Understanding the Structure and Dynamics of Multi-platform Information Operations

Tom Wilson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Kate Starbird (Chair)

David Ribes

Emma S. Spiro

Program Authorized to Offer Degree:

Human Centered Design and Engineering

University of Washington

Abstract

Understanding the Structure and Dynamics of Multi-platform Information Operations

Tom Wilson

Chair of the Supervisory Committee:

Associate Professor Kate Starbird

Department of Human Centered Design & Engineering

Information operations—efforts to distort the information ecosystem through methods such as the dissemination of disinformation in efforts to influence opinions or actions of individuals, governments or publics—are long-established methods of intelligence agencies and the military. However, the advent of social media has led to an era of renewed opportunity: the features and affordances of social media platforms, such as the interconnected social networks, abundance of user data, ability for any user to produce and disseminate content, and algorithm-driven feeds means that they are easier to deploy, and to deploy at scale.

Disinformation (and information operations as the process of *disinforming*) disrupt decision-making, erode trust in institutions such as the media and government, and cumulatively undermine democratic

processes. Recent conceptualizations of information operations as they occur on social media, such as the Russian Internet Research Agency's efforts to interfere in the 2016 US Presidential Election, focus on the explicit coordination of inauthentic accounts. However, in this dissertation I will demonstrate that information operations are more complex—the distinction between the explicitly coordinated and organic aspects are blurred and the activities far more nuanced. Furthermore, despite increasing awareness and recent research there remain gaps in our understanding, including how strategic information operations and disinformation campaigns leverage the wider information ecosystem, taking shape on and across multiple websites and social media platforms.

From a perspective of human-computer interaction (HCI), and in particular computer-supported cooperative work (CSCW), this dissertation aims to contribute to this body of research by presenting four distinct but related studies that collectively build a comprehensive understanding of the structure and dynamics of multi-platform online information operations, including the integration of government-controlled and alternative media, the roles of different social media platforms (i.e. how they are used in complementary ways), and the online collaborative 'work' that brings these aspects together and sustains the information operation over time. Across the studies I adopt a mixed methods approach, combining quantitative and qualitative research techniques to ask questions of digital trace data (evidence of online user activities and interactions) collected from Twitter. I systematically examine both the online conversation on Twitter plus the wider ecosystem of news-publishing websites and social media platforms linked-to in tweets.

The broad contribution of this work is the revelation and exploration of online information operations, specifically how information operations take shape on and across multiple websites and social media platforms, and the collaborative 'work' required to sustain them. Theoretically, based upon the empirical findings presented, I reconceptualize information operations as not solely the product of inauthentic coordinated behavior (orchestrated bots and trolls) but as also involving authentic (as in real, sincerely

participating) actors that loosely collaborate in the production, synthesis, and mobilization of disruptive narratives that serve the strategic goals of the information operation. Methodologically, I contribute to a growing body of work in the area of disinformation and online activism that utilizes a mixed-method approach to research. And, based on my examination of a multi-platform information operation, I end with implications for social media platform policy.

TABLE OF CONTENTS

TABLE OF CONTENTS 6

LIST OF FIGURES 13

LIST OF TABLES 15

CHAPTER 1: INTRODUCTION

BACKGROUND 18

AIM OF THIS DISSERTATION 23

RESEARCH QUESTIONS 23

STRUCTURE OF DISSERTATION 25

INTEGRATION OF PREVIOUSLY PUBLISHED WORK 28

CHAPTER 2: RELATED WORK

SOCIAL MEDIA 29

SHIFTING FORMS OF NEWS PRODUCTION AND CONSUMPTION 31

ONLINE ACTIVISM 33

MISLEADING INFORMATION 34

MISINFORMATION 34

DISINFORMATION 35

INFORMATION OPERATIONS 36

DISRUPTIVE NARRATIVES 39

SUMMARY OF RELATED WORK 40

CHAPTER 3: RESEARCH CONTEXT

THE SYRIAN CIVIL WAR 41

SYRIA CIVIL DEFENCE (THE WHITE HELMETS) 42

KEY EVENTS 43

CHAPTER 4: DATA & METHODOLOGY

TWITTER DATASET OVERVIEW **45**
DATA COLLECTION 46
METHODOLOGY **47**
CRISIS INFORMATICS 48
DIGITAL TRACE ETHNOGRAPHY 50
SUMMARY OF METHODOLOGY 51
LIMITATIONS OF THIS INVESTIGATION **52**
DATA LIMITATIONS 53
METHODOLOGICAL LIMITATIONS 54
ETHICAL CONSIDERATIONS OF THIS RESEARCH **55**

CHAPTER 5: INFORMATION OPERATIONS AS COLLABORATIVE WORK

AUTHOR PREFACE **56**
ABSTRACT **58**
INTRODUCTION **58**
RELATED WORK **60**
INFORMATION OPERATIONS AND THEIR USE ON SOCIAL MEDIA 60
SHAPING NARRATIVES AS A FORM OF INFORMATION OPERATIONS 61
INFORMATION OPERATIONS AS COLLABORATIVE WORK 62
BACKGROUND **63**
SYRIAN CIVIL WAR 63
EVENT BACKGROUND 64
METHODOLOGICAL APPROACH **65**
DATA 65
TWITTER ACCOUNT CODING 66
STATEMENT ABOUT ACCOUNT ANONYMIZATION 66
PAPER ORGANIZATION 67
STRUCTURE OF THE INFORMATION SPACE (MACRO LEVEL) **67**
RETWEET NETWORK GRAPH 67
Blue Cluster: An Active, Cohesive Group of Pro-Syrian Government Journalists and Activists *67*
Orange, Yellow and Green Clusters: The "Mainstream" Media *70*

Red Cluster: A Small Cluster of Anti-Syrian Government Activists *70*
Other Clusters *70*
MOST RETWEETED USERS 71
CONTESTED FACTS, CONFLICTING NARRATIVES (MESO LEVEL) **72**
TEMPORAL ANALYSES 72
CONTENT ANALYSIS 74
PHASE A: Controversy Surrounding Release of Zain Advertisement. *74*
PHASE B: Pre-interview Speculation followed by Response to the Interview *75*
PHASE C: Release of Post-Interview Articles in the Mainstream and Alternative Media *79*
PHASE D: Post-interview — Anti-media sentiment *80*
TECHNIQUES OF INFORMATION OPERATIONS (MICRO LEVEL) **82**
INTRODUCING UNCERTAINTY INTO THE INFORMATION SPACE 82
DISCREDITING INFORMATION PROVIDERS 84
Calling Out Fake News and Propaganda *84*
Targeting Attacks *85*
Suggesting Unproven Affiliations *86*
DISCUSSION **87**
INFORMATION OPERATIONS AS COLLABORATIVE WORK IN AN ONLINE CROWD 88
COLLABORATIVE OPERATING: COLLECTIVE EFFORTS IN INFORMATION OPERATIONS 89
Assembling "Facts" into Undermining Narratives *89*
Challenging the Credibility of Information Sources *90*
LIMITATIONS AND FUTURE WORK **91**
CONCLUSION **91**
CHAPTER 6: THE WHITE HELMETS CONVERSATION ON TWITTER
INTRODUCTION **92**
DATA **93**
METHODS **94**
OVERVIEW OF THE WH CONVERSATION **94**
RETWEET NETWORK VISUALIZATION **95**
TEMPORAL SIGNATURE BY COMMUNITY **98**

CONTENT AND NARRATIVE ANALYSIS BY COMMUNITY 100
THE BLUE (PRO-WH) COMMUNITY: SUPPORTIVE OF THE WHITE HELMETS AND THEIR ACTIVITIES 101
Tweets from the @SyriaCivilDef (the White Helmets) 101
Random tweets from the Blue community 103
Most-retweeted tweets from the Blue community 107
Summary of narratives from the Blue (pro-White Helmets) community 110
THE RED (ANTI-WH) COMMUNITY: CRITICAL OF THE WHITE HELMETS AND THEIR ACTIVITIES 110
Tweets from influential Red-community journalist @VanessaBeeley 110
Random tweets from the Red community 114
Most-retweeted tweets from the Red community 118
Summary of narratives from the Red community 121
SUSPENDED ACCOUNTS BY COMMUNITY 122
***K*-CORE ANALYSIS 123**
LOOKING BEYOND TWITTER TO THE WIDER INFORMATION ECOSYSTEM 124
SUMMARY 126
CHAPTER 7: THE ROLE OF GOVERNMENT-CONTROLLED & ALTERNATIVE MEDIA
AUTHOR PREFACE 127
ABSTRACT 128
INTRODUCTION 129
BACKGROUND 130
THE WHITE HELMETS AND THE SYRIAN CIVIL WAR 130
ONLINE PROPAGANDA & DISINFORMATION IN 2017 131
INFORMATION WARFARE AND ONLINE DISINFORMATION 131
INFORMATION OPERATIONS AND NON-GOVERNMENTAL ORGANIZATIONS (NGOS) 132
CONDUCTING RESEARCH ON INFORMATION OPERATIONS 132
METHODS 133
DATA COLLECTION AND PROCESSING 133
DETECTING ARTICLE SIMILARITY 133
CONSTRUCTION OF "PATHS" 134
INTERPRETATIVE, MIXED-METHOD ANALYSIS 134

NOTE ON DATA AND PRIVACY 136

FINDINGS 136

SALIENT NARRATIVES WITHIN ARTICLES ABOUT THE WHITE HELMETS 136

Pro-White Helmets Narratives 136

Anti-White Helmet Narratives 137

CONTENT-SHARING DOMAIN NETWORK GRAPH 138

CLUSTER A: AN ASSOCIATED PRESS NEWS CLUSTER 139

CLUSTER B: THE ALTERNATIVE MEDIA ECOSYSTEM 140

Central Hubs in the Content Re-sharing Ecosystem 142

Government-Funded News Outlets 143

Central Source Domains 144

Alternative News Aggregators 145

CLUSTER C: A PERIPHERAL HUB: REFRAMING MAINSTREAM CONTENT FOR THE ALTERNATIVE ECOSYSTEM 146

CONTENT REMIXING PRACTICES AND ECHO EFFECTS 146

DISCUSSION: ALTERNATIVE MEDIA ECHO-SYSTEM 147

EXPLICIT CRITIQUE OF MAINSTREAM MEDIA 148

SUPPORT OF RUSSIAN GOVERNMENT 148

SHARED CONTENT ACROSS IDEOLOGICALLY DIVERSE SITES 148

FUTURE WORK TO UNDERSTAND THE DRIVERS AND IMPACTS OF THE INFORMATION ECHO-SYSTEM 149

LIMITATIONS 150

ACKNOWLEDGMENTS 150

SUMMARY AND TAKEAWAYS 150

REFERENCES TO PROBLEMATIC OR MISLEADING CONTENT 151

CHAPTER 8: MULTI-PLATFORM INFORMATION OPERATIONS

ABSTRACT 152

INTRODUCTION 153

BACKGROUND 154

THE SYRIAN CIVIL WAR 154

SYRIA CIVIL DEFENSE — THE WHITE HELMETS 155

METHODS 156

Twitter Data Collection 157
Characterizing the communities in the WH discourse on Twitter 158
The pro-WH (blue) and anti-WH (red) communities on Twitter 159
Looking beyond Twitter to understand cross-platform activities 160
Examining Twitter content integration through retweet cascades 160
FINDINGS 162
The WH conversation on Twitter 162
Top-5 most linked-to social media platforms 164
YouTube 166
Steemit 171
Facebook 173
Medium 175
LiveLeak 177
Multi-platform concurrent introduction of an evidence collage 178
DISCUSSION 181
Sustaining narratives and working to set the agenda 182
Resilience-building using alternative infrastructure 183
CHAPTER 9: DISCUSSION
EXTENDING THE CONCEPTUALIZATION OF INFORMATION OPERATIONS BEYOND THAT OF COORDINATED AND INAUTHENTIC BEHAVIOR 188
The blurred distinction between active measures and activism 190
Positing online information operations as a form of collaborative work 191
EXAMINING INFORMATION OPERATIONS AS TAKING SHAPE THROUGH A HYBRID MEDIA SYSTEM 192
Hybrid news production 192
Timely interventions through fluid opportunity structures 193
A pitfall of hybridity: Losing control of the message 194
Extending the conceptualization of the hybrid media system to consider the mainstream-alternative dimension 195
Alternative of the mainstream, but not independent 195
Alternative tech as a form of stability 196
Alternative but still relying on the mainstream 197

