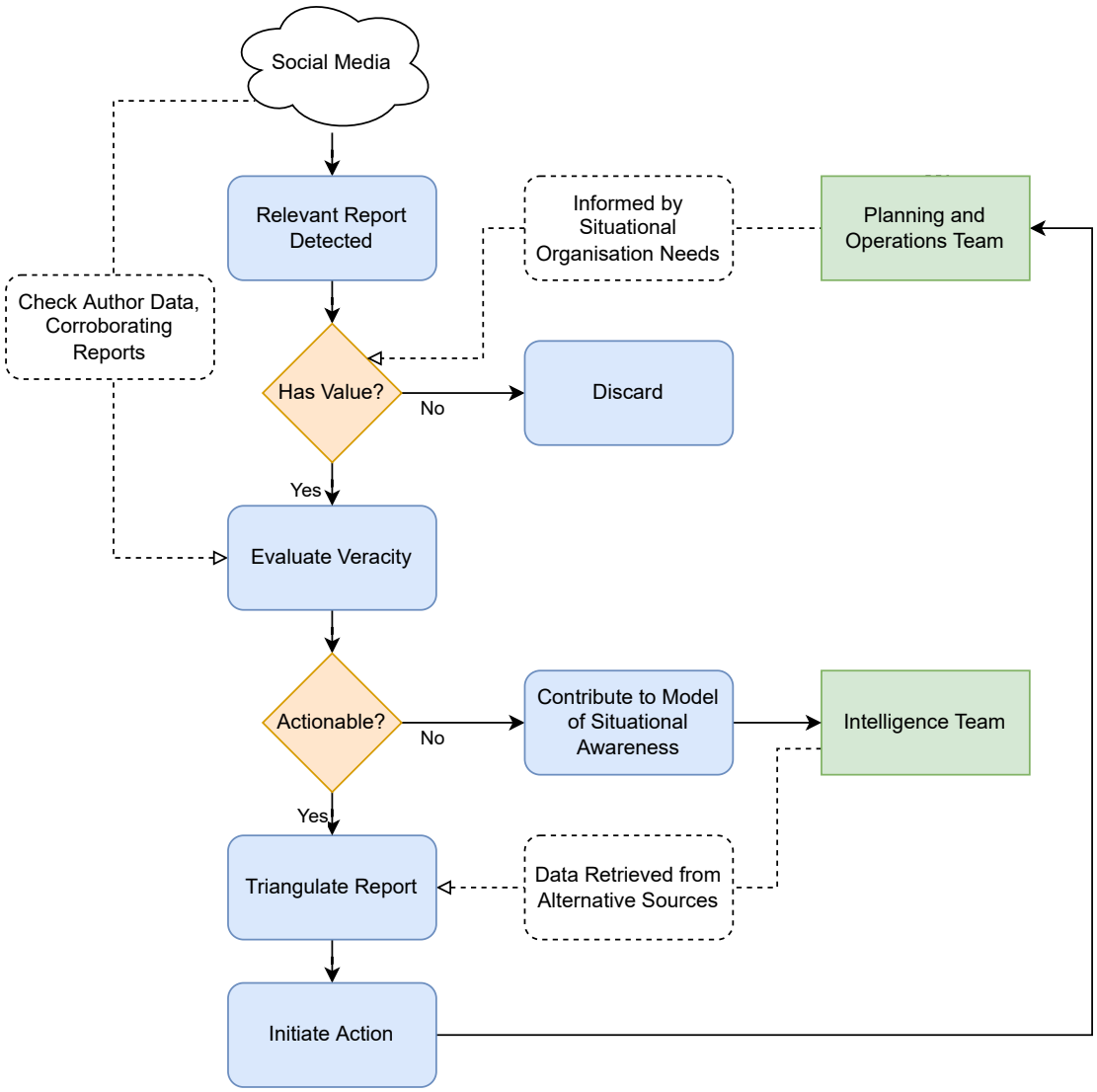


Figure 3.8: Flowchart for social media report interpretation by intelligence officer.

The applications (both potential and realised) of using incidental data were developed into the following five categories, provided with illustrative excerpts of data:

1-V Detect Emerging Events

Detecting emerging events using online conversation was regularly raised as a desirable outcome of social media intelligence. In the following excerpt, P2 shows prior knowledge of social media patterns and expresses a desire to implement intelligence strategies based on these data:

> *'Events in which we're interested are often first discussed or even organised on public platforms, so we'd like to notice that as it happens.'* **(P2)**

By viewing the users of an online platform as a set of distributed sensors, abnormal patterns of behaviour may indicate the emergence of a disaster event. While this application has been shown to outperform traditional sensors in specific cases of large-scale event detection (Sakaki et al. 2010; Earle et al. 2012; Crooks et al. 2013; A. T. Chatfield et al. 2013), the minor time advantage it may provide was not seen to justify the high cost required for constant observation. Rather, the value of event detection was seen to be most relevant in the detection of sub-events within an actively-monitored event. This sentiment is illustrated by the following interview excerpts, in which the participants dismiss the need for enhanced early alert systems and refocus social media intelligence on sub-event detection:

> *'Our early alerts are fast enough, these things are hard to miss.'* **(P7)**
> *'What we can't do is watch everything happening inside [an event].'* **(P3)**
> *'It'd be useful to us to see what's happening in the local groups at the site.'* **(P8)**

For example, police forces found it useful to detect signs of agitation within a mass protest (P2, P3; note: these and certain other data have not been included) and humanitarian organisations monitored the coordination of ad-hoc civilian initiatives which may conflict with their own responses (P8, P15). In these cases, the emergence of the sub-event is effectively a social product of the platform itself — relying on its exponential dispersal characteristics to engage with an audience. These messages are, by design, relatively easy to detect and therefore comprised a key target of the preexisting social media intelligence protocols. In the following field note, P9 inspects trending keywords to detect those which may be related to the event they are monitoring:

> **Field Note - Day 1:**
> `P9 observes multiple Twitter streams on TweetDeck. Trending keywords (as identified by Twitter) are viewed on a secondary browser window. The participant explores the streams of these keywords and adds a selection to the streams in the primary window.`

Once this process emerged as a recurrent interaction, the participant was asked to describe their motivation. Their decision to rely on trends identified by Twitter was made due to time and expertise constraints, as documented by the note:

> **Field Note - Day 3:**
> The interaction documented on day 1 describing the observation of trending keywords has emerged as a common process, conducted several times each day. The participant explained that disaster-related keywords (or 'hashtags') are commonly classified by Twitter as trending due to a sudden surge of interest, and therefore using Twitter's trending list is a useful proxy for keyword identification. Using this proxy avoids the need for a manual examination of live data to identify trends, for which the participant lacked both the time and expertise.
>
> NOTE: This strategy leverages the curational power of 'the crowd' to minimise manual filtering, but is exposed to the selection biases of the crowd and Twitter's undisclosed classification algorithm.

Trending topics emerge from noisy streams due to the high occurrence of topical keywords in the messages of a large population of users. Events which existed independently from the broader public consciousness were more difficult to detect, often represented by only a small number of reports. The sudden collapse of a building or the outbreak of a fire are highly-localised events, the reports of which may not evoke the same level of 'social contagion', thus limiting the extent of their presence within the data. An observer cannot expect to systematically discover these emerging events in a timely manner without more robust techniques.

2-V Situational Awareness

Traditional intelligence input vectors used by disaster response organisations may include, for example, satellite imagery, on-the-ground observers, or drone footage. Supplementing these sources of information with the perspectives of users posting on social media was seen as a valuable potential resource by all participants, with several key advantages:

- The network of users represents a distributed sensor system already in place at the site of interest: data is available immediately and updated frequently. While the distribution of 'human sensors' is not geographically uniform, it may closely follow population density (though demographic factors may introduce a significant representation bias). The concept of humans as sensors is based in academic discourse rather than practical implementation (see for example Goodchild (2007)), however, one participant made the direct comparison:

  > *'Each person becomes a sensor for us, provides us updates of the conditions on the ground. We can't trust it in the same way but it helps paint a picture.'* **(P8)**

- A significant event will often lead to many users generating related reports. The overlapping data allows an observer to verify the content and avoid relying on a single point of observation. Corroboration was seen as the most natural form of verification by respondents from all organisations. P16 described this as *'seeking supporting accounts'* and P10 saw volume as *'the law of averages'* diluting the effect of incorrect data.
- There is typically no financial cost to access the data. More intensive protocols may incur costs due to a need for greater data access or specialised software, however this is not a consideration for early-stage implementation. The limited availability of funding for new projects was a commonly-cited issue (P4, P7, P8, P9, P16), thus a low cost of entry is a valuable trait.
- Many messages included attached photo or video media, providing useful first-hand perspectives whilst limiting the potential for misinterpretation or falsification compared to simple textual reports. Compared to text, these media also take less time to interpret by the operator. Participant 9 describes their preference for images over text from a verification standpoint in the following excerpt:

  > *'It's easy to look at an image and think "that's how it is". If a civilian is telling us about it, we have to think "how well do they know what they're talking about?"'* **(P9)**

Using data in this way informs what is referred to in academic literature as *situational awareness*, and is defined by Vieweg, A. Hughes, et al. (2010) as 'the idealised state of understanding what is happening in an event with many actors and other moving parts, especially with respect to the needs of command and control operations'.

3-V Rumour Correction

Rumours emerge quickly following the uncertainty introduced by a disaster event Procter, Vis, et al. (2013) and Zubiaga, Hoi, et al. (2015). Misinformation causing confusion and harm spreads rapidly through social networks, which are designed to spread information in a manner that resembles viral patterns of behaviour.

While intentional misbehaviour accounts for some cases, the majority of instances illustrated by participants were the product of innocent confusion and miscommunication. A common case was described in which users propagate messages which (incorrectly) claim that resources, such as food and blankets, were being distributed at a given location. The origins of these errors were likely to be genuine mistakes and the messages were shared as acts of good intent. The distinction is of little relevance: the survival and reproduction of a harmful message rely upon the actions of 'good faith' actors who propagate the message whilst remaining unaware of its inaccuracy. Intercepting rumours and providing accurate information from a position of authority (such as a verified authoritative account) is therefore effective at correcting this behaviour and limiting its harmful effect. P2 describes how this practice is used to prevent the spread of misinformation:

> *'If we're able to put a voice of authority in the chain early, a lot of the misinformation stops spreading.'* **(P2)**

As with a biological virus, early intervention is most effective at limiting contagion: detecting the emergence of problematic misinformation was considered a valuable aspect of social media intelligence. This was most commonly brought up by participants working as police (P1, P2, P3), a natural outcome given police that work with anthropogenic events grounded in social interaction and rumour proliferation (e.g. protests, riots, and illegal gatherings).

4-V Identify Urgent Needs

Identifying those in urgent need of aid is a fundamental challenge facing disaster responders. Direct requests for aid are typically made using an emergency telephone number, though requests for aid made on social media platforms have been increasingly observed as access to the requisite technology improves. During a large-scale upheaval, overburdened phone lines and interrupted network coverage may further encourage victims to use online messaging in place of phone calls. This phenomenon was acknowledged by P7, a control centre manager, in the following quote:

> *'We've had cases where people ask for help on Twitter before calling dispatchers. It's hard to know whether it's a matter of network coverage and wait times or it's just more habitual, instinctual to use these apps first.'* **(P7)**

As a relatively new phenomenon, procedures to receive and act upon this data were, where they existed, immature and inconsistent between participants. The process of requesting aid online also lacks the unity of an established emergency telephone number, leaving authors to intuit to whom to direct their message. Messages are often addressed to an inappropriate division of the organisation or simply appeal to the broader online audience and rely on 'the crowd' to direct the message to the correct channel. An intelligence practitioner may also extract an 'urgent need' from a message not intended for the organisation (that is, as incidental information).

P15 described this class of information as *'incredibly useful and actionable, but just too rare to notice'*. Identifying these messages from within the data therefore remains a fundamental challenge. These messages constitute a very small proportion of the streams of data observed during a disaster and were therefore considered inadequate justification for implementing social media intelligence. While the organisations with established intelligence protocols agreed on the value of these data, it was viewed as an incidental benefit due to the low frequency of this type of message (P3, P8, P9).

Furthermore, respondents from three organisations (P1, P6, P14) expressed an inability to act on this class of data, caused primarily by procedural limitations. For some organisations, the lack of verifiability limited the capacity of their dispatchers to allocate a portion of their limited resources based on the message. P1 (police) described the format as being *'just too easy for bad actors to abuse'*. For others, a mismatch between a slow-moving bureaucracy and rapid technological development was cited as a cause for their inability to use the data in this way. P14 explained this as *'waiting for [their organisation] to catch up'*. For these organisations, even where a message is addressed directly to their account, avoiding the challenge of detection, it cannot provide information that is directly actionable.

This category of 'data value' was therefore considered less relevant than those described above. While it represented highly useful information to a proportion of participants, its value was outweighed by the other value categories due to its low frequency and the resulting difficulty of detection.

5-V Public Engagement

The persistent, one-to-many model of broadcasting provided by social media platforms makes them ideal tools for public messaging. Maintaining an organisational account closely resembles, or is identical to operating a personal account, and thus requires little training or specialist knowledge from operators who may already be familiar with the platform. Issuing updates and alerts to the public is standard practice for many response organisations and social media is easily integrated with the existing suite of communication protocols.

Supporting two-way communication requires a more committed investment of resources to handle the volume of requests, and where established, was managed by a team trained in answering public queries. These 'public engagement' teams operated outside of the intelligence pipeline. While they were able to pass on key reports to intelligence specialists where appropriate, they dealt with directed messages rather than incidental information. The purpose of these teams was

described by P7 as to satiate the public's need for information and prevent them from consuming unreliable information from other sources:

> *'People will ask us questions on our [Facebook] page. If we don't respond quickly, there's a risk they then use less reliable sources.'* **(P7)**

Such message-centric approaches are unable to systematically detect broader patterns of behaviour that may provide useful insights and therefore represent only a supplementary aspect of social media intelligence.

3.3.2 Challenges to Social Media Intelligence in Disaster Response

Whilst all participants acknowledged the potential value of incidental information to their organisation, the adoption of more mature intelligence protocols was limited by a number of sociotechnical factors. An examination of the most significant issues presented an opportunity for technical interventions to improve the viability of integrating social media data with existing disaster information systems and intelligence protocols. These are presented as four primary themes, ordered by descending significance as identified by the participants.

6-C Datum Volume and Noise

The volume of data generated on popular social media platforms exceeds the capacity of any human-led team to process without implementing automated filtering techniques. This was a commonly-identified issue summarised as follows:

> *'There's plenty of things we could use on there but there's just no easy way to find them. We don't have the resources to sort through the junk.'* **(P16)**

Platforms such as Twitter allow an observer to filter Tweets by keyword, though even a subset of data created in this way can comprise thousands of Tweets per minute. A human observer will therefore observe only a small fraction of this subset of data, the subsampling method of which is effectively determined by the software used to collect and view the data. For example, the news feed algorithms of Twitter

and Facebook prioritise posts with high levels of engagement from other users, a metric that may not align with the type of posts sought by disaster responders.

While there was not a consensus on what constitutes the most desirable classes of message, there was a clear preference for those authored by users 'on the ground' at the site of the event under observation, as shown in the following excerpts:

> *'We want to hear what people at the site are saying.'* **(P11)**
> *'Someone on the other side of the world is probably not going to know anything that we don't.'* **(P4)**
> *'A lot of [undesirable] stuff we see is from people watching from a distance.'* **(P15)**

A filtered data stream based on keywords related to an event includes discussions from users around the world as they use the same keywords. In the case of a highly-publicised disaster, these low-value data quickly overwhelm the valuable messages, which become time-consuming to detect. The data is therefore characterised by a poor *signal-to-noise ratio*, requiring too great a commitment from already overburdened personnel to process beyond shallow analyses.

Signal-to-noise ratio is a term borrowed from communication science and loosely adapted here to discuss the quality of an information stream. *Noise* is used to refer to messages which do not contain useful information (*signals*). The presence of noise obfuscates the stream, making useful information more difficult to detect. A dataset that has a high proportion of useful messages is characterised by a high signal-to-noise ratio.

Sampling, filtering, and presenting publicly available social media posts at a rate suitable for consumption is a key challenge to the effective use of these data. The design of a data pipeline should be informed by the specific needs of emergency responders rather than relying on broad curatorial methods instituted by source platforms.

7-C Datum Veracity

Data were considered less trustworthy when collected from social media compared to other sources. This presents a critical risk to the allocation of limited resources,

where an incorrect decision could lead to death. Evaluating the *opportunity cost* of each decision is a natural instinct of disaster response personnel, and is summarised with respect to datum veracity by P7:

> *'If we go out there and the data was wrong, that's a wasted resource. We can't trust what we're seeing there.'* **(P7)**

This uncertainty is characteristic of any data obtained online, where the credentials of an author are unverifiable. Reports made online were considered *'less reliable'* (P6, P13) and more likely to be the *'product of mischief'* (P2) than those made over a telephone hotline, though this assumption was untested and appeared to be based in intuition. Indeed, the provenance of an online message is often more verifiable where the author's history can be evaluated, and such investigations were often performed when critical messages were discovered online. In the following excerpt, P5 describes a verification process that is triggered based on the perceived importance of a datum.

> *'If a message is important enough we'll check out the history of the account.'* **(P5)**

Verification based on the characteristics of the author was more difficult for police observers where the point at which such investigation becomes surveillance is not clearly defined, limiting the manual verification they can perform. P3 raises the concern of breaching surveillance protocols but does not firmly state what behaviour is classed as surveillance:

> *'Watching the behaviour of an individual can be classified as performing surveillance.'* **(P3)**
> *'There are strict rules which set out who we can surveil.'* **(P3)**

While some concern existed over *vandalism* and deliberate misinformation, cases of this were considered rare and identifiable. The greater issue was seen as the unreliability of reports made by untrained observers. For example:

> *'If you're not trained, the dust kicked up by a truck can look like smoke coming from a fresh fire.'* **(P5)**

Mistaken reports from instances such as this are naturally equally common for reports made via telephone, however, the two-way conversation of a phone call allows the operator to engage with the caller and identify common cases of misreporting. Many social media posts are enriched by, or comprised entirely of images and video footage. These formats significantly reduce (though do not eliminate) the risk of misreporting by preventing the introduction of human biases which would otherwise emerge in human-formulated text (note that the sampling strategy of these datasets is still determined by human factors). A hierarchy of *perceived* message veracity by source was developed and is illustrated in figure 3.9.

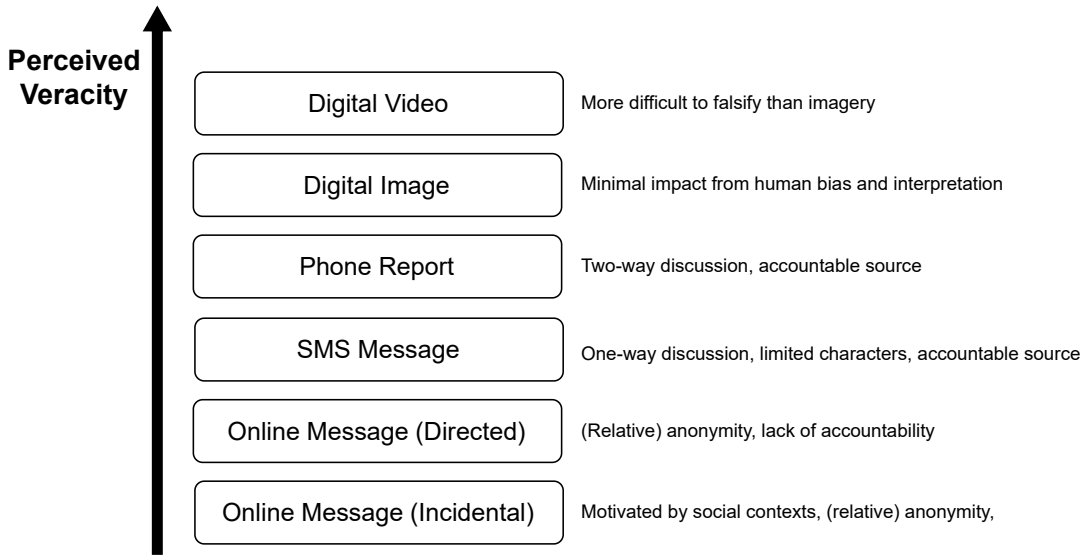


Figure 3.9: Perceived source veracity

The impact of erroneous information (from any source) is mitigated through verification from alternate data sources. This process is described as *'method triangulation'* (P9), drawing from navigation terminology where two bearings are required to fix a point in 2-dimensional space. The role of triangulation in research is discussed in Patton (1999). In disaster response, data sources may include, for example, satellite imagery, UAV footage, deployed personnel, or reports from the public. Therefore, while a report from social media may alert the intelligence officer to an emerging issue, supporting data from a source *other than social media* is

required to elicit action. P9 describes a proactive verification process that may be invoked for high-potential data:

> *'A piece of data I send into the system gets flagged for verification, and this can come from existing data or the other teams might want to redirect aerial imaging to verify something important.'* **(P9)**

This condition of triangulation is equivalently applied to more robust sources of information, thus while the lack of veracity in social media data is acknowledged, it changes little for existing data protocols that have inbuilt data verification procedures.

8-C Organisational

A number of challenges based in organisational and cultural practices were identified. Whilst these were not directly addressable by software-driven solutions, they carried implications for research design and are therefore presented briefly here.

- Skepticism of the value of social media data: Disaster response personnel are highly trained and acutely aware of the challenges facing even qualified sources in the accurate reporting of disaster information. This perspective can conflict with the proposal that untrained members of the public are able to provide accurate and informative reports, precluding even preliminary investigations into the potential of the data they generate. For example, P13 acknowledged the value of social media data but remained skeptical that their organisation would recognise this:

  > *'I think there's a lot in there, but it would take quite some work to convince the other groups.'* **(P13)**

- Procedural limitations: Policies dictating under which conditions certain actions may be taken are common within larger organisations. For safety-critical systems, strict adherence to established policies reduces exposure to the risks of human error and litigation. As a relatively new phenomenon, the growth of social media has often outpaced the rate at which new policies can be implemented. This limitation was raised by P10:

> *'Even if we saw [a message calling for help], we wouldn't be able to use it. We have rules that set out what can trigger a response.'* **(P10)**

9-C Integration

Facilitating communication between groups is an ongoing challenge of disaster information system literature. A disaster control centre comprises multiple member organisations and many inter-connected information systems. Introducing an additional tool to observe social media presents another input that may not integrate well enough with existing software and organisational protocols to justify its use. For example, the extra work to facilitate communication between software was seen as a costly exercise:

> *'There are so many platforms out there that it becomes a job just to send information between them.'* **(P6)**

The challenges of testing and installing a new piece of software into these complex ecosystems can be great enough to preclude even an initial consideration. While there are sophisticated analysis products on the market and developed in the literature, there was little inclination from participant organisations to investigate their potential. P9 identified the time required to test new solutions and to retrain staff as factors precluding such an investigation:

> *'Testing all those tools is something we don't have time for.'* **(P9)**
>
> *'Our team doesn't have time to learn new systems; we need to keep things simple.'* **(P9)**

Furthermore, many disaster information systems are safety-critical and follow strictly-defined implementation protocols. Drawing intelligence from highly-volatile and unproven social data should therefore be conducted independently of existing systems whilst presenting output in a format that conforms to established data models.

3.3.3 Supplementary Observations

This section presents supplementary observations which do not constitute a category as defined above. These findings are included where they bear relevance to the software-driven solutions developed in later chapters.

The Human-in-the-Loop Information Model

Disaster information systems adopt a *human-in-the-loop* (HITL) model, integrating the input of human operators throughout the data management cycle. The domain expertise of human operators and their understanding of the situational needs of the organisation are in this way embedded in the system. The goal of an effective automated data management solution should therefore not be to replace the need for human input, but rather to enhance their ability to perform.

Under a HITL model, the output of an algorithmic intelligence tool is interpreted and validated by a human operator before it can inform decision-making. In terms of automated processing, this eases the importance of maximising *precision* in algorithm design — a human operator with domain expertise performs a filtering pass to eliminate *false positives* from the output before data progress within the system. In such a model, the algorithm is performing the role of *curation* rather than strict classification.

Precision is the measure of the proportion of true positives within the set of positively classified cases. In applied terms, it represents the signal-to-noise ratio of the filtered data stream, where a 'signal' is a datum considered useful by the intelligence officer. Anthropogenic data are typically highly variant and therefore maximising precision comes at a significant cost to *recall*. That is, a model which maximises precision may become so specific that a substantial number of positive cases are mistakenly dismissed. A more detailed explanation of these measures and the implications of the HITL model are provided in chapter 8.

Signals and Noise

The high volume of data and the poor signal-to-noise ratio (SNR) therein were identified as major challenges preventing more widespread use of social media data. More accurately, the challenge may be considered purely as a signal-to-noise issue. A high volume of data would not pose a problem if each datum contained useful information: such a stream would allow an operator to incrementally extract value as their capacity to do so allows. The objective therefore lies in increasing the SNR

such that the time spent conducting manual observation of the stream is justified by the value of the information identified within.

The ratio at which this observation is justified varied based on characteristics unique to each organisation and event instance. The key factors were classified into two broad classes. The *organisational tolerance* for noisy data is determined by, for example, the availability of intelligence personnel, or the (lack of) quality of information available from other data sources. The *operational impact* of noise is the measure by which noise in the data affects its value as intelligence. An increase in the ratio of noise in a dataset, for example, increases the time required of an analyst to identify meaningful data.

Defining Useful Messages

The distinction between signal and noise cannot be strictly delineated: the informative value of a message is governed by the (changing) needs of the reader. A message requesting aid is not useful to an intelligence operator seeking to develop situational awareness, and a situational update is not useful if written in a language the operator cannot read.

Attempting to formally define what constitutes signal or noise is therefore not possible in any generalisable manner. Instead, a proxy characteristic can be used. Given the findings presented above, an intuitive example is the locality of an author with respect to the disaster under observation: a large-scale disruptive event is likely to affect all those within an area, and therefore many of the messages published by *local* authors will contain information useful for situational awareness. More importantly, authors who are not within the area of disruption are very unlikely to produce meaningful information.

Eliminating these *non-local* users from the dataset can therefore significantly reduce the amount of noise, resulting in a signal-rich stream of 'on the ground' updates resembling the sensor network described in 1-V. As geographic data is not typically provided in social media streams, filtering data in this way requires methods by which location may be inferred.

Preexisting Software

Participant organisations with existing social media observation protocols used web-based software provided by the source platforms,[3] accessed using an authenticated account. While more sophisticated solutions exist, the interest to evaluate them was limited due to the integration issues described in 9-C. Users' existing familiarity with the standard platforms also minimised training requirements. The public communications teams provided an exception to this observation and used third-party social media management software to consolidate platforms for broadcasting and responding to messages. The processes of these teams, however, were not integrated with those of intelligence officers.

3.4 Discussion

This chapter has documented the design, conduct and findings from a qualitative study of disaster response organisations with respect to their use of social media as a source of intelligence. Four key challenges facing qualitative disaster research were identified: timing a study to coincide with an event, accessing participants, generalising findings to other events, and issues relating to interdisciplinarity. A qualitative research protocol addressing these challenges was presented, in which the *goal-directed task analysis* method is adapted for the disaster response domain.

A study of sixteen participants within eight disaster response organisations was conducted to develop an understanding of the factors defining how social media data may be used as intelligence during disaster response operations. Data from interviews and observational studies were analysed to develop nine major themes documenting the key areas in which social media data may provide value and the primary challenges limiting their use. Secondary observations are provided which supplement the major themes with contextual data.

The findings presented in this chapter were used to inform the design of novel algorithmic approaches documented in subsequent chapters. These computational

[3]For example, the timeline interfaces presented on twitter.com, tweetdeck.twitter.com, and facebook.com

methods were developed to filter social media data based on the requirements of the participant disaster response organisations. Most importantly, four key cases in which social media data was seen to be informative were identified (in descending order of potential) as follows: detecting emerging events, developing situational awareness, identifying and correcting rumours, and identifying urgent needs. Each of these cases requires a uniquely defined, though potentially overlapping, subset of Twitter data:

- **Detecting emerging events** typically involves analyses of aggregated data and monitoring for emerging topics within geographic clusters.
- **Developing situational awareness** integrates first-hand accounts from users at the site of the event with existing intelligence sources. Informative messages are evaluated on an individual basis and may include a combination of text, images, and video.
- **Identifying and correcting rumours**. Rumours which warrant correction bear two key characteristics: they are incorrect (or otherwise problematic), and show signs of spreading beyond originating user clusters. Problematic rumours may exist outside the set of local users and therefore the task of rumour correction encourages the observation of a user population not constrained to the local geographic region.
- **Identifying urgent needs** requires data that shares similar characteristics to that which develops situational awareness. That is, the messages comprising this dataset originate from users local to the event and are evaluated on a per-message basis. The categories which qualify as urgent (and serviceable) needs will vary based on the observing organisation.

The viability of algorithms based on these datasets was predicated upon the degree to which the relevant datasets are present and identifiable within the larger Twitter stream. For example, if categorising users based on their locality to an event is not possible using metadata or automated analysis, then the effectiveness of cases requiring local user data is limited as messages from non-local users dilute the dataset presented to the human observer. An analysis was conducted in the following

chapters which evaluated Twitter data observed during disaster events to determine how well the characteristics of the data supported the cases documented above.

The most limiting issue identified by participants was the amount of data that is generated on Twitter and the corresponding resources required to manually filter and evaluate streams for useful messages (6-C). This finding aligns with perspectives from the literature (Stieglitz, Mirbabaie, et al. 2018). Therefore, a primary consideration is the application of algorithmic classification to filter data streams into subsets more densely populated by informative messages. The efficacy of this approach is dependent on the degree to which informative messages may be distinguished from the uninformative. As the definition of 'informative' is based upon the case for which the data are analysed, the predictive power of their metadata must be tested for each class of behaviour.

The extent to which a lack of datum veracity (7-C) may impact its usefulness will vary by application. For example, aggregated data is more robust to errors and misinformation produced by individual authors than the analysis of individual messages. As discussed above, the risks of misinformation are relatively minor due to existing data triangulation procedures and the human-in-the-loop model. By continuing to embed human input in the filtering process, domain expertise remains within the system and the effect of this challenge is minimised. In terms of algorithmic classification metrics, placing a human at the end of a classification pipeline eases the importance of achieving a high precision score in favour of improving recall. This is discussed in more detail in chapter 6.

Minimising issues of integration (9-C) requires further work which was beyond the scope of this research. As this work was interested in evaluating underlying filtering approaches, prototypes developed for integration in established processes were designed to closely emulate the user interfaces of preexisting software. In this way, the training required of the evaluating user was minimised and the confounding effect of their unfamiliarity with the tool was reduced. This approach extends to the deployment of a complete solution: by adopting the interface of familiar tools, novel software is more easily adopted by organisations that do not possess

the appetite for retraining their operators. A deployment should consider whether sacrifices in usability and functionality are warranted to better emulate existing interface design and reduce deployment friction.

In the following chapters, social media data generated during disaster events are collected and analysed to evaluate how well they fulfil the objectives identified by this chapter. Algorithmic filtering methods are developed and deployed in a prototype for evaluation by the participants of this qualitative study. By integrating the findings of the study presented here with the design of quantitative filtering approaches, the requirements of the participant organisations are embedded within the system. This approach ensures that the outcome of this research may best serve the disaster response community.

3.5 Summary

This chapter presented the process and results a qualitative study that developed a conceptual framework describing the processes by which disaster response organisations develop and integrate intelligence data into response operations during disaster events.

Key methodological challenges facing disaster research were discussed and informed the development of the study design. Sixteen volunteers from eight organisations participated in the study as interviewees, supplementing data collected during a situated observational study of one disaster response organisation conducted over a two-week period. The data were analysed and coded to identify five key areas in which data generated on social media platforms supported disaster response operations and four thematic challenges preventing the more widespread use of these data by participant organisations.

These findings presented opportunities, grounded in the requirements of the domain, to improve the utility of social media data as a source of intelligence to response organisations during disaster events and informed further examination of methods by which the value of these data may be realised to better aid affected populations. Methods of data collection based on these findings are developed

in chapters 4 and 5 and followed in chapter 6 by empirical analyses of Twitter discourse from the perspectives contributed by this chapter.

4

Opportunities and Challenges of Social Media Data for Research

Contents

4.1 Benefits of Social Media Data in Research 92
4.2 Social Media Data Availability for Research 93
4.3 Technical Implications of Using Twitter Data 96
4.3.1 Twitter Data Structure 96
4.3.2 Twitter's Historical Archive and Live Data Streams . . 98
4.3.3 API Rate Limiting 101
4.3.4 Data Temporality 106
4.4 Collecting Useful Datasets from Twitter 109
4.4.1 Keyword and Hashtag Filtering 110
4.4.2 Key Author Identification 111
4.4.3 Geographic Metadata 113
4.4.4 User Influence and Message Diffusion 115
4.4.5 Social Network Community Detection 117
4.5 Research Ethics and Privacy of Public Data 118
4.5.1 Assumed and Uninformed Consent 119
4.5.2 Author Deletion of Data 121
4.5.3 Publication of Data 122
4.6 Discussion 123
4.7 Summary 130

The vast amounts of data generated on social media platforms present novel opportunities to researchers interested in online social behaviour while also introducing unique challenges to data collection practices. Large-scale, quantitative

studies of social media data such as the one proposed by this research are predicated upon the ability to access data programmatically (as opposed to through manual interaction with a user interface). The degree to which rich, useful data can be efficiently collected in this manner varies between platforms and is sensitive to changes to company policies.

In the past, social media platforms provided (managed) access to much of their data through the provision of an *application programming interface* (API). These interfaces allowed developers (and researchers) to create software that directly interacted with the platform and were intentionally designed to encourage the development of third-party applications, thereby increasing the functionality and attractiveness of the platform to users. However, as certain platforms emerged as dominant actors and public consciousness shifted towards matters of data privacy and control, these interfaces became increasingly restrictive or were removed entirely.

This chapter provides a brief account of the history of social media data availability and the effects that key policy changes had on academic research in online social behaviour. Twitter is identified as the social media platform most suitable both for academic research and the proposed disaster intelligence tool and contextualises a discussion of the challenges facing the use of social media data in research (contribution C-1). Finally, methods by which social media data may be identified and evaluated during data collection are reviewed and discussed. The findings of this chapter inform the design of a novel piece of data capture software developed for this research and presented in chapter 5.

4.1 Benefits of Social Media Data in Research

Social media platforms provide a rich source of unique data that presents novel opportunities to researchers and intelligence practitioners. As online discourse becomes more accessible to a wider population, the volume of data available from social media sources continues to grow (chapter 2). These data provide valuable perspectives on human behaviour and represent populations that are diverse both culturally and geographically.

Social media data are often enhanced with descriptive metadata features which facilitate nuanced analyses such as geospatial information, author relationship data, and category tags. While access to social media data is often freely available, key technical challenges are posed to the design of data collection protocols. These include defining suitable collection parameters, developing tools to record data in a suitable format, managing inconsistent data structures resulting from unpredictable human behaviour, and adapting to the changing landscape of features and platforms in the social media domain.

4.2 Social Media Data Availability for Research

Facebook is the largest and most ubiquitous social media platform, with over 1.66 billion daily active users recorded in December 2019.[1] It is not surprising that it has received significant attention from scholars and has dominated the field of research on social media platforms — a review of social network scholarship found Facebook was the subject of 80% of the studies examined (Rains and Brunner 2015). The prevalence of early Facebook-related studies was facilitated by the ease with which researchers were able to access data from the platform (Kosinski et al. 2015). Changes to Facebook's data protection policies, which moved towards better-protecting user privacy, have since limited the ability for researchers to access user data.

Facebook's early *Graph API* (v1.0) facilitated the development of apps that, once authorised by a user, were able to access large amounts of the user's data. Researchers could then create trivial applications (for example, a 'Which celebrity do you most resemble?' quiz or a photo collage creator (Vajda et al. 2011)) that entice users out of curiosity to provide authorisation, unaware that in doing so, their data could be harvested by the app's creator. Nazir et al. (2008) collected a dataset of more than 8 million users by creating three applications that leveraged social features to encourage the use of the apps by members of their users' networks.

[1] https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Fourth-Quarter-and-Full-Year-2019-Results/default.aspx (accessed 2020-02-25)

These designs would, for example, encourage a user to invite their friends to install the app by offering rewards or other social features.

Authorised apps could access the friend network of an authenticated user and return a set of data for each of their friends, thus affording researchers the opportunity to use this data for social network analysis (using, for example, the Netvizz application (Rieder 2013)). This feature was removed in v2.0 of the Graph API, released in 2015, which introduced strict controls to the information returned about an authorising user's friends.[2] Facebook social network analysis research then shifted to observing interactions between users on public 'group' pages where communities discussed posts shared by other group members or interacted with public figures and organisations (for example, Einwiller and Steilen (2015), Akter and Aziz (2016), and Kaur et al. (2019)).

In April 2018 the *Facebook–Cambridge Analytica data scandal*[3] revealed the extent to which private organisations were able to collect the personal data of Facebook users (Wylie 2019). Facebook responded by imposing additional limitations to their API which restricted data access to only moderator-approved apps for *events* and *groups*, thereby removing the ability of external observers to collect data on members of a group or their interactions therein.[4]

As the parent company of the popular photography-based social network Instagram, Facebook announced similar restrictions to the Instagram API,[5] such that it would no longer return, for example, the list of a user's followers and *followees*, therefore reducing its usefulness to researchers. Since 2016, Instagram's developer guidelines have codified the position that API access is intended only for marketing and advertising purposes; non-commercial research is excluded from its acceptable use cases (Perriam et al. 2019).

[2]https://web.archive.org/web/20151001063607/developers.facebook.com/docs/apps/upgrading/ (accessed 2020-02-25)

[3]https://en.wikipedia.org/wiki/Facebook-Cambridge_Analytica_data_scandal (accessed 2022-07-27)

[4]https://www.facebook.com/business/news/restricting-data-access-and-protecting-peoples-information-on-facebook (accessed 2020-02-25)

[5]https://developers.facebook.com/blog/post/2018/01/30/instagram-graph-api-updates/ (accessed 2020-02-25)

As Facebook grows more resilient to outside observation by better protecting the privacy of its users, Facebook data become less useful both to researchers and other external parties. In the context of this research, the challenges of monitoring Facebook activity limit its value in automated disaster response intelligence practices. This is not to suggest that Facebook is no longer a useful resource worthy of monitoring during disasters; only that the quantitative approaches proposed by this research are not feasible given the current state of data access.

While other networks increasingly limit third-party access to their users and data, Twitter's one-to-many broadcasting format, where posts are publicly accessible by default, continues to allow researchers access to its vast amounts of data. Furthermore, integration with third-party developers enables the production of tools which are able to collect and process data efficiently, providing datasets upon which large-scale research projects are conducted. Twitter reported 152 million daily active users in Q4 2019,[6] who publish over 500 million Tweets per day.[7] These user numbers are effectively enhanced by unregistered users who are able to read public content on Twitter without creating an account. Twitter has, therefore, become the *de facto* source of data for social media research and is used in a wide range of studies including analyses of public sentiment during pandemics (Boon-Itt, Skunkan, et al. 2020; Manguri et al. 2020), measurements of user influence and information dissemination (Xu et al. 2014; Guo et al. 2020), and hate speech identification (Poletto et al. 2021; Basile et al. 2019).

The features which make Twitter suitable for academic research inherently support its use in information and intelligence platforms. Chapter 3 proposes the development of a tool that is able to monitor social media data for useful eyewitness information, a use case openly supported by Twitter's public access format. The Twitter platform, therefore, represented a suitably constrained domain within which the feasibility of such a tool could be explored and the effectiveness of social media

[6]https://www.sec.gov/Archives/edgar/data/1418091/000141809120000019/twtrq419ex991.htm (accessed 2020-02-25)

[7]https://blog.twitter.com/official/en_us/a/2014/the-2014-yearontwitter.html (accessed 2020-02-25)

data in disaster intelligence evaluated. The remainder of this chapter discusses the challenges and limitations of using Twitter as a source of research data.

4.3 Technical Implications of Using Twitter Data

Twitter provides the most publicly accessible source of social media data and is built upon a framework of unilateral many-to-many relationships between public accounts where accounts represent, for example, individuals, groups, organisations, or even automated 'bot' users (Martí et al. 2019). This deviates from the model used by other popular platforms in which individual accounts are differentiated from those of organisations, follower relationships require consent, and bots are unwelcome.

Access to Twitter data is segmented by volume and offered at various price points. The sampled live stream most commonly used in research is available without cost and represents only a small fraction of total activity on the platform. Historical searches are constrained for unpaid users such that only a fixed amount of queries may be made per window of time. This section discusses the structure of Twitter datum objects and the methods by which these data may be collected for research and intelligence purposes.

4.3.1 Twitter Data Structure

The fundamental structure of Twitter data is the collection of Tweets created by its users and their related author data. Tweets are short messages of 280 characters (previously 140) which may include URLs, embedded media, or references (mentions) to other Twitter users. Tweet objects are enhanced with metadata which may include (for example) the location from where they are authored, references to Tweets from which they quote, or the number of replies and 'likes' received from other users. All Tweets are accompanied by an authoring user account object which includes similar metadata such as the network of other users which they follow or are followed by. The core structure of Twitter may therefore be considered as a set of user objects which generate a set of Tweet objects. User objects may be connected by

follower relationships and Tweet objects may be linked unidirectionally through conversational *threads*, where they are in reply to or quote from previous Tweets.

In this thesis, the terms *author*, *user*, and *account* are used interchangeably when referring to a Twitter user account for reasons of clarity. This was done to differentiate between users of the data collection and analysis software (developed in chapters 5 and 8) with the Twitter users represented in the software output. The singular term of *user* is also somewhat misleading given that an account may be managed by a group of people, an organisation, or an automated system.

In Twitter terminology, a *friend* is another user account that a given user follows (i.e. a *followee*). On many social platforms friend status represents a bilateral relationship in which both users have accepted the relationship, whereas on Twitter relationships are unilateral. That is, a user does not require the permission of the target user to follow them and therefore may be following a 'friend' who is unaware of their existence (while the user would be listed in the friend's list of followers, the friend has little oversight of this list). See figure 4.1 for a representation of these relationships. The one-way relationship model of Twitter allows users to follow key information disseminators such as public figures and news outlets in the same way that they follow friends and family.

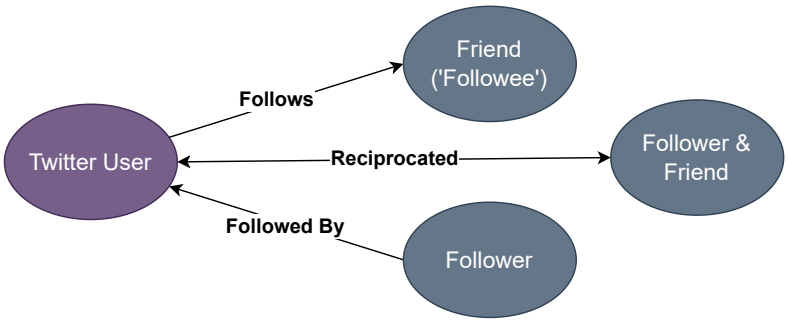


Figure 4.1: Twitter relationship terminology

The data returned from the Twitter APIs are structured around a set of Tweet and User objects. User objects *author* Tweets, *Retweet* or '*favourite*' existing Tweets, and reference other users as *followers* and *friends*. In addition to the message text, Tweets can include hashtags, media (photos, videos, and gifs), URLs,

geospatial data, references to other users (mentions), stock symbols, and polls. These relationships and their cardinalities are illustrated in figure 4.2.

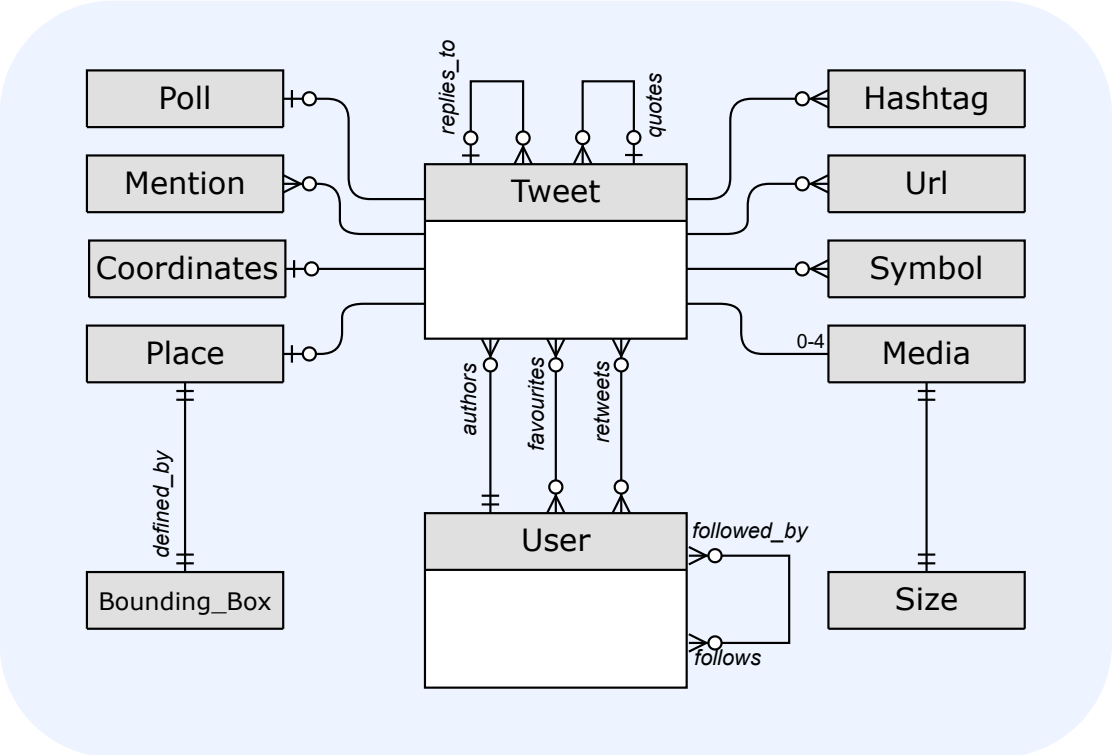


Figure 4.2: Twitter object data structure

4.3.2 Twitter's Historical Archive and Live Data Streams

At the time of this research, Twitter provided public access to its data in two formats — a live stream from which messages were received in the moment of their publication[8] and a historical record available through the Search API which searched against a sample of Tweets published in the past seven days.[9] In 2020, Twitter announced a new API (the Twitter API v2) which introduced new features and access levels.[10] As the data collection in this research was conducted prior to the release of v2, the discussion below concerns v1.1.

[8]https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/overview (accessed 2022-07-23)

[9]https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/overview (accessed 2022-07-23)

[10]https://web.archive.org/web/20200725024734/https://blog.twitter.com/developer/en_us/topics/tools/2020/introducing_new_twitter_api.html (accessed 2022-07-23)

Live Data Streams

Twitter provides its data streams in a tiered system based on sampling proportion: the full stream of Tweets, or *Firehose*, was originally available as an enterprise-level product supplied through third parties partners and then transitioned in 2015 to be offered directly by Twitter through the *PowerTrack API*.[11] The *Decahose*, also referred to as the *Garden Hose* stream, is the second tier of enterprise access and provides a 10% random sample of the Firehose stream. The freely available stream, `GET statuses/sample`, also known as the *Spritzer*, is intended to provide at most a 1% sample of the public Tweet volume at any time,[12] however the precise proportion and sampling method are not formally documented by Twitter (Boyd and Crawford 2012; Olteanu, Castillo, et al. 2015).

The Twitter API platform offers two options for filtered streams: the enterprise-level PowerTrack API can apply filtering arguments to the complete Firehose stream, and the `POST statuses/filter` API is a freely available option that filters the same stream but limits its output to 1% of the total stream of Tweets. If messages matching the filtering criteria exceed the 1% threshold, the stream will return a sample of the results such that the threshold constraint is satisfied, though the sampling methods which Twitter uses for this process are not known (Ruz et al. 2020). In principle, the public stream should return all Tweets matching the filter parameters provided the set of results constitute no more than 1% of total Tweets generated at the time. For a disaster event, there is a reasonable possibility that the 1% limit will be exceeded: Morstatter et al. (2013) observed the spritzer stream to have a coverage of 39.19% of total Tweets for a set of keywords related to the Syrian Civil War in 2011, indicating that eligible Tweets matching the criteria comprised 2.55% of all Tweets produced during that time. As geotagged Tweets comprise less than 2% of all Tweet data, a geographically

[11] https://web.archive.org/web/20150412145411/https://blog.gnip.com/twitter-data-ecosystem/ (accessed 2022-07-25)

[12] https://twittercommunity.com/t/potential-adjustments-to-streaming-api-sample-volumes/31628 (accessed 2022-07-25)
https://groups.google.com/g/twitter-development-talk/c/BHNl_jz6igI (accessed 2022-07-25)

defined Twitter stream will return a full sample for all but the largest bounding boxes (Martí et al. 2019; Morstatter et al. 2013).

Naturally, the coverage of a given dataset collected from the spritzer stream is predicated upon filter parameters and Twitter activity and is not easily measured. During this research, the rate at which data were provided by the spritzer stream exceeded the capacity of the data collection processes to record and therefore the sampled stream was considered sufficient.

The validity of the sampled stream was examined in Morstatter et al. (2013) and found to provide an inaccurate representation of overall activity on Twitter. Both hashtag rankings and topic detection were less effective using free streaming data than an equivalent random sample drawn from the Firehose feed, particularly for low levels of coverage (that is, where the sampled data represented a smaller proportion of the entire dataset). Analysis performed on retweet networks to identify key information brokers in the dataset (measured by degree centrality and betweenness centrality) found that the free streaming sample identified roughly 50% of the key nodes using daily data, with greater accuracy over longer periods. The research showed that the sampling process used by the free stream introduced systemic biases into the data, the extent of which depended upon the overall proportion of data represented in the sample stream. More recent analysis in Leetaru (2019a), however, found a 1% sample collected over seven years to be nearly perfectly correlated with the daily volume of the firehose stream ($r = 0.987$).

In practice, the negative effects of sampling could be mitigated either through paid access to the enterprise stream or *backfilling* data using the *representational state transfer* (REST) API to search the historical archive for Tweets not detected in the live stream. The importance of sampling bias mitigation techniques is predicated on the unknown degree of systemic bias introduced by Twitter's current sampling process and the extent to which this introduces undesirable outcomes in the application of the data. Such analysis requires access to the firehose stream and is therefore recommended as an opportunity for further research in this domain.

Historical Data

The *REST API* accepts requests to retrieve historical data from the Twitter archives. Most notably, keyword searches can be conducted to return a set of relevant Tweets; and Twitter user information, including follower networks, can be requested by username. The public search API, `search/tweets`, allows searches against a sample of Tweets published in the past seven days, while a premium or enterprise account can access Tweets from as early as 2006.[13] The sampling process of the public API is undocumented and may lead to the exclusion of relevant messages, making it less useful for the process of backfilling where full coverage is required. When the API endpoint `GET users/lookup` is used to access individual Twitter user data, including the entire timeline of their Tweets, the public API returns all relevant data and therefore the sampling limitation only applies when attempting to collect Tweets using a keyword search.