Implications for social media platform policy 198

CHAPTER 10: CONCLUSION & CONTRIBUTIONS

SUMMARY OF STUDIES IN THIS DISSERTATION 201

KNOWLEDGE CONTRIBUTIONS 204

Reconceptualized the view of online information operations 205

Established information operations as a form of collaborative 'work' 205

Increased awareness of disinformation and information operations to the HCI and CSCW communities 205

Considered the mainstream-alternative dimension in a hybrid media system 206

Presented a mixed method approach to studies of online information operations and disinformation 206

REFERENCES 207

LIST OF FIGURES

Figure 1.1: Illustrating the (Twitter-centric) information ecosystem as a color-coded network to denote the studies in the dissertation. Solid lines denote information flows as examined in this dissertation. Dotted lines denote links outside the scope of this dissertation—although links exist, I do not have data to fully understand the information flow. The websites listed in the "echo-system" are only a sample for illustrative purposes. 25

Figure 5.1: Network graph of the accounts involved in the Omran conversation. 68

Figure 5.2: Temporal graph showing the volume of tweets and retweets during the Omran conversation (2017-05-28 - 2017-06-16). 73

Figure 5.3: Temporal graph showing various phases in the Omran conversation. 73

Figure 6.1: Temporal plot showing the number of (re)tweets per 24 hour period in the WH conversation on Twitter. Note the data outage in February 2018. 95

Figure 6.2: A retweet network visualization of the White Helmets conversation on Twitter illustrating the two distinct communities involved in the conversation: the anti-WH community (red, on the left), and the pro-WH community (blue, on the right). The nodes (circles) represent Twitter accounts which are sized relatively according to the number of times that account was retweeted. The edges (lines) between nodes represent a retweet between that pair of accounts. 96

Figure 6.3: Temporal plot showing the number of (re)tweets per 24 hour period from the red and blue communities on Twitter. The plot is annotated with some key events regarding the WH and the Syrian conflict. Note the marked increase in activity—particularly from the red anti-WH community—after March 2018, and the data outage (grey area) in February 2018. 98

Figure 6.4: The modularity ratio (measured on a scale of 0-1) was used to color the domain nodes in the domain retweet network visualization. 124

Figure 6.5: A domain retweet network visualization that looks beyond Twitter, showing the surrounding ecosystem of websites and social media platforms that are linked to in the Twitter conversation. 125

Figure 7.1: WH Content Sharing Domain Network Graph. 138

Figure 7.2: Close-up of Cluster A. 139

Figure 7.3: Close-up, expanded view of Clusters B and C; Colored by degree, Yellow=lower; Red=higher. 141

Figure 8.1: A retweet network visualization of the White Helmets conversation on Twitter illustrating the two distinct communities involved in the conversation: the anti-WH community (red, on the left), and the pro-WH community (blue, on the right). The nodes (circles) represent Twitter accounts which are sized relatively according to the number of times that account was retweeted. The edges (lines) between nodes represent a retweet between that pair of accounts. This is a reproduction of the network visualization published in Wilson & Starbird (2020). 159

Figure 8.2: A retweet cascade plot used for examining content integration to Twitter. In this example, the retweet cascades show content introductions from Medium (different distinct URLs) by accounts in the anti-WH red community on the left, and the pro-WH blue community to the right. The pro-WH content is introduced as a tweet (darker blue circle at y=0, and then retweeted by some accounts with larger followings. It receives more than 30 retweets over the course of two hours. Various anti-WH content is introduced, several of which are not retweeted, including the reply just to the left of the pro-WH cascade, which is not retweeted. 161

Figure 8.3: Temporal plot showing the number of (re)tweets per 24 hour period from the red and blue communities on Twitter. The plot is annotated with some key events regarding the WH and the Syrian conflict. Note the marked increase in activity—particularly from the red anti-WH community—after March 2018, and the data outage (gray area) in February 2018. 163

Figure 8.4: A temporal plot showing the number of (re)tweets, per day containing URLs to Facebook, LiveLeak, Medium, Steemit, and YouTube between April 1 2018 and May 15 2018 inclusive. 168

Figure 8.5: Retweet cascades of YouTube videos in the top-25 most retweeted, April 1 to May 15 2018. 171

Figure 8.6: A temporal plot showing the number of (re)tweets, per day, for each YouTube, LiveLeak, and Steemit versions of Tapestry of Terror. This illustrates when the video was recirculated on Twitter. 179

LIST OF TABLES

Table 4.1: Dataset overview of each of the studies in this dissertation 47

Table 5.1: Top-10 most retweeted accounts in the Omran dataset. 71

Table 5.2: Exemplar tweets from Phase A 75

Table 5.3: Exemplar tweets from Phase B (Around B1) 76

Table 5.4: Exemplar tweets from Phase B (Around B2-B4) 77

Table 5.5: Exemplar tweets from Phase C 80

Table 5.6: Exemplar tweets from Phase D 81

Table 6.1: A table summarizing the White Helmets Twitter dataset 94

Table 6.2: Overview of the blue (pro-WH) and red (anti-WH) communities on Twitter 97

Table 6.3: Descriptive Statistics of Twitter account suspensions by community 122

Table 6.4: k-Core decomposition of the Blue and Red communities 123

Table 7.1: Top 10 Most Tweeted Domains in WH Dataset 142

Table 8.1: Overview of the blue (pro-WH) and red (anti-WH) communities on Twitter 162

Table 8.2: The percentage (and count) of reply tweets that were in-network, cross-network, and out-network by each of the communities on Twitter 164

Table 8.3: The number of URLs linking to content on each platform (re)tweeted by the red (anti-WH) and blue (pro-WH) communities on Twitter 165

Table 8.4: Overview of the 25 most (re)tweeted social media posts from each of the platforms, including the narrative stance toward the WH, number of unique content producers, and details of the introductions to Twitter 165

Table 8.5: ToT appeared on 3 social media platforms in this study: YouTube, LiveLeak, and Steemit. This table breaks down the tweets and retweets of each version 178

This page is left intentionally blank

Chapter 1:

Introduction

Background

In recent years there has been increased interest in information operations — actions by state and non-state actors to distort the information ecosystem to achieve strategic goals, through methods such as the dissemination of disinformation and amplification of specific accounts or content (Weedon, Nuland, & Stamos, 2017). Efforts to manipulate public opinion or distort information spaces are not new (e.g. Bittman, 1972; Bittman, 1985; Rid, 2020), but the features and affordances of social media, including the interconnectedness of online social networks and abundance of user data, mean that information operations are now easier to conduct rapidly and at scale (Lucas & Nimmo, 2015).

In addition to achieving their strategic objectives in the short-term, information operations, and in particular the method of *disinforming*—the acts of producing and disseminating disinformation (to mean false or deliberately misleading information)—function to undermine the integrity of the information ecosystem and reduce human agency by overwhelming our capacity to make sense of information: It becomes harder to separate out fact from fiction, and friend from foe (Rid, 2020). Cumulatively and over time ("*slowly, subtly, like ice melting*" (Rid, 2020 p.11)) this causes lasting damage by disrupting the decision-making capabilities of a society (Bittman, 1985) and undermining democratic institutions (Rid, 2020), such as the media, government, and transnational bodies (e.g., U.N., NATO).

Although *information operations* and other key terms will be discussed in more detail in Chapter 2 (Related Work), as a central concept to this dissertation it also warrants additional context here. The point of departure in this dissertation is Facebook's 2017 definition of information operations, which the company published after realizing that their infrastructure had been exploited, and users manipulated, by the orchestrated activities of the Russian Internet Research Agency (RU-IRA). The RU-IRA operated a "troll farm" that had used the platform as a means of attempting to influence the 2016 US Presidential election and to sow discord in the populace. In a report titled *Information Operations on Facebook* (Weedon, Nuland, & Stamos, 2017), the social media company defined information operations as:

> "*Actions taken by governments or organized non-state actors to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, disinformation, or networks of fake accounts (false amplifiers) aimed at manipulating public opinion.*" (Weedon, Nuland, & Stamos, 2017 p.4)

From this definition we learn that information operations are strategic, in that they are working toward specific aims or interests; manipulative, in that they 'poison the well' by introducing unreliable information or generating uncertainty around previously established understanding; and are conducted through collaborations between multiple actors who may or may not be supported by a specific state. Such activities are not a product of social media: Looking back to the early 19th century Lenin imagined how Western freedom of speech could be exploited for the purposes of subversion and the spread of propaganda (Rees, 1984); and during The Cold War the role of the Soviet intelligence services shifted from passive information gathering to the active dissemination of subversive messages in what became known as *active measures* (Bittman, 1972); "*covert or deceptive operations* [with the] *goal to influence opinions or actions of individuals, governments or publics* [through methods such as] *disinformation and forgeries* [and] *political influence operations"* (US Department of State, 1989). Active measures did not cease with the fall of the Soviet Union. After a respite, the advent of social media led to an era of renewed opportunity: the features and affordances of social media platforms, such as the interconnected social networks, abundance of user data, and the ability for any user to produce and disseminate content, meant that they were easier to deploy, and to deploy at scale (Lucas & Nimmo, 2015; Waltzman, 2017). By seeking to extend influence and advance geopolitical goals by changing how people perceive the world (Bittman, 1972; Bittman, 1985), active measures are similar to Facebook's definition of information operations.

In this dissertation, I adopt the term *information operations* to describe this phenomenon. I acknowledge that this term is not perfect: in the US it was primarily used only by the military and intelligence services

as a label for "soft warfare" until it was invoked by Facebook; and there is overlap with related concepts such as *active measures* and disinformation campaigns. But it does help us to get at the nuances of the activities—the work—of influencing large numbers of people. Furthermore, although intelligence agencies and governments have been interested in these issues for decades, academic research in the area of disinformation (particularly as it takes shape on and across social media platforms) is still in relative infancy. This means that there remains some fluidity as terms, boundaries, and definitions are discussed and refined going forward, and this dissertation may contribute to those debates.

Indeed, as will become apparent from the empirical evidence presented herein, I extend upon the definitions provided above, arguing that by focusing only the highly coordinated activities overlooks the collaborative 'work' of online crowds, including communities of authentic and real individuals, which often play a significant role in sustaining an information operation (Starbird, Arif, & Wilson, 2019). The RU-IRA operation was relatively unsophisticated: it comprised an 'army' of bots and inauthentic trolls that conducted its activities from a single location in St Petersburg (United States District Court for the District of Columbia, 2018). As a consequence, the social media companies such as Facebook and Twitter were able to respond, labelling these actions as *coordinated inauthentic behavior* and developing policies and methods to prohibit and enforce against them on their platforms (Twitter, 2020; Facebook; 2020). However, as I will demonstrate in the empirical studies presented in this dissertation (Chapters 5-8), information operations are more complex—the distinction between the explicitly coordinated and organic aspects are blurred and the activities far more nuanced.

That is not to suggest that elements of orchestration or coordination do not exist—information operations may well be initiated through explicit coordination among a core group of state actors such as government agencies, politicians, or government-controlled media, etc. But when one considers the scope of attempting to influence large numbers of people, while information operations may be initiated by a core subset of actors that are explicitly coordinated (and that may continue to play a supportive role), the sustainability of information operations in the long-term requires that other actors—believers and individuals witting or unwitting—be co-opted into the fold. Motivations and intentions for participation may vary, but the important point is that in addition to meeting their personal goals, these actors also contribute toward the collective effort of the information operation through decentralized and uncoordinated collaborative work. This highlights the participatory nature of information operations and speaks to the efficacy of disinformation and *disinforming.* The co-opting of others could be thought of as a form of snowball sampling: current participants 'recruit' future participants through their acquaintances, or to put this into the context of platforms social media such as Facebook and Twitter, through their online social networks.

This view of information operations as integrated into organic online crowds complicates both how we understand these operations and how we might design technologies and policies to address them. However, the dynamics of these collaborations—e.g., how they are seeded, take shape, and sustain themselves—are not well understood. This dissertation aims to contribute to this understanding.

This dissertation is situated within the field of Human-Computer Interaction (HCI), and adopts a perspective of the sub-genre of Computer-Supported Cooperative Work (CSCW), an interdisciplinary field which is broadly concerned with "*how different types of social groups affect, and are affected by, information and communication technology*" (Karahalios, Fitzpatrick, & Monroy-Hernández, 2017). With this perspective, throughout the dissertation I will express an interest in the collaborative work of the online crowd as I examine information operations. I also draw upon work from the interdisciplinary field of crisis informatics, which has developed methods and techniques to use data collected from information communication technologies (e.g., social media) to understand human behavior during crisis events (Palen & Anderson, 2016). I provide a brief overview of my methodology and the influence of crisis informatics below, but revisit in more detail in Chapter 4.

Throughout my research I adopt a mixed methods approach, combining quantitative and qualitative research techniques, including methods developed within the field of crisis informatics (Palen & Anderson, 2016), the investigation of online rumors in the context of crisis events (e.g., Maddock et al. 2015), and digital trace ethnography (Geiger & Ribes, 2011). My research primarily relies on *digital traces*—evidence of online activity that is logged and stored digitally (Freelon, 2014). Digital traces are rich resources of information, particularly when leveraged through methods such as digital trace ethnography (Geiger and Ribes, 2011; Friedberg, 2020), which places the traces in the context in which they were generated. Through these methods I ask questions of the digital traces, reconstruct events, and reveal action. For example, I can reveal communities of accounts that frequently retweet one another; analyze activity temporally; trace content across social media platforms and websites; and follow specific actors as they engage with content and each other.

However, there are also things that I cannot see through this approach, which limit my overall view of information operations. For example, unlike more traditional forms of ethnography that may rely on in-situ participant observation and perhaps involvement within a culture, my engagement was with the digital traces that were generated through their online activities and not with the participants themselves. A consequence of this is that I do not emic first-hand accounts of the actors' motivations, intentions, or perspectives of their involvement in an information operation. Furthermore, because I observed (as opposed

to becoming an active member of the community), I do not have the 'insider' perspective: I was not privy to the hierarchy of actors involved or 'backchannel' communications (that may or may not exist), which could suggest explicit coordination among actors not visible through the public digital traces that I examined. This is to say that like all research, my view is limited by the data and research method I used, but throughout this dissertation I use the empirical evidence I do have—gathered through observation and analysis—to build a comprehensive and nuanced understanding of online information operations, including the actors (activists, journalists, politicians), artifacts (articles, blog posts, videos, social media posts, tweets), and technologies (websites, social media platforms) that leverage and are leveraged as part of an information operation.

The research presented within this dissertation is situated in the context of the Syrian Civil War and the online battle for public opinion. I focus on two case studies: the reemergence of Omran Daqneesh on Syrian state television almost one year after his family's home was destroyed in an airstrike conducted by the Syrian government or their allies (Study 1, Chapter 5), and; the local volunteer group of first responders known as Syria Civil Defence, more often referred to in the media as the 'White Helmets' due to their distinctive white headgear (Studies 2-4, Chapters 6-8). In both cases I am interested in how information operations take shape on (and in later studies, across) social media platforms and websites, and the collaborative activities (the 'work') that take place to meet the strategic goals of the operation and sustain it over time. Additional background context about the Syrian Civil War and the White Helmets is provided in Chapter 3.

This dissertation makes several knowledge contributions to HCI, CSCW, and researchers involved in studies of disinformation and online manipulation. Empirically it demonstrates how the various aspects of the information ecosystem are leveraged in information operations, and the role of heterogeneous assemblages of actors in doing so. Theoretically, and building upon the empirical observations, it calls for a reconceptualization of information operations beyond that of coordinated and inauthentic behavior to also acknowledge the role of authentic (as in real people, sincerely participating) that loosely collaborate in the production, synthesis, and mobilization of preferred narratives across websites and social media platforms. This reveals how emergent collaborations form to conduct the "work" of online information operations, which extends prior conceptualizations of online activism and online volunteerism. Practically, the findings have implications for the possible remediations available to social media platforms that are further discussed in Chapter 9. The knowledge contributions, among others, are expanded upon in Chapter 10.

Aim of this Dissertation

The aim of this dissertation is to build a comprehensive understanding of the structure and dynamics of multi-platform online information operations, including the integration of government-controlled and alternative media[1], the roles of different social media platforms (i.e., how they are used in complementary ways), and the online collaborative 'work' required to sustain the operation.

Research Questions

This dissertation is built around three main research questions, each of which focus on a different facet of the information ecosystem[2]: Twitter (the social media platform central to this dissertation); a network of government-controlled and alternative media websites, and; other social media platforms from which content is shared to Twitter. Each of the main research questions (RQ1, RQ2, and RQ3) is broken down into more specific questions which serve to guide the inquiry of each study.

[1] I refer to the "alternative media" as a specific set of websites within this information space that describe themselves as alternative to, and independent of, what they view as the agenda-setting power of the "mainstream" mass media, although in many cases they are consistently closely aligned with—or instances of—government-controlled media. I use the term in the context of this research and do not make claims about the alternative or independent media more broadly. Alternative media will be covered in more detail in "Shifting forms of news production" in Chapter 2.

[2] Nardi and O'Day (1999) describe *information ecology* as "*a system of people, practices, values, and technologies in a particular local environment.*"(p. 50). The people, practices, tools, and values of an information ecology are interrelated and closely bound together in the system. Interrelationship and dependencies within the system mean that if one aspect changes (e.g. new tools, updated practices) then there are effects elsewhere in the system as adaptations take place. McLuhan (2003) used the term "cross-fertilization" (p.58-59) to describe how media within an information ecology work together, for example, the integration of one media with another (e.g. newspapers appearing on television broadcasts; in contemporary times tweets appearing in articles published online). I opt for the term *information ecosystem*, favoring it over *information ecology* because of the more active connotations: the flexibility and the fluidity. The term *ecosystem* articulates the interrelationships between the components in an environment while also highlighting the ongoing interactions and adaptations that are constantly taking place within the system.

RQ1 — How do information operations manifest on Twitter? (Study 1, Study 2, Study 4)

RQ1a: To what extent does the emergent online crowd 'work' together as part of the information operation? (Study 1, Study 2, Study 4)

RQ1b: What 'work' are the online crowd engaged in? (Study 1, Study 2, Study 4)

RQ2 — How do government-controlled and alternative media websites shape and propagate competing narratives? (Study 3)

RQ2a: What narratives are presented by websites linked-to in the Twitter conversation? (Study 3)

RQ2b: What are the content-sharing practices of distinct media domains involved in fomenting, shaping, and propagating competing strategic narratives? (Study 3)

RQ3 — How are multiple social media platforms leveraged in complementary ways to achieve the strategic goals of information operations? (Study 4)

RQ3a: How are social media platforms used in multi-platform information operations? What complementary role do the different platforms play? (Study 4)

RQ3b: How do accounts on Twitter 'work' to integrate content from other social media platforms as part of a multi-platform information operation? (Study 4)

As they are presented within the dissertation, the studies represent an increasing level of complexity, both in terms of the temporal scope and the extent of the inquiry: from 3 weeks to 13 months; and from a single platform (Twitter) (Study 1, Study 2), to a network of websites (Study 3) and social media platforms (Study 4) linked to in the Twitter conversation. This illustrates how I work to build up the broader view of the information ecosystem.

Structure of Dissertation

The research questions are addressed across four distinct but related empirical studies that collectively build a comprehensive understanding of the structure and dynamics of multi-platform online information operations. Figure 1.1 provides a high-level network representation of how the components of this (Twitter-centric) information ecosystem fit together, and how my proposed studies align with the various aspects through use of a colored key.

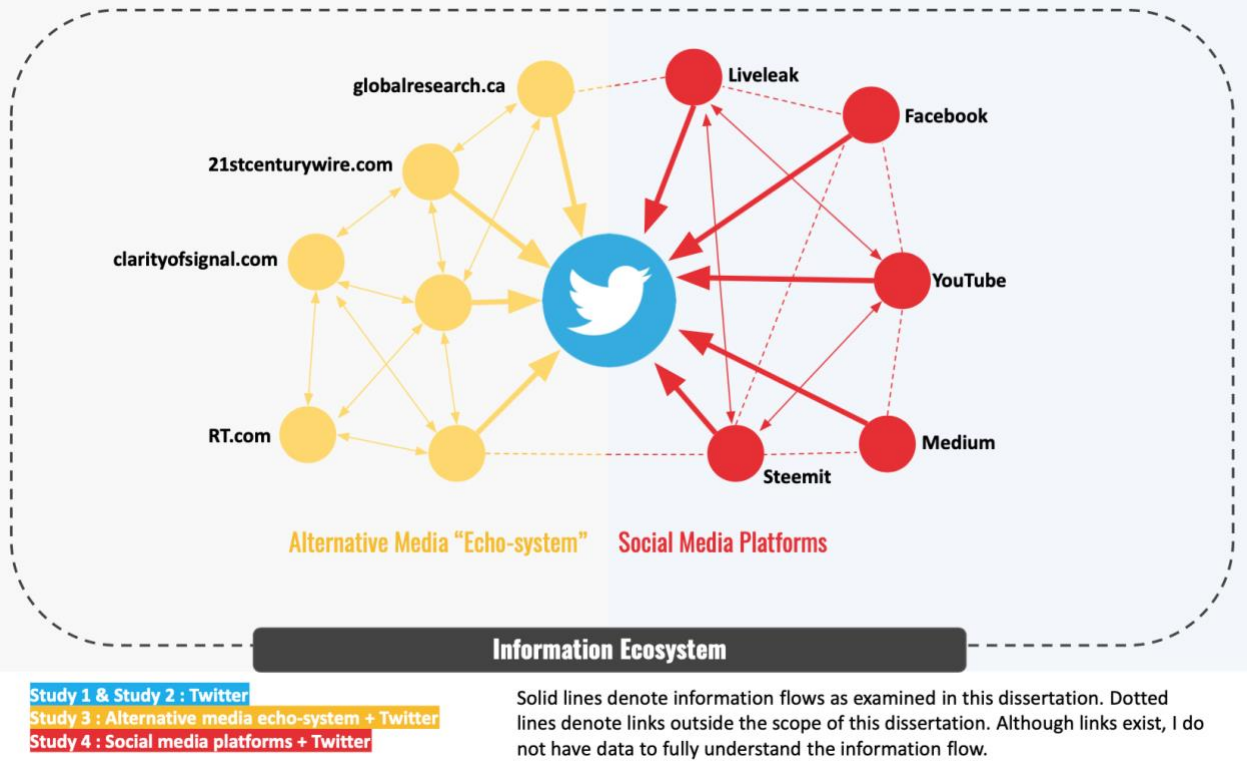

Figure 1.1[3]:Illustrating the (Twitter-centric) information ecosystem as a color-coded network to denote the studies in the dissertation. Solid lines denote information flows as examined in this dissertation. Dotted lines denote links outside the scope of this dissertation—although links exist, I do not have data to fully understand the information flow. The websites listed in the "echo-system" are only a sample for illustrative purposes.