For studies based on historical data, many researchers use Twitter archives which are collected and hosted by third parties (F. B. Keller et al. 2020; Murakami et al. 2016).[14] These collections are usually based on the 1% *spritzer* or 10% *garden hose* samples and are useful when examining past events. In 2020, the Twitter API v2 introduced an *Academic Research Track* which provided free access to the full historical archive dating back to 2006 for academic research.[15] A notable distinction between these sources is that externally-archived data may include Tweets that have since been deleted or hidden, whereas the Twitter API full-archive search will not return this content.

4.3.3 API Rate Limiting

In order to prevent API abuse, online services institute a limit on the number of requests an application may make for a given time window (for Twitter, this is 15 minutes). When developing an app that uses Twitter APIs, the developer must

[13] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/overview (accessed 2022-07-23)

[14] For example, at https://archive.org/details/twitterstream (accessed 2022-08-08)

[15] https://developer.twitter.com/en/products/twitter-api/academic-research/product-details (accessed 2022-07-23)

request an *access token* from Twitter, identifying the app to the Twitter API and allowing limit enforcement. When an app exceeds its limits, further requests are denied until the current time window expires and the limits are reset. When a Twitter user allows an app to use their Twitter authentication details (that is, they log in to Twitter through the app), the app may make requests on the user's behalf, making use of the user's API request allowance in addition to its default allowance.

These limits apply to the Twitter REST APIs which are used to search historical data or return information about a user. Table 4.1 shows a sample of the API v1.1 limits that were relevant to this research.[16] The *app auth* values apply to public access tokens available freely to developers. Higher limits are available to enterprise accounts, however, these were not used for this research. The Twitter API v2 released after this research introduced minor changes to these values.

Endpoint	*Requests*/ Window (user auth)	*Requests*/ Window (app auth)	*MaxObjects* per Request
GET followers/ids	15	15	5000
GET followers/list	15	15	200
GET friends/ids	15	15	5000
GET friends/list	15	15	200
GET search/tweets	180	450	100
GET statuses/lookup	900	300	100
GET statuses/user_timeline	900	1500	200
GET users/lookup	900	300	100

Table 4.1: Twitter API rate limits

The fourth column in table 4.1 lists the maximum number of data objects returned per request for each API endpoint. The effective object rate limit may

[16] https://developer.twitter.com/en/docs/rate-limits (accessed 2022-08-09)

therefore be represented as:

$$max_objects_per_window = max_objects \times requests_per_window \quad (4.1)$$

For example, the user lookup function returns, at most, $300 \times 100 = 30,000$ user objects per 15-minute window using an application access token. With respect to social network analysis, the API is most limited when attempting to retrieve a user's friend and follower list. A single request for a follower list returns a maximum of 5000 user IDs (or 200 user objects) and therefore obtaining the follower list for a user with over 5000 followers requires multiple requests, of which a token is permitted 15 per window. This limitation is a key challenge facing Twitter network analysis and is arguably why research in this area focuses primarily on Tweet and Retweet networks, which do not face this limitation, rather than follower and friend networks (for example: Choudhary and U. Singh (2015), Smith et al. (2014), X. Wang et al. (2012), and Itakura and Sonehara (2013)).

Overcoming API Rate Limits

Scraping

Users of an API service are typically required to attach authentication data (provided by the API provider) to their request parameters, allowing the API provider to track and limit the number of requests made by each user. Web scraping is an alternative to using an API that does not require authentication and is therefore resistant to rate limiting. In essence, a web scraper interacts with a website in the same way that would a human using a web browser. It can therefore be difficult for a web host to differentiate visits made by an automated scraper from those of legitimate users. Human users, and therefore scrapers, are able to access virtually all historical Tweets, whereas the API will only return Tweets from the last seven days. Scraping can therefore enable access to a larger dataset than is available to the API. Hernandez-Suarez et al. (2018) describes a method by which they use a generic web scraper, Scrapy,[17] to collect Tweet datasets using keyword filtering. Their tests

[17] https://scrapy.org/

show that their scraping tool is able to collect larger datasets than possible with the API and can execute queries in less time. TWINT[18] (Twitter Intelligence Tool) is a more feature-rich scraper that is also able to query user follower networks. As shown in table 4.1, accessing follower networks using the API is heavily constrained by rate limiting and therefore scraping presents an attractive alternative.

The availability of API endpoints is reliant entirely on the host organisation, which may restrict or remove API access without warning. Examples such as Facebook, described above, show the fragility of third-party software that relies on API access, forcing researchers in this 'post-API age' to increasingly turn to web scraping to conduct research (Freelon 2018).

Social media providers do not typically approve of scrapers accessing their content. By circumventing authentication (and therefore rate limiting policies), automated scraping can place a heavy burden on the host's servers and is often used by third parties to harvest sensitive user data to be used for marketing campaigns or phishing (identity fraud) attacks. Hosts often enact policies to limit the effectiveness of scraping, effectively creating an 'arms race' where common scraping methods are blocked, then scraping tools are updated, and so on. Scrapers can therefore require higher maintenance than well-documented and authorised API methods. In testing performed for this research, a series of requests made with TWINT were identified by Twitter as automated and the IP address from which they were made restricted for a period of time. This is a 'soft' form of rate limiting and can be overcome by using, for example, a virtual private network (VPN) to distribute the queries across a range of IP addresses (Gheorghe et al. 2018), however, this process is neither cheap, simple, nor ethical and was not pursued further.

Request Queuing

As rate limits are defined for a given period (every 15 minutes for Twitter), requests made in excess of the limit can be queued and executed once the allowance is reset. Tweepy,[19] a popular Python wrapper for the Twitter API monitors response codes

[18] https://github.com/twintproject/twint (accessed 2020-03-30)
[19] https://github.com/tweepy/tweepy (accessed 2021-02-28)

returned by the Twitter API for the 'rate limited' error (status code `429`). When encountered, the offending request is added to a queue and executed when the time window resets and the limits are refreshed. This is a straightforward approach to handling rate limits without losing requests, however, it does not remove the limitation: collection of a large set of data will simply create a long queue of requests that may take hours or days to resolve.

Cycling User Authentication Tokens

Twitter API requests return header data which include the number of requests remaining for a given window (`x-rate-limit-remaining`) and the time before the window ends, at which point rate limiting resets (`x-rate-limit-reset`).[20] Therefore, by modifying a Twitter wrapper such as Tweepy to monitor these values, a custom piece of software can detect when the limit has been reached for a given endpoint before making a request to the API. By cycling through different *user tokens* as each of their allowances is exhausted, the limits depicted in table 4.1 are effectively multiplied by the number of tokens the software has access to. In this way, software authorised by multiple Twitter accounts can collect larger sets of data before reaching the rate-limiting thresholds (Schroeder et al. 2019).



Figure 4.3: Twitter authorisation UI

[20] https://developer.twitter.com/en/docs/rate-limits (accessed 2022-08-09)

User tokens are generated when a Twitter user authorises a piece of software to make requests to the Twitter API on their behalf. The software presents the user with an interface through which they may log in to their Twitter account and grant authorisation (see figure 4.3).

Twitter is capable of monitoring geolocated account activity and performing IP logging and can therefore deny requests from a single physical server (irrespective of multiple account authorisations). This was not an issue encountered during this research, however, a distributed solution may be required for larger collection processes (Schroeder et al. 2019).

4.3.4 Data Temporality

Using the Twitter REST API to collect archived data raises an important consideration in that the accuracy of the resulting historical representation may be affected by interventions made upon the data in the time since its creation. For example, while an API search for Tweets will not return a comprehensive set of matches (as explained in section 4.3.2), any content which has since been deleted or hidden is also excluded. A Tweet can be deleted by an author individually or as the result of an author deleting their entire account and associated Tweets. Twitter may also delete an account where they detect a breach of their terms of service (for example, a bot account or abusive user). Tweets also become hidden from the public when a user sets their profile to private.

Further, metadata associated with the author may change between the time at which the Tweet is published and the moment of data collection. A preliminary analysis of behaviour during disasters was performed during the early stages of this research and revealed instances where users temporarily changed their profile location field to the affected area during the event as a show of support. These phenomena will not be represented in user data collected after the event (as profile locations are reverted). Friend and follower networks are also constantly changing; the latter without any intervention from the subject user.

While the approximate historical representation which the REST API provides may be adequate for research analysing Tweet content, the deletion of data between the time of publication and collection may cause systemic anomalies in the data where, for example, spam or sensitive content is less likely to survive until collection. An author-centric analysis as proposed by this research is more susceptible to these temporal effects, particularly with respect to network analysis, as these features are particularly volatile. Change history is not available through the Twitter API, which provides only the most recent representation of a user account and their relationship network.

Studies using these features therefore mandate a data capture approach designed to negate the effects of temporal volatility. The natural solution is to use a live feed — either by implementing software that periodically polls the REST API for recent data or by using the streaming feed to collect content the moment it is published. Author data can then be retrieved on an as-needed basis (i.e. as close to the time of Tweet publication as possible) using the REST API and both data objects stored in a local archive.

Media objects attached to Tweets are hosted by Twitter and represented in Tweet messages as URLs denoting file locations. Therefore, the deletion of a Tweet also causes the deletion of the media file. As media objects can provide rich contextual information for a Tweet, or may be the only content of a post, to best preserve a dataset these media files should be archived in addition to their parent Tweet objects. Naturally, the same risk of deletion applies to external URLs; links to media on the Instagram social network, for example, are common in Twitter datasets as an alternative to embedding media within a Tweet directly and these Instagram posts are susceptible to the same risk of deletion by their authors.

Documenting temporal features is critical in analyses of emergent social behaviour (Vieweg, Palen, et al. 2008) and by archiving data as it is first detected the effect of deletion or modification is controlled (to the extent where the original datum is captured beforehand). However, these acts of alteration themselves impart useful information and are therefore desirable to observe: message modification

(while not a feature currently supported by Twitter) may add or correct information, and deletion may carry particular contextual implications which are not captured by the static archival method. For example, a conversation thread that continues after the original message is deleted or modified may change in context, particularly where new participants enter the conversation after the alteration.

Therefore, changes to information state provide useful temporal data which are not represented in typical static archives. Furthermore, identifying such changes allows an archive to better align with the ethical principle of respecting deletion by the author (discussed in section 4.5.2). Detecting alteration events is not inherently supported by the Twitter API and therefore an *alteration aware* archival method requires a process to periodically request from Twitter the current state of a data object for comparison with the locally archived entry.[21] The frequency of this process is constrained by the computational load, which increases linearly with the eligible dataset, and the API rate limits (section 4.3.3).

The key temporally volatile features for Twitter data, where authors are unable to edit Tweets once published, are listed below. The degree of relevance for each item is predicated on the goals of the study and should be considered with respect to the constraints discussed above.

- Author-defined profile features (for example, screen name, profile location, list of followed accounts).
- List of followers of an account.
- Account privacy (affects author profile and authored Tweets).
- Tweet engagement counts (quote, reply, retweet, favourite).
- Deletion of Tweet or account by author.
- Deletion of Tweet or account by Twitter.
- Label added to Tweet by Twitter (e.g. misleading content[22]).

[21] Twitter does broadcast *compliance messages* informing clients of certain changes, such as Tweet deletion.

[22] Twitter defines misleading content as 'claims that have been confirmed to be false by external, subject-matter experts or include information that is shared in a deceptive or confusing manner'. Their current policy is available at https://help.twitter.com/en/resources/addressing-misleading-info (accessed 2022-07-15)

4.4 Collecting Useful Datasets from Twitter

The Twitter platform provides a valuable and unique source of information to disaster response organisations that is vast, free, and timely in moments of severe information scarcity. However, the strict informational demands of responders and the low signal-to-noise (SNR) ratio of data collected from Twitter lead to high processing costs that are often untenable during periods of high intensity (challenge 6-C in section 3.3.2).

Computational methods designed to detect and collect Twitter data presented potential to autonomously increasing the ratio of useful messages derived from Twitter such that the informational requirements of disaster response organisations were met. This process comprised three key stages: first, a data collection protocol was defined to capture Twitter messages relevant to each observed event and supplementary features supporting further analyses (chapter 5); second, classification models were designed to eliminate data deemed uninformative to disaster response organisations (chapters 6 and 7); and finally human actors within disaster organisations interpreted the results of the filtered dataset based on the organisational demands of the current state of response operations (chapter 8).

The goal of the approach was therefore to reduce the rate of noise within the datasets and provide a smaller, more relevant feed to the human operator who then exercised final judgement on individual datum items before introducing them into existing information systems. In this way, the burden on the output precision of the computational classification methods was eased such that an acceptable rate of false positives could be eliminated by the human user, following the human-in-the-loop model discussed in section 3.3.3.

Building Twitter disaster event datasets suitable for the analysis and classification conducted in this research relied upon effective detection and filtering methods. Capturing a significant proportion of relevant messages and authors was a core requirement for an effective characterisation of online behaviour and introduced several key challenges. The primary techniques used for message detection in social media analyses are discussed below.

4.4.1 Keyword and Hashtag Filtering

Retrieving Tweets through keyword and hashtag matching is a technically straightforward process that returns a live filtered stream of messages containing any analyst-defined terms. Hashtag filtering is a common method for selecting data in event-based Twitter analyses (Bruns and Stieglitz 2012; A. Gupta, Joshi, et al. 2012; A. Gupta, Lamba, et al. 2013; Sakaki et al. 2010) and provides an effective method for recording data pertaining to a given event that is well-supported by the API.

The emergence of unique hashtags in response to a disaster event (Bruns and J. E. Burgess 2011) creates a data collection opportunity that minimises the introduction of noise. That is, a stream filtered by an emergent event-driven hashtag such as `#HurricaneHarvey2017` is more directly related to the event than a general term such as `#Hurricane`. Streams filtered by keywords and hashtags are therefore the primary method of interacting with the live stream for data collection during an event.

Naturally, a hashtag stream captures any message matching a tag, including those authored from outside of the affected area. Given that disaster events often evoke high levels of public interest and discourse, these *uninformative* messages (from the perspective of disaster response intelligence) comprise a large proportion of the data feed and therefore require further layers of filtration to eliminate.

Effective hashtag selection requires continual attention as new tags emerge within online communities that must be added to the list of monitored terms. Detecting relevant tags is a key challenge of the keyword filtering approach to data collection and poor selection may result in the exclusion of large proportions of online communities as they form around undetected tags.

Co-detection is a hashtag identification process in which a detected message containing a known tag is observed to contain additional new tags. These new tags are recorded and, once a given incidence threshold is reached, added to the list of monitored terms. This method provides a simple and effective collection of content relevant to the event and is a common approach within the literature (Tonkin et al. 2012; Bruns, J. Burgess, et al. 2011; Kogan et al. 2015; Procter, Vis, et al. 2013),

however, relying on the co-detection process requires accurate identification of initial hashtags, the set of which may not encompass all relevant behaviour on Twitter. Co-detection may work to grow the hashtag list, but this is not guaranteed, and due to the live nature of streamed data, will only detect Tweets containing the newly identified tags after the point of co-detection (though provisions for *back-tracing* newly identified keywords are possible using the REST API).

Furthermore, Tweets may not contain hashtags at all: many authors do not use hashtags, retweets often remove hashtags due to space constraints[23] (Bruns, J. Burgess, et al. 2011), and authorities typically do not use hashtags in their messages (Procter, Crump, et al. 2013). Therefore, while datasets collected using the hashtag detection method may show high values of precision (that is, few unrelated messages), they may also provide a relatively low recall of the total relevant messages and a potentially unrepresentative sample of online discourse.

Selection bias and the omission of untagged data have little impact on research interested in broad trends and qualitative characterisation of online behaviour, but become a problem for analyses that evaluate individual messages and users, such as the practical application of disaster intelligence. Undesirable systemic bias may be introduced where particular classes of author accounts are incorrectly measured or neglected entirely based on their choice of language, leading to the generation of a data stream that is not representative of the underlying population and thus not desirable as a source of information from which disaster response decisions are made. An approach designed to identify and act upon eyewitness reports during a disaster event can therefore not rely on this method of data capture alone, and rather, consider it in conjunction with other approaches discussed below.

4.4.2 Key Author Identification

While an appropriately-defined keyword filtering process provides a feed of data related to the event under observation, its reliance upon the content of each individual message may lead to a failure to gather the complete set of relevant

[23] This behaviour is now rare since a platform update stopped counting hashtags towards the character limit.

messages published by a notable author. The Twitter account from which an informative message is published has a demonstrated interest in the event and therefore exhibits a higher-than-average chance to publish further informative messages. For example, citizens in the affected area provide useful eyewitness reports throughout the event period, and new accounts are often created in response to a disaster event for information dissemination (Mason and Power 2015). The complete set of informative messages published by an author account may not be accurately represented in a keyword-filtered stream where the selected keywords are not included in every message and therefore an author-centric approach is more appropriate.

An author-centric approach to data collection builds upon the keyword stream, in which authors of streamed Tweets are monitored and evaluated throughout the collection process. Where an author account is judged to be a key information distributor, their historical Tweet feed is requested using the Twitter API and added to the dataset (Kogan et al. 2015; Kwon et al. 2015; Starbird and Palen 2011). The author is then added to a secondary Twitter stream such that any future messages published by the account are also captured. The parameters by which an author is classified for inclusion are important to prevent large amounts of unnecessary data being captured and may include, for example, the number of times their messages appear in the keyword stream, the proportion of their total messages which classify as useful, or whether the account is geographically located within the affected area (Kogan et al. 2015).

This method is predicated upon the expectation that an author account that uses an event-based keyword or hashtag in a message is more likely than average to have an interest in the event and thus publish further relevant messages during the event period. While this position may hold overall, the *proportion* of streamed data representing these accounts is unknown and therefore further author evaluation methods are required to prevent the collection of large amounts of uninformative data.

4.4.3 Geographic Metadata

Disaster events are, by their nature, geographically constrained, and therefore messages which are published from within a geographic proximity to a disaster event are more likely than average to be posted by authors affected by the event (Vieweg, A. Hughes, et al. 2010). Studies of Twitter may limit data collection to particular geographic locations (Kogan et al. 2015) or weight information from geographically proximate accounts more favourably in order to leverage this effect (Kwon et al. 2015; Xu et al. 2014).

The metadata of a Tweet may optionally include geographic coordinates provided at the time of the message being sent, however, in practice only a small percentage of users enable this feature. A dataset gathered in Leetaru (2019b) found only 0.08% of Tweets to contain geographic metadata while data collected in Morstatter et al. (2013) contained 1.45%. Values between 0-2% are reported by Twitter[24] and common in the literature (Laylavi et al. 2016; Gu et al. 2016; Kogan et al. 2015; Starbird, Muzny, et al. 2012; Z. Cheng et al. 2010). Coordinate data may describe the precise coordinates of the publishing device (drawn from its GPS module) or the centroid of the Twitter *Place* location assigned to the Tweet. The geographic resolution of a Place object ranges from a single point to an entire country and is defined by a rectangular geographic bounding box. In 2015, Twitter announced that Tweets that included a Place object no longer recorded the coordinates of the publishing device unless the user explicitly elected to share their exact location (Tweet coordinates otherwise represented the Place object's centroid), leading to a considerable drop in the proportion of messages containing specific coordinates (Leetaru 2019b). The discrepancies between the reported proportions of geotagged Tweets in the literature are explained by trends in user behaviour shaped by such policies and the decision to include or exclude place-tagged Tweets in the analysis.

Geospatial data, when available, are typically recorded at the time and place of message publication and therefore may not directly describe the contents of

[24] https://web.archive.org/web/20210615125858/https://developer.twitter.com/en/docs/tutorials/tweet-geo-metadata (accessed 2021-06-15)

the message — for example, an author may report an earlier observation once they have reached shelter (Gao et al. 2011). Therefore their message will include the coordinates of the device at the time of publication rather than moment of observation and creation of any attached photo or video data (Palen 2014). Authors may also choose to intentionally disable geotagging as a measure to protect their anonymity during sensitive events (such as riots or protests) or to preserve battery life in situations where access to power is unreliable (due to the high power consumption of GPS modules). Conversely, where authors are intentionally attempting to spread geographically relevant information, geotagging may be used more frequently. This phenomenon leads to the emergence of detectable behavioural patterns where accounts involved in disaster events exhibit greater or lesser propensities than usual to post geotagged information, thus becoming identifiable to a suitably trained system (Palen 2014).

Location information can be self-designated by a Twitter account in its profile information field, however, this field does not always include valid and interpretable data (Hecht et al. 2011). Users may fail to update the field when they change address, become displaced during mass-disruption events, or adjust their location to intentionally spread misinformation.

In practice, user location is determined through manual content analysis of account streams (Kwon et al. 2015; Starbird, Muzny, et al. 2012) or estimated using machine learning tools (Jurgens et al. 2015; Hecht et al. 2011). These methods are time-consuming, unreliable, and suited to specific research applications. For the purposes of this research, it was sufficient to process explicit data where available and otherwise rely on secondary features to characterise user accounts. In other words, the implicit data which signifies a user to be local to a disaster event concurrently signals their importance as an information source and therefore imputing their location becomes redundant.

4.4.4 User Influence and Message Diffusion

Social media messages that contain information valuable to disaster response organisations are typically directed towards a broader class of social media user and are *incidentally* valuable as disaster response intelligence. The most informative messages published during a disaster event are therefore more likely to be propagated within online communities, thereby signalling their value through their rate of *diffusion* within the online social network. Furthermore, an account that consistently generates content exhibiting high rates of diffusion may be a key source of eyewitness information and therefore valuable as a focus of further investigation. Though the efficacy of diffusion-based message detection is a function of the degree to which the content of popular messages aligns with the information requirements of the response organisations (where popularity is an emergent property established by the online community), diffusion monitoring presents a valuable approach to the task of message identification.

The extent to which an author is able to spread information to other accounts may be described as their *influence* (Quercia et al. 2011; Simmie et al. 2014). The influence of an author represents the degree to which they affect the behaviour of others and can be characterised by a number of measures. In general, *influencers* are loosely defined as authors who disproportionately impact the spread of information within a network (Goldenberg et al. 2009; E. Keller and Berry 2003; Weimann 1994), though this definition is broad and may fail to capture different types of influence in a network (Bakshy et al. 2011). For example, an author sharing information within a small network of friends may have a low overall influence relative to a popular celebrity account or news outlet, however, their behaviour could have a disproportionately high impact on their local network. These *local influencers* often produce high-quality, unique, *eyewitness* information and are therefore of more interest to disaster response organisations than traditional high-influence accounts, such as public figures Tweeting messages of support.

Given that useful information is propagated through the network by other Twitter accounts, and may thus be considered to exert influence upon them, metrics that

measure account influence characterise and identify key account nodes within the network. Research in this area is common and often attempts to identify *tastemakers* and *opinion leaders* through *total diffusion* or *connectedness* (Cha et al. 2010; Dubois and Gaffney 2014). Typically, research focuses on the popularity of an account by measuring the propagation of their messages (representing quality of content as judged by the network) (Bakshy et al. 2011; Xu et al. 2014), and their degree of centrality on the network (i.e. potential reach through their network) (Choudhary and U. Singh 2015; Cha et al. 2010; Simmie et al. 2014; Dubois and Gaffney 2014; A. Gupta, Joshi, et al. 2012; Kundu et al. 2011; Weitzel et al. 2012).

The influence or popularity of material, as defined by online communities, may not align with the information requirements of response organisations. During the London Riots, Tweets that became popular were often reflections on media pieces or humourous in content (Tonkin et al. 2012) and were therefore not considered useful to responders. A filtering method cannot rely solely on traditional measures of popularity and influence to discover meaningful information and should examine other metrics which favour meaningful contributors.

Messages containing information useful to disaster responders are unlikely to achieve comparatively high levels of diffusion in a network when judged by the measures described above. These messages often exhibit characteristics of *hyper-locality* and *hyper-temporality* (Palen 2014) and therefore be of limited interest to the broader network. Given the temporally discrete nature of a disaster event, a key disaster information broker may show unusually high message diffusion during the event compared to their usual behaviour (i.e. their diffusion δ), whereas, for example, a celebrity or news account may have a higher overall diffusion, but a smaller diffusion delta during the event period. Meaning may also be derived from data such as the *distance* a message is diffused through a network, or the proportion of re-tweeters who aren't followed by the creator (a possible signal of relevance to the local community). Measuring these values requires unique data collection methods which capture temporal metrics.

4.4.5 Social Network Community Detection

Social network graphs contain valuable information describing how online communities emerge and communicate during disaster events. Connections between nodes representing Twitter accounts describe a range of social interactions including the sharing of Tweets, communication between users, and the formation of follower relationships.

The tendency for people to associate with others with whom they share similar traits is known as *network homophily* (Newman 2018; McPherson et al. 2001), and leads to a clustering effect observable within the social graph. *Community detection* algorithms divide a network into a number of communities which, given the presence of homophilic mixing, are defined by a common set of latent characteristics. Where these traits include physical proximity (i.e. geoproximate homophily), community structure may effectively inform account locality classification (Z. Liu and Y. Huang 2014). Given that local accounts are more likely to be sharing eyewitness reports, this approach provides a valuable solution to the challenge of datum volume.

Identifying community structure is a computationally difficult task and an ongoing area of research in social network analysis (Kanavos et al. 2022; Ruiz et al. 2021; Abdelsadek et al. 2018; Silva et al. 2017). A key challenge facing community detection for social network data is that algorithms used to furcate networks are very sensitive to graph structure and social graphs are typically noisy (Bakillah et al. 2015). Therefore, traditional community detection methods performed on graphs derived from social media data can produce unreliable results and exhibit poor generalisability when applied to networks recorded during other disaster events.

Bakillah et al. (2015) demonstrates an enhancement to a fast-greedy optimisation of modularity (FGM) community detection algorithm by integrating text-based similarity measures and thus leveraging the added context available in Twitter network data. Deitrick and W. Hu (2013) includes supplementary features such as sentiment analysis, retweets, and hashtags as edge weights and shows that doing so increases the modularity of the classification (and therefore the *distinctiveness* of the community structures). This approach, reflected in later work (Inuwa-Dutse

et al. 2021; Mohotti and Nayak 2018; Hanteer et al. 2018), demonstrates the value of including additional features in a mutually informative way for community detection and supports community detection as a method of isolating information-rich communities for further analysis.

Community detection and other social network analyses naturally rely on the collection of suitable network data. Graphs used in existing work are commonly constructed using propagation characteristics (e.g. retweets and replies) or message content (e.g. common use of a hashtag) due to the ease with which such data may be collected (Murthy and Longwell 2013) (as such information is contained within the original Tweet message). Follower relationships provide meaningful data which have been used for geoinference (Funes et al. 2021; Bakillah et al. 2015) and were common in earlier Twitter research (Xu et al. 2014; X. Wang et al. 2012; Mendoza et al. 2010), though Twitter has since made these data more difficult to access. Furthermore, follower relationships are not naturally included in the data structure of a Tweet object and are temporally volatile as users form or dissolve connections. Therefore, capturing follower network data for graph analyses necessitates the development of targeted data collection protocols.

4.5 Research Ethics and Privacy of Public Data

Studies that collect, analyse or publish data derived from users of Twitter must consider the ethical implications of their work — authors of Tweets may not realise the extent to which third parties can access their data and therefore may share sensitive information which they consider private. A user whose network consists of a small number of immediate friends and family may be surprised to learn the updates they directed towards these people are being read and analysed by researchers or other organisations. Furthermore, people affected by a disaster may use the platform to seek help, update friends and family on their well-being, or share information about physical injuries and conditions. When people are at their most vulnerable, their consideration of privacy may diminish. Research that

leverages these sensitive data must respect the authors' right to privacy and the unique conditions under which the messages were published.

4.5.1 Assumed and Uninformed Consent

The principle of informed consent is a cornerstone of ethical guidance in contemporary research involving human participants (Webb et al. 2017). Fundamentally, it requires informed consent to be given by research participants at the point of data collection. The term informed consent comprises two concepts: to be *informed*, the subject must have been told what participation will involve before the research takes place; and to give *consent*, the subject must have been provided with a genuine opportunity to agree or not agree to take part (United States National Commission for the Protection of Human Subjects of Biomedical & Behavioral Research 1978). In the case of research involving data collected from Twitter, informed consent is typically presumed based on the user's acceptance of Twitter's Terms of Service, which allows for the use of their data for research by both Twitter and external bodies (such as academics). This assumption of consent relies on the user's clear understanding of the Terms of Service, which may not be valid for users who rarely read nor understand website terms and conditions (Fiesler, Lampe, et al. 2016; Luger et al. 2013; Reidenberg et al. 2015). Many online data researchers do not perceive informed consent as relevant for collecting public data such as Tweets (Bruckman 2014; Vitak, Shilton, et al. 2016) and there is a lack of consensus among university institutional review boards in the U.S.A. as to whether the use of publicly available data meets the criteria for research involving human subjects as per the Code of Federal Regulations (`45 CFR 46.101`) (Vitak, N. Proferes, et al. 2017). In the United Kingdom, discussion of informed consent typically differentiates between the collection and publication of sensitive data, where the latter is more carefully controlled (CUREC 2019).

The user data and Tweets collected during the course of this research are publicly available, however, research suggests that authors are often unaware of the potential for the use of their information by researchers and are unlikely to

provide informed consent if asked directly (Williams et al. 2017; Zimmer 2010; Evans et al. 2015), particularly for research which uses their entire Tweet history or user information (Fiesler and N. Proferes 2018). This tolerance towards research changes, however, based on the group using their data: users are more sympathetic towards researchers in academic institutions than private organisations (Evans et al. 2015).

Users are able to self-manage their privacy on Twitter (thus preventing their inclusion in data collection methods), however, it is unreasonable to assume that their settings accurately represent their desires (Crawford and Finn 2015; Solove 2012). Furthermore, they may not be aware of how the platform sells their information to third parties or allows access to their data via programmatic collection tools, thus the importance they place on managing privacy settings may not reflect the control they wish to have over their data. Finally, the aggregation of multiple sources of data may reveal insights that a user is unable to anticipate (Crawford and Schultz 2014). For example, observing the times at which a user publishes Tweets may allow for the inference of their location when combined with other features. These possibilities expand as machine learning methods become more sophisticated and data sources become richer. As the ways in which data may be collected and combined evolve, it becomes more difficult to assess whether any given piece of information may reveal sensitive insights (Solove 2012). It is, therefore, the responsibility of the researcher, who is cognisant of these risks, to act in a way that respects user privacy.

The context within which sensitive information is being collected must be considered: the privacy preferences of a person depend on their circumstances, and their choices shift depending on their situation (Crawford and Finn 2015; Solove 2012; Nissenbaum 2009). Tolerance for the use of personal data in research is dependent upon the purpose of the work (Fiesler and N. Proferes 2018; Evans et al. 2015); these findings support the natural intuition that research seeking to better assist the people affected by a disaster is a more acceptable application of public data than market research by a corporation or surveillance by a government. Ethical frameworks commonly balance the value of the research with the risks to the subjects. The Belmont Report, which stands as a significant source of ethical

standards in research, discusses the principle of *beneficence* as minimising harm while also maximising benefit to participants (United States National Commission for the Protection of Human Subjects of Biomedical & Behavioral Research 1978). The Best Practice Guidelines for Internet-Based Research published by the Oxford University Central University Research Ethics Committee (CUREC) state that the type of consent obtained should be 'proportional to the risk of the research to participants' (CUREC 2019).

Given that obtaining informed consent at the scale required for this work was highly impractical, the concept of beneficence informed the ethical justification of this research. The goal of this work was to directly benefit the users of Twitter and took users' acceptance of Twitter's terms of service as a form of consent while working carefully to minimise the risks of harm or re-identification.

4.5.2 Author Deletion of Data

It is common for Twitter users to delete, modify, or make private ('protect') information that they had previously made public (Stowe et al. 2018). Naive data collection methods which create temporal representations of Twitter data therefore will not represent the current state of the data; researchers wishing to strictly respect a user's right to be forgotten would need to update their dataset as these modifications to the data are made by the user. Using deleted content is also against Twitter's Developer Agreement[25] and is a point of concern in the CUREC Best Practice Guidelines for Internet-Based Research (CUREC 2019). Understandably, it also makes users highly uncomfortable when considering how their data may be used in research (Fiesler and N. Proferes 2018). Maintaining the consistency between a dataset and Twitter's current state would quickly exceed the request capacity of Twitter's API, and therefore any collected dataset will exist as a temporal artefact potentially containing data that an author may otherwise expect to have been comprehensively deleted. This dissonance must be taken into account when considering approaches to data publication. In terms of safeguarding the

[25] https://developer.twitter.com/en/developer-terms/agreement-and-policy (accessed 2020-02-21)

integrity of this research and its subjects, any individual data points that were due for publication or identified as highly influential on the results of the analyses were compared to the live Twitter database to check for deletion. In this way, deleted or protected data was identified on a case-by-case basis.

4.5.3 Publication of Data

Responsible publication of any data collected from Twitter must respect the users' right to privacy and account for the risks that emerge in aggregated datasets. A typical measure for the protection of user privacy is the redaction of usernames from published Tweets (Stowe et al. 2018; Zimmer and N. J. Proferes 2014). Here, usernames are replaced with unique markers that maintain, for example, the relationship between Tweets by the same author. This *pseudonymised* data still bears the potential for 're-identification', particularly when containing many Tweets from a single user. Observing a set of Tweets from a single user, combined with the associated metadata, may provide sufficient context to derive a user's location, place of work, or other sensitive features. Furthermore, simply searching the live Twitter database (or other data caches) for the content of an archived Tweet message may return the original Tweet along with the sensitive data that has been removed from the archive (Webb et al. 2017). Datasets that enrich Tweet objects with their associated user objects and follower networks, such as those collected for this research, further compound these risks by providing even greater context to a data point. Therefore, due to the sensitivity of these datasets, they were not made publicly available.

Due to the anonymising effects of aggregation, the results of the quantitative analysis in this thesis were published in such a way that protects the privacy of the individual subjects. Qualitative discussion, however, required the presentation of individual excerpts of data, which raised concerns over the privacy of the subjects. Bruckman (2002) proposes a method of 'disguise' when publishing sensitive data which includes adjusting the original data by changing details to prevent reidentification. Whether this is possible with Tweets is difficult to

determine, as an advanced search engine may still identify the original Tweet on Twitter's servers as being most similar to the adjusted message, thus identifying the original author (Webb et al. 2017). Twitter's Developer Agreement also requires Tweets to be presented verbatim and with correct usernames, though this is not typically respected in research publications (Fiesler and N. Proferes 2018). As this work focused on quantitative, rather than qualitative, analysis of the collected data, discussion (and therefore publication) of individual Tweets was not critical. Avoiding direct quotes was, therefore, a viable approach that minimised ethical concerns (Fiesler and N. Proferes 2018; Bruckman 2002).

Sections of this thesis that would have otherwise published Tweets verbatim used representative Tweets fabricated by the author (Markham 2012). Fabricated Tweets were created in such a way that they represented the relevant characteristics of a Tweet or set of Tweets, such as the topics discussed or the style of language used, thus providing an illustrative example that facilitates discussion. Qualitative coding required an explanation of the types of content observed within the data for which examples representing each category were generated to facilitate explanation. Using fabricated data raises issues of academic integrity (Webb et al. 2017) — for example, how effectively the researcher is able to represent all facets of the original data and how other researchers may replicate or verify the results of the study when the original dataset is not available. While these concerns are valid, they were considered acceptable downsides to protecting the privacy of the subjects of this study and the methodological approach of this study did not require the granular analysis of individual Tweet messages. Therefore fabrication did not impede the findings or presentation of this work.

4.6 Discussion

The use of social media platforms has increased rapidly in recent years as access to the requisite technology improves and a growing proportion of the population becomes familiar with online communities. As a result, the breadth and depth of the data generated on these platforms have continued to expand and present

increasingly valuable prospects to the research of online social behaviour and applications of open-source intelligence.

Recent scandals and data breaches drove an increasing public awareness of data harvesting practices and demand for more protective privacy measures. In response, many social media platforms have severely restricted access to user data once provided openly to third-party operators. While such protections were a positive outcome for user privacy rights, the landscape of social media research rapidly evolved as sources of subject matter became more scarce.

Throughout these industry trends, Twitter has remained a platform built upon the concept of publicly available material (though private messaging options exist) and has therefore become a key source of data for the study of online social behaviour. Furthermore, public access to its data has positioned Twitter as a prime candidate for third-party intelligence practices in which organisations observe behaviour on the platform to identify content that may inform their decision-making processes.

Working Without Informed Consent

While it is made clear to users of Twitter that any messages published are (by default) made public, there remains a complex ethical dilemma. Modern research which uses data derived from human subjects is classically beholden to the expectation that *informed consent* is obtained. Studies using large amounts of online public data are often unable to procure this degree of consent due to the volume of data involved and the difficulty in making contact with each author.

When creating a Twitter account, a user is required to accept the Twitter terms of service, which state that publicly published data may be used in research by third parties. Given research has repeatedly established that users do not typically read nor understand terms of service documents (Fiesler, Lampe, et al. 2016; Luger et al. 2013; Reidenberg et al. 2015), their acceptance of these terms fail to constitute sufficient confirmation of consent. While users are able to make their Tweets private, the feature may not be obvious to many users and this is not the default setting. Furthermore, the decision to enforce privacy may not be

sufficiently informed where a user is not adequately aware of the ways in which their data may be discovered or used by third-party data collection processes. Therefore, the public state of an account cannot be taken as an informed position eschewing privacy with respect to third-party access.

The justification for using Twitter data in this study was therefore based upon the assumption of *expected consent* and the social good represented by this research. That is, the tolerance for the use of personal data in research is a function of the purpose of the work (Fiesler and N. Proferes 2018; Evans et al. 2015), which in this case may be considered as a project of social good designed to directly help, among others, the participants from which the data was derived.

While the requirements of various research ethics boards were met by this approach to data collection, further constraints and protections were implemented to respect the rights of the authors to privacy. A key policy implemented in this work was the decision to avoid the publication of original Tweets. Users of Twitter are able to delete or hide their publicly available Tweets at any point after creation and the publication of these messages creates permanent datum artefacts beyond their control to remove. Furthermore, data collected from human subjects living through disaster events may represent particularly sensitive perspectives and therefore further care was taken to protect the vulnerable users who contributed to the data used in this work. User account names were replaced with dummy names and where used to exemplify a concept, Tweets were synthesised such that they captured the relevant features of the data they illustrated.

Data Access and the Impacts of Sampling

The Twitter platform is fundamentally designed to provide public access to its recently published data through its web interface, a feature that has led to its dominance as a source of information used by disaster response organisations with existing social media intelligence protocols (see chapter 3). Access by third-party software is provided through the Twitter API such that large amounts of data may be accessed programmatically. Freely available API access is a rare feature

in the social media landscape and the reason why Twitter has become a popular focus of online social network research.

Software which interacts with the API is subject to constraints that limit the volume of data made available. While Twitter does offer a paid level of access not limited by volume, the price is prohibitive and very rarely used by researchers. The freely available live data stream is capped at 1% of total Twitter activity, and where selected filtering parameters define a set of data that exceeds this proportion, a sample is returned. While typical sets of filter parameters are unlikely to meet this threshold, the global interest in key disaster events could quite naturally lead to a volume of matching messages exceeding 1% of activity.

The method by which the live stream is sampled to provide the subset of data is not clear and existing research on how representative the sampled feed is of total matching activity shows conflicting results (Morstatter et al. 2013; Leetaru 2019a). Furthermore, there is an intrinsic participation bias in online social media data where technological, cultural, or behavioural factors may lead to the underrepresentation of particular demographic groups. As no verifiable personal details were retrieved when collecting user data, such biases could not be rigorously investigated and present an opportunity for further research.

The implications of latent sampling and participation biases were a critical concern during this work and must be considered in any application of social media data for intelligence. Where resulting data are used to characterise a population or initiate rescue operations, the exclusion of vulnerable demographics introduces serious and undesirable outcomes. Care was taken throughout this work to avoid the assumption of representation when analysing the Twitter data and the disaster intelligence platform developed in chapter 8 was designed to embed an awareness of these risks into the workflow.

Requesting historical Twitter data is rate limited such that each authenticated user is provided an allowance of requests per 15-minute window. The request limits are defined separately for different classes of data: the volume of Tweets available within the allowance is much higher than the number of users for which follower

network data may be requested. These differences shape the scope of analyses conducted in Twitter research — while large volumes of data are required to perform network analysis of social data, the rate limitations imposed by Twitter prevent the collection of suitable datasets that are based on follower relationships.

Web scraping is a common solution used to collect data from sites that do not offer API access and has been used to circumvent Twitter's data access policies (Hernandez-Suarez et al. 2018). Scraping was not implemented in this work for two key reasons: first, scrapers are typically in breach of the terms of service of social media platforms and therefore do not align with the data access policies to which the users of a platform have agreed. Second, scrapers are not robust to changes in interface design; modern web platforms deploy updates many times per day and small changes may break the ability of a scraping tool to collect consistent data. While a supervised data collection protocol could adapt to such modifications, an intelligence system must be robust to these changes without requiring ongoing developer intervention.

API constraints are formulated as a limited allowance assigned to an authenticated account with which requests may be made. Therefore, by distributing requests between multiple authenticated user tokens, the per-user limits are effectively increased. Implementing a distributed method requires the development of custom software that is able to securely ascertain the authentication tokens of multiple user accounts and cycle between them as the request allowances of each account are exhausted. This requires privileged access to users' Twitter accounts and represents a form of crowdsourced data collection for which users may be encouraged to temporarily provide their authentication tokens to an intelligence system as a cooperative act during a disaster event.

Data Collection Protocols

The methods by which Twitter data are identified and collected are governed by the features offered through the Twitter API, however, there remains significant scope to the design of data collection protocols and consequently, the characteristics of

the resulting datasets. The applications in which the data will be implemented form the key considerations of the collection design. The volume of a collected Twitter dataset is constrained in two ways: through the rate and sampling limits imposed by Twitter as discussed above; and as a product of the sheer volume of data available and rate at which the collection system is able to write entries to a database. Astutely defined filtering parameters are therefore a critical aspect of an effective collection method and directly influence the quality of any following analyses.

Keyword and hashtag filtering is a common approach to interpreting the live stream or searching the historical archives. Well-defined keywords may ensure that matching messages are related to the event, however, as disaster events often capture large amounts of public attention, a significant proportion of the resulting messages are likely to be captured from the surrounding public discourse by users not local to the event (and thus not sharing informative eyewitness accounts).

Hyper-local keywords and hashtags which emerge within local communities can be used to limit the introduction of unrelated noise in a filtered dataset, though anticipating the choice of these words is rarely possible and therefore this approach requires ongoing analyses of discourse within local communities. Furthermore, authors sharing valuable information may not be attuned to the evolving patterns of online discourse and therefore not include the relevant trending tags in their messages. Finally, an author identified to have posted a useful eyewitness account has a reasonable chance of remaining in the affected area and therefore continuing to share useful perspectives, some of which may not contain terms matching the filter parameters. A user-centric approach in which such users are actively monitored is therefore more appropriate than a purely content-based detection protocol given the geographically focused nature of disaster events.

The set of users comprising the focus of a data collection process grows throughout the event period and therefore an effective method of ongoing user detection is required. An intuitive approach builds upon a keyword-filtered stream such that the set of monitored users is comprised of authors whose messages bear a match value above a given threshold or occur most frequently within the stream.

More sophisticated methods interpret network data and identify key users within a communication network based on measures of their social influence, such as the rate of diffusion of their messages within the social network.

These approaches may prove effective at controlling the level of noise (i.e. unrelated material) within the data, however selecting for users who match a predefined set of behaviours or are more popular than others introduces an undesirable sampling bias which may exclude key eyewitness users. Given the hyper-locality of a disaster event, user influence is unlikely to represent an effective predictor of message usefulness to disaster response organisations. The eyewitness accounts sought by response organisations may be published by any user local to the event, regardless of their choice of keywords or social influence. Therefore, more general locality measures are preferable to ensure the capture of the largest possible proportion of relevant data.

Geographic coordinates attached to a Tweet can be reliable, though are only observed in a small proportion of Twitter data (less than 2%). Geographic locality may be inferred using secondary attributes such as the user-defined profile location or time zone fields. A more promising approach is based on the concept of *geoproximate homophily*: a pattern of behaviour that suggests that a person is more likely than average to create a social connection with someone within geographic proximity. While this assumption holds intuitively in the offline world, the global nature of online discourse is likely to reduce the effect within Twitter follower networks and therefore further research is required to measure the effect.

Where geoproximate homophily leads to an observable pattern of behaviour in online social networks, the relationships between users become valuable predictive features for locality inference: given a known set of local users, the locality of a candidate user may be evaluated based on their nodal proximity within a relationship graph. New user discovery is facilitated through the traversal of edges connecting local nodes with unknown nodes, thus allowing a form of inference not based on keyword filtering. Measuring the extent to which geoproximate homophily influences the formation of online social relationships is an open research question

that was examined in this thesis as a precursor to developing methods of location inference using network data.

A network-based approach to data collection is robust to variations in message content and not predicated upon the ongoing identification of emerging keywords and hashtags. In regions where multiple languages are used, an insufficiently diverse keyword filter may introduce a significant sampling bias. The network approach is less vulnerable to this effect and more easily generalised to regions in which unfamiliar languages are used. The user-centric method of geoinference is therefore an attractive solution to the challenges of data identification, yet relies upon the unknown extent to which Twitter user locality is captured within follower/followee network data (i.e. the degree of geoproximate homophily). Evaluating this phenomenon presented an opportunity for novel research and required the live capture of user network data such that the rate limitations imposed by Twitter, which most significantly affected this class of data, were effectively managed.

4.7 Summary

This chapter has presented an introduction to the field of social media research with respect to data accessibility and highlighted the vulnerability of the academic landscape to changes in the privacy policies of social media platforms. The popularity of Twitter as a subject for social media research was shown to be a product of the accessibility of its data and the breadth of its user base; characteristics which situated the platform as a useful candidate for disaster response intelligence analyses and therefore the focus of this research.

An introduction to the environment within which Twitter data is made available described the challenges and limitations of using Twitter as a subject of research. In particular, the issues of datum volume and sampling bias were considered with respect to the research goals. These limitations effectively constrain the scope of methods viable in Twitter research and informed key software design decisions in chapter 5.

A set of analytical methods through which Twitter data may be filtered and collected was then compared and discussed. The user-centric collection method was determined to be the most suitable approach for the purposes of this research based on the characteristics of target data identified as desirable to disaster response organisations in chapter 3.

Finally, the ethical implications of using Twitter data in research were explored. In particular, the inability to obtain informed consent from the human authors contributing to social media platforms was identified as an open challenge of social media research. A set of datum management policies was proposed that respect the privacy of the authors whose data contributed to this research (chapter 5). Significantly, the decision was made to present only synthesised Tweets in the analyses and discussions of chapters 6, 7, and 8 to protect against the risk of author deanonymisation.

The network analysis proposed by this approach introduced the requirement for the design of novel data collection software able to capture suitable network data subject to the rate limit constraints imposed by the Twitter API. The design of this system is documented in chapter 5 and followed by an account of the disaster event datasets recorded during this research.

5

The CrisisData Software

Contents

5.1 Software Design 134
5.1.1 Requirements Specification 134
5.1.2 Technology Stack 136
5.1.3 Collection Logic 137
5.1.4 Database Structure 139
5.1.5 Interface Design 141
5.1.6 Post-Processing 143
5.2 Data Collection 144
5.2.1 Collection Parameters 145
5.2.2 Reflections on the Data Collection Process 150
5.3 Discussion 156
5.4 Summary 160

The scope and validity of social media research are fundamentally defined by the methods and tools used to create the datasets upon which the research is based. The previous chapter explored methods of data collection suitable for user-centric research (section 4.4) and highlighted the requirement for novel data capture techniques to enable an approach based on social network analysis and geoproximate homophily (section 4.4.5).

Collecting social network data describing relationships between Twitter users was a difficult task that was not supported by existing data collection tools. The primary

reason this feature had not been developed was likely due to the severe rate limits imposed by Twitter which restrict the ability of an observer to collect a volume of network data suitable for research (section 4.3.3). Furthermore, due to issues of data temporality (section 4.3.4), network-centric analyses of Twitter users mandated the real-time collection of network data such that the reliance on traditional rate-limited methods of accessing historical records was not a valid approach for this research.

Custom software was developed to collect data from the sources provided through the Twitter API to a degree of fidelity that supported the analyses proposed by this work. Most significantly, a method to capture temporally valid user network data was developed which implemented a token cycling feature to increase the rate limit allowances by distributing requests between multiple Twitter users (see section 4.3.3).

A real-time approach captured rich user data at the time of Tweet publication and therefore created an accurate representation of the social network as it existed during the event and minimised the effects of data availability decay (Reuter, A. Hughes, et al. 2018). Recorded data were periodically compared to the current state of the source data as returned by the Twitter API and monitored for changes in state. Through this approach, the data not only represented a more accurate record of user networks than was possible using traditional methods but was further enriched with records of feature changes such as evolutions in follower/followee networks and changes made by users to their profile location fields. These temporal effects were analysed for their predictive value in chapter 6.

This chapter presents an account of the development of novel data collection software informed by the findings of the qualitative study in chapter 3 and the method discussion in chapter 4 (contribution C-5). A description is provided of 10 datasets collected during this research, comprising over 1.4 million Tweets authored by 381,000 unique Twitter users. A reflection on the data collection processes provides useful contributions to the further development of social media data collection software.

5.1 Software Design

The network analysis approach proposed in chapter 4 required a live collection of Twitter data which included the follower/followee networks of detected authors. While existing tools were available for Twitter analysis, they were not found to be capable of a live data capture approach that captured user network data.

Common solutions focused primarily on recording Tweet streams based on keyword filters (for example DMI-TCAT (Borra and Rieder 2014), COSMOS (Burnap et al. 2015) and the Twitter Explorer (Pournaki et al. 2020)). Where available, network data typically documented message diffusion patterns (such as retweet behaviour) or topic coincidence networks, due to the feasibility of obtaining these data given API rate limitations. These designs were both a result of, and cause for, existing Twitter studies that had rarely considered user data and temporal features in their analyses.

Custom software able to capture data in this manner was therefore required for this research, the development of which constituted a key contribution of this research. The software was designed and documented such that accessibility was maintained for academics interested in the study of social media data and made available on Github[1] under the GNU GPLv3 license[2] which permits the free use, modification and distribution of the code under the condition of source code disclosure.

The following section discusses the key design considerations of the software development process and is followed by a reflection on the live deployment of the software during 10 disaster events.

5.1.1 Requirements Specification

The requirements for the data collection software were informed by the findings from the qualitative study (chapter 3) and the opportunities and constraints of social media data discussed in chapter 4 and are listed below.

[1] https://github.com/rosscg/crisis-data (accessed 2022-08-28)
[2] https://www.gnu.org/licenses/gpl-3.0.en.html (accessed 2022-10-01)

Functional Requirements

R-1 **Accept keyword parameters from operator**

The software operator may define multiple keywords by which the raw data streams are filtered. These may be single words, phrases, or Twitter hashtags deemed relevant to the event under observation. Keywords may be defined throughout the data collection period.