[3] This approach presents a Twitter-centric view of the information ecosystem: My 'seed' data is from Twitter, and this determines the flows that I can see. In other words, the information ecosystem may look different if I used a collection of Facebook or YouTube data as my starting point, or collected data from all platforms in parallel.

The rest of this dissertation is organized as follows: The remaining parts of this (Chapter 1) will provide describe the integration of previously published work in this dissertation and outline the contributions of this work;

- **Chapter 2 (Background and Related Work)** will provide definitions and background to key concepts that are central to this research;

- **Chapter 3 (Research Context)** provides some background context within which this research is situated, specifically the Syrian Civil War and Syria Civil Defence (more commonly referred to as The White Helmets);

- **Chapter 4 (Data and Methodology)** describes the central dataset used within this dissertation, the methodological underpinnings of this work, and some of the limitations of the data and my approach.

The four empirical studies are then presented in separate chapters (i.e., Studies 1-4 are presented in Chapters 5-8). Each chapter includes background, methods, and findings for that study, as well as a discussion that points to the larger implications of the research:

- **Chapter 5 (Study 1—Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community)** addresses RQ1, describing how information operations manifest on Twitter through examination of a case study of a specific episode of the Syrian Civil War: the reemergence of Omran Daqneesh (The "Aleppo Boy") almost a year after his family's home was destroyed in an airstrike. Through a mixed methods analysis of tweets and their associated metadata, this work reveals a network of clustered users that contribute to a contested and politicized information space surrounding Omran's story (RQ1a); the presence of undermining narratives that serve to disrupt the established understanding and confuse the audience; and techniques used when promoting, defending, or undermining narratives (RQ1b);

- **Chapter 6 (Study 2—The White Helmets Conversation on Twitter)** contributes further to RQ1**,** presenting findings from the long-form, detailed investigation of the WH conversation on Twitter including the identification of communities of Twitter accounts that represent the two "sides" to the conversation, a temporal analysis of these communities' activities; and a qualitative look at the content being shared about the White Helmets. Toward the end of the chapter, an analysis of URLs

shared within tweets reveals, from the perspective of Twitter, the extent of the WH discourse on other websites and social media platforms;

- **Chapter 7 (Study 3—The Role of Government-Controlled and Alternative Media in Shaping Disruptive Narratives)** addresses RQ2 by examining articles and websites linked-to in the Twitter conversation surrounding the White Helmets. The research identifies a set of salient narratives about the WH (RQ2a) and reveals a network of alternative media websites that propagate disruptive narratives about the role and function of the group. Analyzing the content-sharing practices among these websites revealed that they appeal to diverse audiences, meaning that the websites within this "echo-system" were able to draw in audiences from diverse communities into a set of common narratives, critical of the White Helmets (RQ2b);

- **Chapter 8 (Study 4—Multi-platform Information Operations: Mobilizing Narratives and Building Resilience through both 'Big' and 'Alt' Tech)** addresses RQ3 by looking beyond the WH conversation on Twitter to other social media platforms that are linked to in tweets and retweets about the White Helmets. The findings reveal a network of social media platforms from which content is produced, stored, and integrated into the Twitter conversation (RQ3a). They also highlight specific activities that sustain the disruptive narratives and attempt to influence the media agenda. Underpinning these efforts is the work of resilience-building: the use of alternative (non-mainstream) platforms to counter perceived threats of 'censorship' by large, established social media platforms (RQ3b);

- **Chapter 9 (Discussion)** will provide a summary of the empirical research conducted, before building on some of the key findings from across the studies;

- **Chapter 10 (Conclusion)** will provide a summary of the dissertation, present the contributions of the work, and offer some final remarks.

Integration of Previously Published Work

This dissertation incorporates research that has previously been published and includes the contributions of a co-author or co-authors. I am the first author of three of the studies: Study 1[4] (Chapter 5) which was published in *Proceedings of the Association of Computing Machinery (PACM) on Human Computer Interaction Computer Supported Cooperative Work*; Study 2[5] (Chapter 6) which is a derivative of a paper and its methodological appendix published in the *Harvard Kennedy School Misinformation Review* 1(1); and Study 4[6] (Chapter 8) that has been submitted as a conference paper with a co-author. I am a co-author of Study 3[7] (Chapter 7) and have been granted permission by the first author, Kate Starbird, to include that work in my dissertation. In Chapter 7 I have included additional research and findings that were outside the scope of the original publication but are relevant to the larger theme of this dissertation. I use a sans-serif font to indicate the text that was not included in the original publication. Chapter 6 is a derivative of the original publication, incorporating findings and insights from both the article and its accompanying methodological appendix.

Chapters 5-8 include an author preface to situate that work within the themes of this dissertation. Furthermore, I include information about how to cite material from that study, including a pointer to the previously published work. For Chapters 5 and 7, the previously published work should be considered the original and the version within this dissertation a derivative.

[4] Wilson, T., Zhou, K. & Starbird, K. (2018). Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), pp.1-26.

[5] Wilson, T. & Starbird, K. (2020). Cross-platform Disinformation Campaigns: Lessons Learned and Next Steps. The Harvard Kennedy School (HKS) Misinformation Review.

[6] Wilson, T. & Starbird, K. (In Review). Multi-platform Information Operations: Mobilizing Narratives and Building Resilience through 'Big' & 'Alt' Tech. Submitted to Proceedings of the ACM on Human-Computer Interaction Computer-Supported Cooperative Work

[7] Starbird, K., Arif, A., Wilson, T., Van Koevering, K., Yefimova, K. and Scarnecchia, D. (2018). Ecosystem or Echo-system? Exploring Content Sharing across Alternative Media Domains. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 12, No. 1)

Chapter 2:

Related Work

My aims for Chapter 2 are twofold: to provide definitions and context for the key terms that appear across the research in this dissertation (e.g., social media *platform*, misinformation, disinformation, information operations, and disruptive narratives), and; to describe key components of what I term the information ecosystem, and how these components (and the activities associated with them) fit together. Related work for the specific studies can be found in the relevant section of chapters 5-8.

Social Media

A central aspect to the research in this dissertation is social media, in particular the microblogging and social networking service Twitter that appears across all studies. In this section I discuss the technological and societal advancements that led the emergence of social media (as platforms).

The foundations of social media were made possible by the emergence of Web 2.0, a set of protocols that would link and index the "*fragmented, countless permutations*" of the internet (DiNucci, 1999 p.32). Web 2.0 provided a flexible framework that did not place technical demands or specifications on specific elements, offering "*a transport mechanism, the ether through which interactivity happens*" (DiNucci, 1999 p.32). In effect, Web 2.0 standards allowed the internet to evolve from a collection of static web pages to an interconnected and interactive environment, enabling user-generated content and facilitating a participatory culture as individuals were now able to express themselves online and participate in the "commons" of cyberspace (Horrigan, 2007). Benkler (2006) argued that the increasing accessibility to the means of producing media meant that the internet was becoming a networked public sphere "*a complex*

ecosystem of communication channels that collectively offer an environment that is conducive for communication...an important venue for discussion and debate over matters of public interest [and] an alternative arena for public discourse and political debate" (Benkler et al., 2015 p.5). In contrast to earlier conceptions of the public sphere[8], which were dominated by large media entities and subject to government control, the networked public sphere was perceived as open to wider participation: less dominated by large media entities and subject to a reduced level of government control—rebalancing power by providing individuals with a venue and the means to promote their viewpoints (Benkler et al., 2015).

The emergence of what we now refer to as *social media* was driven by these technological and social advancements, which changed the conditions of social engagement and had an effect on both institutional structures and individuals' interactions (Van Dijck & Poell, 2013). There was a shift from the common top-down or broadcast model of information diffusion to networks of networks in which individuals were able to communicate in "all directions". Individuals were not limited to being the passive receivers of messages but were now able to take a more active role as reactors, (re)makers, and (re)distributors (Varnelis, 2012).

Social media services such as Facebook and Twitter were built upon social networks—social structures in which people with common interests or relationships are joined. Such services allow individuals to build up their social network of friends, family, and anyone else using the service, and engage in social interaction, including the sharing of user generated content. As increasing numbers of consumers (and producers) began using it to share and disseminate news and information (discussed further in the following section), the role of the social media evolved. The social media companies worked to balance the rights to free speech (particularly in the US) while also avoiding a gatekeeping[9] role (in the traditional broadcast media sense), which could lead to them being labelled as publishers and subject them to a set of regulations similar to that of the traditional broadcast media, possibly limiting their growth. While government regulation such as the Communications Decency Act (CDA) of 1997 offered them some legal protections from what their users may say and do on their infrastructure (Electronic Frontier Foundation, n.d.; 47 US

[8] For example, Habermas' conceptualization as the domain of social life in which public opinion is formed and as a space that mediates the relationship between society and states (Habermas, 1964).

[9] Gatekeeping refers broadly to "*the process of controlling information as it moves through a gate or filter*" (Barzilai-Nahon, 2008 p. 1). The term was originally (in 1943, during World War II) used to describe the role of the housewife or maid in selecting what was served for dinner (Lewin, 1943) but later in the 20th century was applied to the role of the media in selecting or filtering news based upon 'newsworthiness' (Shoemaker et al., 2001). More recently, in the Web 2.0 environment, some authors have evolved the concept further, for example introducing "audience gatekeeping" to describe the greater role played by individuals as they are able to "*pass along already available news items and comment on them*" (Shoemaker & Vos, 2009).

Code § 230), the social networking and media companies also strategically defined their place within the information ecosystem through use of the term 'platform'—drawing parallels with the earlier conceptualizations of the public sphere as giving individuals a space from which to have their voice heard. Referring to YouTube, but in terms that could be applied to others social media services such as Facebook and Twitter, Gillespie (2010) points to the computational, architectural, figurative, and political meanings of "platform": "*The more conceptual use of 'platform' leans in all of the term's connotations: computational, something to build upon and innovate from; political, a place from which to speak and be heard; figurative, in that the opportunity is an abstract promise as much as a practical one; and architectural, in that YouTube is designed as an open-armed, egalitarian facilitation of expression, not an elitist gatekeeper with normative and technical restrictions"* (Gillespie, 2010 p.352).

In summary, social media platforms allow individuals to network with others and to communicate in 'all directions' both within and outside of their social network. At this point social media platforms are prevalent in society, used by 70% of Americans to connect with one another, engage with news content, share information and entertain themselves (Pew Research, 2019). Furthermore—as will be covered in later chapters—social media platforms are not only integrated within many aspects of daily life but also integrated with one another, enabling cross-platform interactions. While acknowledging the multitude of uses for social media, in the following subsections I will focus on usage as it is relevant to this dissertation: their use in news production and dissemination, online activism, and information operations.