R-2 **Accept geospatial parameters from operator**

In addition to keyword filtering, geographic coordinates may be provided which define a bounding box area covering the site of an observed event.

R-3 **Capture streamed Twitter data based on defined parameters**

Tweet and author data which match the selected keywords or, where coordinate data is included, fall within the defined geographic bounding box are detected by the system and recorded for later analysis.

R-4 **Hydrate streamed data with additional features**

Data provided by the Twitter stream is enhanced with additional features including the relationship network of the author.

R-5 **Present visual representation of live data capture**

The ongoing data capture process is represented visually in an appropriate manner such that the operator may monitor activity and adjust parameters as appropriate.

Non-functional Requirements

R-6 **Handle API rate limits with constrained queueing**

The rate at which data may be requested from the Twitter API is constrained by rate limits (section 4.3.3). Where these limits are consistently exceeded, overflow requests are to be discarded rather than queued to avoid introducing a lag between the point of publication and data capture that increases as an event matures.

R-7 **Handle errors without user intervention**

The system should be robust to minor unpredicted errors without user intervention. These may include, for example, connectivity or data formatting issues that are otherwise non-critical.

5.1.2 Technology Stack

A robust and comprehensive software specification was developed which satisfied the dual goals of this work: to support future research of social network data and to provide the foundation of a prototype intelligence tool for disaster response organisations. The technology stack was chosen with a view towards extensibility and reliability, and is shown in figure 5.1.

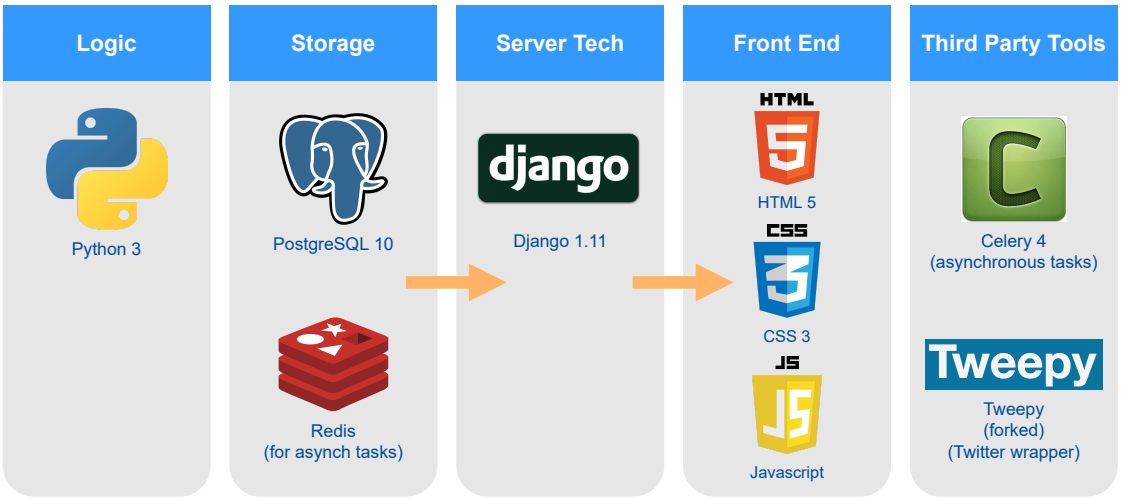


Figure 5.1: CrisisData technology stack

- **Django** is an open-source web framework. A web interface was chosen to support the eventual deployment of an intelligence tool as a web-based prototype, thus simulating the existing web-based solutions observed in chapter 3.
- **PostgreSQL** is an open-source relational database that was implemented once preliminary data collection events generated data at a rate higher than the default Django SQLite database was able to record.
- **Celery and Redis** executed asynchronous requests to the API without disrupting core processes. Streamed Tweet data objects were added (via

Redis) to a queue of tasks awaiting further *hydration* with historical data requested from the API. Queued tasks were managed by *Celery workers* operating in parallel with the primary processes monitoring data streams and hosting the web interface.

- **Tweepy** is a Python wrapper for the Twitter API. A *forked* (customised) version was developed which implemented authentication token cycling.

5.1.3 Collection Logic

The collection software monitored the Twitter stream for messages which either matched any term from a set defined by the operator, or were published within a geographic bounding box encapsulating the area under observation (and contained coordinate information). Detected messages were written to a local database with their associated author data and follower/followee network details were requested from the REST API and recorded. Requests for author network data were quick to encounter rate limiting as described in section 4.3.3. These were therefore run as asynchronous parallel tasks which were able to make requests to the API independently of the main data monitoring processes, thus taking advantage of Tweepy's `wait on rate limit` functionality to effectively queue API requests without interrupting the main data collection process.

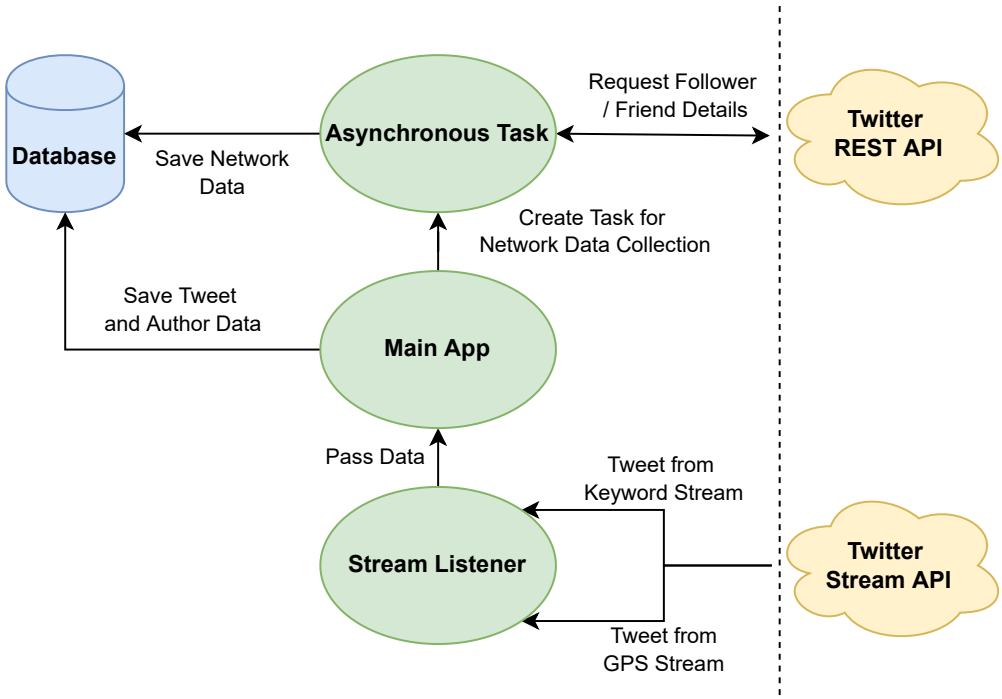


Figure 5.2: Application framework

A simple representation of the app is shown in figure 5.2. The logic implemented to write database records for streamed Tweets and associated metadata is demonstrated in algorithm 1. Note that as Twitter relationships are *directed* and *unilateral*, the methods to record friend and follower relationships bear slight

differences which are ignored in this representation of this algorithm for brevity.

Algorithm 1: Recording Tweet from stream

```
Data: Tweet Object from Stream
Result: Tweet, user, and user network recorded to database
def recording_tweet_from_stream(tweet):
    record_tweet(tweet)
    record_user(tweet.author)
    author_network_id_list = follower and friend ids from REST API
    for networked_user_id in author_network_id_list:
        if networked_user_id in database:
            if relationship[author, networked_user_id] not exists:
                record_relationship(author, networked_user_id)
        else:
            networked_user_data = user data from REST API
            record_user(networked_user_data)
            record_relationship(author, networked_user_id)
    return
```

5.1.4 Database Structure

Data were stored in a relational database which allowed for the use of powerful and efficient analysis techniques using standard SQL queries and was a robust solution that supported scalability and modification. A conceptual entity relationship diagram is shown in figure 5.3 which describes the basic database schema used to store Tweet and User objects. Note that this iteration does not include all available objects from figure 4.2: functionality to record media data was not implemented in the early rounds of data collection, however, these can be extracted using the REST API at a later date if the Tweet has not been deleted or hidden. Coordinates were stored directly within the Tweet class as they were one-to-one relationships. Polls and stock symbols were not considered relevant and were therefore discarded.

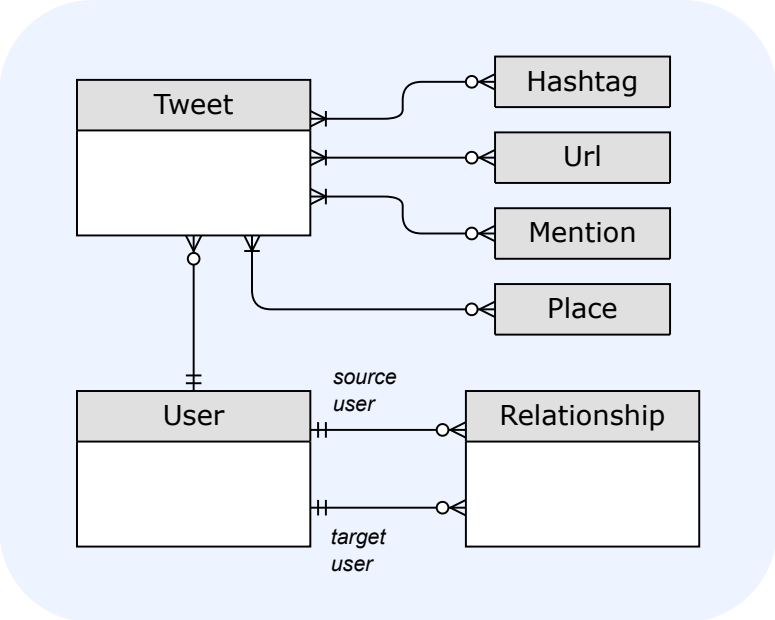


Figure 5.3: Tweet and user database framework

User objects stored both *ego* users, or those that authored detected *Tweets*, and *alters*, which were the followers and friends of an ego user. Alter User objects contained far less information than ego User objects to reduce the burden on the database and the required volume of API calls. The relationship between an ego and alter was modelled through a directed self-referential relationship. In this solution, the relationship was intermediated by the *Relationship* object, which stored metadata such as the time at which the relationship was observed.

Hashtag, *URL* and *Mention* entities were extracted from the Tweet object and stored as distinct objects. These entities could occur in multiple Tweets and storing them as separate objects supported complex analytic queries such as monitoring hashtag coincidence patterns or URL diffusion behaviour. A recent analysis of two emergency events has shown that URLs were present in 50% of Tweets (Meesters et al. 2016). As a consequence of this behaviour, content analysis should go beyond the text of the Tweets and examine associated information (Francalanci and Pernici 2018).

5.1.5 Interface Design

The software was developed to provide an intuitive interface through which data collection parameters could be defined and provide visual feedback describing the current state of the data collection process to the operator for the purpose of validation. The elements were designed modularly to support extensibility for future research projects and to form the basis upon which the prototype disaster response intelligence tool was built in chapter 8.

The key features of the system interface included methods to input tracked keywords within two priority classes, visual representations of collected data as text streams and geospatial maps, and summaries of commonly shared hashtags, user accounts, and URLs which were used to detect trends and adjust collection protocols.

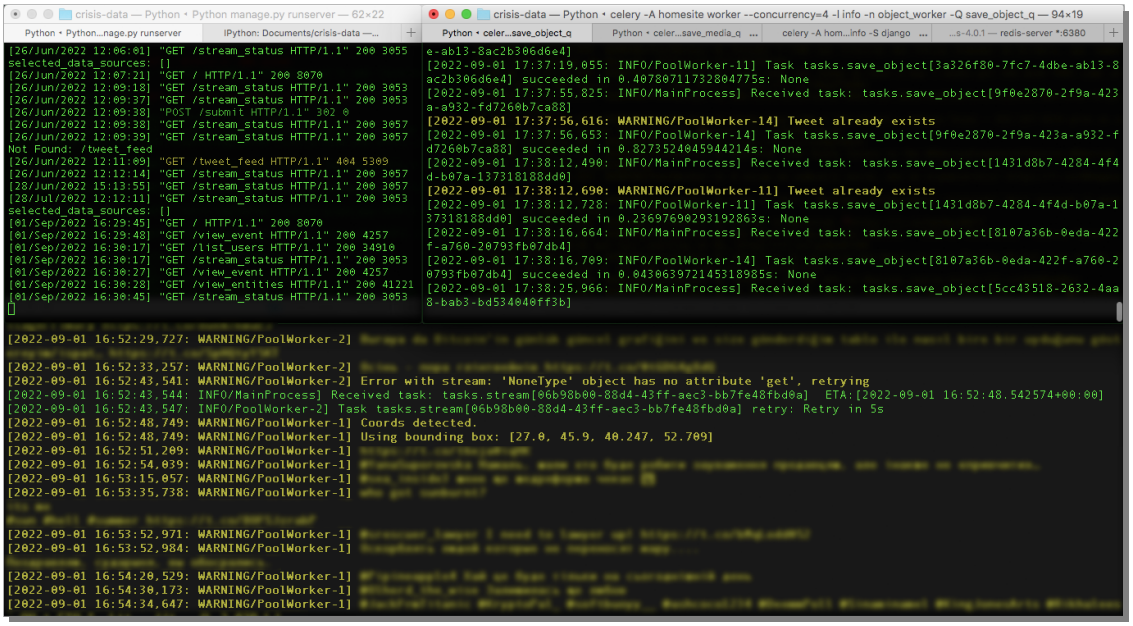


Figure 5.4: Data collection in progress. Top left: web server process, top right: Celery workers handling task queues, bottom: Tweet stream output. Tweet text has been blurred to preserve anonymity.

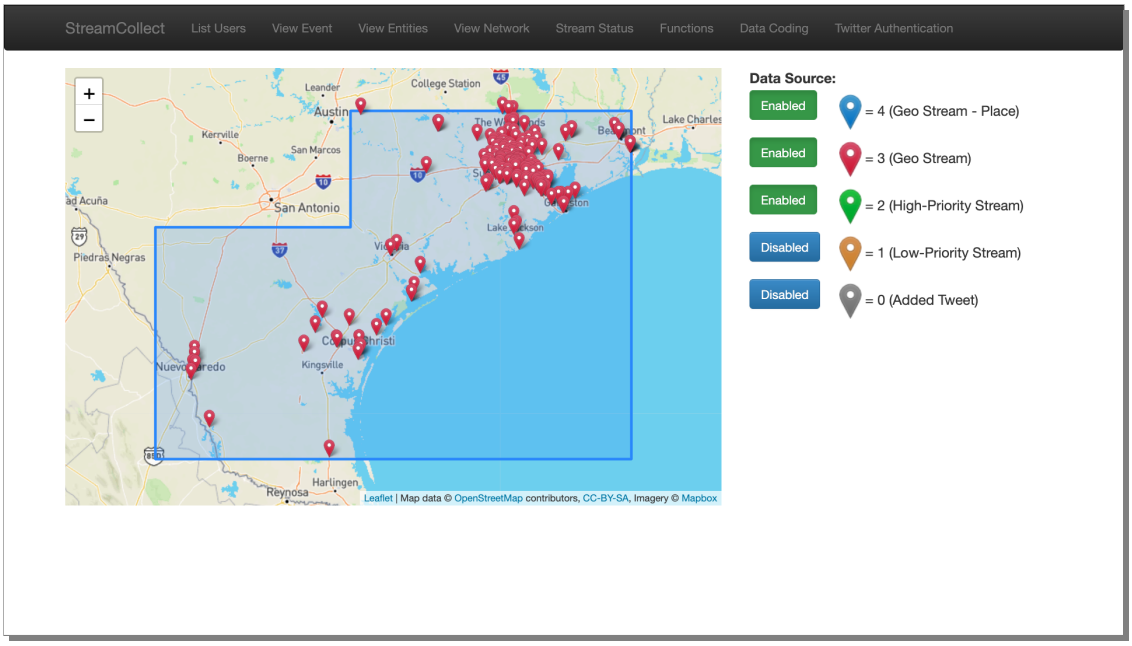


Figure 5.5: A geospatial visualisation of data collection parameters and Tweets containing geospatial data facilitated active monitoring of the data collection process. Note that the geo-aware Tweets shown on this map comprised less than 2% of the total dataset.

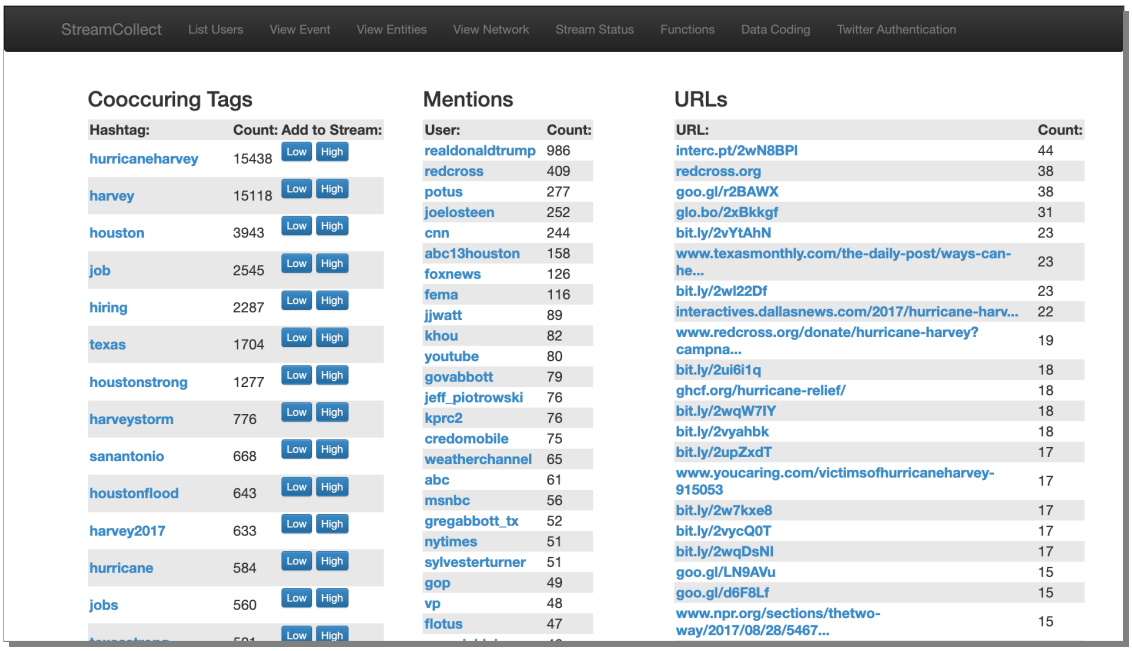


Figure 5.6: Commonly shared hashtags, users, and URLs were summarised for further analyses. For example, untracked co-occurring hashtags were able to be identified and added to the tracked keyword set.

5.1.6 Post-Processing

The data were further enriched by a number of processes which were run after each collection period had ended. These were performed at the end of the process due to both computational and API limitations and were therefore motivated by research curiosity rather than intended practical effect, given the live nature of the disaster response application.

- **Update Relationship Data**

  The network data for each user were collected at the moment they were first detected by the software. User networks change frequently and characterisation of these changes was a consideration in classifying user types. For example, a newly-created bot account may grow its network at an inorganic rate as it is followed by other bots in an attempt to present credibility. As requesting network data from the Twitter API was severely rate limited, updating this information frequently was problematic. Therefore the updated network for each detected user was recorded after the collection period had ended to provide a second point of observation.
- **User Screen Names Verified and Checked for Deletion**

  As Twitter allows users to change their screen name at will, the account of each detected user was queried to check for changes made to the screen name field since the initial observation. The result of this query additionally indicated whether the account had been deleted, suspended, or made protected (private) since the initial observation.
- **User Timelines Downloaded**

  All Tweets authored by detected users between the start and end points of the data collection period were accessed using the historical API and, where not already present, added to the database. Tweets recorded in this manner were marked as distinct from Tweets detected in the filtered streams during the original collection process. Complete sets of user Tweets published during the collection period were known as *timelines* and were informative in evaluating

user locality and informativeness. As this collection drew data from the historical archive, Tweets that had been deleted before the request was made were not captured.

- **User Data Enriched for Follower and Friend Records**

  A small set of data were recorded for each friend and follower of a detected user. The name, screen name, location, account creation date, and follower and friend counts were requested from the API and added to the database.

- **Network Metrics Generated**

  Centrality metrics were calculated for each user based on the network data of the observed set of users. Centrality measures how connected or influential a node is within a network (a deeper discussion is provided in chapter 7). The *degree*, *betweeness*, *load*, *eigenvector*, *Katz* and *closeness* measures of centrality were calculated for each user and added to their record. The network upon which these metrics were calculated comprised relationships between detected users and therefore ignored the followers and friends of each user that were not themselves directly detected, as this would have increased the complexity of the network beyond the capabilities of the hardware to manage.

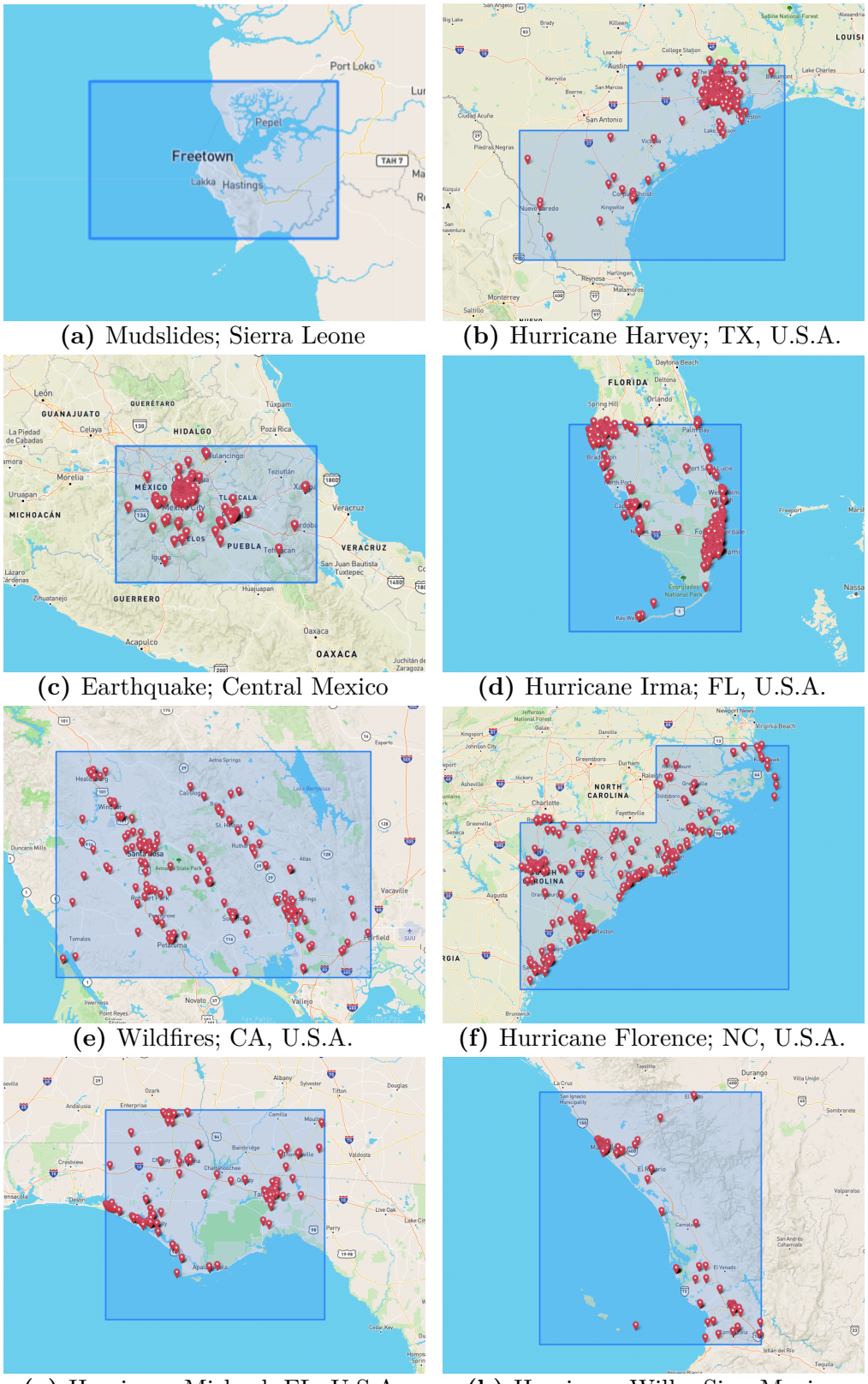
5.2 Data Collection

Ten disaster events occurring between 2017 and 2018 were selected based on their suitability for this research as determined by the disaster suitability taxonomy developed in section 2.1.2. The Twitter discourse surrounding each event was recorded by the CrisisData collection system for use in this research.

Early events acted as *pilot studies* for the system and led to useful insights which were incorporated into the software design for future events. This iterative approach to development allowed for the software to capture more detailed and relevant data as it matured and new features were added. The richness of the data from these studies therefore improved between events as functionality was added or adjusted based on ongoing observations of results. A discussion of the evolution of the software is included below in section 5.2.2.

5.2.1 Collection Parameters

The primary data for the observed events were sourced from a combination of the Twitter keyword and location streams, though geotagged Tweets made up only a small proportion of the sets. Keywords were chosen on an ad-hoc basis based on popular or intuitive terms, with some emergent phrases added after the collection had begun. Geographic bounding boxes were defined around the affected area of each event to capture Tweets that included coordinates or were tagged with a Twitter *Place* object overlapping the defined area. The parameters which defined the scope of each data collection event are detailed in appendix B and visual representations of the geographic bounding boxes are provided in figure 5.7 on page 146.



(a) Mudslides; Sierra Leone **(b)** Hurricane Harvey; TX, U.S.A.

(c) Earthquake; Central Mexico **(d)** Hurricane Irma; FL, U.S.A.

(e) Wildfires; CA, U.S.A. **(f)** Hurricane Florence; NC, U.S.A.

(g) Hurricane Michael; FL, U.S.A. **(h)** Hurricane Willa; Sin., Mexico

Figure 5.7: Bounding boxes used for data collection of Tweets containing geospatial information. Presented with a sample of detected Tweets containing geospatial data.

As the collection system was based on the live Twitter stream, data published immediately following a sudden-onset event (such as an earthquake) were not recorded. This gap represented the period between an unanticipated event first occurring and the activation of the collection software. Dataset metrics are presented in table 5.1, and show the period of data collection for each event, the number of Tweets recorded by the software, the corresponding number of authors (ego users), and the total alter accounts recorded (that is, the total unique followers and friends of the set of authors).

The data collection software was installed on an Apple Macbook Air with a 1.7GHz i7 processor, 8GB of RAM and 500GB SSD. While a cloud server would have been an ideal platform for this application, a local machine was chosen primarily due to budgetary constraints. Running the software on a local machine allowed close monitoring of performance and rapid release of patches as bugs were discovered during collection periods. The primary limitation of the local deployment was the speed at which the machine could record data: the volume of content provided by the filtered Twitter streams regularly exceeded the capacity of the machine to record, therefore a proportion of the available data was discarded. A suitable cloud deployment of the software could be configured to scale as stream activity increases and therefore collect a more comprehensive set of data.

Early exploratory analysis of the data was performed during each collection period. Monitoring and evaluating the behaviour of the data as it was produced was necessary to ensure early versions of the software were performing as intended. Preliminary insights developed during these analyses helped to characterise the data from each event and assess their suitability for further study.

Event	*RecordingPeriod*	*Tweets*	*Authors*	*Followers* & Friends
Mudslides, Sierra Leone	14–16/08/2017	13,842	10,056	4.8 mil
Terror Attacks, Barcelona, Spain	17–18/08/2017	2,698	2,517	1.4 mil
Hurricane Hato, Hong Kong	24/08/2017	784	649	0.4 mil
Hurricane Harvey, Texas, U.S.A.	25/08–02/09/2017	46,872	30,270	12.9 mil
Earthquake, Central Mexico	19–21/09/2017	14,920	11,095	4.3 mil
Hurricane Irma, Florida, U.S.A.	10–17/09/2017	42,167	23,847	10.1 mil
Wildfires, California, U.S.A.	10–24/10/2017	36,534	17,154	7.0 mil
Hurricane Florence, North Carolina, U.S.A.	13/09–03/10/2018	907,835	226,845	157.4 mil
Hurricane Michael, Florida, U.S.A.	11–18/10/2018	335,709	41,666	37.6 mil
Hurricane Willa, Sinaloa, Mexico	23–30/10/2018	53,823	17,207	47.4 mil

Table 5.1: Dataset metrics

The first two events recorded were chosen primarily as 'proofs of concept' and tested the performance of the software under stress. While these events did not entirely satisfy the criteria for a suitable disaster event as discussed in section 2.1.2, at the point in the study when these events were observed it was unclear whether a more suitable event would occur within an actionable timeframe. These two datasets were therefore not studied as rigorously as later events once a general sense of their structure was derived.

The Sierra Leone dataset had a small pool of Twitter users, limiting the potential of useful reports being observed from local users. The developing telecommunications infrastructure also constrained the Internet access of local users, and may have

been less resilient to disruption from disaster events than more developed nations. Initial exploratory analysis of the data revealed a high proportion of the messages detected as authored either by media outlets or those observing from abroad. These sources were not considered useful in the context of this work and therefore the validity of this class of event and dataset was limited.

The direct impact of the Barcelona attacks was limited (relatively) to a smaller number of people than other events and was brought under control by authorities in a short time. Again, discourse consisted primarily of external commentary and media reports, with a small set of eyewitness accounts appearing and quickly circulating. As a geographically focused event that occurred in a short time span (hours rather than weeks), the period of uncertainty for emergency response organisations was brief. The capacity for the public to report information on social media, and the period in which that reported information remained undiscovered (and therefore a candidate for detection by the proposed software) was limited, therefore this kind of terror attack was not a useful candidate for analysis.

The six hurricane events proved to be more suitable for this approach: as slow-onset events, preparing the recording software and selecting appropriate keywords to track were straightforward processes based on cursory observations of online discourse. Recording began as each hurricane made landfall and continued for approximately a week in order to capture behaviour in both the onset and recovery phases (the recording of Hurricane Hato was terminated early to record the landfall of Hurricane Irma). The hurricanes, and the consequent rainfall, affected large geographic areas for several days, creating many severe states of information shortage. As a result, Twitter became a popular tool for sharing updates and organising community relief efforts. The Californian wildfires shared similar characteristics, though far fewer Tweets were produced during the collection period. The hurricane events were therefore chosen as the primary subjects of further analyses conducted in chapters 6 and 7.

5.2.2 Reflections on the Data Collection Process

Each period of data collection identified new limitations of the collection software. Fortunately, the issues encountered did not impede dataset collection nor compromise integrity, and emergent bugs were handled during the collection period by actively monitoring the process. An iterative approach to software development was adopted to enable the improvement of the data collection methods between each event: major issues discovered during each live collection were documented and addressed at the end of the activation period. In this way, the software was continually improved to better capture data during live events and was able to adapt to emergent behaviour and external environmental changes. The key findings developed during this process are discussed below and presented as contributions to the further development of social media data collection tools.

Volume of Streamed Data

During peak periods of activity, the amount of data returned by the keyword stream exceeded the ability of the system to record it. This generated a long queue of tasks that were not sustainable as the collection continued, and therefore overflow data was discarded. Requesting and recording the follower/followee network for each user proved to be the most time-consuming step in the collection process as it awaited a response from the Twitter API and then created a large number of objects to represent each follower and friend relationship. As network analysis was a core part of this research, these requests were a mandatory component of the process, requiring execution at the moment of detection to capture a valid temporal representation of a user's network. The speed at which data could be recorded was a product of both the software design and the hardware upon which it was run. As the choice of hardware was limited for this study, efforts were focused on making the software more efficient. The following interventions reduced the load on the system:

- **Keyword prioritisation:** keywords were categorised as high and low priority and run in separately filtered streams. The low-priority stream included

general keywords which were used globally and therefore had a low signal-to-noise ratio (for example, `#hurricaneharvey`). The high-priority stream included phrases that were considered much more likely to be used by people affected by the event (for example, `#HarveyCleanup`). The streams were then configured to only record data when the system had resources available, over which data from the high-priority stream took precedence. This approach handled a subset of the detected data (based upon computing power) without missing valuable data from the less common, but more 'precise', high-priority keywords.

- **Manual filtering heuristics:** During hurricane events that affected the United States, a large number of consolatory messages were observed from people around the country (and the world). A common theme of these messages involved sending 'thoughts and prayers', or calling on others to pray for the affected population. As messages containing the term 'pray' were very rarely observed to contain any useful information, these were automatically disregarded by the system. While this effectively produced a less representative dataset, it led to a significant reduction in the level of noise and therefore allowed for more precision in the recorded data. Sources that were detected to automatically post geolocated job listings were also filtered out as these had no relevance to the event yet made up a large proportion of the data. These measures were considered acceptable given that the intended purpose of the datasets was not to create representative examples of all Twitter behaviour, but to observe the behaviour of useful contributors.
- **Excluding users:** Popular user accounts would regularly appear in results due to their high message output or the size of their network. The database records for these accounts could become quite large and taxing on the system, and would rarely represent a local user sharing useful information. Rather, they were most often celebrities and politicians, news agencies, or bot accounts, and therefore considered safe to ignore. This approach was in line with previous work (Rahimi et al. 2015; Kwak et al. 2010; Davis Jr. et al. 2011). A fixed

value was chosen to filter these accounts: if the author of a streamed Tweet had over 5,000 followers or friends (usually a public figure or news outlet), or had published over 10,000 Tweets (usually a spam or automated account), their message was discarded. The details of each discarded user were recorded to allow for the evaluation of these thresholds after the recording period.

- **Reduced API calls:** API requests were slow and often exceeded the rate limit thresholds, thereby creating a backlog of requests. Measures were taken to minimise the amount of API calls necessary during the collection period.
- **Reduced database writes:** Creating a new user object for every follower and friend of a detected user placed a significant load on the hardware. The existence of these objects was not required during the data collection period, therefore this process was removed and the details for each follower/followee network were stored directly within the original user object for the duration of the collection period. User network data were turned into the relevant user objects once the collection period had ended.
- **Improved database infrastructure:** A PostgreSQL database was implemented which was able to handle faster access requests than the default Django SQLite database.

These measures improved the performance of the software and coverage of the resulting dataset. While spikes in volume still caused some data to be discarded, it was not required that all streamed data was recorded. The limitations of the publicly available stream imposed by Twitter (section 4.3.2) preclude any research requiring complete datasets and therefore those collected in this study represent samples of Twitter data. It is important to note that as more data was discarded during high-activity periods due to performance limitations, a risk of creating an artificial volume ceiling was introduced and therefore observations of patterns in datum volume were potentially compromised.

Excluding specific keywords or users created a biased dataset. This was considered acceptable for this research, which did not attempt to document a

representative sample of Twitter. Filtering data to identify a specific sub-class (in this case, the eyewitness users) often involves pre-processing steps such as discarding irrelevant sub-classes. The methodology described above was, in effect, an implementation of pre-processing. Discarding unnecessary sub-classes of data at the collection stage freed computational resources to record more instances of the targeted sub-class.

Higher volumes of data could be recorded by more efficient hardware than that used for this research. Cloud environments such as those provided by Amazon Web Services,[3] Microsoft Azure[4] or Google Cloud[5] are able to scale on demand (at a cost) and could handle the spikes in volume that occur at the onset of an event. Future data collection projects which require more comprehensive collections of data should consider using these services.

Activation

The software required manual intervention to activate and begin data collection. An operator must detect an event, observe and select appropriate keywords to track, and then interact with the system to begin recording, leading to a loss of the data published between the event occurring and system activation. This issue primarily affected sudden-onset events, where key data was created in the moments following the event (for example, an earthquake or terror attack).

These losses can be alleviated with a *backfill* function, by which the data published between the incidence of an event and the activation of the software is requested from the REST API to supplement the streamed data. The effectiveness of this approach is limited by the coverage of missing data provided given the parameters used for the archive requests and the proportion of data that remains publicly available. Additionally, the state of certain features (such as follower/followee networks) may change before they are recorded by a backfilling function. Therefore,

[3]https://aws.amazon.com/
[4]https://azure.microsoft.com/
[5]https://cloud.google.com/

the representation that is recorded in this manner effectively supplements the streamed data, but its limitations must be considered when conducting analyses.

An alternative solution to the activation challenge is the deployment of a system that is permanently online and monitors the general Twitter stream to automatically detect an event, allowing it to autonomously activate the data collection process. The requirement for the operator to remain on alert for events is thus removed, but a dedicated server and an effective method for autonomous keyword selection are introduced. Event detection is not a straightforward process and remains an active area of research (Hasan et al. 2018; Alsaedi et al. 2017; Hasan et al. 2019; Y. Han et al. 2020).

Hurricanes, which were the most commonly recorded event in this research, are slow-onset events that were not affected by this issue. The time and location of landfall was known well in advance, thus the recording period encapsulated the entire duration of the event without resorting to backfilling methods.

Keyword Selection

Keywords were selected to match as large a proportion of related data as possible whilst excluding unrelated data. Key and trending phrases used in discussions of an event were identified by observing relevant Twitter discourse. In some cases, the tracked keywords were updated during recording as new terms and phrases emerged. This process was conducted on an ad-hoc basis and did not capture all relevant terms. For example, commonly misspelt versions of popular event hashtags often trended but did not coincide with incidences of the correctly spelt terms (this is discussed further in section 6.1). The manual selection of keywords was also a source of data bias which introduced the risk of excluding population clusters (Murzintcev and C. Cheng 2017).

A more robust solution to keyword selection could systematically detect key phrases as they emerge within the discourse and update the tracked list accordingly, thus removing the need for manual intervention by the operator. Candidates for inclusion in this approach include hashtags that co-occur with already-tracked

keywords and are therefore likely to also be related to the event and keywords that are trending in, or specific to, an affected geographic region. Olteanu, Castillo, et al. (2015) proposes a method to build topic-specific lexicons from which a set of keywords can be generated based on the event class. A pre-defined lexicon eliminates the requirement for operator input before collection can begin, however, if not combined with co-occurrence detection methods, may fail to capture emergent keywords that are specific to an event.

Imperfect Temporal Data

Throughout an event which lasts several days, a user may make changes to their profile data. The software used for these collections recorded user data as it existed when the first detected message was published. Further Tweets detected from the same author were associated with the existing user data record irrespective of subsequent changes in user data. This policy reduced the load placed upon the software system and the number of required API requests, however, changes in user data may contain useful and descriptive information.

Network data for all ego users were updated at the end of the recording periods of each hurricane event to evaluate the significance of changes in user follower networks. This procedure was severely restricted by the API rate limits and therefore impossible to perform during the collection period.

Tweet and user deletion events were not intrinsically captured by the data collection software. A deletion-naive process introduces the risk of overrepresenting classes of messages which are, for example, deleted by the author for containing errors or removed by Twitter as spam material. A process to verify the continued existence of each recorded Tweet and user at the end of a data capture period was developed to address this problem and better respect the privacy rights of the authors (section 4.5.2). The capacity to perform these requests was constrained by Twitter's rate limits (section 4.3.3) and therefore became less viable for larger datasets.

Persistence of Media

While Tweet and user information were recorded to a local database and therefore protected from the effects of deletion of the source, external media to which Tweet messages linked were not preserved in this way. Media attachments comprised a large proportion of the observed messages and therefore the impact of source removal bore significant implications for later analyses.

An initial study of one dataset revealed a large number of Tweet messages which had been cross-posted from Instagram and therefore included photo and video media objects. A non-trivial proportion of these Instagram posts had since been deleted and therefore it was often impossible to interpret their original meaning. As Instagram posts often included geographic coordinates, they were a potentially valuable source of eyewitness information which could be easily detected. A feature to protect these data by creating local copies of attached media was implemented for later datasets.

5.3 Discussion

Many software solutions have been designed to collect data from Twitter and are used for both research and commercial applications. The access to Twitter data, upon which these systems rely, is provided through an API where the primary resource is the Tweet message object. These data are most typically requested by supplying a set of keywords to which message text is compared and returned where a match is found. This intuitive method of message discovery aligns with common approaches to text search and is complemented by a geographically-filtered search. The limitations of these methods, however, impose constraints upon the datasets which may be collected from Twitter and therefore the scope of analyses viable using Twitter data.

The network analysis approach proposed in chapter 4 is grounded in the concept of geoproximate homophily and the predictive power of friendship networks in locality classification. This method was therefore predicated upon a dataset that included relevant network data describing these relationships. Collecting Tweets from authors introduced into the dataset through relationship traversal exemplifies an alternative

approach to message detection not naturally supported by the Twitter API. A custom piece of data collection software was therefore required to create these datasets and support the novel analyses conducted in later chapters, demonstrating the degree to which official methods of data access can influence the research landscape.

Developing Custom Collection Software

The system designed and documented in this chapter was built to provide a methodological contribution to fields of research using Twitter data. It has been released as open-source software under the GNU GPLv3 license[6] and is structured to support extensibility for unique project requirements. Designing data collection tools for social media research requires an ongoing discussion and awareness of environmental disruption, best fostered by a collaborative and open-source approach to development.

The specifications upon which the system was built are vulnerable to modifications made to data structure or policy by Twitter, as illustrated by the release, during this work, of the second version of the Twitter API which both introduced and removed endpoints through which data could be accessed. Adapting to this type of changes typically requires only minor maintenance of the codebase and therefore benefits from encouraging third-party modification.

The CrisisData software directly facilitated the collection of the network-aware Twitter datasets used in this research. The value provided by the inclusion of network data is explored in chapters 6 and 7, which demonstrate the importance of designing custom data capture tools for research. The scope of future research using Twitter data need not be constrained by the traditional methods of data collection — the development of customised data collection software, exemplified by this work, should be considered as an integral component of Twitter data analysis.

[6] https://www.gnu.org/licenses/gpl-3.0.en.html (accessed 2022-10-01)

Considerations of Data Bias

While the datasets collected using this software documented network features not typically captured by existing data collection tools, they remain subject to a number of key limitations which shaped the scope of the analyses for which they were useful.

It remains unclear to what degree the sampling method used by the Twitter *spritzer* stream introduced data bias when sampling the complete *firehose* stream. Existing research on Twitter's sampling bias is conflicting and presents an opportunity for further research to inform studies within a wide range of domains that use the freely available data stream.

Selection bias was introduced through the definition of keywords by which the Twitter stream was filtered. Analyses that attempt to characterise a population or facilitate the administration of aid efforts should consider the implications of these biases and the risks of excluding sub-groups within the population. This was exemplified in an early data collection event for which misspellings of popular keywords were not captured, thus introducing an element of *literacy bias*.

Event Collection

The most commonly occurring events observed during this research were hurricanes (six of ten datasets recorded). The similarity of the datasets collected from these events provided a useful opportunity to evaluate the generalisability of the research outcomes without introducing confounding effects caused by hazard type. The suitability of the hurricane data is described within the disaster taxonomy framework (formulated in section 2.1.2):

Hazard Type: Of the ten datasets collected, only a single case was the product of an endogenous agent (Barcelona terror attacks). Therefore, it was not possible to determine the significance of the agent as the cause of differentiating characteristics.

Speed of Onset: As slow-onset disasters, hurricane events encouraged information-sharing behaviour and lower rates of rumour proliferation. The anticipatable nature of landfall events also allowed for a more reliable definition of data collection parameters such that early data was not lost.

Duration: While the lifespan of each hurricane was relatively short once landfall had been made, the subsequent rainfall and flooding which followed was the primary cause of danger. This sustained period of disruption led to conditions in which online discourse became an important source of public information and a focal point of community coordination. During short-term events, such as the Barcelona terror attacks, response organisations had comparatively brief periods of informational uncertainty and therefore social media discourse provided very limited informative potential.

Magnitude: The hurricanes affected large geographic areas beyond the scope of response organisations to directly monitor, thus reports from situated populations provided useful supplementary information.

Population, Culture, and Technology: The hurricanes studied during this research period made landfall in densely populated areas with high levels of technological infrastructure and literacy. Twitter discourse therefore naturally represented the affected population and contrasted with the Sierra Leone event, in which the majority of detected Twitter data originated from external observers. Whether the lack of data observed in Sierra Leone was the result of infrastructural limitations or cultural habits was unknown and demonstrated the regional limitations of this approach (though these may change as technological advancements are made).

Four of the six hurricane events were based in the United States of America, and therefore the primary language used by affected communities was English. While message content was not directly analysed in this research, the structural characteristics of a language may bear significant implications in how an online

platform is used. For example, the 280-character limit imposed by Twitter is less constrictive to orthographies that contain logographic elements (such as Japanese).

5.4 Summary

Natively supported methods of data access constrain the scope of eligible research analyses. Therefore the development of custom software should be considered when formulating research questions such that these constraints are challenged. This chapter has demonstrated the value of implementing customised data collection software for the study of social media data and provided documentation of the system developed during this research to solve the challenges of capturing Twitter user network data.

Twitter discourse was recorded during ten disaster events and used for analyses presented in chapters 6 and 7. An ongoing reflection upon the data collection events provided useful insights informing further iterations of system development and the design of similar tools in future research.

The software system developed in this chapter was designed to facilitate extensibility and forms the basis upon which the intelligence platform prototype was developed in chapter 8. The source code has been released as open-source software under the GPLv3 license and, in conjunction with the documentation provided in this chapter, comprises a key methodological contribution to the fields of research which examine social media data.

6
Quantitative Analysis

Contents

6.1 Preliminary Observations 162
6.1.1 Misinformation 162
6.1.2 Geospatial Data 163
6.1.3 Message Content 164
6.1.4 Implications to Method Design 165
6.2 Data Coding 165
6.2.1 Tweet Coding 166
6.2.2 Schema Validation and Inter-Coder Reliability 169
6.2.3 User Coding 171
6.3 Object Analysis 172
6.3.1 Recall and Precision 172
6.3.2 Comparison of Codes and Ground Truth Data 175
6.3.3 Tweet Sources 181
6.4 Discussion 186
6.5 Summary 188

In this chapter, Twitter data from two Hurricane event datasets were analysed with a view to making them more useful as a resource with respect to the needs identified in chapter 3. The primary outcome of this study was to characterise online discourse and evaluate the extent to which it provided value as intelligence to disaster response organisations, framed in RQ_2 as:

> RQ_2 —How can publicly available social media data provide meaningful intelligence to disaster response organisations during disaster events?

A preliminary observation was conducted to develop a base understanding of the data, from which several key insights were made. A sample of the data was then coded in two dimensions: first, based on message content, then with respect to the locality of the author to the event. These analyses provided novel insights into communal behaviour on social media platforms during disaster events and highlighted characteristics that inform the design of future studies based on similar datasets (contribution C-1).

Quantitative measures verified the preliminary observations and identified key features useful for algorithmic classification. Two methods for locality inference were proposed and evaluated using the coded locality values: geocoding values of the user profile location field and the detecting geotagged Tweets in the user's historical timeline. The composite of these predictors was shown to provide accuracy sufficient to be used on the uncoded set of data for locality inference (contribution C-4).

An examination of the *source* value of Tweet metadata established a correlation between the software from which a Tweet is published and the class of message it represents. This finding supports the integration of Tweet source evaluation with traditional methods for location inference and message classification.

6.1 Preliminary Observations

Analysis of the Twitter data began with an inductive, data-led examination of a randomly selected sample of 5,000 Tweets collected during the hurricane events of 2017 (see table 5.1). This unstructured examination identified key behavioural patterns in the online discourse and provided context for further structured studies conducted in section 6.3 and chapter 7. A selection of key observations is presented here for illustrative purposes and as a substantive contribution of this chapter to the models of understanding of Twitter behaviour.

6.1.1 Misinformation

Misinformation within the data was loosely categorised into three main classes. The most common form of misinformation observed was the result of user error or

naiveté wherein false information was mistakenly reported and spread by innocently-motivated behaviour. These messages contributed to the rumour-mongering effect identified by disaster practitioners as an area of focus (3-V) in chapter 3. *Malicious* material defined messages created to advance an agenda or viewpoint — for example, falsified reports of looting were commonly used to support a particular political stance. Finally, instances of *vandalism*[1] were often shared for humorous or irreverent reasons and included a well-known recurrent falsified image representing a shark swimming on a flooded highway (figure 6.1).



Figure 6.1: Misinformation — a shark swims on a flooded highway in a falsified photo which first appeared in 2011 during Hurricane Irene (Peschak 2007; Haley 2021).

6.1.2 Geospatial Data

Geotagged Tweets were infrequent and most commonly generated automatically either by advertising accounts (e.g. recruiting or real estate) or as *cross-posts* from Instagram. Cross-posting is automated behaviour in which messages published to another platform are automatically posted to Twitter on behalf of the

[1]The term 'vandalism' is used here to differentiate between two classes of intentionally-spread misinformation: '(malicious) deception' is misinformation created to achieve a specific goal for a stakeholder, whereas 'vandalism' captures a class of misinformation in which the author attempts to spread a rumour motivated by a sense of 'mischief'. A more detailed discussion of these distinctions is not relevant for the purposes of this study but is explored in L. Wu et al. (2019)

user. The scarcity of geotagged messages aligned with the findings of previous work: that disproportionate representation of automated messages within the geotagged streams further diminished the potential for this class of data to provide meaningful intelligence.

In some cases, people who were not in the affected region chose to act in ways that appeared as if they were, apparently as a show of support. This behaviour manifested as users setting their profile location to (for example) Houston, tagging messages with Houston's coordinates, or posting photos from an earlier time in which they were in the city. These messages introduced noise that degraded the validity of geographic features, though the degree to which data is corrupted by this behaviour requires further study.

6.1.3 Message Content

Messages of condolence were generally more common than any other class of message. In the Hurricane Harvey discourse, the terms to send one's 'thoughts and prayers', or to 'pray for Texas' appeared often. Messages using event hashtags for pure self-promotion were rare (i.e. using any trending tag without regard for its meaning). The majority of this behaviour may have been excluded automatically given that the data collection process disregarded high-activity users.

Misspelling was often observed, and common misspellings began to trend within the Twitter community. While `#HurricaneHarvey` was a dominant hashtag, 'Hurricane' was commonly misspelled as 'Hurrican', 'Huricane', or 'Hurricaine', leading to trending tags such as `#HurricanHarvey`, which were not anticipated by the researcher nor added to the tracked tag list. The systemic pattern in these specific versions of misspellings may be grounded in cultural or socio-economic norms present in the southern U.S.A., though a deeper understanding of this phenomenon requires further research. Where systemic patterns exist, a system that does not detect and track common misspellings may introduce into the dataset a participant bias, thereby diminishing its *representativeness*.

During the hurricane events, individuals were commonly observed posting images of the conditions at their location. While coordinates were rarely included, posts were often contextualised by text regarding their location (e.g. 'The view from my apartment downtown: [photo]'). These messages exemplified material supporting the conceptualisation of Twitter as a geographically distributed sensor network (Crooks et al. 2013) and motivated an author-centric approach to classification.

6.1.4 Implications to Method Design

This analysis verified the presence of message classes that presented value to disaster response organisations within online discourse such as eyewitness reports and images. The effects of misinformation were found to be within acceptable bounds such that issues of data validity were minimised. The low informative value associated with messages containing geographic metadata limited the effectiveness of methods using native geospatial features. The focus of the subsequent analyses was therefore placed upon methods of author location inference based on message content and author features.

6.2 Data Coding

Data from Hurricane Harvey were chosen for an initial round of Tweet content coding. The purpose of this approach was to develop a robust coding schema that captured all data while maintaining the perspective of 'useful information' discussed in chapter 3. The schema was developed to distinguish between the classes of content which were identified as 'useful information' and those which were not, as this distinction supported the requirements of the response organisations. Additionally, the coding schema provided insight into the behaviour of online communities so that they may be better understood by response organisations. Therefore, while this study took an author-centric approach to analysis and classification, the coding of individual Tweets was desirable in order to classify these behaviours.

6.2.1 Tweet Coding

The final schema used for Tweet content coding was based on the findings of the requirements analysis from section 3.3 which identified key objectives for disaster response organisations using social media data, including 'identifying urgent needs' and 'developing situational awareness'. The schema was then developed iteratively through an evaluation of messages from both the location and keyword streams. A general understanding of the content was developed through manual inspection of the data from which a draft schema was developed. This schema was then applied to a random selection of Tweets to test its comprehensiveness and adjusted as needed.

The schema was designed to be both exhaustive and exclusive (Stowe et al. 2018), however, due to the author-centric goal of this research, distinguishing between various categories of 'unuseful' messages was of lower importance. The primary goal of this study was to differentiate between useful and unuseful authors; a distinction which may be redefined as *local* and *non-local* users (see chapter 3). A Tweet which may be considered independently unuseful, yet suggests that the author is local to the affected region, may be a useful determinant when identifying the locality of the author. Therefore, classes distinguished between 'local' and 'non-local' where possible.

This first round of coding produced the set of classes in the following list. The definitions were chosen with consideration to the end goal of the system, which was to identify local users and those who are a useful source of information. The classes were considered in hierarchical order. Therefore, where a message satisfied multiple codes, it was assigned to the code highest on the list. This approach allowed the schema to prioritise higher-value messages without requiring mutually exclusive classes or multiple codes per message.

Aid Request: A request for immediate aid directed either towards the authorities or other users. Messages which referenced a request for aid (e.g. through a retweet) were included here, as they serve the same purpose from the perspective

of a responder. Requests for donations were not included as they were considered neither urgent nor actionable by responders.

Ground Truth: A report which gives the reader insight into the conditions within the affected area. This was most often achieved through the sharing of photos and video, though text and other sources were also valid. Users that posted this kind of content were assumed to be at the site of the event, therefore messages that linked to external sources were not relevant, however, messages which retweet or quote another Tweet were valid.

Information for Local Users: Information posted by any user which was directed towards people in the affected area. This included messages which provided updates on conditions (e.g. road closures) or pointed to external resources (e.g. a link to FEMA evacuation protocol).

Note: while a weather update would fall under this definition, it was decided after coding had begun to exclude these messages as they constituted a high volume of the data, the content of which could be easily found without using Twitter. Weather updates warrant their own class in subsequent schema iterations, however for this dataset were classified into the next class.

Information for Non-Local Users: Information that was not directed specifically to those in the affected area. This class encompassed a wide range of messages which included links to donation drives, political discussion (related to the event) and statistical reports.

Thoughts and Emotions from Local Users: This class of message was exemplified by expressions of humour, grief, or mourning; general discussion; and other content that was not considered informative. The judgement on the locality of a user was based on the content of the message or the information on their Twitter profile.

Thoughts and Emotions from Non-Local Users: As above, where the location of the author was not clear. This class therefore could include local users who were not identified as such.

Unrelated to Event: Spam and automated messages, Tweets that used the hashtag for exposure but were not discussing the event, and unrelated use of the keyword or hashtag.

2,500 messages were coded using the schema, the results of which are presented in table 6.1 and figure 6.2. The table shows the percentage of messages in each category, with a total of 16.4% split between the two 'high-value' categories of 'aid request' and 'ground truth'. While it is important to remember this dataset is not entirely representative due to the filtering techniques implemented in chapter 5, it demonstrates that 'ground truth' messages form a large enough proportion of the dataset to be valid candidates for automatic detection.

Code	Tweets	Proportion
Aid Request	35	1.4%
Ground Truth	375	15.0%
Information for Locals	144	5.8%
Information for Non-Locals	577	23.1%
Thoughts & Emotion - Local	202	8.1%
Thoughts & Emotion - Non-Local	603	24.1%
Unrelated	564	22.6%
	2500	100.0%

Table 6.1: Tweets by code — Hurricane Harvey

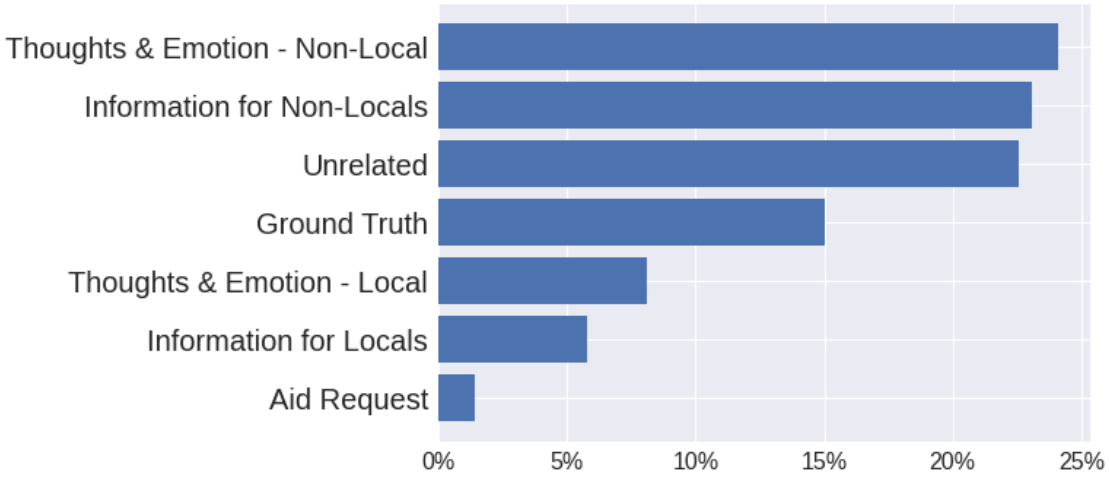


Figure 6.2: Tweet codes as proportions of coded data — Hurricane Harvey

6.2.2 Schema Validation and Inter-Coder Reliability

The coding schema was validated by introducing a secondary coder who evaluated 250 Tweets drawn from the coded set. Agreement between coders was then calculated on a per-code basis to identify cases of systemic disagreement. The proportion of observed agreement, p_o, corresponds to the cases in which the coders selected the same category for a given observation. For a contingency table $\mathbf{X}$ with k categories and N total observations, p_o may be defined as:

$$p_o = \frac{1}{N} \sum_{i=1}^{k} x_{ii} \tag{6.1}$$

Agreement (p_o) for the set of 250 co-coded Tweets was calculated as 0.720. The contingency table is shown in table 6.2. While the individual disagreement percentages are useful indicators of where systemic issues may lie with the coding schema, the categories were imbalanced and therefore the cell values should be considered with respect to the total Tweets for each column. More robust indicators of systemic disagreement are observed through row- and column-wise evaluation. For example, table 6.2 shows that disagreements for the 'Unrelated' category appear to be distributed relatively uniformly, whereas within the messages the secondary coder classified as 'Information for Non-Locals', the majority of disagreements (8) were classified as 'Thoughts & Emotions — Non-Local' by the primary coder. This suggests a systemic confusion regarding the distinction between what constitutes 'Information' or 'Thoughts and Emotions', the cause of which may be a lack of clarity in the definition provided to the coders, a fundamental difference in the coders' interpretation of what is (for example) 'informative', or a mismatch between the schema's distinction and whether such a distinction exists clearly within the data.

		Secondary Coder								
		Aid Request	Ground Truth	Information (Local)	Information (Non-Local)	Thoughts & Emotions (Local)	Thoughts & Emotions (Non-Local)	Unrelated	Total	Disagreement
Primary Coder	Aid Request	5	0	1	0	0	0	0	6	16.7%
	Ground Truth	0	33	0	2	0	2	1	38	13.2%
	Information (Local)	0	0	13	0	0	0	0	13	0.0%
	Information (Non-Local)	0	7	7	40	1	2	1	58	31.0%
	Thoughts & Emotions (Local)	0	3	0	1	8	3	1	16	50.0%
	Thoughts & Emotions (Non-Local)	1	8	2	7	8	47	0	73	35.6%
	Unrelated	0	3	2	2	3	2	34	46	26.1%
	Total	6	54	25	52	20	56	37		
	Disagreement	16.7%	38.9%	48.0%	23.1%	60.0%	16.1%	8.1%		

Table 6.2: Tweet code contingency matrix

Interpreting the inter-coder reliability, and therefore the validity of the coding schema, requires a metric based on the proportion of observed agreement corrected for the chance of random agreement. For two raters, the agreement coefficient can be defined in the general form:

$$AgreementCoefficient = \frac{p_o - p_e}{1 - p_e}, \tag{6.2}$$

where p_o is the observed agreement and p_e is the agreement expected by chance. A common metric suitable for nominal multi-class data coded by two coders is Krippendorff's alpha, α. This agreement coefficient was preferred over the commonly-used Cohen's kappa, κ, due to the latter's assumption of coder independence in its definition of chance agreement. Krippendorff (2004) establishes that κ's expected disagreement p_e is a function of the coders' individual preferences rather than a function of the estimated proportions from the sample data and is inflated by a variance between the marginal distributions of the coders. This is documented in Feinstein and Cicchetti (1990) as the second of two paradoxes, which demonstrates that with p_o held constant, κ will be higher with an asymmetrical rather than symmetrical imbalance in marginal totals. In contrast, α defines p_e based on the estimated proportions from the sample data and is therefore unaffected by the marginal distributions of an individual coder.

For k categories, N observations and n_k the combined number of instances in which category k was assigned by either coder, Krippendorff's p_e is defined as:

$$p_e = \sum_k \frac{n_k(n_k - 1)}{2N(2N - 1)} \tag{6.3}$$

The α for the coded Tweet dataset was 0.656 (observed agreement = 0.720). According to Krippendorff (2004), values greater than 0.667 may be considered useful for drawing tentative conclusions, however, the interpretation of the alpha depends upon the application for which it is used. In this case, as a general proof of concept, α was considered high enough to accept the coding schema as sufficiently defined.[2] The primary conflicts between coders were noted for refinement in the coding of future datasets.

6.2.3 User Coding

While there was a range of user types identifiable within the data, for the purposes of training an algorithm to classify the locality of a user, a simple binary classification was sufficient. Therefore a schema of 'local' and 'non-local' was used with the addition of a third code of 'unsure'. The coding process considered information from the user's Twitter profile as it existed at the time of detection, their current profile (where it was still accessible), and the content of their Twitter stream during the period of the event. This included both messages that were detected by the data collection system and those that were collected at the end of the event (i.e. messages authored during the collection period but not initially detected by the software).

1,500 users from the Hurricane Harvey dataset were coded by the primary coder (see table 6.3). The candidates for coding were randomly drawn from the pool of authors for whom at least one Tweet had already been coded. In this way, statistical analyses were able to investigate the relationship between the distribution of Tweet categories and the locality of their authors. A subset of 200 users was coded by

[2]For comparison, Cohen's κ (J. Cohen 1960) was calculated as 0.657, indicating 'substantial agreement' (Landis and Koch 1977)

a secondary coder and α calculated as 0.674 (observed agreement = 0.835). This which was deemed sufficient to accept the classification method as robust.[3]

Code	Users	Proportion
Local	385	25.7%
Non-Local	1084	72.2%
Unsure	31	2.1%
	1500	100.0%

Table 6.3: Users by code — Hurricane Harvey

6.3 Object Analysis

The data collection process described in chapter 5 monitored Twitter streams filtered using two methods: the keyword stream returned Tweets containing any string from a predefined list and the geographic stream comprised both geotagged Tweets for which the coordinates were within the bounding box defined for the event and Tweets with an associated 'Place' object whose bounds overlapped the bounding box. For each author detected during an event, a timeline of Tweets was recorded to provide context to the Tweets which were detected within the streams. While the sampling method used for this larger set of Tweets is therefore based upon the Tweet authors' detection within the monitored streams, it was used in this section as an approximation of the general population for the purposes of analytic comparison.

6.3.1 Recall and Precision

The metrics chosen to measure and evaluate the performance of a classification algorithm are considered with respect to the overall purpose of the algorithm. In the previous section, Krippendorff's kappa was used to measure agreement between coders. Kappa is a distribution-adjusted measure of *accuracy*, and as such it weights all errors equally. In binary classification, more descriptive statistics are used which distinguish between type I (*false positive*) and type II (*false negative*) errors. The confusion matrix presented in table 6.4 illustrates this distinction.

[3] $\kappa = 0.675$, indicating 'substantial agreement'

		True Condition	
		Negative	Positive
Predicted Condition	Negative	True Negative (TN)	False Negative (FN)
	Positive	False Positive (FP)	True Positive (TP)

Table 6.4: Type I and II errors

Recall (or *sensitivity*) is a measure of how well the classifier captures all positive cases, or *the proportion of all true positive cases which were successfully classified as positive*. Formally, it is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{6.4}$$

Naturally, a naive algorithm maximising for recall alone may simply classify all cases as positive, thus achieving a recall of 100%. Performance is therefore measured by using recall in conjunction with a second metric, *precision*, which is a measure of how accurate the positively-classified cases are. That is, *the proportion of positively-classified cases which are true positives*:

$$Precision = \frac{TP}{TP + FP} \tag{6.5}$$

Precision rewards highly selective algorithms and thus balances the recall measure: the naive positive-only classifier which scores a recall of 100% performs poorly on the precision measure. An aggregation of the two metrics may therefore be used as an effective performance measure, known as the *F-score*. The balanced F_1 *score* is the harmonic mean of precision and recall:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{6.6}$$

The relative importance of precision and recall will depend upon the purpose of the classifier and the characteristics of the data. Chapter 3 identified a key challenge faced by disaster response organisations as the management of the high

volume of social media data with a poor *signal-to-noise ratio* (challenge 6-C in section 3.3.2). Curating a sub-sample of these data with a higher incidence of relevant content is characterised by an increase in precision. Where data are inexhaustible, recall approaches irrelevancy. That is, given that the value of the output stream is constrained by the human operators' ability to interpret the data subset in real-time, there is no practical cost to discarding positive cases (that is, reducing recall) once this capacity has been met. Applying differential weighting to recall and precision is modelled in the more general F-score, F_β, in which recall is considered β times as important as precision.

Optimising precision increases the relevancy of the selected subset of data whilst reducing its size. It is desirable for a curation classifier using a 'human-in-the-loop' model to maximise precision, subject to the constraint that the resulting subset of data is not less than the capacity of the human operators to interpret. This concept is illustrated in figure 6.3: as precision is optimised, the size of the subset of data decreases; as recall is optimised, the size of the subset increases. The equilibrium weighting at w_e represents the point at which the subset exactly matches the capacity of the human operators to process. To the left of this point, the classifier has optimised for precision such that the resulting subset is smaller than the processing capacity of the human operators, therefore failing to maximise throughput. To the right, the subset is larger than can be processed by the operators. Thus while a higher recall is temporarily achieved in configurations right of w_e, the additional data are discarded during the human processing stage (represented by the red area).

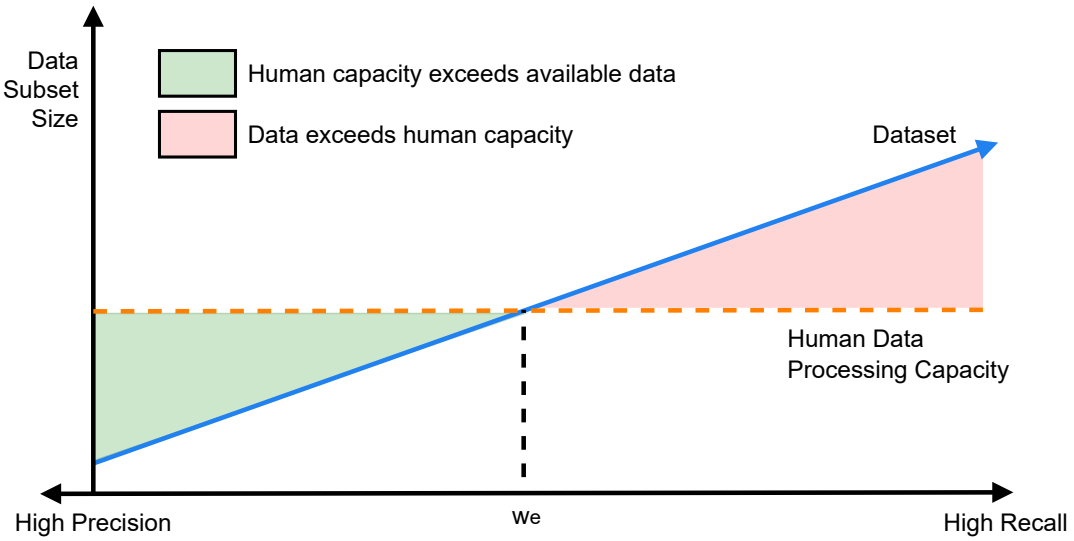


Figure 6.3: Precision and recall equilibrium

The equilibrium point w_e is therefore a function of the characteristics of the data stream, the performance of the selected classification model, and the processing capacity of the human operators. Naturally, all three of these factors will vary between events and organisations and therefore selecting a generalisable F_β is not possible. The performances of the classification algorithms used in this research are therefore discussed in relative terms where comparisons of both precision and recall provide meaningful context to the work.

6.3.2 Comparison of Codes and Ground Truth Data

The coding schema discussed in the previous section was structured around defining the locality of the user. This characteristic was chosen as a priority based upon the assumption that a local user was more likely than others to be sharing useful information and therefore worthy of observation by disaster responders. In this section, the coded locality of a user is compared with metadata provided by Twitter to determine the extent to which these metadata may predict user locality.

Profile Location

The metadata provided by Twitter with each user or Tweet object include geospatial features from which locality may be inferred, the most visible of which is the location field of a user's profile. This optional field is represented as a string of characters with which the user may choose to define their location in a format of their choosing (see, for example, table 6.7). Therefore, the value of the field does not conform to a standardised format, nor is it constrained to geospatial values.

Of the total Hurricane Harvey dataset of 31,932 users, 25,619 had a non-null value in their location field (80.2%). For the coded subset of 1,500 users, this proportion was 79.4% (1,191). To compare this field with the binary-coded locality values, the location strings were first parsed for coordinates using a regular expression[4] and where found, reformatted to decimal degrees. The remaining strings were converted to geospatial coordinates using the geocoding functionality of the Google Maps API.[5] Unparseable strings were discarded and the derived coordinates were tested for whether they fell within the geographic bounding box defined for the event. The results of this test are presented as a confusion matrix in table 6.5.

		True Condition	
		Non-Local	Local
Predicted Condition	Non-Local	662	66
	Local	138	249

Table 6.5: Profile location confusion matrix

The agreement between the metadatum and true condition (as coded) was 0.817, with an α of 0.576 and F_1 score[6] of 0.709. This suggested that while an association appeared to exist, it was not strong enough to represent the true condition. Furthermore, integrating the remaining 20.6% of profiles that did not contain location data would have further degraded the agreement rating. For

[4] `reg_ex = "[nsewNSEW]?\s?-?\d+[\.°]\s?\d+°?\s?[nsewNSEW]?"`

[5] https://developers.google.com/maps (accessed 2021-04-29)

[6] An F-score is a measure of accuracy used in binary classification. The F_1 score is the harmonic mean of *precision* and *recall* and is discussed in the previous section.

example, by classifying all null values as non-local, a proportion of local profiles would be discarded (lowering *recall*). Alternatively, including all null values in the locally-classified set would dilute the positive classifications with false positives (lowering *precision*). The impact of these strategies was based on the distribution within the null-value set and is illustrated in table 6.6.

	Excl. Null Values	Null as Non-Local	Null as Local
Precision	0.643	0.643	0.427
Recall	0.790	0.647	0.771
F_1 Score	0.709	0.645	0.549

Table 6.6: Profile location as locality predictor

These low rates of agreement may have been caused by a number of factors. Firstly, the location field typically represents a user's home (or work) location, which does not necessarily align with the location from which they create their Tweets. Furthermore, the rate at which users update this field may not match how often their location changes; an effect that is amplified during the mass displacement caused by a disaster event. Victims forced to evacuate or responders temporarily moving into an affected area are not likely to update their Twitter profiles, leading to type I (false positive) and type II (false negative) errors, respectively.

Secondly, as a user-set field, there is no form of verification. Users may set whatever location they choose, or indeed, any string of characters (Murthy and Longwell 2013). A selection of observed strings is provided in table 6.7 along with the results of the geocoding test.[7] The outcome of the geocoding relies heavily upon the interpretive power of the geocoding API; for example, the string *'The Land of Sugar'* is (correctly) resolved as *'Sugar Land, Texas'* by the Google Maps API. The examples with a null geocoding value demonstrate instances of the field being used for non-specific or unparseable locations (*'Chicago born: LA living'*), personal expression (*'Working'*), and unrelated information (*'REDACTED@gmail.com'*). Of the 1,191 profiles with non-null location values, 76 (6.4%) returned a null geocoding

[7]Examples containing precise locations such as addresses or coordinates have been synthesised based on observed data to preserve anonymity.

result. From the total of 1,500 users, 74.3% returned non-null geocoded results based on their profile location field, suggesting that geocoding with Google API is a viable approach for profile location parsing and therefore not a significant contributing factor towards the disagreement shown in 6.6.

Location String	Geocoding Result (Hurricane Harvey)
Houston, TX	Positive
12695 SW Freeway Houston, TX	Positive
N 30°28' 40" / w 95°36' 0"	Positive
The Land of Sugar	Positive
Brecksville, Ohio	Negative
U.S.	Negative
Venezuela	Negative
Chicago born: LA living	Null
Working	Null
Somewhere over the rainbow...	Null
REDACTED@gmail.com	Null

Table 6.7: Example profile location strings

Geotagged Tweets

Geotagged Tweet messages include geospatial data in the form of either precise geographic coordinates or reference to a *Place* object. Coordinates are typically read from the GPS sensor of the publishing mobile device. Place objects are selected by the author and represented as four coordinate points defining a bounding box region which can range in granularity from a single building to an entire state or country. The author of a Tweet can select from a number of Places determined to be near their current location, ranging as far as to encompass neighbouring towns. Third-party Twitter platforms can artificially set coordinates for a Tweet, or select a Place that is not within the expected radius of the author, and therefore the veracity of these data cannot be guaranteed. Tweets geotagged with coordinates comprised 1.6% of the data collected from keyword streams during this research. This aligns

with observations from Twitter[8] and other research which note that only 0-2% of Tweets are geotagged (Leetaru 2019b; Morstatter et al. 2013; Laylavi et al. 2016; Gu et al. 2016; Kogan et al. 2015; Starbird, Muzny, et al. 2012; Z. Cheng et al. 2010).

To evaluate the validity of author geocoding using geotagged Tweets, a historical Twitter feed for each detected user was recorded. The requested feed encapsulated the same period as the initial data collection. This process was conducted at the end of the recording period by requesting historical data using the Twitter REST API and thus did not include any deleted or removed messages. The feed of each of the 1,500 coded users was then programmatically inspected for geotagged Tweets. Where found, the coordinates were tested for whether they fell within the bounding box defined for the event. A user was classified as having Tweeted from the local region if any of their geotagged Tweets fell within the bounding box. 467 of 1,500 (31.1%) users were positively classified. This binary user classification was then compared to the coded locality values, the results of which are shown in table 6.8.

	Tweet from Local Region
Precision	0.535
Recall	0.649
F_1 Score	0.587

Table 6.8: Tweet from local region as locality predictor

Tweet Content

A third method of Twitter location inference examines the content of Tweet messages and extracts geographic references. These may be in the form of gazetteer terms[9] or 'location-indicative words' which may then be geocoded using a spatial database or API as described above (see Laylavi et al. (2016)). An early exploration of these methods found them to be ineffective during disaster events due to the increased focus on the affected region from outside observers. The online discussions

[8] https://web.archive.org/web/20210615125858/https://developer.twitter.com/en/docs/tutorials/tweet-geo-metadata (accessed 2021-06-15)

[9] A gazetteer is an entity dictionary used for location recognition.

surrounding the event naturally included local geographic terms, thus rendering uninformative their prevalence within an author's posts.

Multivariate Classification

While the Tweet location classifier is outperformed by the 'Null as Non-Local' profile location classifier (table 6.6), the two positively-classified subsets each represent a different distribution of the true cases. Therefore, the intersection ($\cap$) of these sets (that is, the subset for which both classifiers return positive values; an AND condition) represents a more strictly-defined subset with a higher precision and lower recall than either individual set. While a more precise subset may be desirable, selecting for cases that satisfy both conditions may inadvertently select for a particular (as-yet-unidentified) sub-category of user type and exclude other categories of value. The union ($\cup$) of the sets (the OR condition) has a higher recall than either set and a precision falling between the two. The performance metrics of the two composite sets are shown in table 6.9 alongside the metrics of their component sets. The optimal 'naive classifier' presented by these options is a function of the volume of incoming data and the capacity of the human operators, as discussed in section 6.3.1. The 'Local Profile' classifier performs well on both precision (0.643) and recall (0.647), while the union set captures a much larger proportion of positive cases (0.839) at a cost to precision (0.555). These figures, therefore, provide a suitable baseline from which more sophisticated measures may be compared.

	Local Profile	Local Tweet	Local Profile ∩ Local Tweet	Local Profile ∪ Local Tweet
Precision	0.643	0.535	0.647	0.555
Recall	0.647	0.649	0.457	0.839
F_1 Score	0.645	0.587	0.536	0.668

Table 6.9: Comparison of composite predictors

6.3.3 Tweet Sources

An important distinguishing metadatum of a Tweet is the *source* field, which represents the platform from which the Tweet was published. When creating a third-party application that can interact with the Twitter API, a developer must provide a descriptor string that populates this field. Because many third party applications are designed for specific use cases, this field provides useful information which characterises the motivations for conditions under which the Tweet was created. For example, the source `TweetMyJOBS` refers to a recruitment platform and thus is attached to Tweets advertising job listings. Of the entire dataset of 1,727,438 Tweets, those published by first-party Twitter clients comprised 78.5% (1,355,569). The ten most frequently occurring sources are presented in table 6.10.

Proportion	Source	Description
31.0%	Twitter for iPhone	first-party
20.2%	Twitter for Android	first-party
19.9%	Twitter Web Client	first-party
5.5%	IFTTT	automated process, crossposting
2.7%	Twitter for iPad	first-party
2.5%	Twitter Lite	first-party
2.4%	Instagram	third-party crossposting
1.9%	TweetDeck	first-party
1.7%	Facebook	third-party crossposting
1.6%	Hootsuite	social media manager

Table 6.10: Top 10 Tweet sources

In contrast, of the subset of 15,569 Tweets collected based on their location within the event's bounding box, only 1.7% (259) were published from first-party apps. Tweets *crossposted* by Instagram comprised 76.6% (11,922) of the geotagged Tweets. Crossposting describes the process into which a user can opt such that messages they publish to other social media platforms are automatically posted to Twitter. The

high frequency of Instagram posts in the geographic stream therefore suggests that Instagram posts are much more likely than Tweets to include geographic data, which is preserved during the crossposting process. Excluding those collected from the geographic stream, 30,123 Tweets posted from Instagram were recorded, of which 32.6% (9,818) were geotagged. As Instagram posts are traditionally based upon the publication of a recently-taken photo, the high incidence of geotagging makes this class of message highly useful in supporting the development of situational awareness (2-V in section 3.3.1).

The distribution of sources from which geotagged Tweets were published deviated significantly from the general population. While Instagram posts comprised the clear majority (76.6%), the second most frequent source was a recruitment platform from which 12.5% (1,948) of messages were posted. These job listings typically included coordinates based on the location of the role for which they were advertising (92.6% of all Tweets from this source were geotagged). The coordinates are provided by the source software and therefore do not reliably represent the location of the author, nor are these messages informative during a disaster. Indeed, new job listings were observed within Houston during the peak of disruption as the hurricane made landfall, suggesting an automated publication process. The ten most frequently observed sources within the geographic stream are listed in table 6.11.

Proportion	Source	Description
76.6%	Instagram	third-party crossposting
6.6%	TweetMyJOBS	job listings
6.0%	SafeTweet by TweetMyJOBS	job listings
3.9%	BubbleLife	social media manager
1.7%	Foursquare	check-in app, crossposting
1.5%	Untappd	check-in app, crossposting
0.8%	Twitter for Android	first-party
0.7%	Hootsuite	social media manager
0.6%	Twitter for iPhone	first-party
0.4%	circlepix	real estate listings

Table 6.11: Top 10 Tweet sources — geographic stream

For each source observed within either stream where $n \geq 10$, a random sample of (up to) 50 Tweets was inspected and categorised to develop a taxonomy describing source use cases. Early filtering for these classes can eliminate Tweets from sources known to produce automated or otherwise unuseful content prior to more sophisticated classification. The categories defined for this purpose are shown in table 6.12. Each sample was then manually inspected to characterise the class of information produced by the source. The purpose of this analysis was to identify sources that did not generate content considered informative to disaster response. The qualitative classification was then validated quantitatively by taking the intersection of the sets of Tweets associated with each source and the 2,500 coded Tweets.

The 'relevancy' of each source subset was measured using two metrics. The proportion of 'high-value'[10] Tweets within a subset and the proportion of the total 'high-value' Tweets for which a given source is responsible. These formulae are equivalent to precision and recall (section 6.3.1) and therefore this terminology is preserved. Values exceeding a *relevancy threshold* of 5% in either measure were

[10]That is, coded as 'Aid Request' or 'Ground Truth'

used to classify a source as 'relevant'. Where $n < 10$, the qualitative determination was used. The aggregated measures for source classes are included in table 6.12.

First-party app sources represent 'standard' Twitter activity, of which 11.3% was classified as relevant, comprising 32.4% of the total positive cases. This was exceeded by *crossposting apps* ($pr = 0.318, r = 0.649$), though this category was primarily made up of Instagram posts (734 of 837, $pr = 0.338, r = 0.605$). Instagram posts accept open-ended text input in a similar manner to Twitter and require the author to include a photo. While cases were observed in which text was rendered as an image to meet this requirement, the majority of Instagram posts included photos taken by the author and thus had a higher rate of ground truth content.

Of the remaining sources in the class, relevancy varied based on the purpose of the source apps and therefore filtering decisions must be made on an individual basis. *Social media suites* are designed for use by organisations rather than individuals, though an exception was observed in *Hootsuite* ($pr = 0.100, r = 0.015$), where qualitative inspection revealed the majority of posts originated from individual users. Collectively, the remaining Social media suite sources did not meet the relevancy threshold ($pr = 0.038, r = 0.010$). Data tagged with a source categorised as an *automated app* or *spam* were not seen to contain useful information authored by individual users (0 of 204 and 0 of 87 respectively) and were therefore valid candidates for elimination. The complete results are included in appendix section C.1.

Source Classification	Description	Precision	Recall
First Party Apps	Applications developed by Twitter — these represented the most typical sources used to interact with the platform.	0.11	0.32
Third Party Apps	Applications developed by third party developers to interact with Twitter in a similar fashion to first-party apps.	Insufficient data (n=6)	
Crossposting Apps	Other social media applications with a function to automatically crosspost user content to Twitter.	0.32	0.65
Social Media Suites	Software designed to manage multiple social media accounts for businesses. Typically used by organisations to post news articles or public relations content.	0.06	0.02
Automated Apps	Applications designed to automatically post content such as periodic weather reports.	0.0	0.0
Spam	A secondary classification of automated apps that were designed to generate spam posts. These often included trending hashtags to increase their exposure.	0.0	0.0

Table 6.12: Twitter source taxonomy

Messages from sources failing to meet the *relevancy threshold* comprised 18.9% (2,941 of 15,569) of the geographic stream and 10.3% (3,222 of 31,303) of the keyword stream. By removing these sources, the precision of the coded datasets was improved at a small cost to recall (see table 6.13). The effect was more pronounced for the geographic stream, where a larger proportion of automated messages were observed. Source filtering was therefore a productive approach to curation, though the degree to which the individual source classifications of this research generalise to future events remains to be tested. Additionally, the high incidence of useful content in Tweets crossposted from Instagram encourages further research directly observing Instagram streams.

Stream	Base Precision	Filtered Precision	Filtered Recall
Keyword	0.110	0.120	0.982
Geographic	0.246	0.336	0.992
Combined	0.164	0.194	0.988

Table 6.13: Source filtered relevancy measures

6.4 Discussion

This chapter examined patterns of behaviour in social media discourse to evaluate the extent to which they could be used as a source of intelligence in disaster response. The preliminary analysis documented key characteristics of the data which revealed the informative potential demonstrated in messages published on Twitter.

Misinformation was perceived to be a significant threat to the validity of social media data, however, messages containing misinformation (to the extent that this was able to be determined) were not commonly observed. Furthermore, the eyewitness reports which made up a considerable proportion of the data often comprised photo or video material. Not only were these media formats considered highly informative to disaster response practitioners (see chapter 3), they are more difficult to falsify or misinterpret (as the interpretation of the viewer is not required, in contrast to text reports).

There were, however, observed cases of image-based misinformation: the shark photo shown above is an illustrative example of a falsified image that was spread virally, though cases were also observed in which unfalsified media originating from unrelated disaster events were presented as occurring in the current event (for example, a particular video of a building collapsing into flood waters was observed in four of the datasets collected in this research). This imagery spreads quickly due to its evocative nature, however experienced disaster response operators quickly learn to recognise and disregard the recurrent material.

Location Inference

The location inference approaches developed in this chapter were based on two characteristics of author data. First, the profile location field, which is an optional open text field in which the user can input any text value, was parsed for informative data. A high proportion of users supplied non-frivolous values to the field, though the granularity of results ranged from country to exact home address. The most common case was the name of the user's home city, which was sufficient for locality classification. Other values were queried using the Google Maps geocoding API, which was effective at converting the various formats of location to coordinate data which could then be tested for locality status based on an operator-defined geographic bounding box.

A key limitation of this approach lies in the pricing model imposed by Google (or equivalent geocoding provider). While each request (at the time of the study) cost only $0.005 USD, where large volumes of data are processed high charges may be quickly incurred.

The second method of inference examined the historical *timeline* of an author to detect any instance in which a Tweet was published with coordinate data. The rationale for this approach was based on the observation that authors selectively chose to include geographic data with published Tweets based on the purpose of the message (for example, a Twitter *Place* object may be included when the author is visiting an interesting location). Therefore, though geospatial data may not be included with Tweets detected in the stream, they may be observed in previous messages published by the author.

While both methods of locality inference were found to be effective, the primary limitation of the approach of home location inference for eyewitness detection is the substantive misalignment between the current and home locations of a person. This issue is amplified during disaster events — people who are displaced present cases of false positives (in terms of their current locality as classified using home location) and people who have moved into the affected area to administer aid are not positively classified (i.e. false negative). The extent to which this effect degrades the

performance of this approach is dependent upon the unique characteristics of each event (e.g. the mobility of the population and the informational requirements of the responding organisations) and was not found to introduce a significant problem to the outcomes of this work. An opportunity for further research is presented to measure the extent to which population mobilisation affects locality classification and inform the existing body of work focusing on home location inference.

6.5 Summary

This chapter conducted a preliminary analysis of Twitter datasets to characterise online discourse during disaster events. These findings informed the definition of a coding schema with which 2,500 messages were classified with respect to their value as a source of disaster response intelligence. The proportion of messages providing situational awareness perspectives was sufficient to justify quantitative approaches to detection.

1,500 users were then coded based on their estimated *locality* to the disaster event. The user profile location field and presence of geotagged Tweets in each user's historical timeline were compared to the coded values and determined to provide sufficient accuracy to be used on the uncoded set of data for locality inference.

The `source` field within Tweet metadata was then examined and a correlation demonstrated between the systems of software used to publish Tweets (as defined by the source value) and the distributions of message class. Tweet source was therefore shown to be an informative feature to methods of message curation.

The method of locality inference developed in this chapter provided an approximation of *ground truth* which could be derived for a significant proportion of the total users within each dataset. These values were used in chapter 7 to measure network-based approaches to eyewitness classification and contributed to the system of location inference developed in the chapter.

7

Social Network Analysis

Contents

7.1 Network Homophily 190
7.2 Hurricane Network Datasets 192
7.3 Visualising Network Structure 193
7.4 Verifying Local Clustering using Modularity 195
7.4.1 Monte Carlo Simulation and Statistical Significance . . 200
7.5 Identifying Local Communities 201
7.5.1 Community Structure in Networks 202
7.5.2 Community Detection on Hurricane Datasets 204
7.5.3 Evaluating Community Detection Algorithms 209
7.5.4 Characterising and Differentiating Communities 212
7.5.5 Comparison with Early Graph Structure 214
7.6 Discussion 217
7.7 Summary 221

This chapter examines the relationship between user network data and user locality. The premise of this approach was based on the expectation that relationship formation between users would be correlated with their geographic proximity and therefore provide discriminative power to location inference methods, and by extension, eyewitness classification. This was formulated as RQ_3, which extended the broader investigation of the value of social media data as a source of intelligence raised by RQ_2.

> RQ_3 —To what extent can graph-structured relationship data inform eyewitness classification for social media data?

In the study presented in this chapter, the presence of *assortative mixing* between users in the Hurricane Harvey and Florence Twitter networks was established based on spatial network visualisation, through which the clustering behaviour of local nodes was demonstrated. These findings were then verified formally using quantitative methods and established as statistically significant using a Monte Carlo simulation approach.

The networks were partitioned using a suite of community detection algorithms and the significance of the association between node locality and community membership then established. The purpose of this approach was to facilitate node locality classification based on node membership within a community according to the community's *locality coefficient*. Membership in a given community was shown to represent a useful parameter for node locality prediction (contribution C-4). A comparison of the community detection algorithms identified the Louvain method as most performant for this purpose.

Finally, the community detection method was conducted on subgraphs of the complete event data which represented earlier moments during the data collection period. This study was conducted to measure the validity of the network approach to eyewitness classification in real-time environments during which network data are built over time. The precision of the method was found to stabilise at a sufficiently high level within a few hours of the collection period, thus providing compelling evidence supporting the integration of this approach to eyewitness classification systems.

7.1 Network Homophily

Network homophily, or *network assortativity*, refers to the tendency for people to associate with others with whom they share similar traits (Newman 2018; McPherson et al. 2001). In a graph representing people as nodes and relationships as edges, assortative mixing is observed where a higher-than-expected fraction of edges exists



(a) Homophily by race (vertex shading) **(b)** Homophily by age (vertex shape)

Figure 7.1: Homophily in school friendship networks (Moody 2001)

between nodes of the same type. Where present, the network assortativity effect leads to the creation of node clusters within a graph, known as *communities*.

Homophilic mixing in two US schools is demonstrated by Moody (2001) and presented in figure 7.1. Here, shaded nodes indicate non-white students, node shape represents age group, and edges denote self-reported friendships. The two networks illustrate clear cases of assortative mixing based on the two attributes: in the first example, the friendship patterns suggest that race is particularly salient while in the second, division is based on age.

The analysis conducted in this chapter measured the strength of homophilic clustering based on shared geographic features using social network data from the Hurricane Harvey and Hurricane Florence events. The presence of community structure in these social networks was then evaluated as a predictor for user locality. That is, the assertion that a user is more likely than average to follow other users in their local region was shown to be true, thus supporting the evaluation of network metrics in determining user locality.

7.2 Hurricane Network Datasets

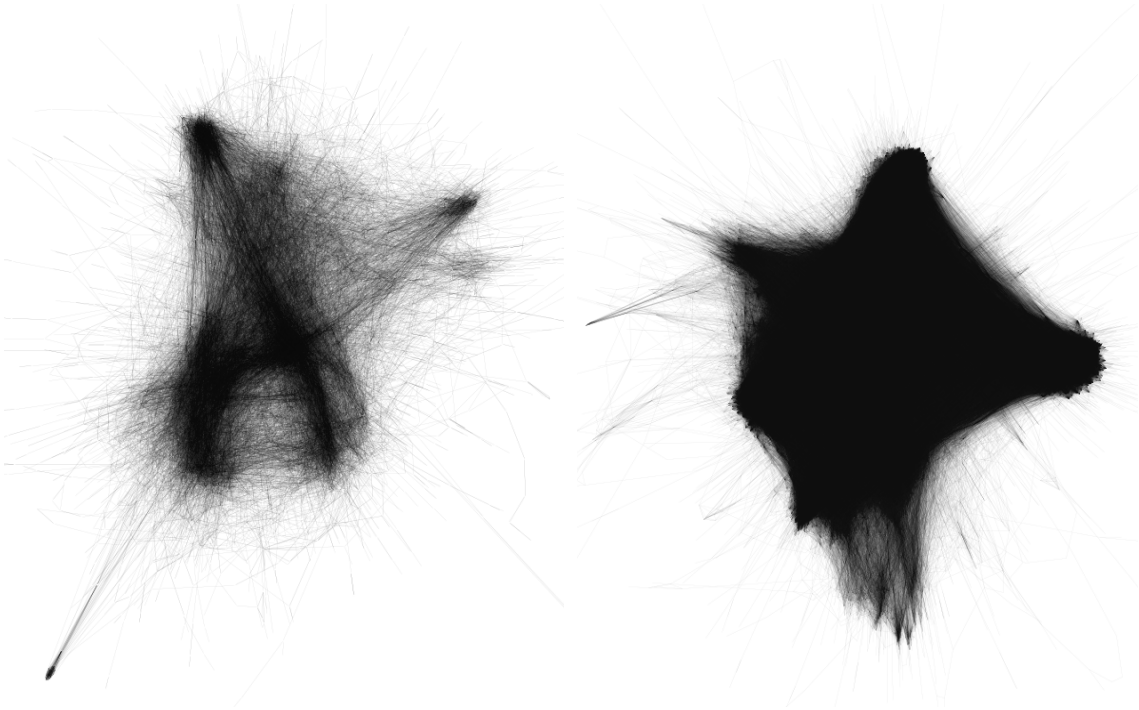
For each Twitter user detected during a data collection period, network data were recorded representing the accounts that followed or were followed by the user. These directed egocentric networks (ego nets) were then merged to create an overall graph containing ego nodes (users detected by the software) and alter nodes (users following or followed by an ego node, where not already classified as an ego node). Two networks were chosen for examination in this section: The Hurricane Harvey dataset which had been enhanced by manual coding, and the Hurricane Florence dataset due to the breadth of its coverage.

The Florence dataset was reduced from the values shown in table 5.1 to include only the first week of data. This subgraph was used to reduce the complexity of the network and focus the dataset on the most relevant period of the event. The size of each graph is shown in table 7.1 with the size of the *largest connected component* (LCC) and the proportion of the whole graph that it represents. The LCC is the subgraph for which all member nodes are connected by an edge (in either direction).

Event	Nodes	Edges
Harvey	31932	101096
Harvey_*LCC*	18,410 (0.58)	76,341 (0.76)
Florence	124,558	3,428,659
Florence_*LCC*	100,343 (0.81)	2,933,280 (0.86)

Table 7.1: Network sizes

For this work, the *bidirected graphs* were simplified to *undirected graphs* such that all relationships between users were considered equivalent. In reality, there exist three states of follower/followee relationship within the graphs (in addition to the null relationship): *following*, *followed by*, and *reciprocal* (see figure 4.1). Undirected graphs were used due to their lower complexity and increased compatibility with community detection algorithms. An extension of this study to include the



(a) Hurricane Harvey **(b)** Hurricane Florence

Figure 7.2: Social network structures

bidirectional features of the original data presents a valuable opportunity for further research.

7.3 Visualising Network Structure

Clustering behaviour was inspected visually by applying the ForceAtlas2 algorithm to the LCC of each graph using the network visualisation software Gephi (Jacomy et al. 2014). ForceAtlas2 is a force-directed layout that simulates a physical system to spatialise a network. Nodes are given a repulsive force while edges attract their nodes. The algorithm converges to a balanced state where *closely connected* nodes are positioned spatially closer than nodes with larger *shortest paths* between them. The shortest path is the minimum sequence of edges that join two vertices. Closely connected nodes are those joined by fewer edges thus suggesting a closer relationship. The results of this algorithm are shown in figure 7.2.

Visual inspection of the follower/followee network for observed users revealed a clear community structure, visible as node clustering in figure 7.2. These structures

demonstrated a degree of homophily (or assortativity); that is, there were certain features influencing the propensity for a user to follow another user with similar (or dissimilar) features. The principle of geoproximate homophily states that geoproximity influences graph structure and, therefore, that node locality may be inferred from the characteristics of the community of which it is a member. Figure 7.3 shows a representation of the same network structure, where coded nodes are visible in colours representing their codes. Uncoded nodes are not shown, though their presence remains relevant in determining network spatialisation. Locally coded nodes show clustered behaviour while non-local nodes are more evenly distributed. The spatial clustering of local nodes corresponds with community clusters shown in the original graph structure. These *local communities* contained a greater proportion of local nodes than other communities, supporting the hypothesis of geoproximate homophily: local users were more likely than average to follow one another. Membership in a local community was therefore shown visually to be a predictive feature for node locality and encouraged the further examination of network metrics in user classification and Twitter stream filtering.

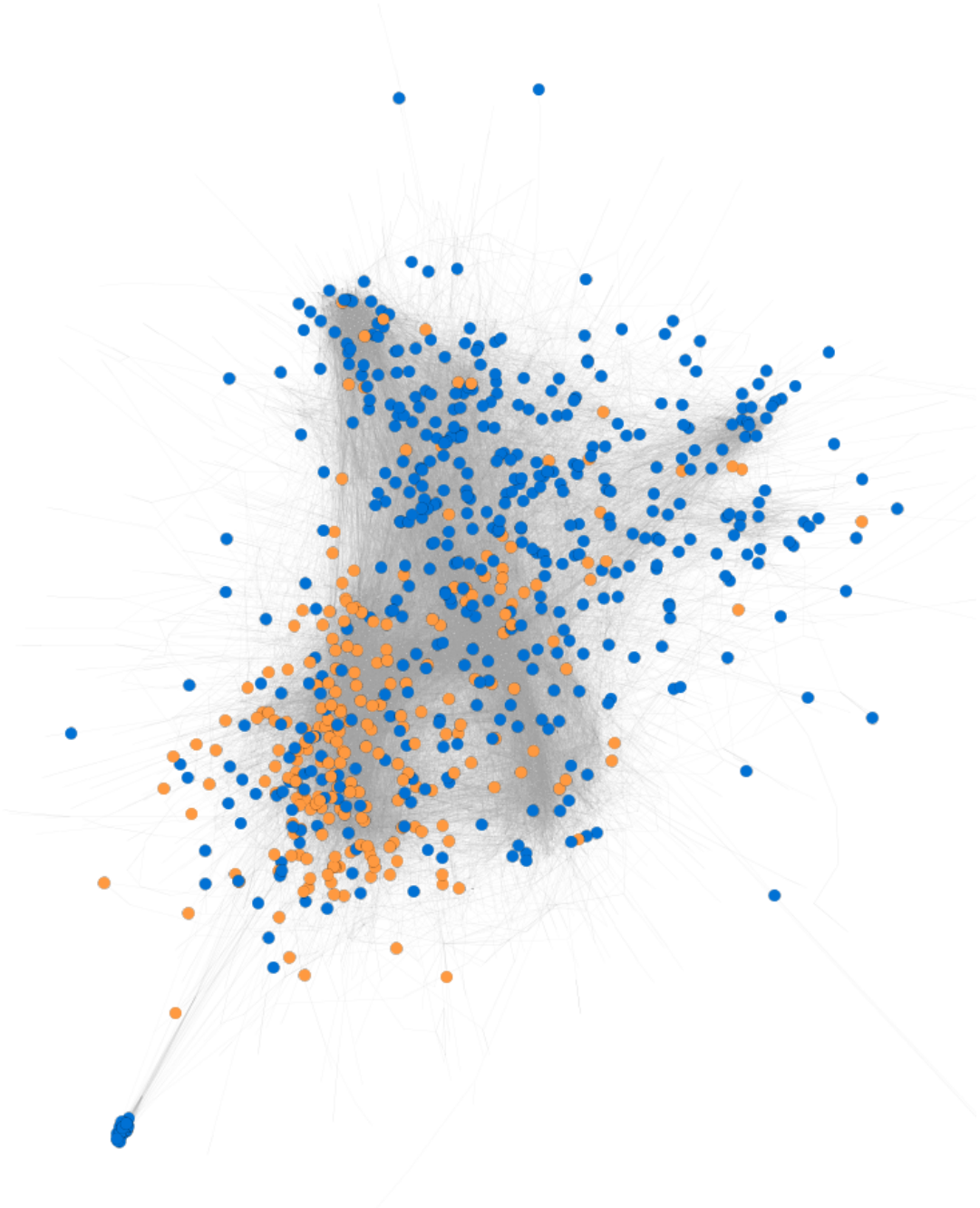


Figure 7.3: Node locality structure — Hurricane Harvey (manual coding). User locations were manually coded according to their locality to the event (section 6.2.3). Orange nodes are local users, blue nodes are non-local users. Uncoded users are shown in grey.

7.4 Verifying Local Clustering using Modularity

In this section, the clustering behaviour observed visually in the preceding section was formally verified. Assortative mixing between nodes based on locality was

established by comparing the fraction of observed edges between nodes with the same label to the fraction of analogous edges existing at random in a graph with a similar degree distribution. For a given graph G represented as an adjacency matrix A, the total number of edges between nodes of the same type is expressed as:

$$\frac{1}{2}\sum_{ij} A_{ij}\delta(g_i, g_j) \tag{7.1}$$

where $\delta(g_i, g_j)$ is the Kronecker delta: 1 if i and j share the same group label and 0 otherwise; and the factor of $\frac{1}{2}$ adjusts for the effect of counting each pair of nodes twice in the sum. Calculating the expected number of edges is based on the *configuration model*; a graph in which the degree of each node is equivalent to that of the observed graph and edges are randomly assigned. Therefore, the probability of an edge existing between nodes i and j can be calculated using their degrees k and the total number of edge stubs in the graph, $2m$:

$$\frac{k_i k_j}{2m-1} \tag{7.2}$$

Here, $2m-1$ is used to exclude the stub currently being observed, however, for sufficiently large values of m this can be ignored. The fraction of edges between nodes with equal labels can then be calculated in the same way as eq. (7.1):

$$\frac{1}{2}\sum_{ij} \frac{k_i k_j}{2m}\delta(g_i, g_j) \tag{7.3}$$

The observed and expected number of edges can then be compared and converted into a fraction by dividing by m. The resulting quantity is called the modularity (Newman 2018; Blondel et al. 2008), denoted Q:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(g_i, g_j) \tag{7.4}$$

Modularity is strictly less than 1 and takes a positive value where there are more edges between nodes with the same label than expected by random chance. Negative values of modularity indicate disassortative mixing. Modularity may be normalised so that it takes a value of 1 in a network with perfect assortative mixing; that is, in a

network in which all edges fall between nodes with equal labels. Such a configuration is called a *perfectly mixed network*, the modularity of which is defined as:

$$Q_{max} = \frac{1}{2m}\left(2m - \sum_{ij}\frac{k_i k_j}{2m}\delta(g_i, g_j)\right) \tag{7.5}$$

The normalised measure, called the assortativity coefficient, is the ratio of the two measures:

$$AssortativityCoefficient = Q/Q_{max} \tag{7.6}$$

This result is an example of a Pearson's correlation coefficient and as such can be interpreted in the same way: a value of 1 denotes perfect positive correlation, 0 no correlation, and -1 perfect negative correlation. In essence, the assortativity coefficient measures the correlation between the chosen characteristics of every pair of nodes that are connected.