Shifting Forms of News Production and Consumption

In parallel to these technological and social advances, the model of news production and consumption that had been dominant through the 20th century was also evolving. The technological progress, coupled with an increasing distrust in the accuracy and objectivity of the mainstream media, led to the emergence in the late 1990s of *citizen journalism* and independent media outlets online. Citizen journalism was characterized by principles such as open publishing and lack of editorial selection process (Meikle, 2002: p.89; Bruns, 2008). These independent media challenged the gatekeeping and agenda-setting roles of the traditional mainstream media as citizen journalists, and individuals more broadly, could be involved in the process of news-making: pursuing the issues they deemed important and creating, publishing, and distributing the resulting content. This branch of media became known as *alternative media* to mean it was situated outside, and independent of, the mainstream media (Kenix, 2011). The term "mainstream media" developed into a pejorative against the agenda-setting power and corporate interests of the traditional, mass media (Kenix,

2011), while the alternative media were viewed as non-commercial and attempting to challenge the dominant power structures within society (Atkinson & Dougherty, 2006; Kinix, 2011).

This "democratization" of news production (Gillmor, 2006) allowed new and different groups of people to have their voices heard, including on social media. Bruns (2008) coined the term *produsers* to describe these networked social media users that were now engaging in productive information use and content creation (Bruns, 2008). Produsage enabled non- and semi-professional "produsers" to produce and curate content for their audiences and perhaps obtain mainstream exposure (Pyrhönen & Bauvois, 2020).

Structural and organizational changes within (older) traditional mainstream media, particularly the press, were also taking place: moving to an online model and incorporating social technologies to disseminate their content led to a reduction in revenue from the sale of print media and advertising within. The widespread adoption of social media in the subsequent decade drastically altered the model of "news"—including who produces it, how it is distributed, who gets paid for it, and how and where people consume it. This progression contributed to what was termed a "crisis in journalism" (Fuller, 2010) as older (traditional, broadcast) forms of mainstream media struggled to adapt their practices and business models to these new and evolving conditions. At the same time, social media was providing individuals with the tools to produce and consume news, and to distribute that within their networks, offering new ways for people to "*observe, filter, disseminate, and analyze news, while at the same time making immense profit and serving as effective channels of distributing news content*" (Witschge et al., 2016).

Jenkins (2006) introduced the term "convergence culture" to refer to the intertwined news production/consumption and social media that was reliant upon the active participation of the public. Under this system Jenkins argued that there were no longer distinct groups of "media producers" and "media consumers", as existed in the more traditional model of broadcast media. Rather, these roles were collapsed into "media participants" because one could be actively involved in both production and consumption at the same time. Due to the volumes of information being produced, Jenkins also viewed consumption as a form of collective intelligence in which "*None of us can know everything; each of us knows something; we can put the pieces together if we pool our resources and combine our skills*" (p.4)— in essence a form of collective sensemaking as individuals come together to form understanding and give meaning. Chadwick (2013; 2017) viewed the underlying premise of convergence culture—that all media would eventually converge on a single platform—as problematic, instead viewing the 'older' (traditional, broadcast) and 'newer' (social) media forms as coexisting within a *hybrid media system* (Chadwick, 2013; 2017). Instead of focusing on the distinctions between these forms, Chadwick's hybrid media system encouraged an

awareness of the blurring of the boundaries—the integration and coevolution of both older and newer media, which, Chadwick argued, are utilized by actors in complementary ways (Chadwick, 2017).

Online Activism

The internet (and later social media) also opened up opportunities for new forms of activism (Vegh, Ayers, & McCaughey, 2003). In *Cyberactivism: online activists in theory and practice* (McCaughey & Ayers, 2003), Vegh, Ayers, & McCaughey, 2003 (2003) define internet activism as "*politically motivated movement relying on the internet*" (p. 71). Online activism can be internet-enhanced or internet-based, depending on the goals and means of the activism (i.e., are the activities only possible online, or are online tools used to coordinate offline action?). Internet activism may involve (among other activities): the development of substitute distribution network as a means of disseminating alternative news and information (e.g., circumnavigating the gatekeeping role of the mainstream media); mobilization (such as putting out calls for action for online/offline activities such as protests); or online attacks ("hacktivism") (Vegh, Ayers, & McCaughey, 2003). Bennett (2003) put forward that the "*Internet and other digital media facilitate the loosely structured networks, the weak identity ties, and the issue and demonstration campaign organizing that define a new global politics*" (p. 143), adding that new media (blogs, websites) offered activists independence from the mass (mainstream) by altering information flows. Freelon, Marwick, and Kreiss (2020) argue that digital media opened up new pathways for activists to reach their target audiences.

Social media platforms further lowered the barriers to online activism, providing the opportunity for widespread, online participation. While some authors have questioned the effectiveness of online participation in the name of activism, coining the term "*slactivism"* to describe "*low-risk, low-cost activity via social media, whose purpose is to raise awareness, produce change, or grant satisfaction to the person engaged in the activity*" (Rotman et al., 2011), others argued online activism complemented offline activism (Freelon, Marwick, & Kreiss, 2020). In addition to producing and disseminating their own content through the new pathways and information flows described above, social media platforms also enabled online convergence and coordination: For example, the online activist networks that coordinated and facilitated the 2010 G20 summit protests in Toronto (Poell, 2013); or the networked social movement's use of communication infrastructure to transition the 2010 Occupy Wall Street protests from distributed encampments into a coordinated movement (Donovan, 2018). Other studies regarding the role of social media in online activism occurred in the wake of the 2011 Arab Spring, for instance documenting how local, affected individuals took to social media to apply political pressure, shape political discourse, and to extend their reach—informing western audiences about events on the ground (Howard et al., 2011). And

Savage and Monroy-Hernández (2015) demonstrated how the online presence of a militia group leveraged social media to facilitate the mobilization of individuals for participation in collective efforts offline. In sum, social media platforms and the shifting forms of news production provided individuals with the means to coalesce, organize, form emergent collaborations, and develop new channels of communications to reach their audiences.

Misleading Information

Misleading information is the umbrella term I am using to describe misinformation (the inadvertent sharing of false information) and disinformation (the deliberate creation and sharing of information known to be false)—terms that I elaborate on below. Note that I do not claim this to be an exhaustive list of the types of misleading information (see Jack, 2017; Wardle & Derakhshan, 2017 for additional terms), but they are the forms that are relevant to the studies in this dissertation.

Misinformation

Misinformation refers to unintentionally incorrect information (Marwick, 2018), misguided or erroneous information (Jowett & O'Donnell, 2015), or unintentionally inaccurate information (Jack, 2017). A consistent property among these definitions is the unintentionally. Misinformation can emerge for various reasons, including at times of uncertainty and ambiguity such as a crisis event when there is a lack of verified information (Maddock et al., 2015), due to cognitive biases (Pennycook et al., 2018), or through acting in haste and publishing misinterpreted or unverified information (Jack, 2017)—a particularly noteworthy example of which is the New York Times' journalists' report of weapons in mass destruction in Iraq. Based upon unverified information that later turned out to be false, this article was cited by the Bush administration as one of the reasons to go to war with Iraq (Posetti & Matthews, 2018).

Misinformation that emerges during crisis events is of particular relevance to this research, if only to help contextualize the distinctions with disinformation (discussed in further detail in the following section). Crisis and mass disruption events can be characterized as times of heightened uncertainty, ambiguity, and a lack of verified information (Comfort, Ko, & Zagorecki, 2004). Drawing upon research in disaster sociology (e.g., Fritz and Mathewson 1957; Dynes 1970; Kendra & Wachtendorf 2003), researchers in the field of crisis informatics (see Chapter 4) have studied how people converge on social media platforms during crisis events (Hughes et al., 2008), to share information (Palen & Liu, 2007), obtain situational awareness (Vieweg et al., 2010), and self-organize (Starbird & Palen, 2011). At times where institutional channels are unavailable and there is an absence of verified information the gaps in information can be

filled by misinformation, or *improvised news*, as members of the public interact, discuss, and form a definition of a situation (Shibutani, 1966). This highly-distributed problem solving activity has been termed *collective sensemaking* (Spiro, Fitzhugh & Sutton, 2012; Maddock et al., 2015; Zeng, Starbird & Spiro, 2016). In this context, misinformation is not produced or propagated with the intent to mislead but is an output of the collective sensemaking process that occurs as individuals collectively attempt to make sense of their current situation.

Disinformation

Disinformation refers to intentionally misleading information. Although the phenomenon of disinformation has been around for centuries, the term disinformation is a translation of *dezinformatsiya*, which can be traced back to Stalin and the Soviet Union. Stalin added the word to the Great Soviet Encyclopedia in 1952, listing it as of French origin and a tool of capitalist oppressors being used against the Soviet Union (Rychiak & Pacepa, 2013). However, earlier in the 20th century Lenin had anticipated that Western freedoms could be exploited for the purposes of subversion and the spread of propaganda (Rees, 1984).

Disinformation describes false information that is intentionally seeded, produced, spread, and amplified to mislead and disrupt decision-making. Bittman (1985) recites the definition found in the Great Soviet Encyclopedia of 1952 as "*the dissemination (in the press, radio, etc.) of false information with the intention to deceive public opinion*" (Bittman, 1985 p.49). Disinformation is more complex and nuanced than simply being incorrect information. Accurate information may be conveyed but shrouded in untruths, for example the provenance of the information or the identity of the publisher (Rid, 2020). In fact, in order to be effective disinformation must be based around a *rational core* of plausible and verifiable information or common understanding—a *kernel of truth*. The *kernel of truth* is then fleshed out with disinformation—half-truths, exaggerations, or misleading information that distort the facts. Disinformation is backed by intention and designed to resonate with the existing beliefs of particular target audiences and, as a consequence, give it greater chance of being accepted as fact—a point that underscores the methodical intentionality (Rid, 2020).

Individuals have long been were targeted to serve as "*unwitting agents*" to disseminate disinformation. For example, journalists were provided anonymous 'tips' (of disinformation) that either aligned with the journalist's existing beliefs or presented the opportunity to publish a 'scoop'. Unaware they were being exploited, the journalist would publish the story, exposing it to a wider audience (Bittman, 1985). In the contemporary age of social media, the lower barriers to the production and distribution of information mean

that it's not just career journalists that can become "*unwitting agents*"—anyone that is exposed to disinformation that aligns with their existing beliefs can propagate to their audience or social network.

Contemporary definitions of disinformation retain the concept of intentionality, for example: intentionally incorrect information (Marwick, 2018); information that aims to sow confusion and proliferate falsehoods (Pomerantsev & Weiss, 2014); news stories deliberately designed to weaken adversaries (Jowett & O'Donnel 2015); and false information that is knowingly shared to cause harm (Wardle & Derakhshan 2017). Across these definitions, disinformation is an artifact and separate from the process of disseminating it— the process of *disinforming*— can be achieved through *information operations*, which will be described in the following section.

To briefly revisit the description of collective sensemaking provided in the previous section on misinformation, it becomes clearer how disinformation, backed up by the intent to deceive, can disrupt this sensemaking process. A difficulty faced when disambiguating between misinformation and disinformation is discerning the intent behind a piece of information and obtaining proof of the intent to deceive. Such insights may not be clear through social media trace data, and while speaking to an individual about their actions (e.g., through an interview) may contribute to understanding, such methods are not always possible or appropriate. An alternative approach, and one used here, is to consider the strategic intent of a campaign rather than the motivation, authenticity, or sincerity of a specific account (Starbird, Arif, & Wilson, 2019).

Information Operations

Information operations refer to long-standing methods that have been assigned various labels such as 'information warfare' (Lucas & Nimmo, 2015), 'weaponization of information' (Waltzman, 2017), 'active measures' (Bittman, 1985), or disinformation campaigns (Starbird, Arif, & Wilson, 2019). Broadly, I view information operations as the process of *disinformation*—the production and/or dissemination of disinformation—but I provide more detail about the term, and its roots, in this section.