The assortativity coefficient for the coded subgraph of 1,500 nodes from the Harvey dataset was 0.542, representing a strong level of homophily based on the coded node-locality label. This value supported the principle of geoproximate homophily: while it was not expected that users follow *only* other local users, the strength of the assortativity coefficient showed a correlation between the locality of a user and their neighbours such that knowledge of social network structures provides useful predictive power in node locality inference.

The coefficient for the network extending beyond the coded subset was calculated based on the geocoded results of the profile location field. This field was found in section 6.3.2 to have an agreement rating of 0.817 with the coded value and was therefore considered a suitable alternative for the analysis of the uncoded dataset. 8,770 unique location strings from the Harvey dataset were geocoded in the previous section. A further 29,507 unique location strings were extracted from the Florence dataset and geocoded using the method described in section 6.3.2. Visualisations based on these classifications are presented in figures 7.4 and 7.5, where similar clustering behaviour to that shown in figure 7.3 is evident.

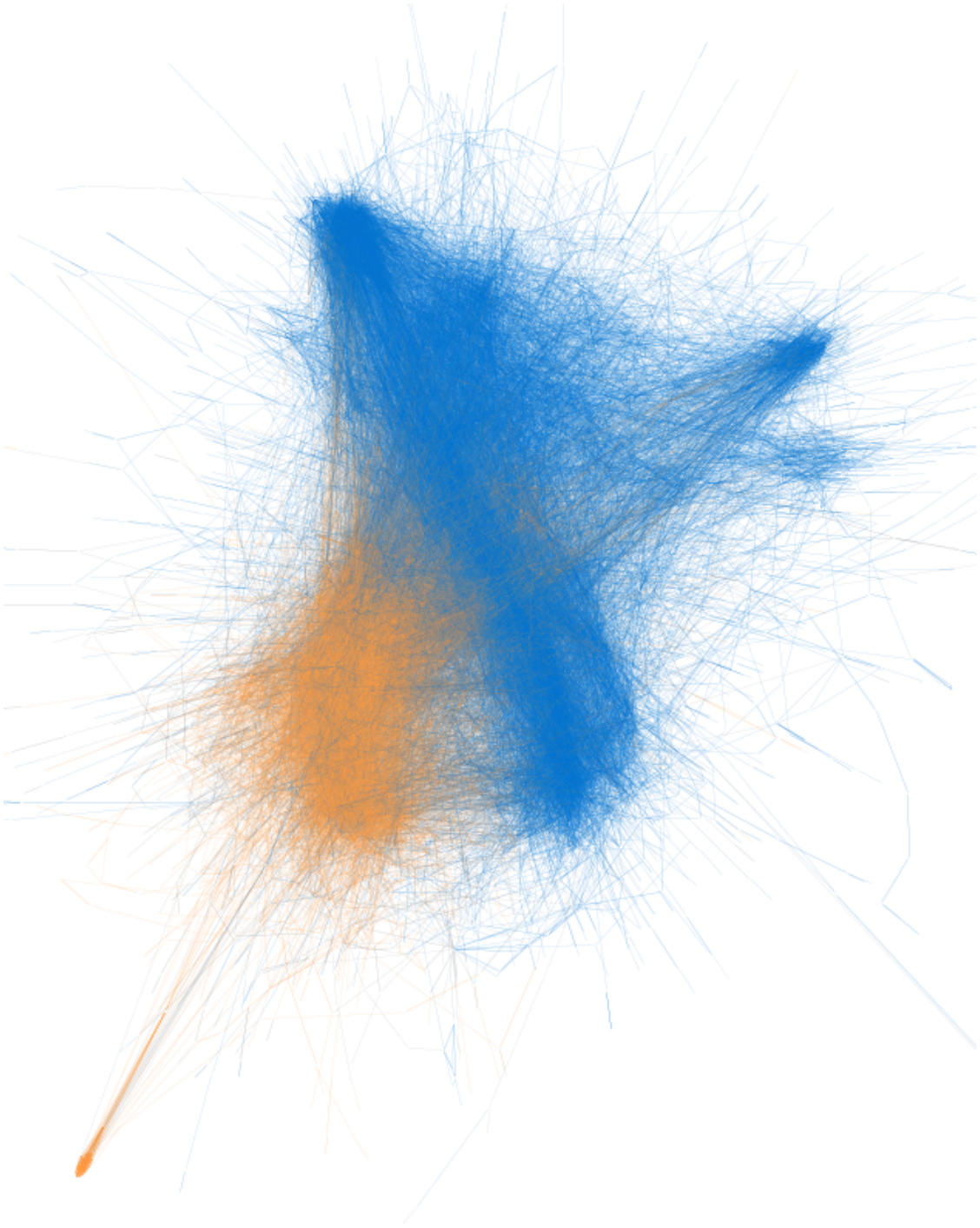


Figure 7.4: Node locality structure — Hurricane Harvey. User locations are geocoded based on profile location field (section 6.3.2) and classified using the bounding box defined for the event (figure 5.7). Orange nodes are local users, blue nodes are non-local users. Uncoded users are not shown.

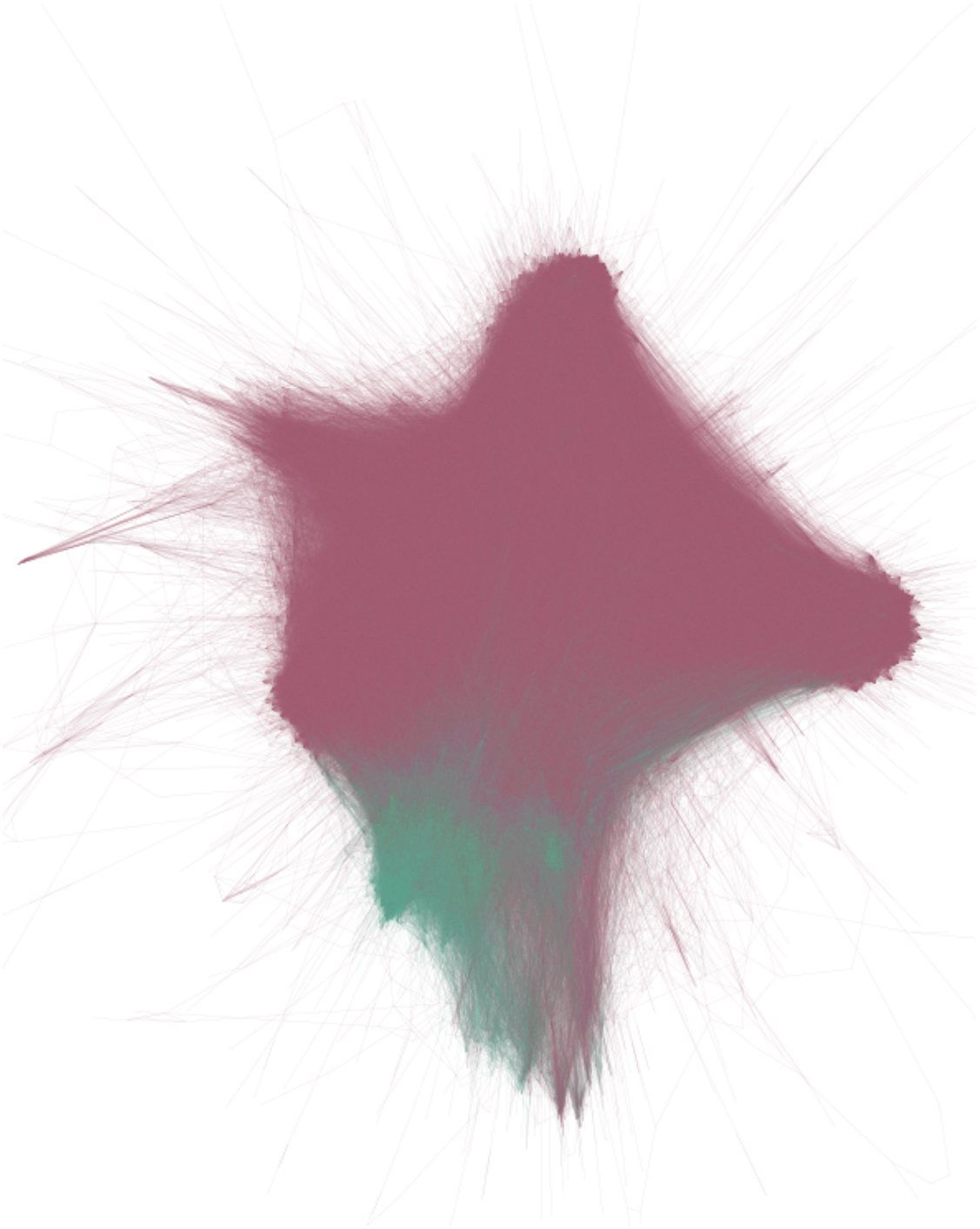


Figure 7.5: Node locality structure — Hurricane Florence. User locations are geocoded based on profile location field (section 6.3.2) and classified using the bounding box defined for the event (figure 5.7). Green nodes are local users, red nodes are non-local users. Uncoded users are not shown.

Each network was filtered to exclude nodes that had not provided profile location data or had returned a null result from the geocoding query. The assortativity coefficients based on locality label were calculated as 0.495 (Florence) and 0.625 (Harvey), showing a strong correlation between the label of a node and the label of its neighbours (table 7.2).

These were lower than the 'true' value due to the implicit assumption that all non-local accounts were equally connected. In fact, there were several distinct clusters of non-local users (as seen in figures 7.4 and 7.5) and therefore the lack of edges between these clusters had an undesirable negative effect on the assortativity value. While the effect of this was not captured in the binary classification used for this research, the magnitude of the coefficients was such that homophily based on locality was sufficiently confirmed.

Event	Nodes_$total$	Nodes_$filtered$	Assortativity
Harvey	31,932	24,855	0.625
Florence	100,343	68,897	0.495

Table 7.2: Network assortativity coefficient

7.4.1 Monte Carlo Simulation and Statistical Significance

The statistical significance of the assortativity values was verified using Monte Carlo simulation in which the computed values were compared to those drawn from an ensemble of randomised networks defining *null models*. The ensemble comprised a set of 100 *configuration models* such that the *degree sequence* of the original network was preserved (that is, the original degree of each node was retained) and edges then randomly assigned. This approach was chosen over other algorithms to distinguish features accounted for by the degree sequence from those caused by other forces or constraints (Foster et al. 2010).

For each network, the significance of the assortativity measure $r_{observed}$ was quantified by comparing its distance to the average assortativity in the randomised

ensemble $\langle r_{rand} \rangle$, in units of standard deviation $\sigma(r_{rand})$. This metric is known as the z-score and is described by the equation:

$$z = \frac{r_{observed} - \langle r_{rand} \rangle}{\sigma(r_{rand})} \tag{7.7}$$

As the assortativity values of the random ensemble follow a normal distribution (see figure 7.6), the z-score was then used to determine the confidence value of the observed results. The p-value for each assortativity measure was <0.00001 and as such, the values in table 7.2 were accepted as significant. The phenomenon of geoproximate homophily was therefore empirically established: Twitter users were more likely than average to form follower/followee relationships with users within their geographic vicinity.

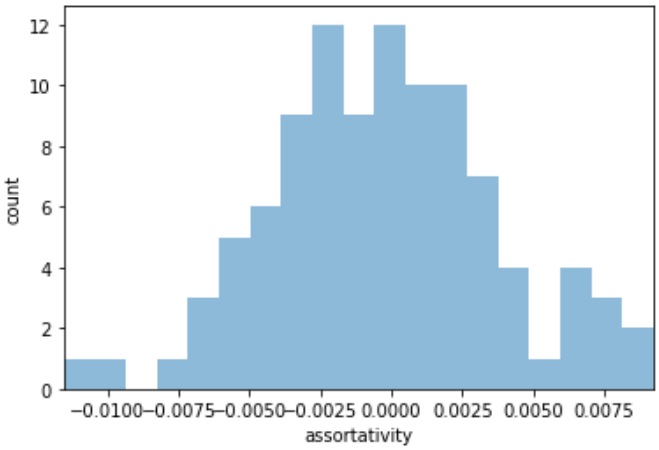


Figure 7.6: Assortativity of 100 configuration models — Hurricane Harvey

Monte Carlo simulation was chosen over the jacknife method described in Newman (2003) due to the computational complexity of the latter on networks with many edges.

7.5 Identifying Local Communities

In the analyses presented in previous sections, the latent binary attribute of `user locality` was represented by hand-coded values and the proxy attribute `profile location field locality`. While the proxy showed a strong correlation with the

latent attribute, the parseable non-null fields comprised only 74.3% of the data and required expensive API calls to an external geocoding service. Furthermore, the correlation may not be robust across different datasets where language or culture influence the way in which the unconstrained text field is used. Establishing predictive attributes based on network structure provides more efficient metrics for user classification which apply to a larger proportion of the dataset and support the discovery of new local nodes through neighbourhood exploration. The assortative mixing established in section 7.4 showed that user locality was correlated with the clustering behaviour observed in figure 7.2 and therefore that identifying the cluster to which a node belongs could meaningfully inform locality classification.

In this section, the characteristics of the network structure are analysed to synthesise node features correlating to the dependent `user locality` variable. Clustered sets of connected nodes are formally partitioned from the main graph as community subgraphs such that inter-community edges are minimised. Node locality homophily is then measured within each subgraph to identify and characterise *local communities*. Membership within a community defined as 'local' is then evaluated as a predictive variable in node locality classification. This approach had two main objectives: first, to verify the observations from section 7.4 showing the correlation between user locality and community membership; and second, to test the viability of detecting and classifying local community subgraphs as a means to classify individual nodes. Finally, the methods established in this analysis were tested on earlier versions of the datasets to verify the validity of the approach in a live situation where network data is incrementally collected and rapid classification desired.

7.5.1 Community Structure in Networks

In network science, community structure exists where the nodes of a network may be grouped such that members of a community are more densely connected internally than with nodes from other communities. That is, two nodes are more likely to be connected if they are both members of the same community. For example, figure 7.7 shows a network of collaboration between scientists (nodes)

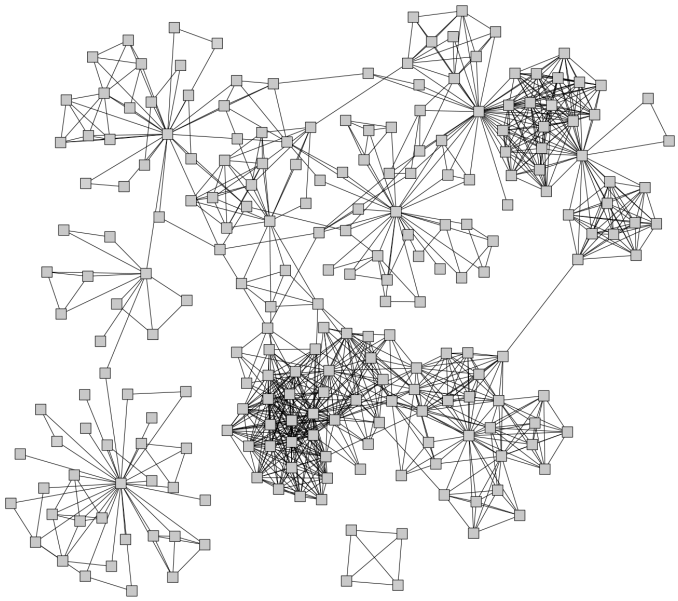


Figure 7.7: Network of coauthorships in a university department (Newman 2018)

where edges represent paper coauthorship. In this example, the community structure, or clustering, corresponds approximately to the structure of the formal research groups at the university. Even without knowledge of the department's structure, a visual inspection of this network reveals the existence of distinct research groups which vary in the density of their internal connections. While visual inspection of much larger graphs, such as those shown in figure 7.2, is less useful, community structure can be verified using a range of algorithmic approaches.

Community detection algorithms are used to find the natural fault lines along which a network separates (Newman 2018; Blondel et al. 2008). A graph is partitioned into a number of community subgraphs based on an optimisation function defined by the algorithm. For example, in section 7.4, the concept of modularity was introduced as a measure of how likely connections were between nodes of similar types. *Modularity maximisation* is a class of community detection algorithm which seeks to assign community labels to nodes in a graph such that the total modularity with respect to community label is maximised (Newman 2006). *Label propagation* is an alternative approach in which labels are randomly assigned to nodes and then iteratively propagated until all nodes bear a label matching

the majority of their neighbours. While these two approaches are popular and have well-implemented libraries, community detection remains an active area of research with many viable approaches. A more detailed discussion of community detection algorithms is, however, outside the scope of this work. The algorithms applied in this study were selected based on their performance and complexity and serve to demonstrate the community-detection approach to locality classification. A practical implementation of these methods in a live environment naturally warrants a more thorough comparison of available algorithms using a diverse set of datasets and presents an interesting area for further research.

7.5.2 Community Detection on Hurricane Datasets

A set of seven community detection algorithms was selected for analysis and applied to the LCCs of the Harvey and Florence graphs, thus assigning to each node a set of seven nominal community labels. For all but one case, the total number of communities was not passed as a parameter to the algorithm and therefore the final count was a function of the structure of the network. The *fluid* algorithm required that the number of community partitions be predetermined, for which the total number of communities (n) designated by the *Louvain* algorithm was used for comparison. As a loose optimisation strategy, the fluid algorithm was repeated using multiples of n. The relevant factor is appended to the algorithm name in the tables below.

The assortativity coefficients were calculated (with respect to each set of community labels) to evaluate the effectiveness of each community algorithm in partitioning the graph based on community structure. An assortativity value of 1.0 represents a set of community labels for which all edges exist between nodes from the same community (i.e. no inter-community mixing) and a value of 0 where community labels are assigned at random. Naturally, the highest achievable value for a given network is constrained by its structure and the number of community labels assigned by an algorithm. The assortativity coefficient is therefore a measurement of how well the community labels defined by each algorithm capture community

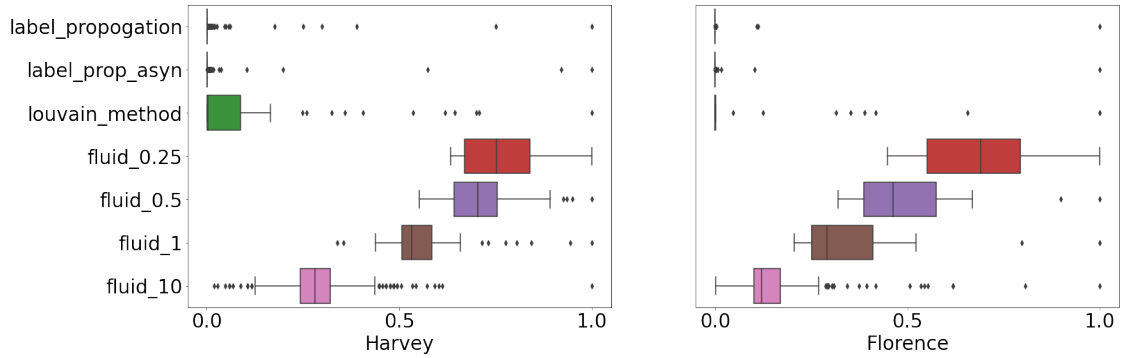


Figure 7.8: Distribution of community size as a proportion of maximum size by algorithm. The label propagation methods created a small amount of very large communities which comprised most nodes and thus limited the effectiveness of the labels. In contrast, the community sizes of fluid_0.25 are so evenly distributed that they lost the ability to capture structure. Both effects are best illustrated on the larger Florence dataset.

structure within the graph and was used here to compare each algorithm and select a candidate for use in the following stages of research.

The results of the community detection algorithm comparison are shown in table 7.3. While the assortativity coefficients for the *label propagation* algorithms were similar to Louvain, they showed significant variation in both the higher number of communities defined and the more highly-skewed distribution of their sizes. This effect is visualised in figure 7.8, which shows the distribution of community sizes for each method. A small number of *giant communities* were created by the label propagation methods which encapsulated the major clusters of nodes and remaining nodes were labelled as sets of minor communities of size 2-3. In the figure, these large communities are shown as black dots — outliers whose size exceeds the bounds of the right whisker.[1] The Louvain and fluid methods defined more balanced and smaller sets of communities, segmenting the largest clusters of nodes in the network to a greater extent than label propagation.

The community labels as assigned by the Louvain algorithm are presented visually for each network in figure 7.9. These examples illustrate how community detection may furcate a network into segments aligning with the clustering behaviour

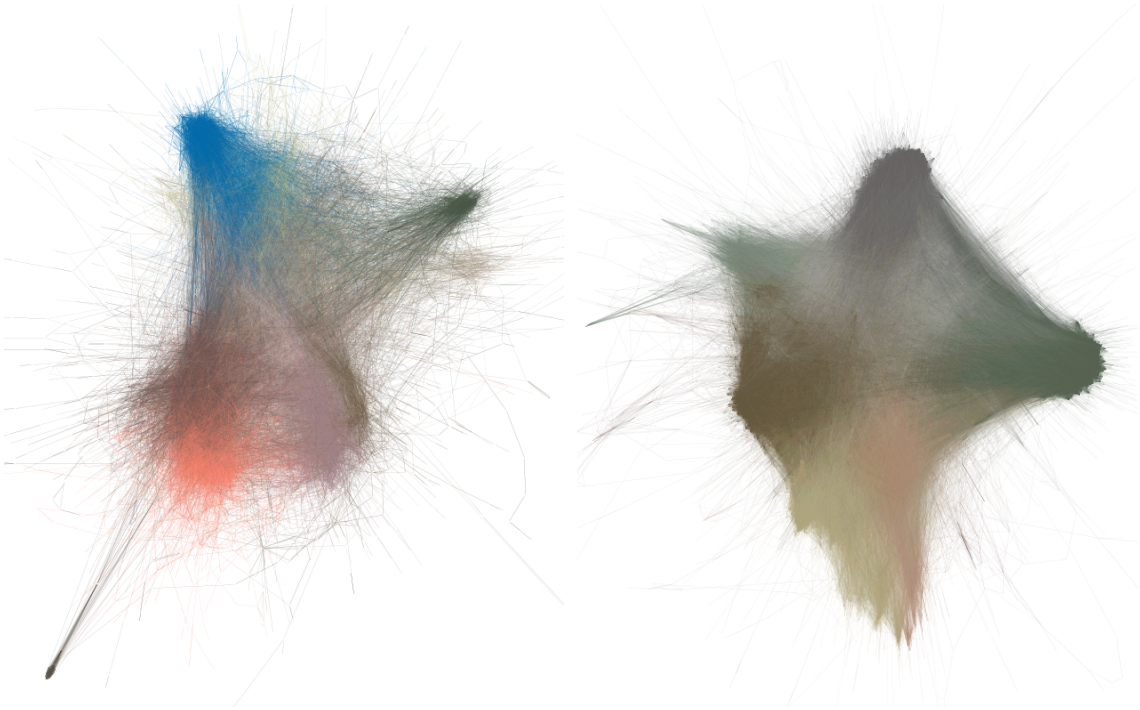
[1]The bound for the right whisker is defined as: $Q_3 + 1.5(Q_3 - Q_1)$ where Q_1 and Q_3 represent the 25th and 75th percentiles respectively (Tukey et al. 1977).

identified visually in section 7.3. In figures 7.4 and 7.5, local users were shown to exhibit clustering behaviour. The significance of using community detection algorithms in locality prediction is predicated on how well the community labels they assign align with this behaviour. If a given set of community labels captures a large proportion of local users in a network it represents a meaningful predictor for locality. The usefulness of the predictive set naturally depends on the extent to which it captures all local users and the rate of errors within the captured set (i.e. the false negative and false positive rates). These thresholds are derived from the tolerances identified by end users in chapter 3 and are discussed in a later section.

Algorithm	Assortativity Coefficient	Total Communities	Largest C	Smallest C	Median C Size
Harvey					
label_propogation	0.67	1816	3565	2	3
label_prop_asyn	0.71	1897	3968	2	3
louvain_method	0.73	63	2682	3	6
fluid_0.25	0.60	15	1585	1005	1194
fluid_0.5	0.55	31	818	452	575
fluid_1	0.52	63	519	176	276
fluid_10	0.38	630	103	2	29
Florence					
label_propogation	0.82	774	65558	1	2
label_prop_asyn	0.88	785	70954	1	2
louvain_method	0.75	37	25287	3	7
fluid_0.25	0.71	9	13161	5888	9079
fluid_0.5	0.67	18	9004	2886	4165
fluid_1	0.60	37	6508	1335	1898
fluid_10	0.34	370	1513	2	182

Table 7.3: Results of community detection algorithms. An assortativity coefficient of 1.0 represents a community structure with no inter-community edges.

The relationship between node locality and community membership was con-

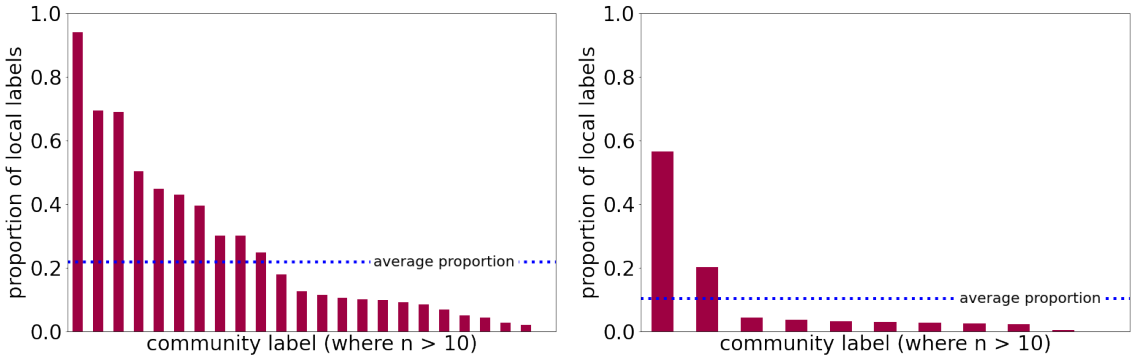


(a) Hurricane Harvey **(b)** Hurricane Florence

Figure 7.9: Community network structure — Louvain algorithm. Each colour represents a distinct (arbitrary) community label.

firmed for all community algorithms using chi-squared tests ($p < 0.001$). This was unsurprising given the assortative mixing established in a previous section, however, it further confirmed the effectiveness of the community detection algorithms at partitioning the graph into local and non-local communities. The proportion of local profiles in each community is presented visually in figure 7.10. Each bar represents a community label assigned by the Louvain algorithm, for communities with a size greater than ten. The y-axis marks the number of local profiles within the community as a proportion of the total local and non-local profiles (i.e. excluding non-labelled profiles). The average proportion constant is the equivalent proportion for the entire graph and is therefore the expected value for each community where an association between locality and community label does not exist.

Clearly, *local* and *non-local* communities exist within each network structure as shown by the significant deviations from the mean of the locality proportion observed in several communities (figure 7.10) (though not all communities are so easily distinguished). The community label assigned to a node was therefore shown



(a) Hurricane Harvey **(b)** Hurricane Florence

Figure 7.10: Proportion of local profiles per community — Louvain algorithm. Large differences in locality proportions illustrate the phenomenon of local and non-local communities.

to have predictive power for node locality, given that some information describing the overall locality of the community is known. In practice, the community label may supplement the locality label assigned by third-party mapping APIs and provides an informative alternative where the locality label was not able to be otherwise assigned (for example, where the profile location field was empty). Such inference relies upon not only the community label, but also knowledge of how 'local' the community associated with the label is computed to be. The method thus far demonstrated in this analysis suggests the following approach:

1. Once a node is detected, where profile location data exists, its locality is estimated based on a query to the API of a third-party mapping service (see section 6.3.2).
2. Detected nodes are incrementally added to an existing graph object representing the event under observation.
3. The community structure of the graph is periodically re-calculated using the chosen community detection algorithm.
4. A *community locality coefficient*, l, is calculated for each community group based on the average locality of their labelled member nodes.
5. The coefficient is then propagated to all member nodes and passed as a parameter to a selected locality classification algorithm.

7.5.3 Evaluating Community Detection Algorithms

In the previous section, a set of community detection algorithms were compared with respect to their performance at partitioning each network into communities with minimal inter-community edges. While the assortativity coefficients calculated for each algorithm demonstrated sufficient performance at this task (table 7.3), the characteristics of the resulting community structure varied significantly (figure 7.8). Algorithm performance, with respect to its ability to classify local users, was a function of how well its resulting structure matched the latent structure of local nodes within the network — that is, the extent to which a given subset of communities captured the local nodes in the network. This section demonstrates how performance was quantified to inform the selection of an algorithm for further analysis.

The strength of an association between the nominal variables of *community label* and *node locality* was measured using Cramér's V (ϕ_c) (Cramér 1946) with the bias correction from Bergsma (2013). ϕ_c is based on the chi-squared statistic and returns a value between 0 and 1, where 1 indicates a perfect association and 0 no association between variables. In the context of this work, a perfect association would indicate a condition in which all local nodes were assigned communities containing only other local nodes. The ϕ_c values for each algorithm on both networks are shown in table 7.4 and figure 7.11.

While the label propagation methods were the highest performers on the Harvey network, they showed very low ϕ_c values for the much larger Florence dataset. Both methods defined a strongly dominant community for the Florence dataset, containing over 65% of the network's nodes (see figure 7.8). The locality distribution of these dominant communities closely matched the distribution of the population[2] and therefore did not provide meaningful distinguishing information. Furthermore, the label propagation community structure consisted of many communities containing 1-3 nodes (see the total communities and median community size values in table 7.3). Given that using community labels as predictive features requires a characterisation

[2]the total proportion of local nodes in the Florence network was 0.104, the proportions in each largest community were: $label_propagation = 0.122$ and $label_prop_asyn = 0.106$

of the communities based on known node values, smaller communities provide less meaningful predictive value as they contain fewer nodes to inform their characterisation.

The Louvain algorithm achieved consistent results between the datasets and created a set of communities smaller in number and more balanced in size than the label propagation methods. The community structure assigned by Louvain was therefore more receptive to community characterisation. The baseline fluid algorithm ($fluid_1$) achieved similar results with more evenly distributed community sizes, though as the total communities must be provided as an input parameter, these results rely upon the output of the Louvain method. Deriving the optimal number of communities for the fluid algorithm is possible without this dependency on Louvain by searching through a range of values and may result in an algorithm that outperforms Louvain, however, the values shown here provide sufficient support for the use of community labels in locality prediction and therefore such optimisation is outside the scope of this work. Based on the results presented here, the Louvain method was selected for use throughout the rest of this work.

Algorithm	Harvey	Florence
label_propogation	0.66	0.13
label_prop_asyn	0.67	0.19
louvain_method	0.63	0.55
fluid_0.25	0.45	0.33
fluid_0.5	0.52	0.53
fluid_1	0.53	0.56
fluid_10	0.38	0.61

Table 7.4: Cramér's V (ϕ_c) value measuring the strength of association between node locality and community label for each community detection algorithm.

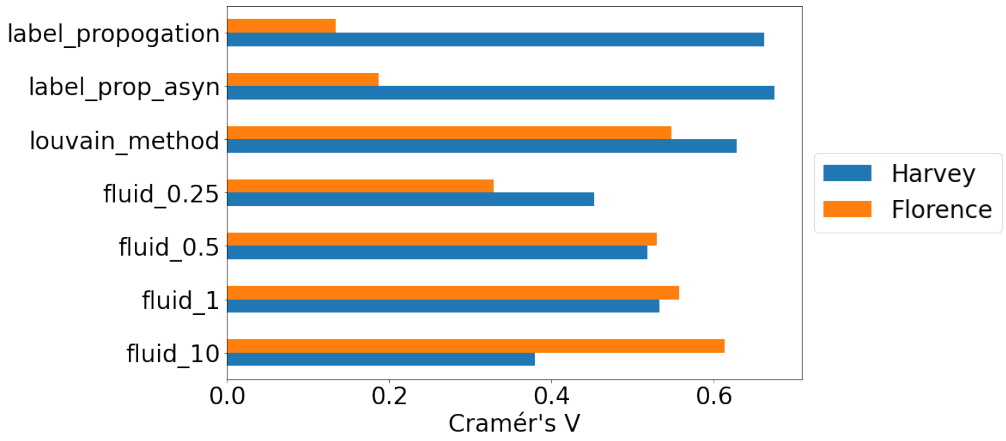


Figure 7.11: Cramér's V (ϕ_c) value measuring the strength of association between node locality and community label for each community detection algorithm.

Comparing Algorithm Runtime

As the method proposed by this work relied upon repeatedly running community detection algorithms on growing networks in live environments, algorithm runtime was compared. Each algorithm was run on 10 configuration models for each dataset and the average time calculated. The normalised results are shown in table 7.5 and figure 7.12. While Louvain was slower, the relative difference was lower for the larger network and was not considered significant enough to influence the selection.

Algorithm	Harvey (n=31,932)	Florence (n=100,343)
label_propogation	0.41	0.33
label_prop_asyn	0.37	0.49
louvain_method	1.00	0.79
fluid_0.25	0.34	0.63
fluid_0.5	0.33	0.62
fluid_1	0.33	0.71
fluid_10	0.34	1.00

Table 7.5: Normalised runtimes for each algorithm based on 10 configuration models.

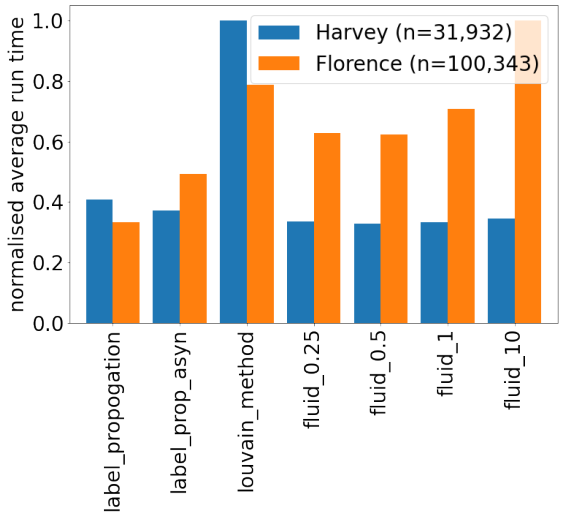


Figure 7.12: Normalised runtimes for each algorithm based on 10 configuration models.

7.5.4 Characterising and Differentiating Communities

Thus far, the method proposed for classifying node locality required *a priori* knowledge of the locality of a subset of nodes to compute the set of community locality coefficients. This feature was the resolution of API calls based on user profile locations, where supplied. As such calls incur a cost in both financial terms[3] and as a delay in data processing, community network structures were examined for alternative identifying features correlating to their proportion of local nodes.

The rationale for this approach was as follows: given the assumption of geo-proximate homophily — that Twitter users are more likely than average to form follower/followee relationships with users within their geographic vicinity — and the high proportion of local users within the collected datasets; it was surmised that local clusters would be more densely connected than other communities and therefore measures of inter-community connectivity would positively correlate with the proportion of local nodes. For example, the node clusters in figure 7.7 show distinct characteristics — while some are centred around a single central node, others are more inter-connected. In this section, these metrics were quantified and compared for the Harvey and Florence networks.

[3]Google's geocoding API charges 0.005 USD per request

The network structures of the communities (as defined by the `Louvain` algorithm) were characterised using the metrics: size (proportional), degree centrality, eigenvector centrality, degree assortativity coefficient, average clustering coefficient, and average degree; calculated using their implementation in the Python `NetworkX` package.[4] Correlations between these metrics and their community locality proportions were then tested using Spearman's Rho. For each set of metrics for both event networks, no significant correlation was found between these community structure descriptors and the proportion of local nodes they contain (see table 7.6). While such relationships may yet exist (and indeed, intuition suggests they do), they may be obscured by the imperfect classification provided by the community detection algorithm or incorrect selection of characterisation metrics and therefore warrant further research. Furthermore, while local users may show clustering behaviour in accordance with the principle of homophily, other latent attributes will be over-represented in the data, forming similar clusters which may confound the analysis. For example, members of the disaster response community may use terms detected by the keyword-filtered stream and thus appear in the network as clusters representing professional relationships (in contrast to geographic proximity).

Metric	r_rho_harvey	p_rho_harvey	r_rho_flr	p_rho_flr
assort_deg	-0.044	0.813	0.439	0.153
avg_cent_deg	-0.006	0.973	-0.444	0.128
avg_cent_eigen	-0.058	0.755	-0.511	0.074
avg_deg	0.058	0.751	0.036	0.907
clustering	0.159	0.385	-0.553	0.050
size_prop	0.093	0.614	0.361	0.225

Table 7.6: Spearman's rho correlation coefficient for various community metric variables and community locality.

[4] https://networkx.org/

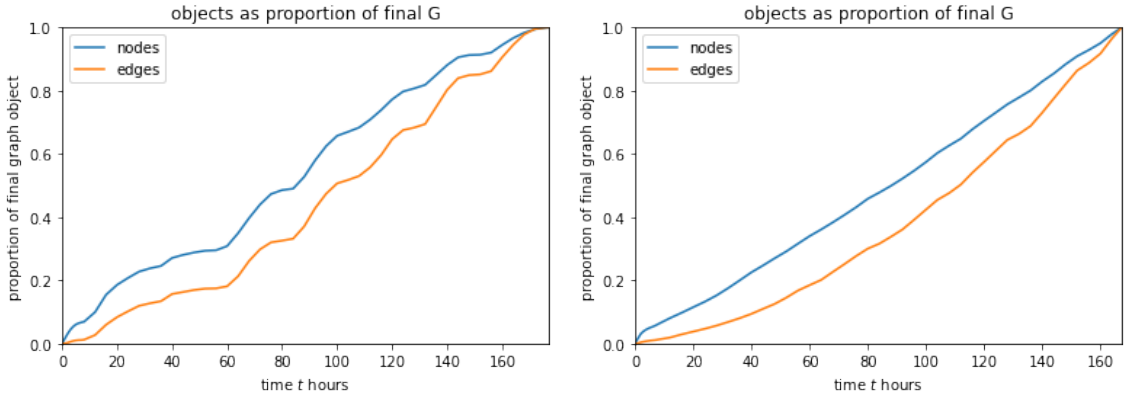
7.5.5 Comparison with Early Graph Structure

The method of locality prediction proposed by this section was shown to be effective on *mature* networks — that is, the state of the network after several days[5] of data collection. For the purposes discussed in chapter 3, classification must occur in a live environment and therefore take a less complete graph object as input. The extent to which the smaller datasets (as they existed at earlier time intervals) affected the predictive power of the community detection method was therefore evaluated to test the applicability of the locality prediction method to live scenarios with evolving input graphs.

For each graph, a set of subgraphs were created representing the graph object as it existed at time t, where t was a four-hour window of time. The growth rates are shown in figure 7.13. As the data collection process was constrained by API limits, the linear growth of recorded nodes was expected and roughly matches the rate allowed by these constraints. Some cyclical behaviour was observed in the Harvey dataset which aligns with the 24-hour cycle (i.e. users are less active during the night) though this trend was obscured in the larger Florence dataset. As edges were recorded only where a relationship existed between the newly observed node and previously observed nodes already within the graph, the exponential growth rate was expected: as the graph grows, a greater proportion of the relationships for each new node are recorded. In theoretical terms, the maximum number of edges in a simple graph with n nodes is $\frac{n(n-1)}{2}$. Therefore, given that community structure is defined by the presence of edges within a graph, the available information informing this structure grows exponentially as more data are observed.

To measure this effect, the community detection process documented in section 7.5.2 was replicated on each time-windowed subgraph. The correlation between community membership and node locality was evaluated using chi-squared tests and an association was established for all values of t ($p < 0.001$). The strength of the association at each interval was calculated using Cramér's V (ϕ_c) (Cramér 1946). The results of these tests, shown in figure 7.14, did not deviate significantly

[5] $t_{harvey} = 8$, $t_{florence} = 7$



(a) Hurricane Harvey **(b)** Hurricane Florence

Figure 7.13: Growth of graph size measured at four-hour time intervals. The smoother trend of the Florence dataset is due to its larger size. Note: due to the collection methodology, growth rate does not represent rate of activity in the total population.

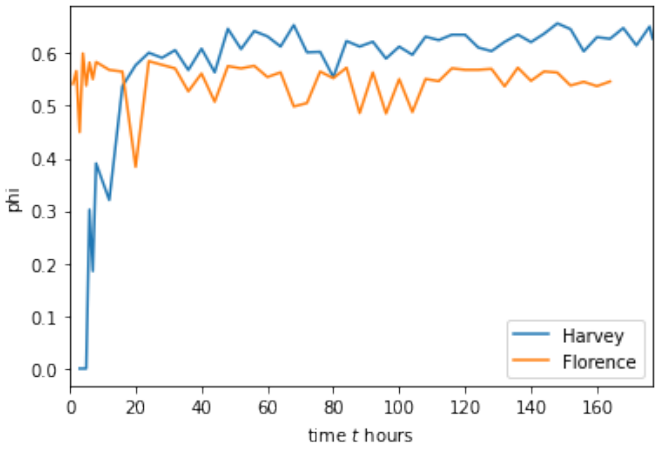


Figure 7.14: Cramér's V (ϕ_c) value measuring the strength of relationship between community label and node locality for subgraphs as extant at time interval t.

from the ϕ_c value for the final network, showing that sufficient network structure existed for node classification in early representations of the network.

Figure 7.14 shows that a stable ϕ_c representing a strong correlation was observed from approximately 24 hours for both events. It is important to note that when generalising this method to other events, this point of stability, which is based on network size, is determined by how quickly the observed network grows. Given that the data collection methodology did not capture all observed data due to technological constraints, achieving similar ϕ_c stability sooner is possible where

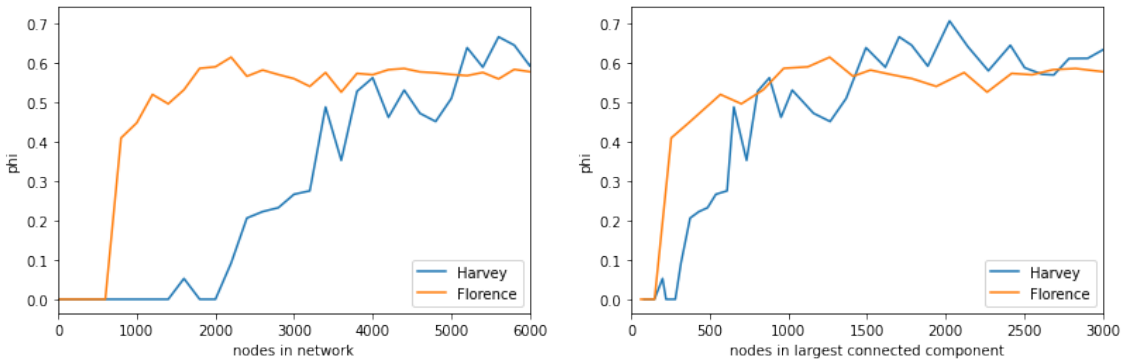


Figure 7.15: Cramér's V (ϕ_c) value measuring the strength of relationship between community label and node locality for subgraphs at given node counts.

data are captured at a faster rate (and inversely, later where data is collected more slowly). Factors influencing this rate of collection include, for example, the technological limitations of the hardware and software used for observation, the unique characteristics of the event under observation, the methodological decisions such as which keywords to track, and the limits imposed by the data providers (which may change at any time).

ϕ_c stability is shown as a function of network size in figure 7.15. While Florence reached a stable level of ϕ_c at a smaller total network size than Harvey, the second subfigure demonstrates comparatively equal performance based on the size of the largest connected component. As community detection was performed on only the largest connected component, nodes belonging to other components have no bearing on the value of ϕ_c. The ϕ_c disparity shown for total network size illustrates the different rates at which the largest connected component formed within each dataset.

For both event datasets, a stable ϕ_c, denoting a strong association between community membership and user locality, was observed in smaller subsets of each network representing points within the first hours of each event. Thus the community detection method proposed in this section was shown to be viable using smaller, less-mature network models as would be observed in the early stages of a live disaster response environment.

7.6 Discussion

This chapter confirmed the hypothesis which stated that the phenomenon of geoproximate homophily was present within the Twitter datasets collected in chapter 5 and showed that network-informed techniques present promising approaches to eyewitness classification and detection. A key advantage of the network-based approach is that its discriminative power is based on fundamental concepts of human behaviour (i.e. preferential attachment) and is therefore robust to challenges facing text-based approaches, such as changes in language patterns and generalisation to other languages. An abstraction of the classification approach developed in this chapter is illustrated in figure 7.16.

Confirming Geoproximate Homophily

Homophilic behaviour was observed visually within the follower/followee Twitter networks using the ForceAtlas2 spatialisation algorithm. This method of spatialisation iteratively updates node positions on a plane by simulating a physics-based environment in which nodes sharing follower/followee connections bear an attractive force and unconnected nodes are repulsed. The result is a two-dimensional map with which a human observer may identify key network characteristics such as clustering and community structure. The clustering behaviour was matched with the hand-coded and geocoded locality attributes (figures 7.3, 7.4, and 7.5), demonstrating an alignment between node locality and the cluster to which it belongs.

These findings were verified quantitatively using the measures of *modularity* and *assortativity*, which confirmed the visual observation that nodes showed a preference for sharing an edge with nodes of the same type (i.e. same locality). The statistical significance of these findings was measured using a set of *configuration models* — reconfigurations of the observed networks in which connections are made at random, thus representing the population set against which the null hypothesis was tested. Clustering behaviour, or *assortative mixing* was confirmed for both the hand-coded and geocoded locality attributes, supporting the hypothesis that

the locality of a node may be inferred based on knowledge of its neighbours and surrounding network structure.

Node Locality Propagation

These findings suggested two network-based approaches to node locality classification. First, where the locality attribute is known for a given subset of nodes (for example, using the geocoding approaches discussed in chapter 6), the locality of the remaining nodes may be reasonably inferred using the attributes of their neighbours propagated throughout the follower/followee edges of the network. This approach also allows for the discovery of new local nodes — where the number of locally-coded nodes sharing a follower/followee relationship with an undocumented user account reaches a threshold value, the undocumented user is passed to the data collection pipeline for evaluation. There are a number of reasons a local user may remain undetected by an observing system (perhaps they simply aren't using the tracked keywords), and therefore this network-based discovery improves the recall of the detector.

Community Locality Propagation

The second classification approach considers the overall network structure and the position of the candidate node within it. In a random dataset, nodes are equally likely to share connections with one node as another. Therefore in a non-random social network, node clusters represent pre-existing homophilic connections based on latent attributes. For example, where the data collection methodology is based upon a geographically constrained disaster event, a natural common attribute may be geographic location (given that local users are more likely to be discussing the event than average).

The effect is such: given that local users are more likely than average to share a common connection (homophily) and that there is a larger than average proportion of local users in the network, the dominant clusters within the network will contain a high proportion of local users and therefore the position of a node within the structure of its network represents a useful attribute informing classification. In illustrating this concept, the key assumption was made that local users were

the only attribute overrepresented in the data, and therefore solely responsible for homophilic behaviour. In fact, many communities were likely to have been overrepresented (consider the emergency response community) and thus, based on the assumption of assortative mixing, also represented as clusters of nodes. Whether these latent attributes exist to the extent that they confuse the node cluster classification approach depends upon the unique characteristics of the observed event and collection methodology.

Community detection algorithms were used to partition the networks into clustered subsets (*communities*) which could then be characterised based on their *locality coefficient*. Community detection approaches create sets of partitions that minimise the number of edges that exist between communities. Tightly connected clusters are distinguished from unrelated nodes, providing a basis from which the homophilic labelling was derived. A suite of algorithms was tested and compared in terms of their performance at classifying clustered subgraphs and in separating nodes based on their locality attribute. The evaluation of these algorithms was thorough, yet there remains significant scope to evaluate alternative algorithms or use datasets from a wider range of events. The Louvain algorithm was identified as the most performant, sufficiently demonstrating the suitability of a network-based approach to node locality classification. The subsequent analyses therefore implemented this algorithm, noting that the findings could naturally be further improved through optimisation of this stage of community detection (community detection algorithm optimisation is explored in Kanavos et al. (2022)).

Evaluating Temporal Constraints

Early states of a graph contain fewer user nodes and consequently considerably fewer edges that define the network structure. An analytic approach that is only valid on network datasets after several days of data collection would be unsuitable for the requirements of the response organisations and therefore any candidate method must be validated on earlier data states. A set of subgraphs were created for each dataset representing their state at given points in time, on which the community detection

analysis was repeated and evaluated. Similar acceptable levels of association were observed for graph sizes reached within the early hours of data collection.

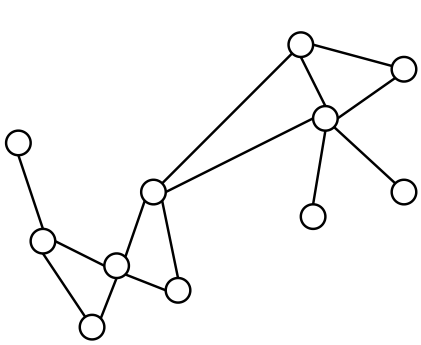
As this threshold is a function of graph size, an informative structure may be achieved even earlier through optimisation of the data collection methodology. The most significant volume constraints in these datasets were the speed at which the hardware could process data and the API rate limiting imposed by Twitter, both of which were imposed due to budgetary limitations. Data that exceeded these constraints were discarded and therefore a funded implementation of this approach collecting data at a much faster rate could achieve suitable graph sizes very early in the disaster event life cycle.

Eyewitness Identification Through Edge Traversal

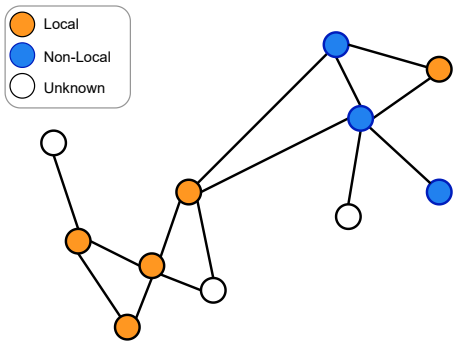
Finally, the findings of this analysis present exciting opportunities for further enhancement of the data collection methodology. For example, both datasets represented a set of users who either used a tracked keyword or posted a message which included geographic data, however, there are likely to be a large number of *false negative* local users who do not meet these criteria and are therefore never candidates for detection.

By observing the network data of detected local users, a data collection methodology may be designed which adds a third avenue of detection: where multiple local nodes are observed to share a relationship with a user that has not been detected (and thus not added to the network), a network-informed data collection worker may inspect the foreign node as a candidate for positive classification.

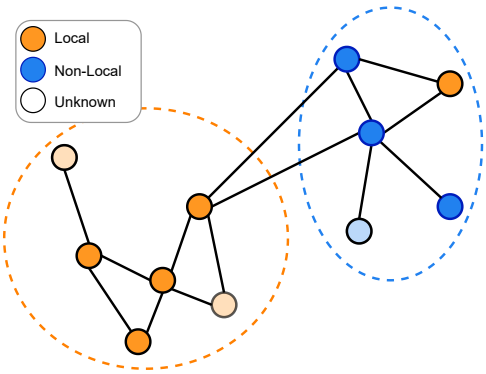
As assortative mixing was shown to exist based on the locality attribute, the locality distribution of nodes detected in this way will be higher than the population mean, thus increasing data recall. The extent to which this approach captures otherwise-undetected positive cases (and for what cost in increased false-positive cases) is an interesting avenue of future research.



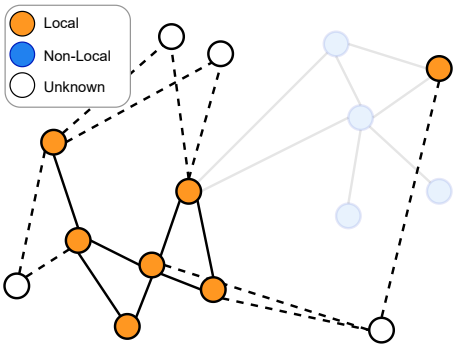
(a) User relationship data is captured and represented as a graph object.



(b) Location inference is conducted using profile location data and Tweet history geotags (chapter 6).



(c) Network is partitioned into community clusters. Nodes with unknown localities are classified by parent community locality coefficient.



(d) Non-local nodes are discarded. Common relationships from local nodes to external nodes introduce new candidates for classification.

Figure 7.16: User centric network based method of binary locality inference.

7.7 Summary

This chapter has demonstrated the value of network-based analyses for locality classification of Twitter users during the Hurricane Harvey and Hurricane Florence events using follower/followee relationships. It has confirmed the presence of assortative mixing (homophily) based on the node locality attribute and shown that community detection algorithms are effective at partitioning networks such that *local communities* may be identified and used to classify the locality of their member nodes.

The method was tested on earlier states of network data, demonstrating its effectiveness in real-time classification applications during which graph objects are constructed gradually. User-centric network informed eyewitness classification was

therefore shown to be an effective tool for disaster response intelligence curation. These results compelled the consideration of network data collection, which has not been observed in existing intelligence systems.

The statistical validation of the network-based approach presented here is complemented by a qualitative validation study in the following chapter which integrates the conceptual framework of disaster response organisations developed in chapter 3 with the findings from this study through the development of a software prototype system.

8

Validation of User-Centric Network Based Locality Inference

Contents

8.1 Digital Tools for Disaster Response Analytics 225
8.1.1 Software Categorisation 226
8.1.2 Limitations of Existing Tools 228
8.2 Prototype Development 229
8.2.1 Requirements Specification 230
8.2.2 Software Design 232
8.3 Evaluation 237
8.3.1 Participant Selection 239
8.3.2 Method 240
8.3.3 Results 243
8.4 Discussion 248
8.5 Summary 254

The software requirements of disaster intelligence analysts are as varied as the events to which they respond, however, it was clear during this study that many of these requirements stood apart from those of a typical social media user and were therefore unmet by traditional digital tools. The study in this chapter analysed how digital tools are used by disaster response practitioners for social media analysis and identified key domain requirements not adequately supported by off-the-shelf packages. A prototype system was developed to implement these requirements

and introduce the data classification algorithm developed in chapters 6 and 7 as a solution to the issue of datum volume and the ratio of useful material therein, identified in chapter 3 as a key obstacle limiting the value of social media streams for disaster intelligence (contribution C-6).

A qualitative study was conducted to validate the requirement specification and data curation approach implemented by the classification algorithm using data provided by disaster response practitioners. Six participants from the initial qualitative research (chapter 3) were invited to participate in the study, thus maintaining consistency between the perspectives provided in both datasets and further validating the findings developed in chapter 3. The relatively low number was the result of the disruption caused by the COVID-19 pandemic, which severely impacted the working environments of participants' organisations. Given the diversity of domains in which participants operated, this number was considered sufficient to provide meaningful feedback given the constraints.

A dataset of Tweets was synthesised from existing Twitter hurricane datasets and used to simulate the conditions of a disaster response event. A pre-existing system which did not implement a classification mechanism was populated with the same set of data and represented an uncurated feed for comparison. Participants were invited to interact with each system by completing a series of tasks simulating typical processes and then asked to describe their experiences. The data generated in this study were drawn from the observation of participants using the software and interviews conducted upon the completion of the simulated tasks.

Participant feedback expressed a positive view of the prototype software and underlying classification algorithm and showed a preference for its novel design over that of pre-existing systems which failed to adequately meet their requirements. These response data provided valuable perspectives on interface design for the domain of disaster intelligence and demonstrated that the author network geoinference approach to data classification effectively improved the value of social media data as intelligence during disaster response operations by increasing the proportion of valuable information within the observed dataset. This study therefore

validated the academic contributions presented in previous chapters and extended the qualitative contribution of chapter 3, concerning software requirements for disaster response organisations, with reflective data and novel perspectives on user interface design for disaster information systems.

The feedback data collected during this study demonstrated the findings and conclusions of the previous chapters to align with the needs of disaster response organisations and thus represent a validated approach to the integration of social media data as intelligence in disaster response.

8.1 Digital Tools for Disaster Response Analytics

Digital tools which interact with social media platforms are a common feature in disaster response processes (see chapter 3) and provide intelligence analysts access to rich feeds of data generated by a wide range of geographically-distributed authors. By monitoring these feeds, an analyst may observe eyewitness accounts from the scene of an event or identify emerging events, thus supplementing traditional sources of intelligence with online social data.

A study of the software currently used by, or available to, disaster response practitioners identified the core products used for social media analyses by these organisations and established that most software solutions were designed for a general audience or the domain of brand management, and were effectively co-opted by the disaster response analysts as a product of best fit (Marbouti and Maurer 2016).

The qualitative study of disaster response organisations in chapter 3 documented the characteristics of social media data used for disaster response intelligence and concluded that the software requirements unique to the domain of disaster response were not adequately met by the products currently available on the market. Consequently, the tools that these organisations use for social media analysis are varied and a dominant product meeting the requirements of the disaster response domain has not emerged.

This section summarises the key design principles used in existing software products and notes the dimensions in which they fail to meet the domain requirements of the intelligence analyst. These principles inform the design of the prototype software developed later in this chapter and present opportunities to explore how domain-informed design choices can improve the value of social media data in disaster response intelligence.

8.1.1 Software Categorisation

The social media intelligence software available to disaster response organisations ranges from simple consumer-facing web interfaces to complex suites of analytic tools. The systems documented in section 2.2.3 were broadly classified into four illustrative categories and are summarised in table 8.1.1.

General Clients — systems designed for the general user and typically available as free web applications. Examples includ the standard home pages of social media platforms and alternative clients for Twitter such as TweetDeck,[1] which are made possible through the Twitter API. As these systems are designed for use by the general public, embedding them within existing intelligence processes requires minimal investment or training and were therefore popular in organisations that had not yet recognised social media data as a valuable resource. As products designed to allow users to create and consume social content, their analytic capabilities are extremely limited.

Social Media Managers — suites of software which allow users to monitor multiple kinds of social media account. Typically used by brand managers and marketing personnel, their analytic tools are designed for brand-related purposes — for example, to monitor sentiment in relation to a brand name, or identify emerging competition. These trend analyses are structured around *aggregations* of data rather than providing the individual perspectives useful to disaster response. While these suites require paid licenses, larger response organisations operating with a public

[1] While TweetDeck was bought by Twitter in 2011, it was originally developed by a third party as an alternate interface to the platform.

relations team typically used social media managers and therefore there was little financial cost in extending access to the intelligence team.

Custom Software — software developed specifically for an organisation. This class of software was observed to have been implemented in various law enforcement organisations, where regulatory constraints place strict requirements on analysis software. For example, a participant from a police organisation discussed how the laws defining the point at which observing the public is considered *surveillance* (therefore requiring authorisation) limited the kinds of social media analyses they could conduct. The features of these systems were generally kept confidential, though it is reasonable to assume that they provide significant improvements in functionality over the general tools previously mentioned. Bespoke software is comparatively expensive and therefore limited to organisations that have a considerable budget and compelling use cases which justify the expense.

Academic Software — systems developed as products of academic research which targeted the domain of disaster response. While the initial purpose of these tools is typically to facilitate a study, a number of projects deliver a viable analytic tool that is published as open-source software. The lack of ongoing support for these projects and their relative obscurity limited their availability to practitioners, and no references to such systems were made by participants in this research.

	General Clients	Social Media Managers	Custom Software	Academic Software
Analytic Features			●	
Domain Specific	○			●
Free	●	●		●
Multi-platform	●	●	●	

Table 8.1: Feature comparison of analytic software categories.

8.1.2 Limitations of Existing Tools

The evaluation of the capabilities of available social media analysis tools was grounded in the conceptual framework derived from the ethnographic study of disaster response organisations (chapter 3) and the characteristics of Twitter data identified through empirical studies (chapters 6 and 7). Several key limitations of current systems were documented and exemplified areas in which a domain-informed software design process could better align software features with organisational requirements. These findings informed the design of the prototype discussed in the following section and are presented here as a contribution to related development:

Insufficient support for author inspection — Inspecting the profile of an author was a fundamental step in evaluating the veracity of a Tweet. For authors determined to be local to the observed event, the feed of their previously published Tweets provided further informative material and was a valuable source of data. The process of author inspection was therefore one of the most commonly performed tasks in disaster response intelligence, though was not actively supported by existing tools which prioritised the consumption rate of messages and aggregative visualisations. Due to the high volume of messages generated during a disaster, an embedded process of author inspection which does not require the analyst to navigate away from the main interface would improve the rate at which data can be processed.

Lack of author discovery — The phenomenon of *geographic homophily* was established in chapter 7 which showed that author accounts local to an event were more likely than average to form relationships with other local authors. Given a known local author, further local authors which may not have been otherwise detected can be identified by inspecting relationship networks. This approach of author identification does not fall within the typical use cases of the traditional users of these tools and the process is therefore poorly supported.

Inadequate visualisations — Visualisations provide useful summaries for large volumes of data. Where such figures existed (i.e. within the *social media manager* category of software), their focus was on queries more relevant to brand managers than disaster responders. For example, depictions of demographic data

and sentiment trends were common but not considered useful by the disaster response analysts who required figures supporting tasks such as geoinference.

Inability to export data — Many disaster information systems combine information from multiple sources. For example, the verification process of *method triangulation*[2] may combine satellite imagery, official news reports, and verbal statements to establish a ground truth. A novel data source must therefore integrate with these systems by passing relevant data into a pipeline using a suitable format.

Limited data filtration parameters — Defining advanced rules to filter data feeds is useful for a number of tasks. For example, many undesirable automated or business accounts can be eliminated by filtering by Tweet source (i.e. the software used the publish the Tweet — see section 6.3.3). Where filtering rules were available, the selection of parameters was limited and failed to support queries relevant to the domain.

8.2 Prototype Development

A prototype software system was created to implement the classification methodology discussed in the previous chapter and validate its findings. The software provided an interpretable representation of a social media intelligence tool which was presented to emergency response practitioners for comparison with existing tools. As the goal of the comparison was to validate the underlying data filtering mechanisms, the initial software specifications were designed to otherwise emulate existing tools and therefore did not aim to significantly implement new features. A second system which introduced new features based on use cases derived from findings of the previous chapters was developed to address the key limitations identified in section 8.1.2 and explore the potential for novel visualisation tools to serve analyst processes.

[2]A method of verification which combines data from multiple sources. Discussed in section 3.3.2.

8.2.1 Requirements Specification

The core requirements for the prototype were drawn from the findings of the qualitative study from chapter 3, which documented several key design considerations, and the study of Twitter data from chapters 6 and 7, which identified the patterns of behaviour within online public discourse providing valuable information to disaster response intelligence. The elicitation of requirements was built into the qualitative methodology using an unstructured interview format in which the participants were asked to describe the functionality of a desirable social media intelligence platform. The interview data were formalised into a set of requirements based on the processes discussed in J. M. Corbin and A. Strauss (1990) and A. L. Strauss (1987) and further shaped by the findings of the Twitter data studies (chapters 6 and 7).

As the prototype implemented an automated decision-making system that was essentially invisible to the user, the requirements defining the implementation of the filtering algorithm were not naturally discussed by the participants and were therefore synthesised by the researcher based on data collected from interviews and observations (Vogelsang and Borg 2019).

A number of requirements were related to features of a comprehensive and mature social media intelligence platform and were therefore not relevant to the design of this prototype. This was a natural outcome of an unguided requirement elicitation process in which participants were asked to discuss their needs in a broad sense, without knowledge of the particular aspects of the workflow which the resulting intervention was designed to improve. The key requirements which were identified as relevant to the design of this prototype are listed below.[3] A summary of table 3.3, which lists the key values and challenges of using social media data for disaster response intelligence, is provided for reference in table 8.2.

[3]The comprehensive list of requirements developed for this project is not relevant to this research discussion and therefore not presented.

ID	Thematic Category
1-V	Detect Emerging Events
2-V	Situational Awareness
3-V	Rumour Correction
4-V	Identify Urgent Needs
5-V	Public Engagement
6-C	Datum Volume and Noise
7-C	Datum Veracity
8-C	Organisational
9-C	Integration

Table 8.2: A summary of table 3.3 listing key values and challenges of social media data for disaster response intelligence.

Functional Requirements

R-1 **Deliver situational awareness**

The core functionality demonstrated by the prototype was to provide the operator with a source of situational awareness drawn from reports created by Twitter users in the area under observation. This feature was grounded in the values identified during the qualitative study and described in section 3.3.1 as 2-V, and to a lesser extent, 4-V and 3-V. The class of data which delivers 'situational awareness' was not strictly defined (see section 3.3.3) and included messages containing any combination of text, photo, or video media.

R-2 **Support user inspection**

The intelligence operator could inspect data related to a given message, particularly with respect to the message author. By viewing the profile data of an author in combination with a history of their previous messages, the operator was able to evaluate the authenticity of the author and the relevance of their message to the observed event (see 7-C in section 3.3.2). This human-verification layer supplemented automated filtration methods and embedded human values in the classification process. Furthermore, manual inspection of an eyewitness author's historical feed often yielded further valuable data not identified by the automated collection methodology.

Non-functional Requirements

R-3 **Provide a suitable proportion of meaningful content**

One of the fundamental challenges facing more widespread adoption of social media as intelligence was the inadequate signal-to-noise ratio within the data (challenge 6-C in section 3.3.2). A viable system should therefore present a curated sub-feed of data for which the signal-to-noise ratio is within an appropriate range. The strict values of this range were undefined and dependent on complementary factors including, for example, the format in which data is presented. Addressing this requirement was the focus of the previous chapter.

R-4 **Minimise training requirements**

User operation of the system should be intuitive and require little training (see 9-C in section 3.3.2). This was particularly relevant with respect to a prototype designed for validation, for which users would have limited incentive to learn a new interface, and friction in usability would influence the quality of the validation results. For this reason, initial interface design was based on systems currently in use by participant organisations as recommended in section 3.3.3.

R-5 **Integrate with existing procedures**

Samples drawn from the data and insights made by the operator should be easily integrated with existing intelligence processes (9-C in section 3.3.2). For example, data were validated by comparing reports from alternative sources through a process called *method triangulation* (discussed in 7-C, section 3.3.2), and validated reports were then passed to a unified database for further action. Therefore, the prototype was designed to export data in a standardised format to enable integration with complementary software.

8.2.2 Software Design

The purpose of the prototyping process was not to create a robust tool to be used in practice, but to provide an interpretable framework from which insights could

be drawn from feedback provided by disaster response analysts. It was therefore not necessary for the prototype to innovate significantly upon established design standards, though the data from the initial qualitative study (chapter 3) suggest that the designs of software currently in use by disaster response organisations do not closely align with domain requirements and therefore present opportunity for improvement. User interface specification is a complex process that is not supported by the data collected during this research and therefore the design of the primary prototype follows principles adopted by the software currently used by analysts, as observed in section 8.1. A second prototype was developed in parallel which explored how the inclusion of novel features could better support analyst processes. The two systems are hereafter referred to as $P_{emulated}$ and P_{novel}, respectively.

By modelling the user interface design of $P_{emulated}$ on existing software (primarily, *TweetDeck*), the secondary advantage of user familiarity was leveraged to minimise the training requirements faced by participants. This reduced the extent to which confounding factors influenced the validation data: the study was designed to elicit feedback on the underlying classification algorithm and as such, using an interface which deviated from the control software would obfuscate the underlying determinants of the feedback derived from the study.

In terms of integration with existing disaster information systems, reducing the need for training improves the viability of a new system in disaster response environments, as captured in requirement R-4. The data from which this requirement was derived documented the inability to allocate time to training staff as a factor preventing the adoption of new software otherwise identified as an improvement on current systems (9-C in section 3.3.2). The approach of interface emulation should be balanced with the development of new features when designing software for disaster response organisations to minimise the training burden and improve the viability of a novel system.

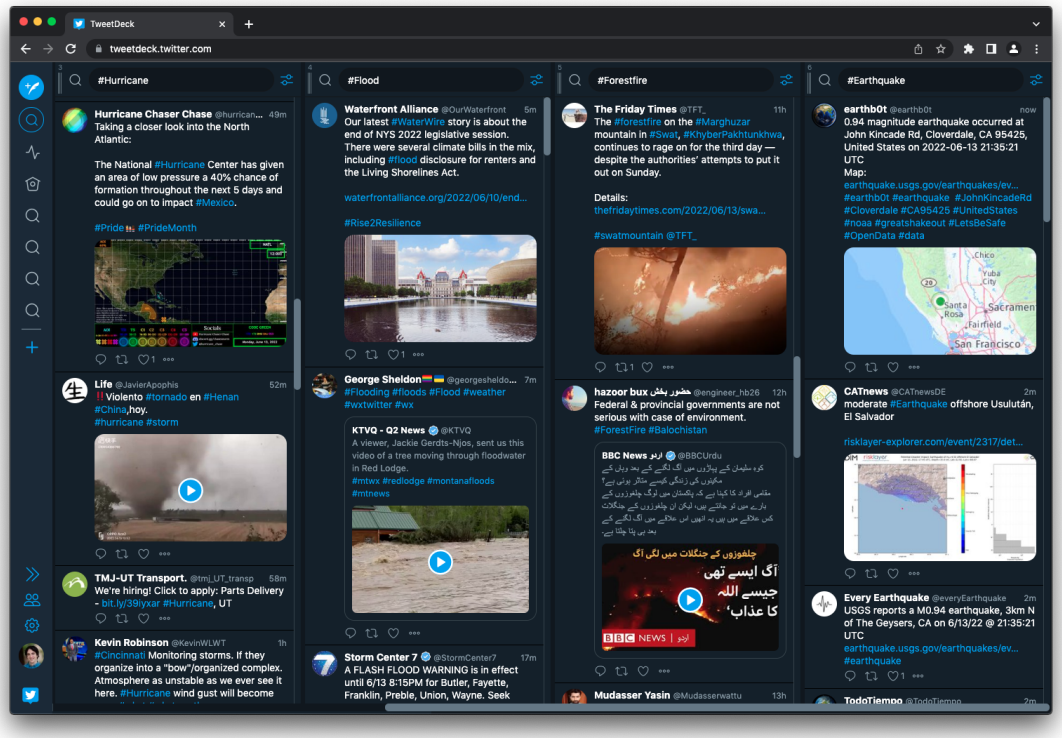


Figure 8.1: An example of an existing system, TweetDeck[4], which demonstrates a four-column view of various Twitter feeds.

P_{novel} implemented a number of new features and acted as an alternative to the design of $P_{emulated}$. The features were selected to extend the existing system and support the tasks defined by functional requirements R-1 and R-2. Participants were therefore presented with three systems: an existing familiar system currently used by analysts, $S_{TweetDeck}$, a new system resembling the familiar interface, $P_{emulated}$, and an alternative advanced system which introduced novel features, P_{novel}. All three systems were populated with the same simulatory dataset and the latter two identically implemented the data classification algorithm proposed in the previous chapter (requirement R-3). The purpose of presenting the advanced interface of P_{novel} was twofold: first, to validate the findings of section 8.2.1, wherein interface-based requirements were defined, and second, to elicit feedback which could further inform the interface design of disaster information tools. P_{novel} was therefore used as the primary prototype of the study, while $P_{emulated}$ provided

[4]https://tweetdeck.twitter.com/

an alternative which was used to focus feedback on the underlying classification decisions as discussed in section 8.3.2.

The design of P_{novel} aimed primarily to streamline the workflow of the analyst. The earlier study in chapter 3 identified processes that were not adequately supported by existing software and therefore limited task efficiency. An instructive example of this was *author inspection*: when viewing a Tweet, the analyst typically examined the profile of the author to evaluate the veracity of the Tweet content and to identify other useful Tweets published by the author. P_{novel} streamlined the process of author inspection by displaying author information on the same page as the Tweet feed, allowing the analyst to quickly view author data without leaving the main feed interface or opening multiple windows (requirement R-2).

A number of visualisations were developed for P_{novel} which summarised event data, including time series charts showing Tweet volumes for each day and histograms aggregating both the type of Tweet (i.e. original post, retweet, reply or quote) and the originating software from which they were published. These figures were implemented to support common analyst queries related to the judgement of author veracity. For example, a simple first-pass procedure for bot detection observed in chapter 3 involved the inspection of daily posting patterns to detect users whose history did not align with a typical circadian cycle and therefore were likely to be automated or multi-user accounts. The `hour / day of week` heatmap shown in figure 8.2 demonstrates an example of this circadian behaviour, where a period of low activity is clearly defined during the nighttime hours. Note that the selected data in the chart aggregate all users and therefore while the low activity period of 'sleep' is apparent, it is not as clearly defined as it may be for a typical single human user.

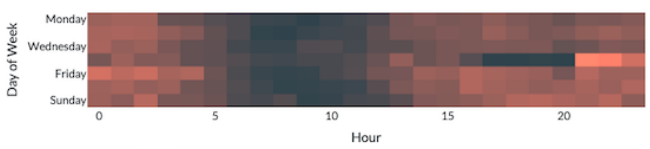


Figure 8.2: The heatmap figure shows a clear circadian cycle and can be useful for detecting automated accounts.

An interactive map[5] displayed the originating points of geolocated Tweets to provide a visual layer of verification of the data collection parameters and support emergent processes such as the identification of unexpected geographic clusters of activity (requirement R-1). Finally, a visual representation of the follower/followee network used by the classification algorithm was added to explore ways in which these representations of data could inform the construction of human-defined queries (with respect to requirement R-2) by observing analyst interactions with chart data during the validation study. A screenshot of P_{novel} populated with synthesised data is presented in figure 8.3.

The figures were populated with the results of an underlying database query defined by analyst-selected values. The default output of the query was a subset of data sampled randomly from the database, of a variable sample size based on the capability of the hardware (typically between 2,000-3,000 Tweets). Further constraints could be added to the sampling query by selecting data within the graphs — for example, the analyst could specify a selection of data from a single author, sub-location, or given range of dates by selecting the relevant areas of the figures or Tweet feed. Constraints could be defined additively using a shift-select action. After each selection event, the database query would be reformulated and executed, and the figures and Tweet feed updated using the new set of results. This allowed the analyst to focus their investigations by constraining the displayed data based on their categories of interest. Selected data could be exported as a JSON[6] text file; an open-standard file format common for data interchange (requirement R-5). By adhering to an open standard, compatibility with third-party software and communication protocols was maintained.

The dashboard interface of P_{novel} was built using the Dash framework from Plotly,[7] an open-source Python package that extends the Plotly Python graphing library.[8] Dash was chosen for its implementation of interactive figures with live

[5]Map data obtained from OpenStreetMap (https://www.openstreetmap.org/)
[6]https://www.json.org/
[7]https://dash.plotly.com/
[8]https://plotly.com/python/

updating; a feature that allowed the analyst to conduct user inspection while maintaining the single window constraint discussed above. Data were stored in a Postgres[9] database using the same structure as the data collection software (developed in chapter 5) such that the two systems were entirely integrated. In practice, fresh data recorded by the data collection process were periodically requested from the database by the dashboard application and figures updated accordingly, though for the purposes of the study, this process was simulated using a synthesised dataset with a predetermined schedule defining the moment of 'detection' (that is, the point at which the data became available to the system).

Requirement	$S_{TweetDeck}$	$P_{emulated}$	P_{novel}
R-1 Deliver Situational Awareness	○	○	●
R-2 Support user inspection			●
R-3 Suitable signal-to-noise ratio		●	●
R-4 Minimise training requirement	●	●	
R-5 Integrate with existing procedures			●

Table 8.3: Requirement satisfaction matrix for validation software. Note that while all systems satisfy requirement R-1, P_{novel} provides a significantly higher level of support through the presentation of data visualisations.

8.3 Evaluation

The findings presented in the previous chapters of this thesis were evaluated in this study using a set of three pieces of data analysis software: an unadulterated existing system ($S_{TweetDeck}$ — figure 8.1), a replication of this system which implemented the classification algorithm from chapter 7 ($P_{emulated}$), and a custom dashboard designed using requirements derived in section 8.2.1 using the same algorithm (P_{novel} — figure 8.3). Each system was populated with an identical set of data simulating a live event and presented to a group of disaster response analysts for evaluation.

[9]https://www.postgresql.org/

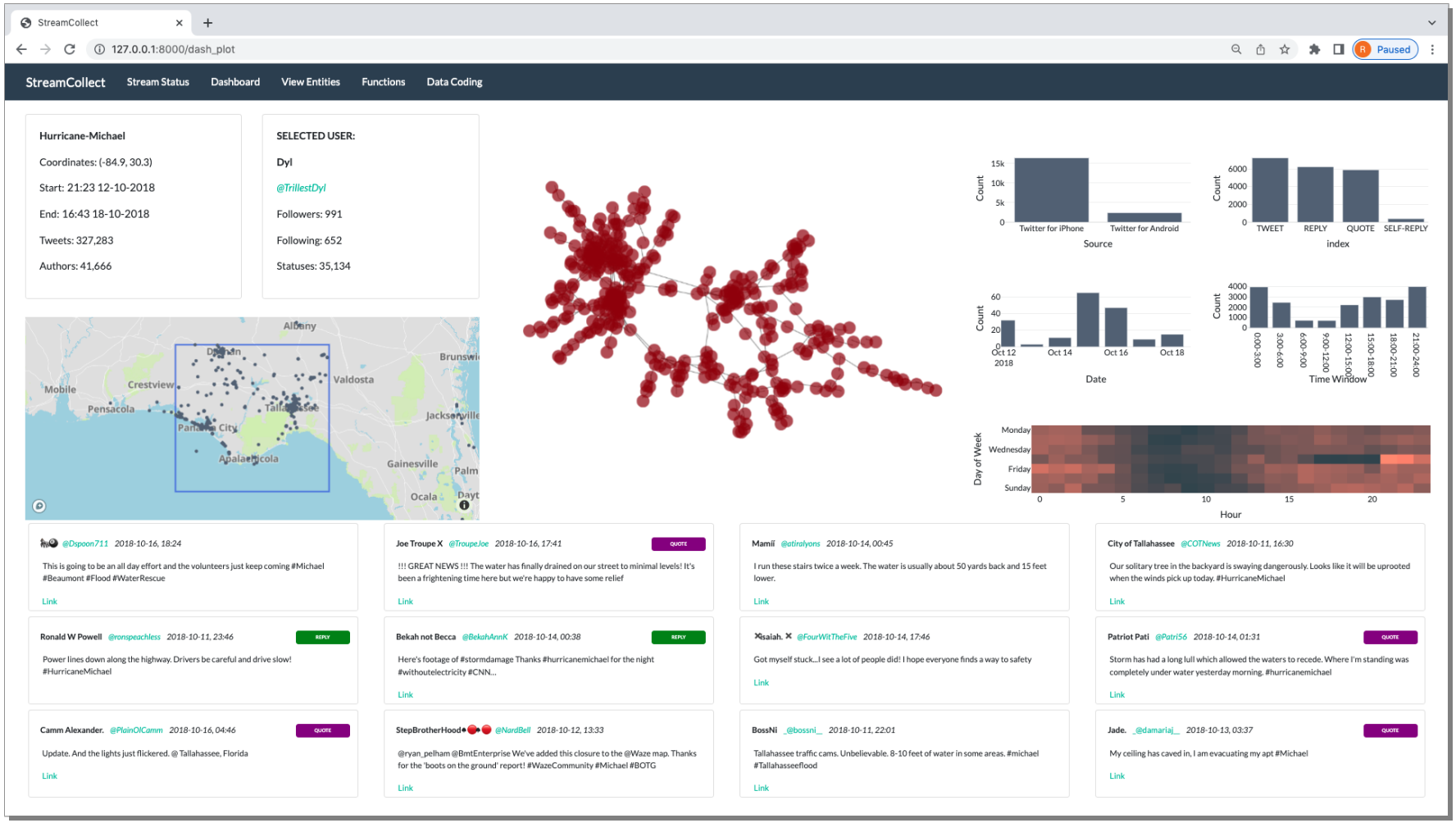


Figure 8.3: The event interface of the prototype showing author follower network, various data visualisations, and a sample of (synthesised) messages.

The data collected during this study demonstrated a clear preference of participating analysts for the curated data feed and provided valuable insights informing future design practices. The proposal of this thesis to implement a classification algorithm based on author geoinference was therefore validated by these participants as an appropriate method to address the key challenges facing disaster response analysts when using social media data as intelligence and provides a grounded perspective from which future validation studies may be conducted. Furthermore, the strength of requirements specification processes grounded in disaster response perspectives is established through the realisation of improvements to interface design derived from domain-informed requirements.

8.3.1 Participant Selection

Participants for the study were drawn from the disaster response organisations approached during the primary qualitative study of chapter 3. The initial recruitment method is described in section 3.1.4 and selection of the subset of participants who provided the evaluative data of this study was based on how closely their role aligned with that of the *disaster intelligence officer*, for whom the prototypes were designed. Reintroducing these prior participants ensured that the perspectives embedded within the evaluative data were aligned with the requirements captured in the earlier study.

This study included a total of six participants encompassing the same breadth of domain as the initial cohort. A summary of participant roles is provided in table 8.4 in which participant labels are maintained from the earlier table 3.2. Names remain redacted as discussed in section 3.1.5.

Domain	Participant	Position
Police	P3	Superintendent
Fire Department	P6	Team Manager
Emergency Response	P9	Intelligence Officer
	P10	Deputy Head
	P13	Specialist
Humanitarian Organisation	P16	Manager

Table 8.4: Subset of research participants from the study in chapter 3 who contributed to the evaluation study.

8.3.2 Method

The feedback elicitation process involved observing the completion of guided tasks using each software system followed by a semi-structured interview. Participants were first invited to reflect on the perspectives they had offered in the previous study and were then presented with prototype software that implemented the classification algorithm while modelling the interface of an existing system with which participants were familiar ($P_{emulated}$), thus minimising the influence of user interface design in feedback data.

A period of familiarisation with the interface was followed by a set of guided tasks simulating typical use cases based on a synthesised dataset (table 8.5). The primary outcome of this stage of the study was to expose the participants to a range of experiences from which they could draw when providing feedback during the subsequent interview. The tasks were designed to encompass a range of intelligence-gathering processes modelled on data collected during the earlier observational study (chapter 3). Goals were defined broadly, allowing the participants to align the tasks with processes familiar to their own experiences.

Following observation of the guided tasks, semi-structured interviews were conducted to elicit feedback from the participants. The interviews began by discussing the previous tasks with respect to how well the core functionality of

the prototype supported intelligence-gathering processes. A representative piece of software currently used by response organisations ($S_{TweetDeck}$) was then initialised using the same synthesised dataset and participants were asked to repeat the tasks to provide a baseline from which comparisons could be made.

Due to the latent nature of the data filtering processes used in the prototypes, initial feedback was naturally based on more apparent features such as interface design. To counteract this behaviour, guided questions focused participant attention on the quality of the content presented and the comparison to the uncurated data. The general format of the interview questions is included in appendix A.1.2.

Finally, the novel prototype software developed in section 8.2 (P_{novel}) was presented and the tasks repeated a third time. Once the primary tasks were completed, participants were encouraged to perform their own analyses and experiment with the novel features of the software. As a secondary outcome, therefore, this directed observational study enriched the existing dataset with new observations focused on analyst interactions relevant to this study.

Task
Add tracking keywords to instantiate event monitoring.
Define geographic region to instantiate event monitoring.
Observe data feeds for items of interest.
Examine Tweet datum identified as *useful*:
– Inspect author details.
– Evaluate veracity of content.
– Identify content related to observed Tweet.
– Identify and evaluate related authors.
– Repeat for 2-3 Tweets of various formats.
Examine Tweet datum identified as *unuseful*:
– Determine whether author is an automated account.
– Remove author from dataset.
Conduct participant-defined analyses.

Table 8.5: List of tasks performed by participants during observational study of prototype.

The first round of validation served as a pilot study and informed the refinement of the prototype systems and definition of tasks to be performed by subsequent participants. The most significant changes implemented between these rounds of study were made to the task phrasing and structure to improve clarity. Observations and interviews were conducted remotely as a single session per participant using a combination of video conferencing and screen sharing with each round lasting approximately three hours. While *in situ* observation was preferred for observational research, the study was conducted in 2020, during which time travel was restricted due to the COVID-19 pandemic.

A set of success criteria were defined based on the analyst perspectives discussed in section 8.2.1 and are presented in table 8.6. These criteria ensured that the validation of the research was grounded in the requirements defined by the end users. Note that as results were based on participant feedback, requirement R-5 (*integrate with existing procedures*) was not validated.

Objective	Requirement
O-1 Support for desired analytic queries.	R-1, R-2
O-2 Sufficient quality of data.	R-1
O-3 Sufficient ratio of quality data.	R-3
O-4 Acceptable time-to-complete for analytic queries.	R-1
O-5 Intuitive interface design.	R-4

Table 8.6: List of success criteria defined for P_{novel} evaluation.

Data coding followed the approach of section 3.2, which was based on J. M. Corbin and A. Strauss (1990). As an evaluative study for which the objectives were predefined, a less rigid application was used. *Selective coding* was therefore prioritised to systematically categorise data relevant to the research question while *open coding* was used to identify emergent themes.

8.3.3 Results

The results of the feedback were coded using two key categories: *User Interface*, which described data related to the selection of the figures and visual design of P_{novel}; and *Data Curation*, which captured perspectives comparing the raw data feed with the algorithmically-curated feed. This latter code therefore represented data validating the work of chapters 6 and 7, while the first category validated the findings of this chapter (and in turn, chapter 3) and provided useful insights informing further work in disaster information system design.

User Interface

User interface is, by definition, the most apparent and interpretable feature of software and as such elicited a large amount of feedback. As P_{novel} was designed specifically for the domain of the participants, it naturally provided a more customised interface than existing solutions and therefore better supported their workflow. While the interface design of P_{novel} was exploratory, this feedback provided useful insights which inform future work developing disaster information systems.

In the following quote, P9 highlighted the value of a streamlined workflow, referring to a process in which they observed the filtered feed of data and chose to inspect the author of a given Tweet to evaluate their authenticity:

> *'I like that I can see the [author] information without leaving the [Tweet] feed. Opening lots of windows can create a mess that makes it hard to remember what they relate to.'* **(P9)** (8.1)

Similarly, P3 described the critical importance of author information and consequently the value in streamlining author inspection when analysing the message feed:

> *'The [author] details are arguably more important than the message. Or at least, they're intertwined, and we have to look at them together ... That was much easier with [P_{novel}].'* **(P3)** (8.2)

Feedback favouring existing software was expected given the immaturity of P_{novel}, which naturally does not implement the full set of features available in existing solutions. While the guided tasks were designed to fall within the bounds of implemented features, P3 identified use cases that were not supported:

> *'There are some questions [P_{novel}] can't answer, so I'd have to go back to [their existing software] for that.'* **(P3)** (8.3)

Interface familiarity was highlighted as a key consideration in section 3.3.3 to minimise the burden of retraining personnel on a new system. The significance of this concept was further strengthened in the evaluation data. For example, P13 linked the effect of their unfamiliarity with the new interface to an increase in the time required to perform a task:

> *'Obviously I'm less familiar with [P_{novel}] but I felt more at home with [$S_{TweetDeck}$], so [the task] was faster to do with that.'* **(P13)** (8.4)

The speed of task completion was raised by other respondents during the interview

process and largely seen as having been improved by the features of P_{novel}. The greater efficiency of the novel system was attributed to the domain-driven design, which embedded the unique requirements of the disaster analyst in the interface layout. In the following quote, P9 reflects on the increased efficiency with which they were able to complete a task using P_{novel} compared to the system they currently use:

> *'Trying to work that out would take much longer on the system that I use, so designing a tool that addresses my needs specifically would make my job a lot easier.'* **(P9)** (8.5)

The strategy of minimising the requirement for training by designing an interface resembling existing software was further supported by P10, as shown in the following quote, in which they refer to how $P_{emulated}$ resembled familiar software and discuss a seamless deployment model requiring no retraining:

> *'I can see how you could redesign [$P_{emulated}$] a bit and just drop it in without people knowing. Then you'd start adding your new features behind the scenes.'* **(P10)** (8.6)

A challenge to requirements elicitation for systems implementing novel artefacts is demonstrated in the following quote from P9, wherein they consider the possibility of a new analysis process enabled by the data visualisation presented in P_{novel}. Capturing this class of emergent requirement requires a tightly-coupled design process in which the designer formulates and presents artefacts that may be outside the scope of the initial requirements provided by the participants, and is therefore reliant on a strong domain understanding by the designer.

> *'The [visualisation artefact] made me think that checking the sleeping patterns of the [author] would be a good starting point for bot detection. I wouldn't have thought of that if it wasn't there.'* **(P10)** (8.7)

Data Curation

The primary outcome of this research was the development of a classification algorithm that could filter a raw stream of social media data and present a curated feed of relevant messages to the operator. While this filtration process occurs before any data is presented to the observer, participants were invited to discuss their thoughts on the overall relevancy (i.e. signal-to-noise ratio) of the curated feed and presented with an unfiltered 'control' feed for comparison. All six participants reacted positively to the prototype demonstration, expressing a preference for the filtered feed over the raw feed. These results showed that the implementation of the geoinference algorithm for data curation effectively addressed the primary concern identified in section 3.3.2 of *Datum Volume and Noise* (6-C). P16 describes a noticeable improvement in data relevancy when comparing the two feeds in the following observation:

> *'It's clearly removed a lot of the chatter from the original search. This is mostly useful stuff.'* **(P16)** (8.8)

P10 expressed a similar sentiment, though was unwilling to quantify the improvement. It is informative to understand that as the usefulness of the uncurated data feed varied between domains based on their unique needs, the direct comparison of the feeds was less significant than the *overall usefulness* of the curated feed. Therefore, while the magnitude of the curation may not have been intuitively quantifiable by the analyst (though it *was* quantified in chapter 7), the validity of the curation approach was predicated on the feed achieving a ratio of signal-to-noise palatable to the analyst.

> *'It's hard to tell exactly how much better it is than the [uncurated data], but it's definitely a rich feed and there's a lot of value in there.'* **(P10)** (8.9)

When asked whether the curated feed was more suitable as a source of intelligence

data than systems currently in use, participants expressed an overall favourable response. The following quote from P6 demonstrated this positive reaction whilst noting that further development would be required to achieve a suitable level of functionality:

> *'If it works for other sites, this type of thing would be a great addition to our suite.'* **(P6)** (8.10)

The participant organisations in chapter 3 included a number that had not implemented social media intelligence processes. P13 represented one such organisation and showed an interest in introducing social media intelligence based on their interaction with P_{novel}.

> *'There's definitely useful information here. It's something I'd want to start looking at.'* **(P13)** (8.11)

The issue of *Datum Volume and Noise* (6-C) was re-identified as a key obstacle by P9, who observed that the filtered stream reduced irrelevant content to an appropriate level. In the quote below, the term *spam* has been used to refer to messages deemed unrelated to the intended goals of P9.

> *'It was the spam that kept us from using this type of data. The level here seems manageable.'* **(P9)** (8.12)

Participant P3, a police officer, raised concern over data incorrectly removed from the feed by the classifier (i.e. *false negatives*). This was the only case of negative feedback with respect to curation and was not mentioned by the other participants. It is, however, a key consideration in the data curation approach and is discussed in the next section.

'I would be worried about what I'm missing out on, how much data is being hidden ... Is what I'm seeing representative of the situation?' **(P3)** (8.13)

Outcomes for the success criteria defined in table 8.6 were derived from participant feedback and followed a three-value schema: *passed*, *failed*, and *improved*. *Passed* denoted cases in which P_{novel} was deemed to have satisfied the criterion and *improved* signified that P_{novel} provided a better solution than existing software (exemplified by $S_{TweetDeck}$ but including other systems familiar to the participants). Table 8.7 summarises the results and links the success criteria with the supporting data presented in this section.

Objective	Result	Supporting Evidence
O-1 Support for desired analytic queries.	Improved	8.1, 8.2, 8.3, 8.7, 8.10
O-2 Sufficient quality of data.	Passed	8.7, 8.10, 8.13
O-3 Sufficient ratio of quality data.	Improved	8.8, 8.12
O-4 Acceptable time-to-complete for analytic queries.	Improved	8.4, 8.5
O-5 Intuitive interface design.	Passed	8.1, 8.2, 8.6

Table 8.7: Evaluation of success criteria for P_{novel}. The *result* column marks criteria in which P_{novel} exceeded $S_{TweetDeck}$. The *supporting evidence* column references datum excerpts presented in this section.

8.4 Discussion

The feedback data presented above show a clear positive response to the application of the methods developed in this research. The user interface of P_{novel}, built upon requirements derived from the domain of disaster response, was seen overwhelmingly by participant disaster response practitioners to better support their goals, establishing the importance of a domain-informed requirements elicitation process.

The data curation performed by the author network geoinference algorithm demonstrated a clear improvement to the raw data stream and effectively addressed

a key challenge identified in the earlier study. While the prototypes provided an early-stage implementation of the data filtration model, the results they presented were supported in participant feedback. The reduction of *noise* in the data stream was viewed as an overall favourable outcome, offering an observable improvement to existing solutions and effectively eliminating a key barrier preventing (at least) one participant organisation from using social media data for intelligence. While the classification algorithm represents only a single aspect of what a suitable disaster information system would comprise, its practical potential in social media intelligence has been validated by the feedback presented in this chapter.

User Interface Requirements

While user interface and user experience (UI/UX) research were not the focus of this study, a significant proportion of feedback was coded within this theme, clearly signifying UI/UX as a valuable dimension in which existing software solutions are failing to meet domain requirements. The breadth of feedback coded within this theme demonstrated the extent to which domain requirements deviate from those of the public and are therefore not adequately supported by existing software. A selection of relevant requirements drawn from the feedback data is provided here and represents a novel contribution of this research, informing the further development of disaster information systems. While these data were derived from participants in the domain of disaster response, the findings provide a contribution to related fields in which open-source intelligence (OSINT) is studied.

Data visualisation: While a simple chronological data feed was implemented as the primary method of inspection, participants expressed a need for more informative visualisations which could summarise and identify interesting patterns in the data. For example, the heatmap figure included in P_{novel} (figure 8.2) immediately signalled patterns of behaviour which analysts used in the identification of automated, or *bot*, accounts. While such patterns *could* be distilled into a summary statistic, the lack of a clearly-defined boundary between the behaviour of bot and human accounts required analysts to evaluate a range of characteristics and make judgements based

on intuition. Therefore, visualisations that supported queries relevant to the analyst significantly improved the speed and accuracy of their decisions.

These findings suggest further research be undertaken to identify effective data visualisation and presentation formats using requirement elicitation techniques. Furthermore, the feedback data documented cases in which participants conceived of new analyses which were made possible or inspired by the visualisations presented in P_{novel}. The selection of data visualisation artefacts used in a mature system should therefore not be constrained to processes identified during requirements elicitation, but also introduce formats drawn from similar systems in unrelated domains. This process of experimentation, modelled in the principles of participatory design (Schuler and Namioka 1993), facilitates the emergence of novel use cases enabled through the application of modern technology.

Data manipulation: Data analysis was observed to be an iterative practice in which the analyst constantly explored and revisited data in a process of *sensemaking* (Muhren and de Walle 2009). Participants expressed a desire for improved data manipulation functionality to allow them to better explore incoming data and document emerging themes. Such processes included, but were not limited to, sorting data into subcategories, removing uninformative data from the feed, linking related data, and flagging significant author accounts. Manipulating, linking and organising data visually on a two-dimensional plane resembles the process of *mind mapping* and is exemplified in the popular OSINT platform Maltego.[10] These features allow an analyst to customise the layout of the data to best support the evolving demands of their work, therefore reducing the burden on the developer.

In describing the tasks for which improved data visualisation and manipulation were sought, participants commonly expressed their objectives as *Tweet author inspection* and *Tweet source categorisation*. Both processes were closely aligned with the desire of an analyst to infer the validity of an author and their Tweets, addressing *datum veracity*, identified in chapter 3 as a significant challenge limiting the value of social media data as intelligence (7-C). While more complex methods

[10]https://www.maltego.com/

of author validation have been explored in the literature (Masood et al. 2019; Shu et al. 2017; Sahoo and B. Gupta 2021; Khaled et al. 2018), the observed practice of author inspection demonstrated a validation process defined by the analysts and thus aligned with the unique requirements of their environment. Supporting these datum verification processes through improved user experience design should therefore be explored as a validated solution to the challenge of *datum veracity*.

Speed of inspection: Software systems that have not been designed for disaster response analysts are optimised for user experiences that do not align with analyst needs. Many queries commonly executed by analysts were therefore inelegant, often requiring multiple browser windows to be opened and compared. This interface friction resulted in slower analyst performance and dissuaded them from conducting otherwise informative queries.

In most cases, tasks were judged to be faster to complete with P_{novel} than with the existing system $S_{TweetDeck}$. This was an unsurprising result given that both the list of performed tasks and the features of P_{novel} were derived from the same set of requirements, though one participant observed that some tasks took longer to perform with P_{novel}. While speed of task execution was not timed during this study, and therefore the claim not validated, the simple perception of inefficiency represents a significant limitation that may affect the adoption tolerance of any introduced artefact. While outside the scope of this research, this feedback highlighted the importance of stakeholder involvement in system design to maintain alignment between feature development and end-user needs (further discussion of participatory design principles is available in Simonsen and Robertson (2013) and Baxter and Sommerville (2011)).

Data Classification

The focus of this research was to develop an effective classification algorithm that could process a raw stream of data and present the analyst with a curated feed of messages containing a higher proportion of useful content. The quantitative efficacy of the algorithm was demonstrated in the preceding chapter and the

feedback from the study presented in this chapter validated this approach using domain-based perspectives. Network-based author classification was therefore shown to be an effective technique in improving the utility of social media data for disaster response intelligence.

Concern over missing data, or false negatives, was raised by a study participant from a police organisation. While false negatives existed as a product of the classification algorithm, the impact of these misclassifications is only relevant under one of the following conditions: first, where the observational capacity of the analyst is not satisfied by the positively-classified data and therefore available data is exhausted. In this case, reducing the *precision* of the classification algorithm to increase *recall* is a simple adjustment that can be performed dynamically in response to the rate at which the analyst processes data. Second, where classification contains a systemic bias, data from a protected or sensitive class may become (for example) overrepresented in the set of false negative cases, leading to a failure of the curated feed to provide a representative sample of local author accounts. The significance of this bias is contingent upon the case for which the data are used and is more likely to be of relevance to analysts in law enforcement, however, it remains a fundamental risk and further research is required to identify hidden biases in this approach.

The feedback obtained in this study demonstrated a fundamental agreement between the values expressed by the disaster response personnel and the output of the classification algorithm, validating the technique of author geoinference using network data and the target outcome of developing situational awareness. While a small number of issues were identified during the validation study, the overall consensus of participants was positive, expressing an optimistic view towards the classification approach to improving the value of social media data. Participants from organisations with existing social media intelligence processes observed improvements in efficiency enabled by the prototype software and those from organisations without existing processes acknowledged that the classification algorithm effectively resolved the primary barrier preventing the integration of social media data with their existing systems (6-C). Author-centric network-based

geoinference was therefore shown to be an effective approach to improving the value of social media data to disaster response organisations during disaster events.

Implementation Pathways

Author-focused data curation built upon such an algorithm could be implemented in one of several ways, based on the existing infrastructure and processes of the target organisation. In the most simple form, curation occurs 'invisibly' as part of a data pipeline situated between the raw data feed and the system which takes the feed as an input. This approach ensures that the curated data conform to organisational standards and therefore integrate with existing software. The core challenge of retraining staff (9-C in section 3.3.2) is also avoided, though this introduces a risk of naiveté: that is, where staff are not explicitly aware that a curation process has been put in place, they may fail to take into account the natural biases introduced by curation and thus make invalid or harmful conclusions. The relevancy of this risk will depend upon the application for which the data is used and may be offset by suitable training.

The second challenge to the integrated approach is that the analysis performed by the algorithm is an ongoing process and as such, the value and confidence of a given classification prediction will change as the user network grows over the course of the event. A pipeline implementation therefore requires a heuristic defining the point at which the classification of a datum is considered correct and consequently passed on to the analysis software. This heuristic could be represented, for example, as a threshold value of confidence required to consider a datum as *local*, or a period of time since detection after which a classification outcome is locked in. In either case, the raw data feed is delayed from entering the analysis software, and while classification accuracy improves as the user network grows, data already passed on through the pipeline cannot be updated. This effect may result in an increase in the number of `false positive` cases as `negative` cases are prematurely classified as `positive` and then unable to be updated once passed on to the pipeline. Alternatively, an increase in the number of `false negative` cases may occur where

`positive` cases do not reach the heuristically-defined threshold for advancement and therefore remain undetected. The definition of such a heuristic must be made with careful consideration of the unique characteristics of the relevant use case.

Limitations of Simulation

The method used for feedback elicitation has several key limitations. Firstly, participants performed a series of tasks which were designed to simulate typical processes informed by data from the earlier study in chapter 3. Simulations are a common approach in the study of disaster information systems, where conducting studies during live response events is not feasible, however, they are constrained by a number of factors (Moats et al. 2008; Helsloot 2005). The isolated simulations used in this study were not designed to capture the communication between an analyst and the stakeholders with which they interact during a disaster event and therefore the effect of organisational hierarchy is not comprehensively modelled. In practice, this may lead to a mismatch between what the analyst believes to be a desirable outcome (and is thus expressed in the data) and what is demanded by their organisation. Furthermore, environmental factors such as time pressure and evolving organisational priorities are not simulated and thus the validity of the data relies upon the participants' ability to integrate these effects into their responses.