Within the US, the term information operations can be traced to the US military, originally posited as a doctrinal approach to operationalize the power of information. This 'information capability' was seen as crucial to effective foreign policy in an era characterized by open communication in which individuals could control their information flows (Armistead, 2004). Military planners anticipated that this increased connectivity and new forms of communication would lead to new categories of warfare (Hollis, 2007). Early definitions of information operations comprised an integrated strategy of electronic warfare;

computer network operations; psychological operations; military deception; and operational security (The Joint Chiefs, 2006). Other countries, notably Russia, also began to realize the asymmetrical possibilities of using the contemporary (networked) information space, leveraging technologies and information networks to reduce the fighting potential of the enemy. A prominent Russian General called for research and development to support these activities (Gerasimov, 2013).

Despite being in existence for almost 20 years, the term information operations was seldom used outside of US government documents and intelligence reports until 2017 when Facebook invoked the term in a security bulletin titled "Information Operations and Facebook" (Weedon, Nuland, & Stamos, 2017). In this whitepaper, which outline the company's expanding security focus, Facebook's security specialists laid out the central challenge to the security of the platform as information operations, defining them as:

> "*...actions taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, disinformation, or networks of fake accounts aimed at manipulating public opinion…*" (Weedon, Nuland, & Stamos, 2017 p.4)

From this definition, we learn that Facebook's information operations are strategic, aim to manipulate public opinion, and are conducted by organized groups (on their platform). The adoption (and adaptation) of the term information operations by Facebook is somewhat aligned with "information-related capabilities" (in the US military definition) but also bears resemblance to Soviet *active measures*—"*covert or deceptive operations* [with the] *goal to influence opinions or actions of individuals, governments or publics* [through methods such as] *disinformation and forgeries* [and] *political influence operations"* (US Department of State, 1989).

To learn about Soviet *active measures* we turn to Ladislav Bittman, a high-ranking intelligence officer for the Czechoslovak Intelligence Service before his defection to the United States in 1969. Bittman explains how active measures marked a shift in the role of the intelligence services from *passive* intelligence gatherers to *active* disseminators of subversive messages (Bittman, 1972). This active role was supported by a dedicated apparatus, including a department that focused on the production and dissemination of disinformation (Bittman, 1985), which could support the "*new era in which distortion concisely and purposefully taints the natural flow of information throughout the world*" (Bittman, 1976; p.20). Active measures can also be thought of as antithetical public relations—seeking to undermine, delegitimize, and

denigrate a public image rather than working to improve it (Bittman, 1985). The aim behind such efforts were not necessarily to persuade the audience and bring them around to a particular way of thinking, but rather to 'muddy the waters'—to confuse, distract, and distort information spaces around particular topics (Pomerantsev & Weiss, 2014). Such measures ultimately paralyze the decision-making capabilities of the public as increasing uncertainty leaves individuals unsure of what to believe and how to react (Pomerantsev & Weiss, 2014; Rid, 2020). It should also be noted that these activities were not limited to the Soviet Union: the US intelligence agencies also engaged in what they termed "political warfare" (Rid, 2020).

Facebook's interest in information operations stemmed from its discovery that its platform had been exploited: 470 inauthentic accounts had spent $100,000 on advertising between June 2015 and May 2017 (Stamos, 2017). Similarly, in January 2018, Twitter notified 1.4 million of its users that they may have engaged with suspected "trolls" from the Russian government-funded Internet Research Agency (IRA) during the 2016 US Presidential Election (Twitter Public Policy, 2018). It became clear that the interconnected structure of online social networks and the abundance of user data collected by social media platforms were setting up conditions for contemporary active measures, or what Facebook termed information operations. As revealed in the subsequent investigations and congressional hearings, the IRA's activities on social media platforms such as Facebook and Twitter circa 2016-2017 were highly orchestrated, comprising an 'army' of bots and inauthentic trolls that conducted its activities from a single location in St Petersburg (United States District Court for the District of Columbia, 2018). As such, the social media platforms' definition of information operations was built around "organized actors" and inauthenticity[10], and labelled as *coordinated inauthentic behavior*. Subsequent policies and methods were designed to prohibit and enforce against such activities (Twitter, 2020; Facebook; 2020).

The term information operations and its definition is not without issues: in the US the roots are militaristic, describing a range of activities under information warfare. Jack (2017) sees Facebook's use of the term information operations as positioning the platform as part of an information war (perhaps suggesting it is taking the problem seriously) but argues that all information presented to users on Facebook is curated by black box systems of advertising and engagement metrics—which are as opaque as the unnamed "organized actors" working to distort the information space; there is overlap with related concepts such as active measures and disinformation campaigns; and as will be demonstrated in the empirical research herein, information operations are more complex than defined by Facebook—the distinction between the explicitly coordinated and organic aspects are blurred and the activities far more nuanced. However, use of the term

[10] For more on this see: Arif, A., Stewart, L. G., & Starbird, K. (2018). Acting the part: Examining information operations within #BlackLivesMatter discourse. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-27.

does help us to get at the nuances of the activities—the work—of influencing large numbers of people, looking beyond a single instance of disinformation to a wider campaign of activities. I also view this as a contribution to the field of HCI and what I will term *disinformation studies*: while we still work to fully understand activities and how they relate to one another, there is opportunity for discussion and iteration as terms, boundaries, and definitions are refined.

Disruptive Narratives[11]

Narratives are a communicative tool that allow humans to structure information, giving meaning to actions that facilitate sensemaking. They comprise a system of stories that are interrelated and sequentially organized. Stories represent event units that include information about the *who, where, when, and how* and have a beginning and an end (i.e., are resolved) (Halverson, 2011; Halverson, Corman, & Goodall, 2011). Narratives, on the other hand, are evolving and remain open to interpretation and orient the audience to potential future outcomes (Miskimmon, O'loughlin, & Roselle, 2014).

The open-endedness of narratives mean that they can be used strategically. For example, in political communication, narratives are employed by politicians to shape discussions within international relations, to manage expectations and shape the future behavior of domestic and international actors (Miskimmon, O'Loughlin, & Roselle, 2014). The strategic creating, shaping, and dissemination of narratives can help to meet (geo)political aims—activities that align with the aim of information operations. Indeed, one aspect of information operations can involve the intentional creation, propagation, and shaping of narratives (Chivvis, 2017): *counter-narratives* symmetrically confront an existing narrative and offer an alternative, while *anti-narratives* negativity and cynically spoil the existing narrative without offering an alternative explanation (Rogers & Tyushka, 2017) or by offering multiple, and perhaps conflicting versions of events. For example, after the Malaysian Airlines Flight MH17 was shot down over the Ukraine in 2014, multiple, conflicting, narratives about what happened were propagated by RT and other Russian media (BBC Rudin, 2016). Indeed, the number of conflicting narratives about the event generated so much "white noise" that foreign policy experts argued that the strategy being employed was to "*...convince viewers that nothing was true at all...putting out alternative explanations to muddy the waters and undermine confidence in what Western media and governments say…*" (Luhn, 2015).

[11] This is an expanded version of a section of "Related Work" in: Wilson, T., Zhou, K., & Starbird, K. (2018). Assembling strategic narratives: Information operations as collaborative work within an online community. *Proceedings of the ACM on Human-Computer Interaction*, *2* (*Computer Supported Cooperative Work*), pp.1-26.

In this dissertation, I use the term *disruptive narratives* to describe the presence of multiple, often conflicting anti-narratives that undermine the current understanding of events but do not offer an alternative explanation. Disruptive narratives introduced to the information ecosystem may be based upon disinformation, be of unknown provenance, or be based on information taken out of context, but nevertheless their presence introduces ambiguity and raises levels of uncertainty as individuals attempt to process and derive meaning from competing and perhaps conflicting information. This approach considers information (or disinformation) as evidence, fragments of stories, or story units that form into narratives that can then be strategically shaped or deployed. This moves beyond the truth value of specific information and focuses on how these pieces fit together to serve a particular purpose as part of a campaign (Starbird, Arif, & Wilson, 2019).

Summary of Related Work

In this chapter I have aimed to conduct definitional work by tracing the lineage of various key terms used throughout this dissertation. I have also presented aspects that are key to the information ecosystem within which my studies are based. In summary, I charted the emergence of social media and the shifting forms of news production, illustrating how through technological and societal advances, individuals were provided the ability to produce and disseminate information. I also outlined the use and role of social media in online activism and sensemaking. Finally, I provided background context to two distinct types of misleading information: misinformation and disinformation and described the roots of the term information operations: from a form of "soft power" in US military parlance, to a term that was adopted by Facebook to describe activities more akin to Soviet-era active measures or a disinformation campaign. In the subsequent chapters these terms will feature as I present my research into how information operations take shape within the information ecosystem.

Chapter 3:

Research Context

Each of the studies in this dissertation are situated in the context of the Syrian Civil War: Study 1 focuses on a specific case study, that of the reemergence of Omran Daqneesh with his father on Syrian state television almost one year after his family's home was bombed in an airstrike allegedly carried out by the Syrian government or their allies; Studies 2-4 examine the online discourse surrounding Syria Civil Defense (The White Helmets). This chapter provides additional background context.

The Syrian Civil War

The context of this research is the Syrian Civil War, an ongoing conflict that began in 2012 after Arab Spring protests against the rule of Syrian President Assad turned violent. The conflict descended into a multi-sided, internationalized conflict that involves global powers such as Russia and Iran, which are allied with the Syrian government; the US, UK and other global and regional allies, which support the anti-Syrian government 'rebels'; and ISIS, a now-weakened group that was attempting to establish a caliphate across the region (Gilsinan, 2015; Zaram, 2013). The ongoing conflict has taken many lives and displaced millions more: accurate figures for the number of people killed are disputed, but the Syrian Observatory for Human Rights (2020) put the figure at almost 600,000. The United Nations Refugee Agency in 2018 reported that 5.6 million Syrians were refugees in other countries while a further 6.6 million were internally displaced (United Nations High Commissioner for Refugees, 2018).

Alongside the kinetic war 'on the ground', the Syrian Civil War also manifests itself online as a battle of public opinion. Syria and its allies, including Russia and Iran, take one "side", while the other "side"

includes the US, UK, and other allied nations. The research contained herein examines the online manifestation of the Syrian Civil War, and how this takes shape on and across social media.

Syria Civil Defence (The White Helmets)

Amongst this complicated geopolitical terrain operate Syria Civil Defence, the official name of the group that has become known as the White Helmets (WH) due to their distinctive headgear. The volunteer group provides search and rescue, medical evacuation, and essential services to civilians impacted by the ongoing civil war there. According to the organization's website[12], they abide by the fundamental principles of the Red Cross and work in accordance with Article 61 of the Additional Protocol I, which defines the activities that constitute civil defense: protecting civilians from hostilities or disasters, aiding recovery in the aftermaths of such events, and providing conditions for the survival of the civilian population (Pictet, 1979). WH activities include search and rescue, evacuating buildings, firefighting, medical aid, providing emergency shelter and supplies. In other words, the White Helmets are primarily a humanitarian response organization—working to protect civilians from violence; provide rescue, medical aid, and recovery after violence events; supporting conditions for the survival of the civilian population.

The WH first emerged as loose collaborations between local volunteers fulfilling the role of first responders, rescuing civilians and retrieving bodies in the aftermath of military attacks and airstrikes conducted by the Syrian government or their allies, Russia, in "rebel"-held areas that opposed the Syrian government. After some of the volunteers reached out for financial help from foreign humanitarian groups, James Le Mesurier, a former British army officer, established Mayday Rescue as a means to channel aid to the group. This funding was sourced from the government of the US, UK, Canada, Denmark, Germany, and the Netherlands (Chulov, 2020).