Secondly, the synthesised datasets used in the simulations were based on the hurricane event datasets from which the algorithms were designed. The extent to which the results of this study are generalisable to other types of event requires further research, though participant feedback suggested that the classification method of data enrichment would be useful for the domains in which they were familiar.

8.5 Summary

The study presented in this chapter has identified key requirements relevant to the domain of disaster response intelligence and validated the findings of chapters 6 and 7 based on qualitative feedback derived from disaster response practitioners.

An analysis of existing digital tools was conducted within the context of disaster response using domain data collected during the qualitative study in chapter 3 and the results of an analysis of Twitter hurricane event data in chapter 6. A set of requirements was defined which addressed the areas in which existing tools failed to meet the needs of disaster response analysts and was used for the design of a prototype software system implementing the classification algorithm developed in chapter 7.

A set of six participants was drawn from the pool of disaster response analysts involved in the chapter 3 study to represent practitioners from a range of disaster response domains. The prototype was presented together with a system previously observed to be in use by several participant organisations and each populated with a set of synthesised data simulating a real-world disaster event. Qualitative data were collected using a method of observation, in which participants were requested to perform a range of guided tasks using each system, followed by semi-structured interviews analysing the decisions made during the simulation.

The feedback data strongly validated the author-classification approach to improving the value of social media data in disaster response, with participants confirming that the prototype system effectively addressed their primary concerns with using social media data as a source of intelligence (managing datum volume and improving the signal-to-noise ratio). Furthermore, the data presents a number of insights grounded in the domain of disaster response which support the requirements presented in this chapter and inform the user experience design of disaster information systems.

The primary contributions of this chapter were therefore the situated validation of the algorithm developed in chapter 7 and a documentation of the software requirements of disaster response practitioners, which include novel domain-driven perspectives on user interface design and further enhance the findings of chapter 3. Through the results of this study, the outcomes of the previous chapters were shown to have correctly identified and addressed the requirements of disaster response organisations in improving the value of social media data as intelligence, and a

pathway towards further integrating the models developed in this research has been presented.

9 Discussion

Contents

9.1 Thesis Summary 258
9.2 Research Findings 262
9.3 Research Contributions 266
9.4 Limitations and Challenges 269
9.4.1 Participant Access During Disasters 269
9.4.2 Data Bias and Technology-Induced Privilege of Care . . 270
9.4.3 Psychosocial Safety and Secondary Trauma 271
9.5 Further Research 272
9.5.1 Preparing for Live Deployment 272
9.5.2 Evaluating Generalisability of Findings 272
9.5.3 Analysing Bidirectional Relationship Data 273
9.5.4 Integrating Approaches to Intelligence 273
9.5.5 Expanding System Scope — the Russo-Ukrainian War . 274
9.6 Conclusion 276

The research conducted in this thesis set out to examine how social media data could be used by disaster response organisations as a source of intelligence during disaster events. The approach to data curation proposed by this work was grounded in perspectives drawn from a study of disaster response practitioners, and validated with a situated deployment of prototype software. This concluding chapter summarises the findings and contributions of the research and reflects on the limitations of the sociotechnical framework and methodology. A long-

term vision is then considered for the continued improvement of social media integration with existing intelligence processes with a view to improving the quality of disaster response operations.

9.1 Thesis Summary

Chapter 1 — Introduction

Chapter 1 introduced the research context and motivation. The overall research question was defined and broken down into the four sub-questions around which the thesis was structured. This section lists the major contributions of the research and the broad methodology which linked the studies of each chapter.

Chapter 2 — Literature Review

Chapter 2 developed a disaster suitability taxonomy with which disaster events were evaluated with respect to their potential as sources of disaster response intelligence (contribution C-2). A review of the literature examined the value of social media data as a source of disaster response intelligence from the perspectives of practitioners and identified eyewitness reports of ground truth data as highly informative to key decision-making processes and therefore the target of this research.

An examination of existing approaches to eyewitness identification informed the development of the research questions (discussed below). The primary gaps in existing approaches were (i) the disconnect between system design and organisational requirements and (ii) the reliance on existing methods of location inference using text data. A mixed methods approach was proposed to combine ethnographic studies of disaster response organisations, empirical studies of social media discourse, and quantitative methods of eyewitness identification through location inference.

Chapter 3 — Qualitative Study of Disaster Response Organisations

Chapter 3 conducted an ethnographic study of disaster response organisations to develop a conceptual framework from which the value of social media data as intelligence could be evaluated (contribution C-3). The study design was adapted to

the challenges unique to the study of the disaster response domain such as timing and access to participants during periods of high alert. Data from an observational study of an emergency control centre were supplemented with interview data from sixteen disaster response practitioners. Participants were drawn from eight organisations across four domains: fire, emergency response, police, and humanitarian aid.

The iterative nature of the interview process integrated the inductive, deductive, and verificative phases of analysis, described in A. L. Strauss (1987), to build a conceptual framework to make sense of domain perceptions towards social media as intelligence. Analysis of the data identified five key dimensions in which social media data were seen to be informative and four challenges limiting its integration with existing processes.

These dimensions aligned with the perspectives observed in the literature and confirmed the value of eyewitness reports in decision-making processes and situational awareness. Managing datum volume and noise was established as the primary challenge to integration and motivated the development of quantitative curation methods for eyewitness identification.

Chapter 4 — Opportunities and Challenges of Social Media Data for Research

Chapter 4 reviewed the existing landscape of social media research in terms of data access provided by the primary social media platforms (contribution C-1). Twitter was identified as the most suitable candidate both for this research and for disaster response intelligence systems due to its open provision of data through a public API.

A discussion of the technical considerations of using Twitter data was then presented. The key implications of this analysis were related to the temporal decay of data validity, which compelled the requirement of real-time data collection; and the rate-limiting imposed on data access, which constrained the structure of datasets derived from Twitter and necessitated the development of novel approaches to collecting user relationship data.

Approaches to Twitter data capture were examined in relation to their methodological implications and impacts on the validity of resulting datasets. An author-centric approach to data collection was proposed based on the goal of eyewitness detection and existing methods of author classification reviewed. This methodological review extended the findings of chapter 2 and provided context which informed the design of data collection protocols.

Chapter 5 — The CrisisData Software

Chapter 5 documented the design of the collection software used by this research to capture Twitter data (contribution C-5). The custom software captured the relationship data of Twitter users, thus facilitating novel studies based on network analysis approaches that have not been conducted in prior work due to the limitations of existing tools and methods of data collection. The tool was published as an open-source project to facilitate and encourage further research based on relationship data.

Datasets were collected for eight disaster events and their suitability for quantitative eyewitness identification approaches evaluated. Two hurricane datasets were selected for further analysis. A reflection on the data collection process provided supplementary perspectives to the methodological discussion of the previous chapter.

Chapter 6 — Quantitative Analysis

Chapter 6 conducted a preliminary analysis of the Twitter datasets which informed the definition of a coding schema with which messages were classified. 2,500 messages were coded by a primary coder to quantify the sufficient prevalence within the data of messages which aligned with the intelligence requirements of disaster response organisations (contribution C-1).

1,500 users were then coded with respect to their estimated *locality* to the disaster event. Coded values were compared to two classes of geographically-aware metadata: user profile location, and geotagged Tweets in a user's historical timeline. The composite of these predictors was shown to provide accuracy sufficient to be used on the uncoded set of data for locality inference (contribution C-4). An examination of the `source` field of Tweet metadata was then conducted and

demonstrated a correlation between the systems of software used to publish Tweets and the distributions of message class. Tweet source was therefore shown to be an informative feature to methods of message curation.

Chapter 7 — Social Network Analysis

Chapter 7 introduced the concept of network homophily and established the clustering phenomena of Twitter users in both hurricane datasets. The correlation between clustering behaviour and node locality was established visually using spatial network representations then formally verified by calculating an assortativity coefficient. The statistical significance of the correlation was confirmed using a Monte Carlo simulation approach: 100 configuration models were constructed to represent a random ensemble from which a z-score was derived.

A suite of community detection models were then used to partition both graphs and assign labels to node clusters. The purpose of this approach was to facilitate node locality classification based on node membership within a community according to the community's *locality coefficient*. The coefficient was calculated based on the locality of member nodes for which geographic data were known, as established in the previous chapter. In this way, the ground truth user locality data, as represented by profile location and Tweet geotags, were propagated to unlabelled nodes according to community membership and the principle of geoproximate homophily.

This method demonstrated considerable discriminative power and therefore represented an effective approach to locality classification and eyewitness curation (contribution C-4). It was then repeated on temporal partitions of the event data representing earlier moments within the collection window to validate its efficacy in a real-time application and found to be effective within hours of the commencement of data collection.

Chapter 8 — Validation of User-Centric Network Based Locality Inference

Chapter 8 analysed the features and limitations of existing tools available to disaster response organisations for social media intelligence. The evaluation was

grounded in the conceptual framework derived from the ethnographic study in chapter 3 and the characteristics of Twitter data identified through empirical studies in chapters 6 and 7.

A prototype system was developed which extended the tool in chapter 5 with data analysis and visualisation capabilities and implemented the data curation methods developed in the previous chapter (contribution C-6). Features were based on requirements drawn from the conceptual framework of response organisation practices. An ethnographic validation study was then conducted in which participants from the initial study were invited to compare the system implementing the curation approach with existing software in a simulated disaster scenario. The responses provided by participants validated the approach taken by the research to improve the value of social media as intelligence to disaster response organisations.

9.2 Research Findings

As outlined in chapter 2, the motivation for this research was based on three key factors: (i) the informative potential represented by the rapid growth of public participation in social media platforms, (ii) the inadequate integration of sources of online discourse with existing disaster response intelligence systems, and (iii) the relative scarcity of quantitative user geoinference research using network data. The research questions proposed by this work are presented below with the findings developed in this research.

> RQ_1 —What opportunities and challenges are presented to the intelligence processes of disaster response organisations by social media data?

The ethnographic study conducted in chapter 3 documented four key areas in which social media data were seen by participants to provide valuable informative capabilities: detecting emerging events, providing situational awareness, identifying and correcting rumours, and identifying urgent needs of the affected population.

Event detection using social media data is an established area of research, however initial event detection is not an unmet challenge facing disaster response

organisations, which have access to a range of sophisticated apparatus. Rather, this concept was related to emergent activity within the broader disaster environment such as mass congregation or mobilisation of the affected public in ways that impacted response operations.

Reports from eyewitness authors were acknowledged to be a valuable aspect of social media data and an area in which the intelligence capabilities of online discourse were able to exceed traditional formal sources. For example, imagery published on social media platforms could be observed by an operator earlier than corroborating reports from alternative sources such as drone or satellite imagery. The diverse population of authors provide geographically distributed (though urban-weighted) perspectives of the affected environment.

Rumour proliferation is a phenomenon that is facilitated by social media platforms, and therefore must be addressed within the medium. The majority of instances in which a rumour is spread are innocent, regardless of the intentionality of its conception, and therefore early intervention is effective at preventing viral growth, for which early detection is required.

Requests for aid made via posts to social media take two forms: messages directed to the relevant response organisations through the use of appropriate tags, and those which do not contain the relevant features which would otherwise alert response organisations to their existence. While the former is naturally identified and handled by existing social media communication protocols, detecting the latter form presents an opportunity to improve the quality of aid provided by response organisations.

Four classes of challenge limiting the use of social media data were identified in chapter 3. The high volume of data generated on social media platforms and subsequent demands of manual interpretation and classification were the primary issues identified by respondents. This aligned with the observations of previous work, though is better reformulated as an issue of noise and filtration: as datum volume increases, so too does the volume of useful messages. It is the rate at which the useful messages occur within the data, and consequently, the work required to identify a useful message, that is the primary constraint.

The second challenge was datum veracity, or misinformation as a result of malicious or unintentional behaviour. The impact of misinformation to disaster response intelligence is mitigated by the practice of *source triangulation*, in which all actionable data are verified by reports from complementary sources before acceptance. Methods by which the veracity of social media data may be evaluated can be adopted in software design — for example, features supporting author inspection were implemented in the prototype system developed in chapter 8.

Ancillary challenges were unrelated to characteristics of the data. Organisational constraints included inadequate support from management, in terms of personnel and financing, to implement social media intelligence processes, and legal limitations preventing the initiation of response action based on social media data. Integrating social media messages with existing disaster information systems was seen as a key issue by organisations that did not have the technical means to develop compatible software pipelines.

> RQ_2 —How can publicly available social media data provide meaningful intelligence to disaster response organisations during disaster events?

The degree to which these *perceived* values identified by RQ_1 were supported by empirical patterns of social media discourse was measured in chapter 6. Data relevant to the first three use cases were observed at rates sufficient to support their systematic use in intelligence systems, however the incidence of messages coded as 'aid requests' was below 2%.

In contrast with the first three use cases, an actionable 'urgent need' can be extracted from a single message, and therefore the relative rarity of this message class does not present an insurmountable obstacle to its use as intelligence, however, it raises practical concerns with respect to how well this class of message can be detected, the social cost of failing to detect some proportion of relevant messages, and the extent to which the lower volume of these messages justifies integration with existing processes. These questions require further research.

Within the context provided by RQ_1, which identified data volume as a key constraining challenge of social media intelligence, the degree to which these forms

of data can provide meaningful intelligence is predicated upon the effectiveness of methods by which they can be extracted from the data streams. A network-based approach to curation was examined in RQ_3.

> RQ_3 —To what extent can graph-structured relationship data inform eyewitness classification for social media data?

Graph data describing user following/followee relationships were analysed in chapter 7 and determined to exhibit clustering behaviour which was correlated with node locality. These findings aligned with the principle of geoproximate homophily and demonstrated the potential represented within these data as a contributing feature to eyewitness classification.

The Louvain community detection algorithm partitioned network data into distinct communities for which *locality coefficients* were calculated using the estimated locality of the set of member nodes for which geographic data had been observed. In this way, observed user locality data were propagated to unlabelled nodes according to community membership and the principle of geoproximate homophily. This approach was shown to be effective at locality classification of unlabelled nodes and provided compelling motivation to integrate network-based approaches in methods of eyewitness detection.

The event data were then temporally partitioned to represent the dataset as extant at given points in time during the data collection period. The network-based locality classification method was repeated on the sub-networks to evaluate the effectiveness of the approach during the early stages of a real-time event, during which network data was limited. The classification accuracy was found to stabilise within hours of the commencement of data collection and thus the method was determined to present a promising approach to eyewitness classification during disaster events.

> RQ_4 —How well can a network-based user-centric eyewitness classification approach curate social media data and address the volume constraints of disaster response organisations?

RQ_4 tied together the conceptual framework derived from the study of disaster response organisations, findings from empirical analyses of hurricane data, and the user-centric network approach to location inference and eyewitness classification. A prototype system implementing the classification model was developed in chapter 8 and presented to disaster response practitioners to evaluate the extent to which it provided informative data and addressed the key challenges identified in chapter 3.

A set of synthesised data were generated based on empirical data to simulate a real-time disaster event. The simulated data were used to populate the prototype system and a control system which emulated existing tools and did not institute an underlying process of data curation. A set of tasks modelling disaster response processes was performed by participants, after which feedback was procured through semi-structured interviews.

The prototype system which demonstrated the network-based method of curation was seen by participants as superior to the control system. The curated feed was acknowledged to satisfy the informational requirements of the intelligence practitioners, and thus effectively address the key challenge of datum volume documented in chapter 3. Additionally, the results of this study captured supplementary findings which contributed to the conceptual framework of social media intelligence use by disaster response organisations.

9.3 Research Contributions

The primary motivation of this research was based on the improvement of data derived from social media platforms as a source of intelligence. As public participation in these sites of online discourse increases, so too do the volume and diversity of data available to observers. Technological advancements in hardware miniaturisation improve the capabilities of handheld devices such that camera sensors in phones are able to capture high-quality imagery which, when posted to online platforms, provide valuable situational awareness to disaster response organisations.

While the *potential* for these data to provide informative material grows, practical limitations are imposed by the environment. The structure and character of online

public discourse are in a constant state of flux as cultural trends shape patterns of behaviour. Most significantly, online social platforms rise and fall rapidly in popularity, often as the inevitable result of younger generations favouring emerging products. Adapting to the changing landscape of social media discourse therefore necessitates an ongoing analysis and discussion of the online environments in which it occurs.

Similarly, the methods by which these data can be interpreted and classified evolve, as do the requirements of the organisations to which they present value. Therefore, leveraging social media data to improve the outcomes of affected populations requires a continual examination of the challenges facing the disaster response domain and an iterative development of systems that align the character of online discourse with the requirements of domain practitioners.

To this end, this thesis makes the following contributions to the fields of computer science and crisis response,[1] expanding upon section 1.2.

C-1 **Empirical studies of online behaviour:** Empirical studies of online behaviour on Twitter during ten disaster events contribute valuable perspectives to social media research and inform a conceptual framework of the values presented by these data to disaster response information systems. The focus of the results presented in chapter 6 was based on domain perceptions and included a discussion of the minor obstructive elements of misinformation and a quantification of message category in terms of its role as intelligence.

These analyses extend existing empirical work with results drawn from disaster event data. The novel method of data collection used in this work further augments the empirical perspectives provided by these results with captured relationship data.

C-2 **Disaster suitability working taxonomy:** A *disaster suitability working taxonomy* was synthesised from the literature examined in chapter 2 and contributes a framework within which disaster events may be evaluated with

[1] While the term 'disaster' was preferred in this thesis for scoping reasons discussed in chapter 2, 'crisis' is used here based on academic terminology.

respect to how they shape the character of the resulting online discourse and, in turn, the value of the discourse as a source of intelligence.

C-3 **Conceptual framework for disaster response perspectives of social media data:** A conceptual framework documented the roles in which social media data present value to disaster response organisations and the key challenges preventing their expanded use within the domain. These findings develop existing studies of intelligence officer requirements and introduce perspectives from a broader set of organisations. Ethnographic studies of disaster response organisations are constrained by methodological challenges due to the intensity of response operations and therefore remain an underdeveloped area of research.

C-4 **User locality inference:** Methods were developed by which Twitter data may be curated such that the resulting subset of data contains higher incidences of relevant and informative material. Chapter 3 identified a strong preference of disaster intelligence officers for messages comprising eyewitness reports. A sufficient prevalence of messages falling within this categorisation was confirmed by empirical analyses in chapter 6. Capturing this class of message from within social media data streams was therefore selected as the focus of this aspect of the research.

The first part of this methodological contribution to eyewitness classification and detection was conducted in chapter 6, which evaluated two methods for user location inference: parsing Tweet history for geospatial data and geolocating profile location information. These approaches established home location as a predictive feature for eyewitness detection and informed the substantive methodological contribution made in chapter 7, which developed a novel approach to user locality inference using relationship data and community detection methods. The high accuracy of classification demonstrated by the network method introduces interesting possibilities to social media analyses and comprises a key contribution to computer science research using network relationship data and other fields of social media classification research.

C-5 **Data collection tool:** The analyses conducted in chapter 7 necessitated the development of a novel data collection tool that enhanced data detected using classic filtering techniques with user relationship data from which network analyses could be conducted. Temporally fragile features of network data were captured in real-time using live data streams. The design considerations for this tool were developed in chapter 4 and are accompanied by reflections on the practical implementation of the system in chapter 5. The data collection tool has been released as open-source software and, together with the documentation of the design process, provide a practical contribution to computer science research seeking to use novel sets of social media data.

C-6 **Disaster intelligence prototype:** The development and validation of the disaster intelligence prototype system in chapter 8, which reflected on software development processes grounded in the disaster intelligence domain, provides a secondary practical contribution.

9.4 Limitations and Challenges

All research is conducted within a set of constraints and limitations; while the approaches adopted by this work were effective in addressing the research questions and developing an effective system for social media disaster intelligence, a selection of key challenges encountered during the research are presented below.

9.4.1 Participant Access During Disasters

When disaster events occur, operators within response organisations are often overwhelmed with information and demands on their time. Thus, interacting with participants during live operations is not only inappropriate, but, given the critical nature of their tasks, ethically problematic. The interviews conducted retrospectively in this research introduced epistemological limitations by removing the participants from the disaster response environment both temporally and functionally. The data derived from these studies may have therefore failed to capture unique features such as the effects of stress on decision-making capacity (Fleischman et al. 2006).

The strategy taken in A. L. Hughes and Shah (2016) included becoming certified to act within the responder role to conduct embedded observation less intrusively, though it notes that limitations remained. Given the diverse and distributed nature of participant organisations contributing to this study, such training was not a viable solution.

Equally, the validation study used synthesised data to represent a simulated disaster response event with which participant use of the system was evaluated through overt naturalistic observation. While further context was provided by participant responses, similar limitations were introduced wherein the simulation did not instil the environmental stress and changing demands which may otherwise influence real-world behaviour. Thus, while a true *in situ* deployment of a safety critical system for evaluation is an incredibly complex process well outside the scope of this research, the responses provided by participants must be considered with respect to these abstractive limitations.

9.4.2 Data Bias and Technology-Induced Privilege of Care

Systemic biases in population data collected from social media sources are an unavoidable limitation of empirical studies of online discourse. For example, it has been recognised that certain groups may lack the tools, skills, and motivations to interact with social media discourse (Xiao et al. 2015) and are therefore less likely to be represented in data derived from online sources. The characteristics which define these groups may include low income, low education, elderly, non-native language, or other latent traits; and significant biases have been observed towards those located in urban areas (Vieweg, Castillo, et al. 2014).

Curation algorithms for disaster intelligence systems integrating social media data, to which this research contributes, are highly vulnerable to the risk of embedding undesirable biases within disaster response protocols by presenting data to operators which are not representative of the underlying population. A significant risk vector is thus introduced wherein the views of people who use social

media disproportionately shape the situational awareness models of the response organisations and lead to the administration of privileged rates of aid.

Addressing this limitation requires the introduction of diverse perspectives from a range of disciplines and is therefore beyond the scope of this research, however, data biases represent a fundamental challenge facing social media data in disaster response and must therefore contextualise any consideration of intelligence systems such as that of this research.

9.4.3 Psychosocial Safety and Secondary Trauma

The empirical studies of Twitter data conducted in this work exposed the researcher (and secondary coder) to large volumes of unfiltered data from a diverse set of authors, much of which was created by vulnerable users in conditions that were often distressing or traumatic. Periods of analysis were conducted intensively as the observed events occurred so that adjustments to the data capture system could be made. The researcher was therefore at times deeply immersed in a flood of graphic and disturbing imagery which posed the risk of re-surfacing outside of work in the form of intrusive thoughts and disrupted sleep.

The dangers of *secondary* or *vicarious traumatisation* become significant in situations where the exposure is repeated (K. Cohen and Collens 2013) and are well studied with respect to journalists and newsrooms. Immersive work with traumatic imagery has been recognised by the American Psychiatric Association as a risk vector of post-traumatic stress disorder (American Psychiatric Association 2013). These psychosocial risks were not well recognised within the field of computer science and were further magnified by the isolated nature of DPhil research.

Once recognised, measures of risk mitigation were drawn from established protocols for open-source investigations (OHCHR and UC Berkeley 2022), which included adjusting the viewing environment (for example, using a greyscale filter to view imagery and removing sound from video content), limiting exposure to traumatic material, and developing a support network to identify changes in behaviour. For the health of future researchers, such interventions are highly

recommended as mandatory considerations in the design of any research protocol which introduces risks of exposure to traumatic material.

9.5 Further Research

A number of opportunities for further research were discussed throughout this thesis and demonstrate the emerging research frontiers of computational social media analysis and disaster response intelligence systems. Continued development of the methods adopted by this research will ensure state-of-the-art approaches are aligned with developments in the landscapes of social media discourse and disaster response.

9.5.1 Preparing for Live Deployment

While the software developed in chapters 5 and 8 were designed with a view to be extended into production environments, development was focused primarily on research use. Deploying software to live environments introduces new challenges which were not resolved as part of this research and therefore further work is required to raise the prototype software to a production standard. Key challenges include, for example, scalability, integration, and data privacy requirements, which are best addressed by the inclusion of relevant stakeholder perspectives and developer expertise.

The primary data used in this project were captured from the Twitter API and thus subject to rate limitations which constrain further scalability. Increased access to data through an API typically incurs additional charges which are set by the source platform and therefore relies upon business policies of third-party providers which may change without notice.

9.5.2 Evaluating Generalisability of Findings

The primary datasets used by this research were based on hurricane events occurring within the U.S.A. The similarity of the events provided a useful baseline from which the location inference methods could be developed, however, whether the same features exist within communication networks formed during other types of disaster remains to be explored. Furthermore, the patterns of social behaviour which have

been observed during this research are a product of the platform within which they emerged. The extent to which the same patterns exist within alternate social networks depends upon both platform design and community culture.

Generalising the findings of this work in the dimensions of both event type and data source therefore presents an important avenue for further research and prerequisite to advanced integration with existing intelligence systems.

9.5.3 Analysing Bidirectional Relationship Data

The network analysis in chapter 7 was conducted on an undirected simplification of the bidirectional relationship graph. In this way, all edges were considered as equivalent for the purposes of simplicity and increased compatibility with community detection algorithms. In reality, there is a meaningful difference between the three states of extant user relationship: *following*, *followed by*, and *reciprocal*. The significance of these distinctions and their impact on the results of this work should be explored to further improve the performance of the methods developed by this research.

9.5.4 Integrating Approaches to Intelligence

The network-based classification method developed in this work contributed to a field of location inference research in which a number of approaches exist. While the prototype system developed in chapter 8 did not integrate complementary methods due to its purpose as a framework for evaluating the novel approach, further performance improvements may be realised by introducing text-oriented models.

Furthermore, practical requirements not addressed by this approach present further opportunities to create effective systems. For example, Hiltz, A. Hughes, et al. (2020) notes that in its requirements study, 61% of the experts claimed that it was not useful to have a system that uses data from only one social media platform. Twitter was selected as the singular focus of this work to constrain the scope of the research, though whether a similar approach is suitable for other platforms is predicated upon the availability of network data and a replication of

the methodology of this research to establish equivalent homophilic phenomena. Regardless of method, integrating sufficiently curated data from alternative sources further enhances the informative potential of a system and by extension, its ability to satisfy the needs of response organisations.

9.5.5 Expanding System Scope — the Russo-Ukrainian War

The disaster response perspective taken by this research was motivated by the goal of improving the outcomes for those affected by disaster events, however, the approach to data curation and eyewitness identification presents valuable opportunities to a broader range of domains. For example, social media data curated for eyewitness reports may provide valuable information to journalists, governments, or social science researchers. The application of the findings of this research to non-disaster domains, however, introduces unique requirements and constraints and demands a revisitation of the ethical considerations.

In February 2022, Russia invaded Ukraine in a major escalation of the Russo-Ukrainian War. Naturally, conditions in eastern Ukraine were severely disrupted, placing a high value on information that could inform situational awareness. In terms of this work, the characteristics of this event were a strong match with those discussed in the disaster suitability taxonomy in chapter 2: a long-term, diffuse event which affected a technically developed and highly populated region. The speed of onset was arguably variable by location, based on the movement of the Russian front. The fundamental differentiating factor was, naturally, the hazard cause, though in this dimension the taxonomy breaks down.

Due to the similarity in event characteristics, and as an effort to examine further applications of the approach developed in this research, the data collection system developed in chapter 5 was initialised with parameters relating to the conflict in Ukraine on the day of the initial invasion (24$^{\text{th}}$February 2022) and continues to run at the time of submission of this document (October 2022). As a result, over 2.8 million Tweets and 1.3 terabytes of photo and video material have been collected.

The important distinction between the application of this system to hurricane events and human conflict lies in the intended purpose: eyewitness reports during hurricanes provided valuable data to response organisations, however, within the context of conflict, the 'stakeholders' for which such data would be useful are undefined, nor has the system been designed with requirements of conflict participants in mind.

These technical considerations are preceded by fundamental ethical concerns — first, eyewitness reports of conflict events are highly sensitive: not only do they exploit a vulnerable population during sensitive moments, but they also pose a risk of providing evidence enabling future persecution. Furthermore, given the adversarial nature of the event, the prevalence of misinformation is likely to be much higher and include sophisticated manipulative activity by state actors. Therefore attempts to use these data to present a model of ground truth may inadvertently amplify an agenda embedded within the data by *bad actors*.

The principle of *beneficence* informed the ethical position taken in chapter 5 concerning the use of data created by human subjects (United States National Commission for the Protection of Human Subjects of Biomedical & Behavioral Research 1978). The tenets of minimising harm while also maximising benefit to participants were not considered reconcilable with the application of the system to conflict data. For this reason, the location inference methods were not conducted on the Ukraine dataset.

A collaboration with the United Nations Office of the High Commissioner of Human Rights (OHCHR) to provision collected eyewitness data as evidence for the investigation of war crimes, crimes against humanity, and genocide is currently undergoing an ethical evaluation, and illustrates an application of the technology developed in this work which may be considered to satisfy the principle of beneficence.

9.6 Conclusion

During disaster events, vast amounts of data are generated on online platforms by people in affected areas. A meaningful proportion of these data contain eyewitness reports and other types of information which provide insights into environmental conditions and are therefore valuable to disaster response organisations in developing models of situational awareness. During periods of disruption, response organisations are placed under immense pressure and have access to limited resources. Therefore, analysing the high volumes of data available from social media sources is often not possible, leading to a failure to identify useful information which may otherwise inform response efforts.

By improving systems with which disaster response practitioners can access and interpret data derived from social media platforms, the value of online discourse as a source of intelligence is improved. Social media users represent a geographically-distributed network of observers providing real-time reports which augment existing sources of disaster response intelligence.

While it is evident from the literature that social media data have an enormous potential to provide valuable insight to disaster response practitioners, the unpredictable nature of disaster events and changing landscape of social media discourse provide compelling motivation for the ongoing analysis of new perspectives and approaches.

The systems developed in this research brought together an understanding of the intelligence requirements of disaster response organisations and the patterns of behaviour within online discourse to develop methods by which social media data were made useful as a source of intelligence during disasters. In this way, contributions made by the affected population to online discourse were leveraged to improve the quality of response operations and lead to better outcomes.

Appendices

A

Qualitative Research Documents

A.1 Interview Questions

All interviews used a semi-structured format in which additional questions were formulated by the interviewer based on the responses of the participant and therefore these lists provide general structures of the interviews rather than comprehensive accounts.

Nested list items represent questions conditional upon preceding answers.

A.1.1 Exploratory Study

1. What is the role of your organisation within the domain of disaster response?
2. What forms of intelligence are required by your organisation during response operations?
3. What is your role within the organisation, and how does it relate to intelligence processes?
4. Could you describe whether/how social media data is used within your organisation?
5. Describe the functionality of an ideal system of social media intelligence.

If social media data is used by respondent's organisation:

1. What are the primary values presented by social media intelligence?

2. In what ways do you feel social media data could further augment your existing intelligence systems?
3. What are the primary challenges and concerns facing the use of social media intelligence by your organisation?

If social media data is not used by respondent's organisation:

1. In what ways do you feel social media data could augment your existing intelligence systems?
2. For what reasons are social media data not implemented by your organisation?
3. What would need to change before intelligence from social media data could be used by your organisation?

Interview continued with collaborative process modelling.

A.1.2 Prototype Evaluation

After reviewing their answers provided in the previous study:

1. Would you like to update any of your answers?
2. Has your organisation changed the way in which it uses social media for intelligence since the previous interview?
3. Have you used other tools to analyse social media data since the previous interview?
4. How do you think the value of social media data has changed since the previous interview?

After completion of simulated tasks on $P_{emulated}$ and $S_{TweetDeck}$:

1. Were you able to complete the tasks with $P_{emulated}/S_{TweetDeck}$?
2. What were the most limiting issues of $P_{emulated}/S_{TweetDeck}$?
3. Were there any areas in which you felt the tasks did not align with typical processes from your domain?

After completion of simulated tasks on P_{novel}:

1. Were you able to complete the tasks with P_{novel}?
2. What were the most limiting issues of P_{novel}?

3. Which, if any, of the visualisations in P_{novel} did you find most useful, and how did you use them?
4. Which, if any, of the features of P_{novel} would you use in your current domain?
   (a) Does the existence of these features improve your perception of the value of social media data as intelligence?
5. What key features could most improve P_{novel}?
6. Overall, on which system did you find the tasks easier to complete?
7. Overall, on which system did you find the tasks faster to complete?
8. Overall, which system did you find better supported the analyses you might conduct in your typical work processes?
9. What tasks did you self-define, and how did each system support their completion?
10. Did you perceive any differences between the data presented on each system?
    (a) How do they influence your perception of the systems?
11. How has this experience changed the way in which you perceive the value of social media data as intelligence?
12. Are there any comments you would like to add?

A.2 Participant Consent Form

[see next page]

Department of Computer Science

Ross Gales
ross.gales@cs.ox.ac.uk
DPhil student, Human-Centred Computing

PARTICIPANT CONSENT FORM

CUREC Approval Reference: R49170/RE002

The Use of Social Media During Crisis Events by Response Organisations

Purpose of Study: This study aims to understand how social media data is used within crisis response organisations for the purpose of intelligence-gathering following crisis events. This information will be used to guide the development of further studies on the use social media by crisis-affected populations.

Initial each box

1	I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without any adverse consequences or academic penalty.	☐
3	I understand that research data collected during the study may be looked at by designated individuals from the University of Oxford where it is relevant to my taking part in this study. I give permission for these individuals to access my data.	☐
4	I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.	☐
5	I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	☐
6	I understand how this research will be written up and published.	☐
7	I understand how to raise a concern or make a complaint.	☐
8	I consent to being audio recorded.	☐
9	I agree to take part in the above study. I understand that my responses will not be anonymised; however, the data will not be published in such a way that identifies me or my organisation by name.	☐
Optional:	I agree for research data collected in this study to be given to researchers, including those working outside of the EU, to be used in other research studies. I understand that any data that leave the research group will be fully anonymised so that I cannot be identified.	☐

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature

B
Twitter Event Datasets

B.1 Collection Parameters

B.1.1 Mudslides; Sierra Leone

Collection Period
2017-08-14 14:12 — 2017-08-16 14:16
Keywords
freetown, sierra leone, sierraleone, freetownfloods
Geographic Bounding Boxes
(8.2157, -13.6317), (8.7157, -12.8317)

B.1.2 Terror Attacks; Barcelona, Spain

Collection Period
2017-08-17 18:19 — 2017-08-18 08:38
Keywords
barcelona, ramblas, fuerzabarcelona, todosconbarcelona, ánimobarcelona

B.1.3 Hurricane Hato; Hong Kong

Collection Period
2017-08-24 07:51 — 2017-08-24 11:45
Keywords
typhoon, hato

B.1.4 Hurricane Harvey; Texas, U.S.A.

Collection Period
2017-08-25 23:37 — 2017-09-02 09:30
Keywords

#harvey, #harveystorm, #hurricaneharvey, #corpuschristi
Geographic Bounding Boxes
(26.5486, -99.9590), (29.1197, -97.5021)
(26.5486, -97.5021), (30.3893, -93.9790)

B.1.5 Earthquake; Central Mexico

Collection Period
2017-09-19 20:36 — 2017-09-21 18:57
Keywords
#terremoto, #mexicocity, #ciudaddemexico, #sismo, #sismomexico, #cdmx
Geographic Bounding Box
(18.0101, -100.3601), (20.3446, -96.7072)

B.1.6 Hurricane Irma; Florida, U.S.A.

Collection Period
2017-09-10 01:47 — 2017-09-17 00:41
Keywords — High Priority
sosirma, irmahelp, irmarescue, irmaresponse, irmasos, irmashelters
Keywords — Low Priority
irma2017, irmawatch, flprepares, irmahurricane2017, irmahurricane, hurricaneirma
Geographic Bounding Box
(24.38, -83.23), (28.19, -79.71)

B.1.7 Wildfires; California, U.S.A.

Collection Period
2017-10-10 16:40 — 2017-10-24 22:54
Keywords
#santarosafire #napafire #tubbsfire #sonomafire #santarosafires #napafires #tubbsfires #sonomafires #californiafires #californiawildfires #atlasfire #atlasfires
Geographic Bounding Box
(38.1583, -122.9741), (38.6726, -122.0567)

B.1.8 Hurricane Florence; North Carolina, U.S.A.

Collection Period
2018-09-13 19:19 — 2018-10-03 10:52
Keywords — High Priority
wrightsvillebeach, wilmington, myrtlebeach, cityofmyrtlebeach, northmyrtlebeach, charleston, sosflorence, florencerescue, wilmingtonnc
Keywords — Low Priority
hurricaneflorerence, florence, hurricaneflorence, hurricaneflorence, hurricanflorence, hurricainflorence, hurricaineflorence, hurricaneflorence2018, hurricane, florencenc, florencehurricane, southcarolina, northcarolina, huricanflorence, huricaneflorence, huricane, hurricanceflorence, florencehurricane2018, flooding, tropicalstormflorence
Geographic Bounding Boxes

(31.7161, -81.4328), (34.9624, -78.2680)
(31.7161, -78.2680), (36.4134, -75,1782)

B.1.9 Hurricane Michael; Florida, U.S.A.

Collection Period
2018-10-11 21:09 — 2018-10-18 16:33
Keywords — High Priority
mexicobeach, panamacity, panamacitybeach, michaeltlh
Keywords — Low Priority
hurricanemichael, hurricanmichael, huricanemichael, huricanmichael, tropicalstormmichael, hurricanemichael2018
Geographic Bounding Box
(31.3153, -86.1726), (29.2568, -83.6787)

B.1.10 Hurricane Willa; Sinaloa, Mexico

Collection Period
2018-10-23 15:43 — 2018-10-30 00:26
Keywords — High Priority
islasmarias, nayarit, sanblas, mazatlan, tepic, puertovallarta
Keywords — Low Priority
hurricanewilla, hurricanwilla, huricanwilla, huricanewilla, hurricanewilla2018, hurricanwilla2018, huricanewilla2018, huricanwilla2018, willa, huracánwilla, huracanwilla
Geographic Bounding Box
(23.878, -107.1681), (21.1525, -104.5715)

C

Twitter Source Metrics

C.1 Twitter Source Relevancy — Hurricane Harvey

Source	Total	Precision	Recall
First-Party:			
Twitter for iPhone	404	0.144	0.144
Twitter Web Client	412	0.080	0.082
Twitter for Android	236	0.127	0.074
TweetDeck	35	0.086	0.007
Twitter for iPad	35	0.057	0.005
Twitter Lite	30	0.100	0.007
Twitter for Windows	7	0.000	0.000
Twitter for Mac	2	0.000	0.000
Mobile Web (M2)	5	0.000	0.000
Twitter for Windows Phone	2	0.500	0.002
Twitter for BlackBerry	2	0.000	0.000
Tweetbot for iOS	1	0.000	0.000
TweetCaster for Android	5	0.000	0.000
Third-Party:			
Instagram	732	0.333	0.605
IFTTT	34	0.176	0.015
Facebook	35	0.171	0.015
Foursquare	9	0.222	0.005
Untappd	13	0.000	0.000
Periscope	4	0.750	0.007
Google	4	0.250	0.002

LinkedIn	2	0.000	0.000
WordPress.com	2	0.000	0.000
Social Media Suites:			
Hootsuite	60	0.100	0.015
BubbleLife	56	0.018	0.002
Buffer	14	0.000	0.000
SocialNewsDesk	11	0.091	0.002
Sprout Social	3	0.000	0.000
Botize	6	0.000	0.000
VoiceStorm	1	1.000	0.002
Hearsay Social	3	0.000	0.000
HubSpot	1	0.000	0.000
Crowdfire - Go Big	4	0.250	0.002
GaggleAMP	2	0.000	0.000
Radian6 -Social Media Management	1	0.000	0.000
Cloudhopper	2	0.000	0.000
Automated:			
TweetMyJOBS	97	0.000	0.000
SafeTweet by TweetMyJOBS	104	0.000	0.000
circlepix	3	0.000	0.000
TownTweet	7	0.000	0.000
SeeYourWeather.com Galveston	1	0.000	0.000
Donate a Photo	1	0.000	0.000
Weather Message	1	0.000	0.000
Spam:			
Paper.li	74	0.000	0.000
Error-log	8	0.000	0.000
despa ringtones	4	0.000	0.000

Table C.1: Twitter source relevancy.

References

Abdelsadek, Youcef et al. (2018). "Community extraction and visualization in social networks applied to Twitter". In: *Information Sciences* 424, pp. 204–223.

Abel, Fabian et al. (2012). "Semantics + Filtering + Search = Twitcident Exploring". In: *Proceedings of the 23rd ACM conference on Hypertext and social media*, pp. 285–294.

Adams, David S (1970). "Policies, programs, and problems of the local Red Cross disaster relief in the 1960s". In:

Akter, Sanjida and Muhammad Tareq Aziz (2016). "Sentiment analysis on facebook group using lexicon based approach". In: *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. IEEE, pp. 1–4.

Alessa, Ali, Miad Faezipour, et al. (2019). "Preliminary flu outbreak prediction using twitter posts classification and linear regression with historical centers for disease control and prevention reports: Prediction framework study". In: *JMIR public health and surveillance* 5.2, e12383.

Alexander, David E (2005). "An interpretation of disaster in terms of changes in culture, society and international relations." In: Xlibris Press.

Ali, Raian, Fabiano Dalpiaz, and Paolo Giorgini (2010). "A goal-based framework for contextual requirements modeling and analysis". In: *Requirements Engineering* 15.4, pp. 439–458.

Alsaedi, Nasser, Pete Burnap, and Omer Rana (2017). "Can we predict a riot? Disruptive event detection using Twitter". In: *ACM Transactions on Internet Technology (TOIT)* 17.2, pp. 1–26.

American Psychiatric Association (2013). *Post-Traumatic Stress Disorder. Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Publishing, Arlington.

Aramaki, E, S Maskawa, and M Morita (2011). "Twitter catches the flu: detecting influenza epidemics using Twitter". In: *Conference on Empirical Methods in Natural Language Processing*, pp. 1568–1576.

Ashktorab, Zahra et al. (2014). "Tweedr: Mining Twitter to Inform Disaster Response". In: *ISCRAM*, pp. 269–272.

Auerbach, Carl and Louise B Silverstein (2003). *Qualitative data: An introduction to coding and analysis*. Vol. 21. NYU press.

Avvenuti, Marco et al. (2014). "EARS (earthquake alert and report system): a real time decision support system for earthquake crisis management". In: *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, pp. 1749–1758.

Bailly, Charles and Carole Adam (2017). "An interactive simulation for testing communication strategies in bushfires." In: *ISCRAM*.

Bakillah, Mohamed, Ren-Yu Li, and Steve H L Liang (2015). "Geo-located community detection in Twitter with enhanced fast-greedy optimization of modularity: the case

study of typhoon Haiyan". In: *International Journal of Geographical Information Science* 29.2, pp. 258–279.

Bakshy, Eytan et al. (2011). "Everyone's an influencer: quantifying influence on twitter". In: *Proceedings of the fourth ACM international conference on Web search and data mining SE - WSDM '11*, pp. 65–74. arXiv: 1111.1896.

Barton, Allen (1989). "Taxonomies of disaster and macrosocial theory". In: *Social structure and disaster*, pp. 346–350.

Basile, Valerio et al. (2019). "Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter". In: *Proceedings of the 13th International Workshop on Semantic Evaluation*, pp. 54–63.

Baxter, Gordon and Ian Sommerville (2011). "Socio-technical systems: From design methods to systems engineering". In: *Interacting with computers* 23.1, pp. 4–17.

Bergsma, Wicher (2013). "A bias-correction for Cramér's V and Tschuprow's T". In: *Journal of the Korean Statistical Society* 42.3, pp. 323–328.

Bevensee, Emmi et al. (2020). "SMAT: The social media analysis toolkit". In: *Proceedings of the 14th International AAAI Conference on Web and Social Media*. Vol. 14.

Bharosa, Nitesh, Jinkyu Lee, and Marijn Janssen (2010). "Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises". In: *Information Systems Frontiers* 12.1, pp. 49–65.

Birkmann, Jörn et al. (2014). "Theoretical and conceptual framework for the assessment of vulnerability to natural hazards and climate change in Europe: the MOVE framework". In: *Assessment of vulnerability to natural hazards*. Elsevier, pp. 1–19.

Blondel, Vincent D et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10, P10008.

Boeije, Hennie (2002). "A purposeful approach to the constant comparative method in the analysis of qualitative interviews". In: *Quality and quantity* 36.4, pp. 391–409.

Boersma, Kees, Peter Groenewegen, and Pieter Wagenaar (2009). "Emergency Response Rooms in Action: an ethnographic case-study in Amsterdam". In: *Proceedings of the 6th International ISCRAM Conference* May.

Boin, Arjen, Paul't Hart, and Sanneke Kuipers (2018). "The Crisis Approach". In: *Handbook of Disaster Research*. Springer, pp. 23–38.

Boon-Itt, Sakun, Yukolpat Skunkan, et al. (2020). "Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study". In: *JMIR Public Health and Surveillance* 6.4, e21978.

Borra, Erik and Bernhard Rieder (2014). "Programmed method: Developing a toolset for capturing and analyzing tweets". In: *Aslib Journal of Information Management* 66.3, pp. 262–278.

Boyd, Danah and Kate Crawford (2012). "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon". In: *Information, communication & society* 15.5, pp. 662–679.

Breckenridge, Jenna and Derek Jones (2009). "Demystifying theoretical sampling in grounded theory research." In: *Grounded Theory Review* 8.2.

Britton, Neil R (2005). "What's a word? Opening up the debate". In: *What is a disaster?: New answers to old questions*, pp. 60–78.

Bruckman, Amy (2002). "Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet". In: *Ethics and Information Technology* 4.3, pp. 217–231.

Bruckman, Amy (2014). "Research Ethics and HCI". In: *Ways of Knowing in HCI*. Ed. by Judith S Olson and Wendy A Kellogg. New York, NY: Springer New York, pp. 449–468.

Bruns, Axel and Jean E Burgess (2011). "The Use of Twitter Hashtags in the Formation of Ad Hoc Publics". In: *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*. August, pp. 25–27.

Bruns, Axel, Jean Burgess, et al. (2011). "#qldfloods and @ QPSMedia : Crisis Communication on Twitter in the 2011 South East Queensland Floods". In: *Methodology* Cci, pp. 1–57.

Bruns, Axel and Yuxian Liang (2012). "Tools and methods for capturing Twitter data during natural disasters". In: *First Monday* 17.4.

Bruns, Axel and Stefan Stieglitz (2012). "Quantitative Approaches to Comparing Communication Patterns on Twitter". In: *Journal of Technology in Human Services* 30.3-4, pp. 160–185.

Buckle, Philip (2005). "Disaster: mandated definitions, local knowledge and complexity". In: *What is a disaster?: New answers to old questions*, pp. 173–200.

Burnap, Peter et al. (2015). "COSMOS: Towards an integrated and scalable service for analysing social media on demand". In: *International Journal of Parallel, Emergent and Distributed Systems* 30.2, pp. 80–100.

Burnett, John J (1998). "A Strategic Approach To Managing Crises". In: *Public Relations Review* 24.4, pp. 475–488.

Caragea, Cornelia et al. (2011). "Classifying text messages for the Haiti earthquake". In: *ISCRAM*. Citeseer.

Castillo, Carlos (2016). *Big Crisis Data: social media in disasters and time-critical situations*. Cambridge University Press.

Cha, Meeyoung et al. (2010). "Measuring User Influence in Twitter : The Million Follower Fallacy". In: *International AAAI Conference on Weblogs and Social Media*, pp. 10–17.

Charmaz, Kathy (2014). *Constructing grounded theory*. sage.

Chatfield, Akemi Takeoka, Hans J Jochen Scholl, and Uuf Brajawidagda (2013). "Tsunami early warnings via Twitter in government: Net-savvy citizens' co-production of time-critical public information services". In: *Government information quarterly* 30.4, pp. 377–386.

Chatfield, Akemi and Uuf Brajawidagda (2012). "Twitter tsunami early warning network: a social network analysis of Twitter information flows". In: *Faculty of Engineering and Information Sciences - Papers: Part A*.

Cheng, Zhiyuan, James Caverlee, and Kyumin Lee (2010). "You are where you tweet: a content-based approach to geo-locating twitter users". In: *Proceedings of the 19th ACM international conference on Information and knowledge management*, pp. 759–768.

Chon, Jaime et al. (2015). "Modeling flu trends with real-time geo-tagged twitter data streams". In: *International Conference on Wireless Algorithms, Systems, and Applications*. Springer, pp. 60–69.

Chong, Wen-Haw and Ee-peng Lim (2017). "Exploiting Contextual Information for Fine-Grained Tweet Geolocation". In: *Eleventh International AAAI Conference on Web and Social Media*, pp. 488–491.

Choudhary, Pankaj and Upasna Singh (2015). "A survey on social network analysis for counter-terrorism". In: *International Journal of Computer Applications* 112.9, pp. 24–29.

Ciesielska, Malgorzata, Katarzyna W Boström, and Magnus Öhlander (2018). "Observation methods". In: *Qualitative methodologies in organization studies*. Springer, pp. 33–52.

Cohen, Jacob (1960). "A coefficient of agreement for nominal scales". In: *Educational and psychological measurement* 20.1, pp. 37–46.

Cohen, Keren and Paula Collens (2013). "The impact of trauma work on trauma workers: A metasynthesis on vicarious trauma and vicarious posttraumatic growth." In: *Psychological Trauma: Theory, Research, Practice, and Policy* 5.6, p. 570.

Cohen, Louis, Lawrence Manion, and Keith Morrison (2002). *Research methods in education*. routledge.

Compton, Ryan, David Jurgens, and David Allen (2014). "Geotagging one hundred million Twitter accounts with total variation minimization". In: *2014 IEEE International Conference on Big Data (Big Data)*, pp. 393–401.

Cooney, Adeline (2010). "Choosing between Glaser and Strauss: an example". In: *Nurse Researcher* 17.4, pp. 18–28.

Corbin, Juliet M and Anselm Strauss (1990). "Grounded theory research: Procedures, canons, and evaluative criteria". In: *Qualitative sociology* 13.1, pp. 3–21.

Corbin, Juliet and Anselm L Strauss (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, Thousand Oaks, CA.

Cramér, Harald (1946). *Mathematical methods of statistics*. Tech. rep.

Crawford, Kate and Megan Finn (2015). "The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters". In: *GeoJournal* 80.4, pp. 491–502.

Crawford, Kate and Jason Schultz (2014). "Big data and due process: Toward a framework to redress predictive privacy harms". In: *BCL Rev.* 55, p. 93.

Cresci, Stefano et al. (2015). "A linguistically-driven approach to cross-event damage assessment of natural disasters from social media messages". In: *Proceedings of the 24th International Conference on World Wide Web*, pp. 1195–1200.

Crooks, Andrew et al. (2013). "#Earthquake: Twitter as a Distributed Sensor System". In: *Transactions in GIS* 17.1, pp. 124–147.

CUREC (2019). *Best Practice Guidance - Internet-Based Research*.

Dashti, Shideh et al. (2014). "Supporting disaster reconnaissance with social media data: A design-oriented case study of the 2013 Colorado floods." In: *ISCRAM*.

Davis Jr., Clodoveu A et al. (2011). "Inferring the Location of Twitter Messages Based on User Relationships". In: *Transactions in GIS* 15.6, pp. 735–751.

De Albuquerque, João Porto et al. (2015). "A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management". In: *International Journal of Geographical Information Science* 29.4, pp. 667–689.

Deitrick, William and Wei Hu (2013). "Mutually enhancing community detection and sentiment analysis on Twitter networks". In: *Journal of Data Analysis and Information Processing* 1.August, pp. 19–29.

Denzin, Norman K and Yvonna S Lincoln (2011). *The Sage handbook of qualitative research*. sage.

Dingwall, Robert and Gale E Miller (1997). "Context and method in qualitative research". In: *Context and method in qualitative research*, pp. 1–240.

Do, Tien Huu et al. (2017). "Multiview deep learning for predicting Twitter users' location". In: *arXiv preprint arXiv:1712.08091*.

Doan, Son, Bao-Khanh Ho Vo, and Nigel Collier (2011). "An analysis of Twitter messages in the 2011 Tohoku Earthquake". In: *International conference on electronic healthcare*. Springer, pp. 58–66.

Donner, William and Walter Diaz (2018). "Methodological issues in disaster research". In: *Handbook of disaster research*. Springer, pp. 289–309.

Dubois, Elizabeth and Devin Gaffney (2014). "The Multiple Facets of Influence: Identifying Political Influentials and Opinion Leaders on Twitter". In: *American Behavioral Scientist*, pp. 0002764214527088–.

Dugdale, Julie et al. (2010). "Simulation and Emergency Management". In: *Information Systems for Emergency Management*. Ed. by Van de Walle et al. Vol. 16. Advances in Management Information Systems. M.E. Sharpe.

Dynes, Russell R (1998). "Coming to terms with community disaster". In: *What is a Disaster*, pp. 109–126.

Dynes, Russell Rowe (1970). *Organized behavior in disaster*. Heath Lexington Books.

Earle, Paul S, Daniel C Bowden, and Michelle Guy (2012). "Twitter earthquake detection: earthquake monitoring in a social world". In: *Annals of Geophysics* 54.6.

Ehnis, Christian and Deborah Bunker (2012). "Social media in disaster response: Queensland police service - Public engagement during the 2011 floods". In: *ACIS 2012 : Proceedings of the 23rd Australasian Conference on Information Systems*, pp. 1–10.

Einwiller, Sabine A and Sarah Steilen (2015). "Handling complaints on social network sites–An analysis of complaints and complaint responses on Facebook and Twitter pages of large US companies". In: *Public Relations Review* 41.2, pp. 195–204.

Elbanna, Amany et al. (2019). "Emergency management in the changing world of social media: Framing the research agenda with the stakeholders through engaged scholarship". In: *International Journal of Information Management* 47, pp. 112–120.

Eshghi, Kourosh and Richard C Larson (2008). "Disasters: lessons from the past 105 years". In: *Disaster Prevention and Management: An International Journal* 17.1, pp. 62–82.

Evans, Harry, Steve Ginnis, and Jamie Bartlett (2015). *#SocialEthics: A Guide to Embedding Ethics in Social Media Research*. Tech. rep. November. Ipsos MORI.

Fallis, A.G (2013). "Real-Time Twitter Mining for Earthquake Detection and Response". In: *Journal of Chemical Information and Modeling* 53.9, pp. 1689–1699. arXiv: `arXiv:1011.1669v3`.

Feinstein, Alvan R and Domenic V Cicchetti (1990). "High agreement but low Kappa: I. the problems of two paradoxes". In: *Journal of Clinical Epidemiology* 43.6, pp. 543–549.

Fernandez-Marquez, Jose Luis et al. (2017). "E2mC: Improving rapid mapping with social network information". In: *iTAIS Conference*. Milan.

Fiesler, Casey, Cliff Lampe, and Amy S Bruckman (2016). "Reality and Perception of Copyright Terms of Service for Online Content Creation". In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW '16. New York, NY, USA: Association for Computing Machinery, pp. 1450–1461.

Fiesler, Casey and Nicholas Proferes (2018). ""Participant" Perceptions of Twitter Research Ethics". In: *Social Media and Society* 4.1.

Finch, Kathryn C et al. (2016). "Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns". In: *Natural Hazards* 83.1, pp. 729–760.

Fleischman, Alan R, Lauren Collogan, and Farris Tuma (2006). *Ethical issues in disaster research*.

Fohringer, J. et al. (2015). "Social media as an information source for rapid flood inundation mapping". In: *Natural Hazards and Earth System Sciences* 15.12, pp. 2725–2738.

Foster, Jacob G et al. (2010). "Edge direction and the structure of networks". In: *Proceedings of the National Academy of Sciences* 107.24, pp. 10815–10820.

Francalanci, Chiara and Barbara Pernici (2018). "Data integration and quality requirements in emergency services". In: *Communications in Computer and Information Science* 707, pp. 211–218.

Freelon, Deen (2018). "Computational Research in the Post-API Age". In: *Political Communication* 35.4, pp. 665–668.

Fritz, Charles E and John H Mathewson (1957). *Convergence behavior in disasters: A problem in social control*. 9. National Academy of Sciences-National Research Councel.

Fritz, Charles Edward (1961). *Disaster*. Institute for Defense Analyses, Weapons Systems Evaluation Division.

Funes, Federico M, José Ignacio Alvarez-Hamelin, and Mariano G Beiró (2021). "Designing weighted and multiplex networks for deep learning user geolocation in Twitter". In: *arXiv preprint arXiv:2112.06999*.

Gao, Huiji, Geoffrey Barbier, and Rebecca Goolsby (2011). "Harnessing the crowdsourcing power of social media for disaster relief". In: *IEEE Intelligent Systems* 26.3, pp. 10–14.

Gheorghe, Mihai, Florin-Cristian Mihai, and Marian Dârdal\ua (2018). "Modern techniques of web scraping for data scientists". In: *Romanian Journal of Human-Computer Interaction* 11.1, pp. 63–75.

Giasemidis, Georgios et al. (2016). "Determining the veracity of rumours on Twitter". In: *International Conference on Social Informatics*. Springer, pp. 185–205.

Glaser, Barney G (1992). *Basics of Grounded Theory Analysis*. Sociology Press.

Glaser, Barney G and Anselm L Strauss (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Vol. 1. 4, p. 271. arXiv: `9809069v1 [arXiv:gr-qc]`.

Goldenberg, Jacob et al. (2009). "The Role of Hubs in the Adoption Process". In: *Journal of Marketing* 73, pp. 1–13.

Goodchild, Michael F. (2007). "Citizens as sensors: The world of volunteered geography". In: *GeoJournal* 69.4, pp. 211–221.

Goodman, Leo A (1961). "Snowball Sampling". In: *The Annals of Mathematical Statistics*, pp. 148–170.

Greene, Jennifer C (2007). *Mixed methods in social inquiry*. Vol. 9. John Wiley & Sons.

Gu, Yiming, Zhen (Sean) Qian, and Feng Chen (2016). "From Twitter to detector: Real-time traffic incident detection using social media data". In: *Transportation Research Part C: Emerging Technologies* 67, pp. 321–342.

Guha-Sapir, Debarati, Philippe Hoyois, and Regina Below (2015). "Annual Disaster Statistical Review 2015". In: *Centre for Research on the Epidemiology of Disasters*.

Guo, Lei, Jacob A. Rohde, and H Denis Wu (2020). "Who is responsible for Twitter's echo chamber problem? Evidence from 2016 US election networks". In: *Information, Communication & Society* 23.2, pp. 234–251.

Gupta, Aditi, Anupam Joshi, and Ponnurangam Kumaraguru (2012). "Identifying and Characterizing User Communities on Twitter During Crisis Events". In: *Proceedings of the 2012 Workshop on Data-driven User Behavioral Modelling and Mining from Social Media*, pp. 23–26.

Gupta, Aditi, Ponnurangam Kumaraguru, and Carlos Castillo (2014). "TweetCred : Real-Time Credibility Assessment". In: *International Conference on Social Informatics*. Springer International Publishing, pp. 228–243. arXiv: `arXiv:1405.5490v2`.

Gupta, Aditi, Hemank Lamba, et al. (2013). "Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy". In: *WWW '13 Companion Proceedings of the 22nd International Conference on World Wide Web*, Pages 729–736.

Haley, Craig (2021). "The Origins Of The Notorious "Shark On Flooded Highway" Photo". In: *https://www.thatsnonsense.com/the-fake-highway-shark-photo/*.

Hammersley, Martyn and Paul Atkinson (2007). *Ethnography: Principles in Practice*. 3rd ed. Routledge, p. 275.

Han, Su Yeon et al. (2019). "How Do Cities Flow in an Emergency? Tracing Human Mobility Patterns during a Natural Disaster with Big Data and Geospatial Data Science". In: *Urban Science* 3.2, p. 51.

Han, Yi, Shanika Karunasekera, and Christopher Leckie (2020). "Image Analysis Enhanced Event Detection from Geo-tagged Tweet Streams". In: *arXiv preprint arXiv:2002.04208*.

Hanteer, Obaida et al. (2018). "From interaction to participation: The role of the imagined audience in social media community detection and an application to political communication on twitter". In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, pp. 531–534.

Hasan, Mahmud, Mehmet A Orgun, and Rolf Schwitter (2018). "A survey on real-time event detection from the twitter data stream". In: *Journal of Information Science* 44.4, pp. 443–463.

— (2019). "Real-time event detection from the Twitter data stream using the TwitterNews+ Framework". In: *Information Processing & Management* 56.3, pp. 1146–1165.

Hecht, Brent et al. (2011). "Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles". In: *Proceedings of the 2011 annual conference on Human factors in computing systems* Figure 1, pp. 237–246.

Helsloot, Ira (2005). "Bordering on reality: Findings on the bonfire crisis management simulation". In: *Journal of Contingencies and Crisis Management* 13.4, pp. 159–169.

Herfort, Benjamin et al. (2014). "Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe Flood in June 2013." In: *ISCRAM*.

Hernandez-Suarez, A. et al. (2018). "A Web Scraping Methodology for Bypassing Twitter API Restrictions". In: pp. 1–7. arXiv: `1803.09875`.

Hiltz, Starr Roxanne, Paloma Diaz, and Gloria Mark (2011). "Introduction: Social Media and Collaborative Systems for Crisis Management". In: *ACM Trans. Comput.-Hum. Interact.* 18.4.

Hiltz, Starr Roxanne, Amanda Hughes, et al. (2020). "Exploring the usefulness and feasibility of software requirements for social media use in emergency management". In: *International Journal of Disaster Risk Reduction* 42.

Hiltz, Starr Roxanne, Jane Kushma, and Linda Plotnick (2014). "Use of Social Media by U.S. Public Sector Emergency Managers: Barriers and Wish Lists". In: *Proceedings of ISCRAM* 279.

Holloway, Immy and Les Todres (2003). "The Status of Method: Flexibility, Consistency and Coherence". In: *Qualitative Research* 3.3, pp. 345–357.

Hu, Huiqi et al. (2016). "Crowdsourced POI labelling: Location-aware result inference and task assignment". In: *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*. IEEE, pp. 61–72.

Huang, Qunying and Yu Xiao (2015a). "Geographic Situational Awareness: Mining Tweets for Disaster Preparedness, Emergency Response, Impact, and Recovery". In: *ISPRS International Journal of Geo-Information* 4.3, pp. 1549–1568.

— (2015b). "Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery". In: *ISPRS International Journal of Geo-Information* 4.3, pp. 1549–1568.

Huang, Xiao et al. (2019). "Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture". In: *International Journal of Digital Earth*.

Hughes, Amanda Lee and Rohan Shah (2016). "Designing an application for social media needs in emergency public information work". In: *Proceedings of the 19th International Conference on Supporting Group Work*, pp. 399–408.

Hughes, Amanda and Leysia Palen (2010). "Twitter adoption and use in mass convergence and emergency events". In: *International Journal of Emergency Management* 6.3/4, p. 248.

— (2012). "The evolving role of the public information officer: An examination of social media in emergency management". In: *Journal of Homeland Security and Emergency Management* 9.1.

Hughes, Amanda, Leysia Palen, et al. (2008). ""Site-Seeing" in Disaster : An Examination of On-Line Social Convergence". In: *5th International ISCRAM Conference*. May, pp. 44–54.

Hughes, Amanda, Lise A St. Denis, et al. (2014). "Online public communications by police & fire services during the 2012 Hurricane Sandy". In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, pp. 1505–1514.

Imran, Muhammad, Carlos Castillo, Fernando Diaz, et al. (2015). "Processing Social Media Messages in Mass Emergency: A Survey". In: *ACM Computing Surveys* 47.4, pp. 1–38. arXiv: `arXiv:1407.7071v1`.

Imran, Muhammad, Carlos Castillo, Ji Lucas, et al. (2014). "AIDR: Artificial intelligence for disaster response". In: *Proceedings of the 23rd International Conference on World Wide Web*. April. ACM, pp. 159–162.

Imran, Muhammad, Prasenjit Mitra, and Jaideep Srivastava (2016). "Cross-language domain adaptation for classifying crisis-related short messages". In: *arXiv preprint arXiv:1602.05388*.

Imran, Muhammad, Ferda Ofli, et al. (2020). *Using AI and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions.*

Inuwa-Dutse, Isa, Mark Liptrott, and Ioannis Korkontzelos (2021). "A multilevel clustering technique for community detection". In: *Neurocomputing* 441, pp. 64–78.

Itakura, Kelly Y and Noboru Sonehara (2013). "Using Twitter's Mentions for Efficient Emergency Message Propagation". In: *2013 International Conference on Availability, Reliability and Security*, pp. 530–537.

Jacomy, Mathieu et al. (2014). "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software". In: *PloS one* 9.6.

Jurgens, David (2021). "That's What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships". In: *Proceedings of the International AAAI Conference on Web and Social Media* 7.1, pp. 273–282.

Jurgens, David et al. (2015). "Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice". In: *The 9th International Conference on Weblogs and Social Media (ICWSM)*, pp. 1–10.

Kanavos, Andreas et al. (2022). "Evaluating Methods for Efficient Community Detection in Social Networks". In: *Information* 13.5, p. 209.

Karami, Amir et al. (2020). "Twitter speaks: A case of national disaster situational awareness". In: *Journal of Information Science* 46.3, pp. 313–324.

Karimi, Sarvnaz, Jie Yin, and Cecile Paris (2013). "Classifying microblogs for disasters". In: *Proceedings of the 18th Australasian document computing symposium*, pp. 26–33.

Kaur, Wandeep et al. (2019). "Liking, sharing, commenting and reacting on Facebook: User behaviors' impact on sentiment intensity". In: *Telematics and Informatics* 39, pp. 25–36.

Keller, Ed and Jon Berry (2003). *The Influentials: One American in ten tells the other nine how to vote, where to eat, and what to buy.* New York: Free Press.

Keller, Franziska B et al. (2020). "Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign". In: *Political Communication* 37.2, pp. 256–280.

Kendra, James M and Tricia Wachtendorf (2003). "Reconsidering convergence and converger legitimacy in response to the World Trade Center disaster". In: *Research in Social Problems and Public Policy* 11.1, pp. 97–122.

Khaled, Sarah, Neamat El-Tazi, and Hoda M O Mokhtar (2018). "Detecting fake accounts on social media". In: *2018 IEEE international conference on big data (big data)*. IEEE, pp. 3672–3681.

Killian, Lewis M (1954). "Some accomplishments and some needs in disaster study". In: *Journal of Social Issues* 10.3, pp. 66–72.

Kogan, Marina, Leysia Palen, and Kenneth M Anderson (2015). "Think Local , Retweet Global : Retweeting by the Geographically - Vulnerable during Hurricane Sandy". In:

Kong, Longbo, Zhi Liu, and Yan Huang (2014). "Spot: Locating social media users based on social network context". In: *Proceedings of the VLDB Endowment* 7.13, pp. 1681–1684.

Kongthon, Alisa et al. (2014). "The Role of Social Media Studyuring a Natural Disaster: A Case of the 2011 Thai Flood". In: *International Journal of Innovation and Technology Management* 11.03.

Kosinski, Michal et al. (2015). "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines". In: *American Psychologist* 70.6, pp. 543–556.

Krippendorff, Klaus (2004). *Content analysis: An introduction to its methodology*. 2nd ed.
Kumar, Abhinav and Jyoti Prakash Singh (2019). "Location reference identification from tweets during emergencies: A deep learning approach". In: *International journal of disaster risk reduction* 33, pp. 365–375.
Kumar, Shamanth et al. (2011). "TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief". In: *Fifth International AAAI Conference on Weblogs and Social Media*, pp. 661–662.
Kundu, Suman, C. A. Murthy, and S. K. Pal (2011). "A new centrality measure for influence maximization in social networks". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6744 LNCS, pp. 242–247.
Kwak, Haewoon et al. (2010). "What is Twitter, a Social Network or a News Media?" In: *Proceedings of the 19th international conference on World wide web*. AcM, pp. 591–600.
Kwon, K. Hazel et al. (2015). "A spatiotemporal model of Twitter information diffusion". In: *Proceedings of the 2015 International Conference on Social Media & Society - SMSociety '15* 2015-July, pp. 1–7.
Lachlan, Kenneth A et al. (2014). "Screaming into the wind: Examining the volume and content of tweets associated with Hurricane Sandy". In: *Communication Studies* 65.5, pp. 500–518.
Lamb, Alex, Michael J Paul, and Mark Dredze (2013). "Separating Fact from Fear : Tracking Flu Infections on Twitter". In: *Proceedings of NAACL-HLT 2013* June, pp. 789–795.
Landis, J Richard and Gary G Koch (1977). "The measurement of observer agreement for categorical data". In: *Biometrics*, pp. 159–174.
Lapouchnian, Alexei (2005). "Goal-oriented requirements engineering: An overview of the current research". In: *University of Toronto* 32.
Latonero, Mark and Irina Shklovski (2011). "Emergency management, Twitter, and social media evangelism". In: *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* 3.4, pp. 1–16.
Laylavi, Farhad, Abbas Rajabifard, and Mohsen Kalantari (2016). "A multi-element approach to location inference of twitter: A case for emergency response". In: *ISPRS International Journal of Geo-Information* 5.5, p. 56.
— (2017). "Event relatedness assessment of Twitter messages for emergency response". In: *Information processing & management* 53.1, pp. 266–280.
Leetaru, Kalev (2019a). *Is Twitter's Spritzer Stream Really A Nearly Perfect 1% Sample Of Its Firehose?*
— (2019b). "Visualizing Seven Years Of Twitter's Evolution: 2012–2018". In: *Forbes*.
Ley, Benedikt et al. (2012). "Supporting Improvisation Work in Inter-Organizational Crisis Management". In: *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1529–1538.
Li, Pengfei et al. (2019). "Location Inference for Non-Geotagged Tweets in User Timelines". In: *IEEE Transactions on Knowledge and Data Engineering* 31.6, pp. 1150–1165.
Littman, Justin (2017). *Hurricanes Harvey and Irma Tweet ids*.
Liu, Sophia B. (2014). "Crisis Crowdsourcing Framework: Designing Strategic Configurations of Crowdsourcing for the Emergency Management Domain". In:

Computer Supported Cooperative Work: CSCW: An International Journal 23.4-6, pp. 389–443.

Liu, Zhi and Yan Huang (2014). "Community Detection from Location-Tagged Networks". In: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. SIGSPATIAL '14. New York, NY, USA: Association for Computing Machinery, pp. 525–528.

Lotan, Gilad et al. (2011). "The Revolutions Were Tweeted: Information Flows during the 2011 Tunisian and Egyptian Revolutions". In: *International Journal of Communication* 5, p. 31.

Ludwig, Thomas et al. (2015). "CrowdMonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4083–4092.

Luger, Ewa, Stuart Moran, and Tom Rodden (2013). "Consent for All: Revealing the Hidden Complexity of Terms and Conditions". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. New York, NY, USA: Association for Computing Machinery, pp. 2687–2696.

Madey, Gregory R et al. (2007). "Enhanced situational awareness: Application of DDDAS concepts to emergency and disaster management". In: *International conference on computational science*. Springer, pp. 1090–1097.

Madianou, Mirca (2019). "Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises". In: *Social Media + Society* 5.3.

Manguri, Kamaran H, Rebaz N Ramadhan, and Pshko R Mohammed Amin (2020). "Twitter sentiment analysis on worldwide COVID-19 outbreaks". In: *Kurdistan Journal of Applied Research*, pp. 54–65.

Marbouti, Mahshid and Frank Maurer (2016). "Social Media Use During Emergency Response - Insights from Emergency Professionals". In: *Social Media: The Good, the Bad, and the Ugly*. Ed. by Yogesh K Dwivedi et al. Vol. 9844. Cham: Springer International Publishing, pp. 557–566.

Marcus, Adam et al. (2011). "Twitinfo: aggregating and visualizing microblogs for event exploration". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 227–236.

Markham, Annette (2012). "Fabrication as ethical practice: Qualitative inquiry in ambiguous Internet contexts". In: *Information Communication and Society* 15.3, pp. 334–353.

Martí, Pablo, Leticia Serrano-Estrada, and Almudena Nolasco-Cirugeda (2019). "Social media data: Challenges, opportunities and limitations in urban studies". In: *Computers, Environment and Urban Systems* 74, pp. 161–174.

Martínez-Rojas, María, María del Carmen Pardo-Ferreira, Antonio López-Arquillos, et al. (2019). "Using Twitter as a Tool to Foster Social Resilience in Emergency Situations: A Case of Study". In: *Engineering Digital Transformation*. Springer, pp. 243–245.

Martínez-Rojas, María, María del Carmen Pardo-Ferreira, and Juan Carlos Rubio-Romero (2018). "Twitter as a tool for the management and analysis of emergency situations: A systematic literature review". In: *International Journal of Information Management* 43, pp. 196–208.

Mason, Claire and Robert Power (2015). "Improving social media monitoring and analysis tools for emergency management". In: *MODSIM2015, 21st International*

Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 1195–1201.

Masood, Faiza et al. (2019). "Spammer Detection and Fake User Identification on Social Networks". In: *IEEE Access* 7, pp. 68140–68152.

McLennan, Jim et al. (2006). "Decision making effectiveness in wildfire Incident Management Teams". In: *Journal of Contingencies and Crisis Management* 14.1, pp. 27–37.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a feather: Homophily in social networks". In: *Annual review of sociology* 27.1, pp. 415–444.

Meesters, Kenny, Lars Van Beek, and Bartel Van De Walle (2016). "# Help. The Reality of Social Media Use in Crisis Response: Lessons from a Realistic Crisis Exercise". In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, pp. 116–125.

Meier, Patrick and Kate Brodock (2008). "Crisis mapping Kenya's election violence: Comparing mainstream news, citizen journalism and Ushahidi". In: *iRevolution Blog, October* 23.

Mendoza, Marcelo, Barbara Poblete, and Carlos Castillo (2010). "Twitter under crisis: Can we trust what we RT?" In: *Proceedings of the first workshop on social media analytics*. ACM, pp. 71–79.

Mileti, Dennis S, Thomas E Drabek, and John Eugene Haas (1975). *Human systems in extreme environments: A sociological perspective*. Vol. 26. Institute of Behavioral Science, University of Colorado.

Militello, Laura G. et al. (2007). "Information flow during crisis management: Challenges to coordination in the emergency operations center". In: *Cognition, Technology and Work* 9.1, pp. 25–31.

Miura, Yasuhide et al. (2017). "Unifying text, metadata, and user network representations with a neural network for geolocation prediction". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1260–1272.

Miyabe, Mai, Asako Miura, and Eiji Aramaki (2012). "Use trend analysis of twitter after the great east japan earthquake". In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, pp. 175–178.

Moats, Jason B, Thomas J Chermack, and Larry M Dooley (2008). "Using Scenarios to Develop Crisis Managers: Applications of Scenario Planning and Scenario-Based Training". In: *Advances in Developing Human Resources* 10.3, pp. 397–424.

Mohotti, Wathsala Anupama and Richi Nayak (2018). "Corpus-based augmented media posts with density-based clustering for community detection". In: *2018 IEEE 30th International conference on tools with artificial intelligence (ICTAI)*. IEEE, pp. 379–386.

Moody, James (2001). "Race, School Integration, and Friendship Segregation in America". In: *American Journal of Sociology* 107.3, pp. 679–716.

Moore, Harry Estil (1958). "Tornadoes over Texas: A study of Waco and San Angelo in disaster". In:

Moran, Dermot (2002). *Introduction to phenomenology*. Routledge.

Morrow, Nathan et al. (2011). "Independent evaluation of the Ushahidi Haiti project". In: *Development Information Systems International*, pp. 1–36.

Morstatter, Fred et al. (2013). "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose". In: *Proceedings of the International*

AAAI Conference on Web and Social Media 7.1, pp. 400–408. arXiv: arXiv:1306.5204v1.

Muhren, Willem J and Bartel de Walle (2009). "Sensemaking and information management in humanitarian disaster response: Observations from the triplex exercise". In: *Proceedings of the 6th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Vol. 32.

Murakami, Daisuke et al. (2016). "Participatory Sensing Data Tweets for Micro-Urban Real-Time Resiliency Monitoring and Risk Management". In: *IEEE Access* 4, pp. 347–372.

Murthy, Dhiraj and Scott A. Longwell (2013). "Twitter and Disasters: The uses of Twitter during the 2010 Pakistan floods". In: *Information Communication and Society* 16.6, pp. 837–855.

Murzintcev, Nikita and Changxiu Cheng (2017). "Disaster Hashtags in Social Media". In: *ISPRS International Journal of Geo-Information* 6.7.

Nazir, Atif, Saqib Raza, and Chen-nee Chuah (2008). "Unveiling facebook: a measurement study of social network based applications". In: *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. Vouliagmeni, pp. 43–56.

Neppalli, Venkata K et al. (2017). "Sentiment analysis during Hurricane Sandy in emergency response". In: *International journal of disaster risk reduction* 21, pp. 213–222.

Newman, Mark E. J. (2003). "Mixing patterns in networks". In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 67.2, p. 13. arXiv: 0209450 [cond-mat].

— (2006). "Modularity and community structure in networks". In: *Proceedings of the National Academy of Sciences of the United States of America* 103.23, pp. 8577–82. arXiv: 0602124 [physics].

— (2018). *Networks*. 2nd ed. Oxford University Press.

Nissenbaum, Helen (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Nurse, Jason R. C. et al. (2015). "Tag clouds with a twist: using tag clouds coloured by information's trustworthiness to support situational awareness". In: *Journal of Trust Management* 2.1, p. 10.

Ofli, Ferda, Muhammad Imran, and Firoj Alam (2020). "Using artificial intelligence and social media for disaster response and management: an overview". In: *AI and Robotics in Disaster Studies*, pp. 63–81.

Ofli, Ferda, Patrick Meier, et al. (2016). "Combining human computing and machine learning to make sense of big (aerial) data for disaster response". In: *Big data* 4.1, pp. 47–59.

Oh, Onook, Kyounghee Hazel Kwon, and H Raghav Rao (2010). "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010." In: *Thirty First International Conference on Information Systems*. Vol. 231. St. Louis, pp. 7332–7336.

OHCHR and UC Berkeley (2022). *Berkeley Protocol on Digital Open Source Investigations*. New York and Geneva: OHCHR & Human Rights Center UC Berkeley.

Olteanu, Alexandra, Carlos Castillo, et al. (2015). "CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises". In: *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* 35.2, p. 9. arXiv: 1604.00758.

Olteanu, Alexandra, Sarah Vieweg, and Carlos Castillo (2015). "What to Expect When the Unexpected Happens: Social Media Communications Across Crises". In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, pp. 994–1009.

Palen, Leysia (2008). "Online Social Media in Crisis Events". In: *Educause Quarterly* 31.3, pp. 76–78.

— (2014). "Frontiers in Crisis Informatics". In: *ISCRAM*.

Palen, Leysia and Kenneth M Anderson (2016). "Crisis informatics: New data for extraordinary times". In: *Science* 353.6296, pp. 224–225.

Palen, Leysia and Amanda Hughes (2018). "Social Media in Disaster Communication". In: *Handbook of Disaster Research*. Ed. by Havidán Rodríguez, William Donner, and Joseph E Trainor. Springer International Publishing, pp. 497–518.

Palen, Leysia and Sophia B Liu (2007). "Citizen communications in crisis: anticipating a future of ICT-supported public participation". In: *Natural Hazards*, pp. 727–736.

Palen, Leysia, Kate Starbird, et al. (2010). "Twitter-based information distribution during the 2009 Red River Valley flood threat". In: *Bulletin of the American Society for Information Science and Technology* 36.5, pp. 13–17.

Palen, Leysia, Sarah Vieweg, et al. (2007). "Crisis Informatics : Studying Crisis in a Networked World". In: *Third International Conference on e-Social Science*. Ann Arbor, Michigan.

Palinkas, Lawrence, Michael Downs, et al. (1993). "Social, cultural, and psychological impacts of the Exxon Valdez oil spill". In: *Human Organization* 52.1, pp. 1–13.

Palinkas, Lawrence, Erica Prussing, et al. (2004). "The San Diego East County School Shootings: A Qualitative Study of Community-Level Post-traumatic Stress". In: *Prehospital and Disaster Medicine* 19.01, pp. 113–121.

Parsons, Will (1996). "Crisis Management". In: *Career Development International* 1.5, pp. 26–28.

Patton, Michael Quinn (1999). "Enhancing the quality and credibility of qualitative analysis." In: *Health services research* 34.5 Pt 2, p. 1189.

Peek, Lori A and Jeannette N Sutton (2003). "An exploratory comparison of disasters, riots and terrorist acts". In: *Disasters* 27.4, pp. 319–335.

Perriam, Jessamy, Andreas Birkbak, and Andy Freeman (2019). "Digital methods in a post-API environment". In: *International Journal of Social Research Methodology*, pp. 1–14.

Perry, Ronald W (2018). "Defining Disaster: An Evolving Concept". In: *Handbook of Disaster Research*. Springer, pp. 3–22.

Perry, Ronald W and Enrico Louis Quarantelli (2005). *What is a Disaster?: New Answers to Old Questions*. Xlibris Corporation.

Peschak, Thomas (2007). "White Shark Kayak". In: *https://www.thomaspeschak.com/kayak-great-white-sharks-/*.

Plotnick, Linda et al. (2015). "Red tape: Attitudes and issues related to use of social media by US county-level emergency managers". In: *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). Kristiansand, Norway.*

Polanyi, Michael (1958). "Personal Knowledge towards a Post-Critical Philosophy". In:

Poletto, Fabio et al. (2021). "Resources and benchmark corpora for hate speech detection: a systematic review". In: *Language Resources and Evaluation* 55.2, pp. 477–523.

Potter, Emma (2016). "Balancing conflicting operational and communications priorities: Social media use in an emergency management organization". In: *ISCRAM 2016*

Conference Proceedings-13th International Conference on Information Systems for Crisis Response and Management. International Association for Information Systems for Crisis Response and˜. . ., pp. 1–10.

Pournaki, Armin et al. (2020). "The twitter explorer: a framework for observing Twitter through interactive networks". In: *arXiv preprint*. arXiv: 2003.03599.

Power, Robert and Justin Kibell (2017). "The Social Media Intelligence Analyst for Emergency Management". In: *Proceedings of the 50th Hawaii International Conference on System Sciences*.

Procter, Rob, Jeremy Crump, et al. (2013). "Reading the riots: What were the police doing on Twitter?" In: *Policing & Society* 23.4, pp. 413–436.

Procter, Rob, Farida Vis, and Alex Voss (2013). "Reading the Riots on Twitter: Methodological Innovation for the Analysis of Big Data". In: *International Journal of Social Research Methodology* 16.3, pp. 197–214.

Qu, Y, Pf Wu, and X Wang (2009). "Online Community Response to Major Disaster: A Case Study of Tianya Forum in the 2008 China Earthquake". In: *42nd Hawaii International Conference on System Sciences* January, pp. 1–11.

Quarantelli, Enrico Louis (1984). "Perceptions and reactions to emergency warnings of sudden hazards". In: *Ekistics*, pp. 511–515.

— (1987). "Disaster studies: An analysis of the social historical factors affecting the development of research in the area". In:

— (1993). "Community crises: An exploratory comparison of the characteristics and consequences of disasters and riots". In: *Journal of contingencies and crisis management* 1.2, pp. 67–78.

— (1997). "The Disaster Research Center (DRC) Field Studies of Organized Behavior in the Crisis Time Period of Disasters". In: *International Journal of Mass Emergencies and Disasters* 15.1, pp. 47–69.

— (2000). *Disaster research*. University of Delaware, pp. 681–688.

— (2005). "A social science research agenda for the disasters of the 21st century: Theoretical, methodological and empirical issues and their professional implementation". In: *What is a disaster* 139, pp. 325–396.

Quercia, Daniele et al. (2011). "In the Mood For Being Influential on Twitter". In: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. Boston, MA, pp. 307–314. arXiv: arXiv:1011.1669v3.

Rahimi, Afshin, Trevor Cohn, and Timothy Baldwin (2015). "Twitter user geolocation using a unified text and network prediction model". In: *arXiv preprint arXiv:1506.08259*.

— (2018). "Semi-supervised user geolocation via graph convolutional networks". In: *arXiv preprint arXiv:1804.08049*.

Rains, Stephen A and Steven R Brunner (2015). "What can we learn about social network sites by studying Facebook? A call and recommendations for research on social network sites". In: *New Media & Society* 17.1, pp. 114–131.

Reidenberg, Joel R et al. (2015). "Disagreeable privacy policies: Mismatches between meaning and users' understanding". In: *Berkeley Tech. LJ* 30, p. 39.

Ren, Kejiang, Shaowu Zhang, and Hongfei Lin (2012). "Where are you settling down: Geo-locating Twitter users based on tweets and social networks". In: *Asia Information Retrieval Symposium*. Springer, pp. 150–161.

Reuter, Christian, Amanda Hughes, and Marc-André Kaufhold (2018). "Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research". In: *International Journal of Human–Computer Interaction*.

Reuter, Christian and Marc-André André Kaufhold (2018). "Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics". In: *Journal of Contingencies and Crisis Management* 26.1, pp. 41–57.

Reuter, Christian, Thomas Ludwig, Marc-André Kaufhold, and Volkmar Pipek (2015). "XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters". In: *Chi*, pp. 4093–4102.

Reuter, Christian, Thomas Ludwig, Marc-André Kaufhold, and Thomas Spielhofer (2016). "Emergency services attitudes towards social media: A quantitative and qualitative survey across Europe". In: *International Journal of Human-Computer Studies* 95, pp. 96–111.

Reuter, Christian, Alexandra Marx, and Volkmar Pipek (2012). "Crisis management 2.0: Towards a systematization of social software use in crisis situations". In: *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* 4.1, pp. 1–16.

Rieder, Bernhard (2013). "Studying Facebook via Data Extraction: The Netvizz Application". In: *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*, pp. 346–355.

Robinson, Bella, Robert Power, and Mark Cameron (2013). "A sensitive twitter earthquake detector". In: *Proceedings of the 22nd international conference on world wide web*, pp. 999–1002.

Rodrigues, Erica et al. (2016). "Exploring multiple evidence to infer users' location in Twitter". In: *Neurocomputing* 171, pp. 30–38.

Rodríguez, Havidán, William Donner, and Joseph E Trainor (2018). *Handbook of Disaster Research*. 2nd ed. Springer.

Rogstadius, Jakob et al. (2013). "CrisisTracker: Crowdsourced social media curation for disaster awareness". In: *IBM Journal of Research and Development* 57.5, pp. 1–4.

Rolland, Colette, Carine Souveyet, and Camille Ben Achour (1998). "Guiding goal modeling using scenarios". In: *IEEE transactions on software engineering* 24.12, pp. 1055–1071.

Roshan, Mina, Matthew Warren, and Rodney Carr (2016). "Understanding the use of social media by organisations for crisis communication". In: *Computers in Human Behavior* 63, pp. 350–361.

Rosser, J. F., D. G. Leibovici, and M. J. Jackson (2017). "Rapid flood inundation mapping using social media, remote sensing and topographic data". In: *Natural Hazards* 87.1.

Rudra, Koustav et al. (2018). "Identifying sub-events and summarizing disaster-related information from microblogs". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 265–274.

Ruiz, Jeanette, Jade D Featherstone, and George A Barnett (2021). "Identifying vaccine hesitant communities on twitter and their geolocations: a network approach". In: *Proceedings of the 54th Hawaii international conference on system sciences*, pp. 3964–3969.

Ruz, Gonzalo A, Pablo A Henríquez, and Aldo Mascareño (2020). "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers". In: *Future Generation Computer Systems* 106, pp. 92–104.

Al-Saggaf, Yeslam and Peter Simmons (2015). "Social media in Saudi Arabia: Exploring its use during two natural disasters". In: *Technological Forecasting and Social Change* 95, pp. 3–15.

Sahoo, Somya Ranjan and BB Gupta (2021). "Real-time detection of fake account in twitter using machine-learning approach". In: *Advances in Computational Intelligence and Communication Technology*. Ed. by Xiao-Zhi Gao et al. Vol. 1086. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore, pp. 149–159.

Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo (2010). "Earthquake Shakes Twitter Users : Real-time Event Detection by Social Sensors". In: *Proceedings of the 19th international conference on World wide web*. ACM, pp. 851–860.

— (2012). "Tweet analysis for real-time event detection and earthquake reporting system development". In: *IEEE Transactions on Knowledge and Data Engineering* 25.4, pp. 919–931.

Santos, José Carlos and Sérgio Matos (2014). "Analysing Twitter and web queries for flu trend prediction". In: *Theoretical Biology and Medical Modelling* 11.1, pp. 1–11.

Sarter, Nadine B and David D Woods (1991). "Situation awareness: A critical but ill-defined phenomenon". In: *The International Journal of Aviation Psychology* 1.1, pp. 45–57.

Schempp, Timothy et al. (2019). "A framework to integrate social media and authoritative data for disaster relief detection and distribution optimization". In: *International Journal of Disaster Risk Reduction*, p. 101143.

Schroeder, Daniel Thilo, Konstantin Pogorelov, and Johannes Langguth (2019). "FACT: a Framework for Analysis and Capture of Twitter Graphs". In: *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, pp. 134–141.

Schuler, Douglas and Aki Namioka (1993). *Participatory design: Principles and practices*. CRC Press.

Scrimshaw, Susan C and Elena Hurtado (1987). "Rapid assessment procedures for nutrition and primary health care. Anthropological approaches to improving programme effectiveness." In:

Shaw, Frances et al. (2013). "Sharing news, making sense, saying thanks: Patterns of talk on Twitter during the Queensland floods". In: *Australian Journal of Communication* 40.1, pp. 23–39.

Shklovski, Irina et al. (2010). "Technology adoption and use in the aftermath of Hurricane Katrina in New Orleans". In: *American Behavioral Scientist* 53.8, pp. 1228–1246.

Shu, Kai et al. (2017). "Fake news detection on social media: A data mining perspective". In: *ACM SIGKDD explorations newsletter* 19.1, pp. 22–36.

Silva, Wendel et al. (2017). "A Methodology for Community Detection in Twitter". In: *Proceedings of the International Conference on Web Intelligence*. WI '17. New York, NY, USA: Association for Computing Machinery, pp. 1006–1009.

Simmie, Donal, Maria Grazia Vigliotti, and Chris Hankin (2014). "Ranking Twitter Influence By Combining Network Centrality and Influence Observables in an Evolutionary Model". In: *Journal of Complex Networks* 2.4, pp. 495–517.

Simonsen, Jesper and Toni Robertson (2013). *Routledge international handbook of participatory design*. Vol. 711. Routledge New York.

Sinnappan, Suku, Cathy Farrell, and Elizabeth Stewart (2010). "Priceless tweets! A study on Twitter messages posted during crisis: Black Saturday". In:

Smith, Marc A et al. (2014). *Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters*. Tech. rep. Pew Research Center.

Sokolowski, Robert (2000). *Introduction to phenomenology*. Cambridge university press.

Solove, Daniel J (2012). "Privacy Self-Management and the Consent Dilemma". In: *Harvard Law Review* 126.

Spielhofer, Thomas et al. (2016). "Data mining twitter during the UK floods Investigating the potential use of social media in emergency management". In: *Proceedings of the 2016 3rd International Conference on Information and Communication Technologies for Disaster Management, ICT-DM 2016*.

Sreenivasan, Nirupama Dharmavaram, Chei Sian Lee, and Dion Hoe-Lian Goh (2011). "Tweet me home: Exploring information use on Twitter in crisis situations". In: *International Conference on Online Communities and Social Computing*. Springer, pp. 120–129.

Stallings, Robert A (2003). *Methods of disaster research*. Xlibris Corporation.

— (2007). "Methodological issues". In: *Handbook of disaster research*. Springer, pp. 55–82.

Starbird, Kate, Grace Muzny, and Leysia Palen (2012). "Learning from the Crowd : Collaborative Filtering Techniques for Identifying On-the-Ground Twitterers during Mass Disruptions". In: *Iscram* 2011.April, pp. 1–10.

Starbird, Kate and Leysia Palen (2011). "Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis". In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pp. 1071–1080.

Stieglitz, Stefan, Deborah Bunker, et al. (2018). "Sense-making in social media during extreme events". In: *Journal of Contingencies and Crisis Management* 26.1, pp. 4–15.

Stieglitz, Stefan, Milad Mirbabaie, et al. (2018). "The Adoption of social media analytics for crisis management–Challenges and Opportunities". In:

Stosz, Vice Admiral Sandra (2017). *Conference Presentation*. Washington, D.C.

Stowe, Kevin et al. (2018). "Developing and Evaluating Annotation Procedures for Twitter Data during Hazard Events". In: *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pp. 133–143.

Strauss, Anselm L (1987). *Qualitative analysis for social scientists*. Cambridge University Press.

Takahashi, Bruno, Edson C Tandoc Jr, and Christine Carmichael (2015). "Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines". In: *Computers in human behavior* 50, pp. 392–398.

Tanev, Hristo, Vanni Zavarella, and Josef Steinberger (2017). "Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media." In: *ISCRAM*.

Tapia, Andrea, Kartikeya Bajpai, et al. (2011). "Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations". In: *Proceedings of the 8th International ISCRAM Conference*. May, pp. 1–10.

Tapia, Andrea, Nicklaus Giacobe, et al. (2015). "Scaling 911 Messaging for Emergency Operation Centers During Large Scale Events." In: *ISCRAM*.

Tapia, Andrea and Kathleen Moore (2014). "Good Enough is Good Enough: Overcoming Disaster Response Organizations' Slow Social Media Data Adoption". In: *Computer Supported Cooperative Work (CSCW)* 23.4-6, pp. 483–512.

Tashakkori, Abbas and John W Creswell (2007). *The new era of mixed methods*.

Tatham, Peter et al. (2013). "Humanitarian logistics: development of an improved disaster classification framework". In: *11th ANZAM (Australia and New Zealand Academy of Management) Operations, Supply Chain, and Services Management Symposium*. Brisbane, Australia, pp. 20–21.

Thapen, Nicholas, Donal Simmie, and Chris Hankin (2015). "The Early Bird Catches The Term: Combining Twitter and News Data For Event Detection and Situational Awareness". In: *Association for the Advancement of Artificial Intelligence*, pp. 1–11. arXiv: 1504.02335.

The Institute of Medicine (2004). *Review of the Centers for Disease Control and Prevention's Smallpox Vaccination Program Implementation: Letter Report #6*. Tech. rep. Washington, DC, p. 50.

Tong, Yongxin et al. (2020). "Spatial crowdsourcing: a survey". In: *The VLDB Journal* 29.1, pp. 217–250.

Tonkin, Emma, Heather D Pfeiffer, and Gregory Tourte (2012). "Twitter, information sharing and the London riots?" In: *Bulletin of the American Society for Information Science and Technology* 38.2, pp. 49–57.

Truelove, Marie, Maria Vasardani, and Stephan Winter (2015). "Towards credibility of micro-blogs: characterising witness accounts". In: *GeoJournal* 80.3, pp. 339–359.

Tsou, Ming-Hsiang et al. (2017). "Building a real-time geo-targeted event observation (Geo) viewer for disaster management and situation awareness". In: *International Cartographic Conference*. Springer, pp. 85–98.

Tukey, John W et al. (1977). *Exploratory data analysis*. Vol. 2. Reading, MA.

UNFCCC (2012). *Slow Onset Events*.

United States National Commission for the Protection of Human Subjects of Biomedical & Behavioral Research (1978). *The Belmont report: ethical principles and guidelines for the protection of human subjects of research*. Department of Health, Education, Welfare, National Commission for the Protection of Human Subjects of Biomedical, and Behavioral Research.

Vajda, Peter et al. (2011). "Social game epitome versus automatic visual analysis". In: *2011 IEEE International Conference on Multimedia and Expo*. Barcelona, pp. 1–6.

Van Wassenhove, Luk N (2006). "Humanitarian Aid Logistics: Supply Chain Management in High Gear". In: *Journal of the Operational Research Society* 57.5, pp. 475–489.

Van Ruijven, Theo (2011). "Serious games as experiments for emergency management research: A review". In: *ISCRAM 2011: Proceedings of the 8th International Conference on Information Systems for Crisis Response and Management, Lisbon, Portugal, 8-11 May 2011*. ISCRAM.

Verma, Sudha et al. (2011). "Natural Language Processing to the Rescue? Extracting 'Situational Awareness' Tweets During Mass Emergency". In: *ICWSM*. Citeseer.

Vieweg, Sarah (2012). "Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications". PhD thesis. University of Colorado at Boulder.

Vieweg, Sarah, Carlos Castillo, and Muhammad Imran (2014). "Integrating social media communications into the rapid assessment of sudden onset disasters". In: *International Conference on Social Informatics*. Springer, pp. 444–461.

Vieweg, Sarah, Amanda Hughes, et al. (2010). "Microblogging during two natural hazards events: what twitter may contribute to situational awareness". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, pp. 1079–1088.

Vieweg, Sarah, Leysia Palen, et al. (2008). "Collective Intelligence in Disaster : Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shooting". In: *Iscram* May, pp. 44–54.

Vitak, Jessica, Nicholas Proferes, et al. (2017). "Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards". In: *Journal of Empirical Research on Human Research Ethics* 12.5, pp. 372–382.

Vitak, Jessica, Katie Shilton, and Zahra Ashktorab (2016). "Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community". In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. CSCW '16. New York, NY, USA: Association for Computing Machinery, pp. 941–953.

Vogelsang, Andreas and Markus Borg (2019). "Requirements engineering for machine learning: Perspectives from data scientists". In: *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. IEEE, pp. 245–251.

Wakamiya, Shoko, Yukiko Kawai, Eiji Aramaki, et al. (2018). "Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study". In: *JMIR public health and surveillance* 4.3, e8627.

Wakamiya, Shoko, Ryong Lee, and Kazutoshi Sumiya (2011). "Crowd-Based Urban Characterization: Extracting Crowd Behavioral Patterns in Urban Areas from Twitter". In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. LBSN '11. New York, NY, USA: Association for Computing Machinery, pp. 77–84.

Wallace, Anthony F. C. (1956). *Human behavior in extreme situations: a survey of the literature and suggestions for further research*. National Academy of Sciences.

Wang, Xufei et al. (2012). "Identifying information spreaders in twitter follower networks". In: *School of Comput., Infor., and Decision Sys. Eng.*

Waugh, William L (2007). "Terrorism as Disaster". In: *Handbook of disaster research*. Springer, pp. 388–404.

Webb, Helena et al. (2017). "The ethical challenges of publishing Twitter data for research dissemination". In: *WebSci 2017 - Proceedings of the 2017 ACM Web Science Conference*, pp. 339–348.

Weimann, G (1994). *Influentials, The: People Who Influence People*. SUNY series, Human Communication Processes. Albany: State University of New York Press.

Weitzel, Leila, Paulo Quaresma, and José Palazzo M. de Oliveira (2012). "Measuring node importance on Twitter microblogging". In: *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12*, p. 1.

White, Gilbert, Ian Burton, and Robert Kates (1978). *The Environment as Hazard*.

Williams, Matthew L, Pete Burnap, and Luke Sloan (2017). "Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation". In: *Sociology* 51.6, pp. 1149–1168.

Wilson, John R and Sarah Sharples (2015). *Evaluation of human work*. CRC press.

Wood, Lisa and Monika Büscher (2012). "On Missed Beginnings". In: *On Work, Interaction and Technology: A festschrift for Christian Heath*. June, pp. 1–8.

Wu, Liang et al. (2019). "Misinformation in social media: definition, manipulation, and detection". In: *ACM SIGKDD Explorations Newsletter* 21.2, pp. 80–90.

Wylie, Christopher (2019). *Mindf*ck, Cambridge Analytica and The Plot to Break America*. Random House, p. 288.

Xiao, Yu, Qunying Huang, and Kai Wu (2015). "Understanding social media data for disaster management". In: *Natural Hazards* 79.3.

Xu, W. W. et al. (2014). "Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election". In: *American Behavioral Scientist*.

Yang, Lili, Raj Prasanna, and Malcolm King (2015). "GDIA: Eliciting information requirements in emergency first response". In: *Requirements Engineering* 20.4, pp. 345–362.

Yin, Jie et al. (2012). "Using Social Media to Enhance Emergency Situation Awareness". In: *IEEE Intelligent Systems* 27.6, pp. 52–59.

Zade, Himanshu et al. (2018). "From situational awareness to actionability: Towards improving the utility of social media data for crisis response". In: *Proceedings of the ACM on human-computer interaction* 2.CSCW, pp. 1–18.

Zahra, Kiran, Muhammad Imran, and Frank O. Ostermann (2020). "Automatic identification of eyewitness messages on twitter during disasters". In: *Information Processing and Management* 57.1, p. 102107.

Zahra, Kiran, Muhammad Imran, Frank O Ostermann, et al. (2018). "Understanding eyewitness reports on Twitter during disasters". In:

Zhang, Shanshan and Slobodan Vucetic (2016). "Semi-supervised discovery of informative tweets during the emerging disasters". In: *arXiv preprint arXiv:1610.03750*.

Zimmer, Michael (2010). *Is it ethical to harvest public Twitter accounts without consent?*

Zimmer, Michael and Nicholas John Proferes (2014). "A topology of Twitter research: disciplines, methods, and ethics". In: *Aslib Journal of Information Management* 66.3. Ed. by Dr Axel Bruns Weller and Dr Katrin, pp. 250–261.

Zubiaga, Arkaitz, Ahmet Aker, et al. (2018). "Detection and resolution of rumours in social media: A survey". In: *ACM Computing Surveys (CSUR)* 51.2, p. 32.

Zubiaga, Arkaitz, Geraldine Wong Sak Hoi, et al. (2015). "Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads". In: *Arxiv - Social & Information Networks*, pp. 1–33. arXiv: `1511.07487`.