The group actively promoted their work to garner attention for the plight of the Syrian people caught up in the war: they captured rescues using helmet cameras (Solon, 2017) and were the focus of a Netflix documentary that won an Oscar in 2017 (Larkin & Lewis, 2017). Their video footage documented atrocities committed by Syria and their allies, Russia, that the WH observed when responding to attacks. Such evidence directly challenged Russia's claims that it respected international law and was upholding global stability through its role in the Syrian conflict (Levinger, 2018). In response, and in a bid to discredit the group and undermine their reporting of alleged atrocities, the WH became the targets of a sustained online

[12] http://syriacivildefense.org

disinformation campaign designed to influence public perceptions of the group (Solon, 2017) and in some cases justify physical attacks against them (Grey-Ellis, 2017).

For the most part, this information operation was effective in inundating the information space with negative content about the WH and nurturing a community that was willing and able to further advance online activities against the group. The success of the online operation against the White Helmets was noted by prominent Russian General Gerasimov in early 2019[13] (Kramer, 2019), and as the group's reputation became more tarnished (e.g., by associating them with known terrorist groups), it became easier to justify physical attacks against them (Grey-Ellis, 2017). As of October 2018, it was reported the once 3000 strong group had lost more than 250 members to the violence (Di Giovanni, 2018). In November 2019, James Le Mesurier (founder of Mayday Rescue, the organization providing financial support to the WH), was found dead in Turkey with injuries consistent with falling from a height (Yee, 2019; Loveluck, 2019). The following investigation found that his death was caused by suicide and this has been attributed, at least in part, to the persistent efforts to discredit him through a disinformation campaign (Chulov, 2020).

Key Events

As will be described in more detail in Chapter 4, the research within this dissertation is centered around a Twitter dataset of English-language tweets about the Syrian civil war collected between May 2017 and June 2018. During this period, there were several key events of the Syrian conflict that coincided with an increase in the online conversation, including:

- June 2017: The reemergence of Omran Daqneesh "The Aleppo Boy" with his father in a broadcast interview with a pro-Syrian TV journalist (see Chapter 5 of this dissertation);
- June 2017: The White Helmets send a message of solidarity to London firefighters after the Grenfell Tower fire;
- August 2017: The murder of seven members of the White Helmets in their office in the city of Sarmin in Idlib province;

[13] General Gerasimov also realized the asymmetrical possibilities of leveraging technologies and information networks to reduce the fighting potential of the enemy earlier in the decade in what was termed the Gerasimov Doctrine (Gerasimov, 2013; Kramer, 2019).

- September 2017: The White Helmets are awarded the Tipperary International Peace Award[14] for humanitarian work;
- December 2017: Olivia Solon's article in the Guardian (UK newspaper) outlining the disinformation campaign against the White Helmets, titled "How Syria's White Helmets became victims of an online propaganda machine";
- February 2018: military offensive by the Syrian Army to capture for rebel-held Eastern Ghouta (Rif Dimashq offensive);
- April 7 2018: Heavy airstrikes of the city of Douma that included the use of chemical munitions;
- April 14 2018: Missile strikes on multiple Syrian government targets by the US, UK, and France;
- May 4 2018: Announcement by the US government that they would cease funding the White Helmets.

The online conversation surrounding these events will be revisited in the later empirical chapters (5-8).

[14] The Tipperary Peace Award was created by residents of the Irish town to counter the association between their town and war, an association that was created by the song "*It's a long way to Tipperary*". The annual award is given for humanitarian work.

Chapter 4:

Data & Methodological Approach

In this chapter I first provide an overview of the dataset that is central to each of the studies presented in this dissertation, including its parameters, the method of collection, and its limitations. I then outline the methodological approach I used across this research. Each study (presented in Chapters 5-8) contains a detailed description of the specific methods used for that particular research; therefore, the aim of this chapter is to provide additional details about those elements that are common across studies: the iterative mixed-methods approach combining techniques developed in the field of crisis informatics, the method of digital trace ethnography.

Twitter dataset overview[15]

This research within this dissertation is centered around a Twitter dataset of English-language tweets about the Syrian civil war collected between May 2017 and June 2018. Each of the studies presented in this dissertation use a subset of this data, which vary in temporal scope and keyword selection. Table 4.1 provides an overview of the parameters of the data used in each individual study (for a more detailed description please refer to the data section of the specific study). The studies presented in this dissertation appear in the temporal order that they were conducted, therefore the dataset used increases in temporal scope and size (i.e., number of tweets) as data collection continued. In the following section I describe how the Twitter data was collected and processed before discussing some limitations of the approach.

[15] Adapted from: Tom Wilson & Kate Starbird. (2020). Methodological Appendix to Cross-platform disinformation campaigns: lessons learned and next steps. *Harvard Kennedy School Misinformation Review*, 1(1)

Data Collection

The Twitter data used in this dissertation consists of public tweets posted between May 27, 2017 and June 15, 2018 (or temporally bounded subsets of this data: see Table 4.1 for an overview), which was collected using infrastructure developed at the University of Washington[16].

Using this infrastructure, we accessed the Twitter Streaming API[17] to track (in real-time) various keyword terms related to the Syrian conflict including names of affected areas (`Idlib`, `Ghouta`, `Douma`), terms related to specific events and features of the conflict (`chemical attack`, `barrel bomb`, `sarin`), entities in the conflict (`Assad`, `Syrian Army`, `FSA`, `ISIS`), and the White Helmets (`white helmets`, `whitehelmets`, `Syria Civil Defence`, `SyriaCivilDefence`). By the end of the collection period in June 2018 this dataset contained more than 100 million tweets and their associated metadata[18].

The subsets of data used for each study were defined by specific keyword selection and temporal filtering as outlined in Table 4.1. A final step before analysis was to limit the dataset to the tweets that were designated as English (the language set to 'en' in the tweet metadata). In other words, the focus was on the discourse that was generated by and/or for English-speaking audiences and therefore the subsequent analyses are limited to English-language tweets. This is an important constraint and limitation of this work, discussed further in the section discussion the limitations below.

[16] Full credit for the development of this infrastructure goes to fellow PhD candidate in the Department of Human Centered Design and Engineering John Robinson, my PhD advisor Associate Professor Kate Starbird, and PhD committee member Associate Professor Emma S. Spiro of the University of Washington's Information School.

[17] The API is an Application Programming Interface, which is an interface that allows multiple software applications to communicate to each other. The Twitter Streaming API allows developers to filter the real-time stream of public tweets, for example 'listening' for specific terms within the Twitter conversation and collecting those tweets. More information is available via Twitter's web pages for developers: https://developer.twitter.com/en/docs/twitter-api/tweets/filtered-stream/introduction

[18] More details of the metadata associated with a tweet can be found on the Twitter API pages: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview

Ch.	Study	Research conducted	Temporal scope of data used in research	Keywords[19] (case insensitive)	Number of EN-language (re)tweets[20]
5	(1) Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community	Jun 2017—Mar 2018	May 28, 2017—Jun 16 2017	"omran"	23,906
6	(2) The White Helmets Conversation on Twitter	Jan 2019—Dec 2021	May 27 2017—Jun 15 2018	"white helmets", "whitehelmets"	1,023,547
7	(3) The Role of Government-Controlled & Alternative Media in Shaping Disruptive Narratives	Sep 2017—Jan 2018	May 27 2017—Sep 9 2017	"white helmets", "whitehelmets"	135,827
8	(4) Multi-platform Information Operations	Jan 2019—Dec 2021	May 27 2017—Jun 15 2018	"white helmets", "whitehelmets"	1,023,547

Table 4.1: Dataset overview of each of the studies in this dissertation

In studies 2-4 (Chapters 6-8), this Twitter data was 'expanded', meaning that I treated the Twitter data as 'seed' data as I looked beyond Twitter by following the URLs that appeared in tweets and retweets to collect additional data from a variety of websites and social media platforms. This process, and the data collected, is described in more detail within the Methods section of each of those studies.

Methodology

Throughout this research I use a grounded, interpretive, mixed-method approach. I draw upon techniques developed within the field of crisis informatics (Palen et al., 2009; Palen & Anderson, 2016), including the investigation of online rumors in the context of crisis events (e.g., Maddock et al. 2015) and digital trace ethnography (e.g., Geiger & Ribes, 2011). My approach is iterative and blends both quantitative and qualitative analyses—for example, by generating descriptive statistics and visualizations of the data to see from a higher level the patterns and anomalies, and then using those representations as entry points for a closer examination of the data.

[19] The keyword selected would need to appear in the 'text' of the tweet or retweet, or in the case of 'quote' tweets (retweets with comment), within the 'text' of the embedded quote tweet

[20] I use (re)tweets as shorthand for tweets and/or retweets

Central to my approach are *digital traces*—evidence of online activity that is logged and stored digitally (Freelon, 2014). By employing methods such as digital trace ethnography (Geiger & Ribes, 2011), I am able to place the digital traces within the context in which they were originally generated and ask questions of the data. To contextualize this by way of an example, across the studies I use temporal graphs—visualizing the volume of tweets over time—and I am then guided by features of that visualization, such as peaks in the volume, as events that warrant further qualitative examination to understand what was driving that increase. In other cases, I build an understanding of the structure of the conversation by identifying pairs of accounts that retweet one another, and then through quantitative methods (clustering) I detect communities of accounts before qualitatively examining accounts of interest (for example perhaps those that are most often retweeted). This work was not done in isolation, rather it involved fellow researchers, and we would collectively conduct close up qualitative analyses of the data. Answers to the questions we posed of the data would guide the next iteration of analyses and, over time, the empirical observations would be constructed into grounded (small 't')[21] theories, guided by the constructivist grounded theory methodology (Charmaz, 2014). The grounded theory methodology is a means of constructing theory of the phenomenon of interest that is *grounded* in the data (versus selecting an existing theory or conceptual framework *a priori*) (Olson & Kellogg, 2014). In the following subsections I provide additional details of the core components of this methodology: crisis informatics and digital trace ethnography.

Crisis Informatics

Crisis informatics is a multidisciplinary field combining computer science and social science (Palen & Anderson, 2016) and drawing upon prior research in disaster sociology (e.g., Fritz and Mathewson 1957; Dynes 1970; Kendra & Wachtendorf 2003). As an evolving field, multiple definitions of crisis informatics have been put forward by researchers (see Palen et al., 2020 for a recent, comprehensive list), but one that encapsulates the main focus and the methodologies borne from it is that of Soden and Palen (2018), who define crisis informatics as "*Concerned with the ways in which information systems are entangled with socio-behavioral phenomena connected to disasters...offer[ing] a rich set of research methods and empirical opportunities for examining the consequences of the role of technology in mediating our relations with the world.*" (Soden & Palen 2018). This definition positions crisis informatics as able to draw from and contribute to the field of human computer interaction (HCI), and its sub-genre computer-supported cooperative work (CSCW) (Soden & Palen, 2018).

[21] I credit my advisor Kate Starbird with the terms "little 't' theory" which I borrow here to describe the collective explanatory power of empirical observations about a phenomenon

In essence, crisis informatics is concerned with using data collected from information communication technologies (e.g., social media) to understand human behavior during crisis events. There are multiple strands to this research, for example: examination of the socio-behavioral phenomenon of self-organization during crisis events (e.g. Hughes & Palen, 2009); self-organization during crises (e.g. Hughes et al., 2008; Vieweg et al., 2008); the role of digital technology as a part of the formal emergency management and operations (e.g. Reuter, Marx, & Pipek, 2011), and; issues of veracity, misinformation, and disinformation that emerge during crisis events (e.g. Arif et al., 2017; Maddock et al., 2015). This last area is broadly where my work fits in, and my academic advisor Kate Starbird has been heavily involved in forwarding this research agenda. Drawing on historical research on rumoring (e.g., Allport & Postman, 1947; Shibutani, 1966), the initial focus was on how rumor and misinformation can be introduced (and spread) within the online conversations in the wake of crisis events, which can be characterized as times of heightened uncertainty and ambiguity (Comfort, Ko, & Zagorecki, 2004). As individuals engage in collective sensemaking—the highly-distributed problem-solving activity that occurs to fill knowledge gaps and make sense of an emerging situation (Vieweg et al., 2008; Palen et al., 2009)—rumors and misinformation can emerge (Maddock et al., 2015). More recently the focus has shifted to the structure and dynamics of online disinformation (e.g., Starbird 2017; Starbird, Arif, & Wilson, 2019), particularly how the uncertainty and ambiguity of crisis events can be exploited to pollute the information space with intentionally incorrect or misleading information.

Particularly relevant here is the methodology that has been developed to examine the online participation that takes place on social media in response to crisis events, notably the techniques that facilitate rapid data collection (as the event happens); the use of both quantitative and qualitative research methods so as to go beyond reductionist measures and high-level observations; and the iterative nature of both the research and the data collection (Palen & Anderson, 2016). Indeed, approached in this way, Palen and Anderson (2016) call for approaching "*'big data' qualitatively and even ethnographically*" (p.225) and posit that social media data can be "*observed with a different kind of precision, including temporal sequencing of events and discourse*" (p.225).

To put this into context, a foundational paper from crisis informatics which articulated a methodological approach to identifying and understanding the spread of misinformation online is the Maddock et al. (2015) study of the Boston Marathon Bombing. Through analysis of digital trace data from Twitter, the authors integrated quantitative and qualitative methods to build multi-dimensional signatures of rumors that were circulating on social media during the Boston Marathon bombings. Quantitative measures such as temporal

patterns and lexical diversity were combined with qualitative coding of rumor behavior types. In addition to enhancing understanding of how different kinds of rumors propagate online during crisis events, this approach of recursively integrating qualitative and quantitative methods to analyze social media trace data is an approach that I've applied across my own work—including within this dissertation.

Digital Trace Ethnography

While digital traces can take various forms, depending on their source and available data, they broadly comprise metadata that is generated at the time of online action. To put this into context I will illustrate with a simplified example of posting a tweet to Twitter: The digital trace data for a tweet includes (but is by no means limited to) a unique id of the specific tweet, publicly available information about the author of that tweet (e.g., their Twitter username and bio), and the timestamp of when the tweet was posted. Digital traces (or digital footprints, as they are sometimes known) are 'side effect'[22] of interacting online (Breiter & Hepp, 2018) that are primarily generated to support platform functions. While on the surface digital trace data can appear 'thin' and lacking as a source of social data (Geiger & Halfaker, 2017), through the appropriate methods they can also be used by researchers to describe or understand online behaviors.

The main source of digital trace data in this dissertation is the Twitter dataset of tweets and their associated metadata containing keywords terms related to the Syrian conflict (a more specific breakdown of the keyword terms in each study can be found in Table 4.1). To derive meaning from the digital traces I am guided by trace ethnography (Geiger & Ribes, 2011) as a method to capture documentary practices of technology-mediated and distributed phenomenon that, due to the scale, are otherwise challenging to study.

By adopting digital trace ethnography, the researcher does not view digital traces as a 'means to an end' but is also interested in how the members (of these cultures) themselves use the traces to make meaning in the sociotechnical context and coordinate their activities. For example, in the context of social media use during crisis events, users may filter information based on geotags to find information coming from local sources (Huang et al., 2015). Digital trace ethnography brings together the *emic* (insider view that "attempts to capture participants' indigenous meaning of real-world events" (Yin, 2015 p.12)) and *etic* (the outside view that "represent[s] the same set of real-world events, but from an external perspective" (Yin, 2015 p.12)) viewpoints, leveraging the digital traces to understand how members act within a broader organizational setting. In practice the emic/etic distinction is not a dichotomy—the perspectives are better

[22] That the users of platforms such as Twitter may not necessarily realize the presence of these digital footprints raises ethical considerations of using digital trace data, which are discussed later in this chapter.

thought of as lying on a continuum that can complement one another (Yin, 2015). To achieve this, digital trace ethnography calls for a mixed method approach, drawing on both quantitative and qualitative techniques. By analyzing digital traces with a blend of emic and etic perspectives and quantitative and qualitative techniques, researchers can develop 'thick descriptions' (Geertz 2008) that are placed within the context in which the traces were made (Geiger & Halfaker 2017).

To describe this by way of an example, in an analysis of interactions between bots that edit articles on Wikipedia, Tsvetkova et al. (2017) adopted quantitative methods and operationalized revert counts as a measure of conflict on the platform. Although from an etic (outsider looking in) perspective this is an entirely reasonable approach, it takes the traces of these interactions out of their sociotechnical context and interprets higher revert counts as problematic. In response to the research, Geiger and Halfaker (2017) place these alleged cases of conflict in sociotechnical context to better understand the interactions that took place. Guided by trace ethnography, Geiger and Halfaker (2017) use mixed-methods to interrogate, refine, and extend the initial dataset, using the traces as a starting point for further contextualization and interpretation by following digital traces held in transactional log data. By doing so, the researchers are able to place the reverts in the context in which they were made and reveal that the bot-on-bot reverts are generally not a cause for concern but are part of the regular inner-workings of the platform. By adding the additional contextual detail (understanding why these reverts were made), the authors built a more nuanced understanding.

Summary of Methodology

In this section I provide a short walkthrough of the general methodology, providing examples for context. In Study 1 (Chapter 5), I articulate my methodological approach as taking place at various levels (macro, meso, micro), which I think are useful for illustrative purposes here, although they should be viewed as fuzzily bounded and not an exhaustive list of activities that can take place. For simplicity, and consistency with the data used across these studies, I start with an example of Twitter data.

The macro level describes the higher level view and I am interested in the structure of the data. For example, I may use the digital trace data to build a network visualization based upon some relationship between accounts (e.g., whether they retweet one another). From this I can start to see accounts that frequently interact (e.g., through retweets), and how these may form clusters or communities of accounts. Through more quantitative interrogation of the digital traces, I can begin to understand who are the most active accounts, or what is the most retweeted content. In another example, I may use the timestamps associated

with tweets to build a temporal plot of tweets over time (e.g., per hour), and then use this as an entry point for further analysis.

At the meso level the interest is the narratives. Continuing the above example, a 'spike' in activity (volume of tweets) may warrant further attention. This involves a qualitative analysis of tweets. Patterns and anomalies identified at the macro level can help with sampling; for example, an in-depth analysis of tweets posted during the "spike" in activity may be a starting point. A more detailed examination may reveal that content is being introduced from other websites (based upon URLs in tweets), in which case following these URLs to their destination websites would be pertinent. From this point there may be a necessity to collect more data (e.g., about the website, the content it hosts).

And the micro level is more concerned with building an understanding of interactions between accounts. While the structural properties such as retweets can give some insight, other interactions may take place through the body text of tweets. It is also possible that these interactions become clearer over time, requiring a more detailed temporal analysis (e.g., looking at an accounts' activities over time, or the interactions between multiple accounts over time).

The research methodology I adopt is highly iterative. To use the levels described above as a framework, analysis involves frequently moving between these levels as outputs at one level generate more questions at another. There is also a need for data to get bigger before it gets smaller: as I follow the digital traces through Twitter and onto other websites and social media platforms, additional data collection, and analyses, are required. While my methods sections are written in a way that (I hope) is easiest for readers to understand, the order as presented does not necessarily reflect a linear progression across the analyses.

Limitations of this Investigation

Olteanu et al. (2019) provide a comprehensive discussion of the issues of using social media trace data in research, including general biases; issues at the data source; issues introduced while collecting data; issues introduced while processing data; methodological pitfalls while analyzing the data; and ethical considerations. Rather than revisiting that comprehensive discussion here, I will focus on some points that are particularly relevant to this research, one of which is the ethical considerations of using social media trace data that I will come back to at the end of this chapter. The non-exhaustive list of limitations of this investigation are organized into two categories: those related to the data and those related to my methodological approach

Data Limitations

A broader limitation of the use of social media trace data that I use in this dissertation is that it can offer only a narrow view of the wider discourse surrounding the White Helmets. This discourse is also taking place elsewhere, for example on other websites (some of which I attempt to capture in Study 3), and other social media platforms (Study 4), but also a multitude of other on- and off- line locations that are not included in this dissertation. In other words, the research presented here represents only a subset of the WH discourse. In a similar vein, Tufekci (2014) reminds us that online interactions are just one aspect of human behavior. There is only so much that we can infer from online interactions, which can have various meanings, some of which may be ambiguous or unintelligible from social media trace data. Other noteworthy limitations include:

- **We don't know what we don't get:** Acquisition of social media trace data is governed by the platforms themselves and often comes with limitations. In particular there is often a lack of transparency regarding the sampling strategies used to provide the data, meaning that of the data returned by the API, we don't know how it was sampled from the larger pool of data, nor what was *not* sampled (Olteanu et al., 2019).

- **But we know we don't get everything:** Platforms do not give access to all the data that they possess (Olteanu et al., 2019). Tufekci refers to this as the "missing denominator"—while we may know who clicked (or shared/retweeted) but don't know who saw but decided not to take action (i.e., not to share/retweet). Furthermore, we are not provided insight into how content/users have been promoted/demoted/organized and how that could shape the online discourse. On social media platforms algorithms that rank and filter content influence user behavior, determining what is shown and to whom (Olteanu et al., 2019), which has been termed "algorithmic confounding" (Chaney, Stewart, and Engelhardt, 2018x). These algorithms are proprietary, and the digital trace data used in this dissertation does not provide insight into how content was organized, or if it was promoted or demoted, which may have affected how many people were exposed to the content.

- **Language limitation:** In each study the Twitter dataset was scoped to include only the tweets that were designated as English (the language set to 'en' in the tweet metadata). This limitation is in part due to language constraints of the team involved in the research; however, it also encapsulates a specific and strategic element, particularly of the White Helmets online discourse. Although not their first language, the White Helmets made significant and sustained efforts to communicate in English to western and global audiences. For example, they run a Twitter account dedicated to

English-language content and they have taken photographs of themselves holding signs with English-language statements. In 2016, they were featured in an English-language documentary on Netflix. Similarly, a significant portion of the anti-White Helmets campaign has occurred within English-language discourse, attempting to counter the White Helmets' communications by targeting the same or similar audiences. This research therefore focuses on the aspects of the White Helmets discourse that were designed for—and taken up by—English-speaking audiences.

- **Infrastructure outage (**Relevant to studies 3 and 4): Between February 7 and February 24 2018, the collection infrastructure that was developed to automatically collect tweets from the Twitter API by 'listening' for specific keywords suffered an outage. I am therefore unable to account for Twitter activity during this time. This is a point I raise in the data and methods section of Study 2 and 4 (Chapter 6 and Chapter 8).

These issues notwithstanding, social media data can provide useful insights into phenomena that have previously been difficult to study, such as information operations and disinformation, particularly when examined with an iterative and mixed method approach that values placing digital traces in context and strives to develop richer descriptions.

Methodological Limitations

Like any methodology, there are limits to the approach I used in this research. By using digital traces and interrogating these using a blend of quantitative and qualitative research techniques, I can reveal the temporal practices, routines, and coordination. I can follow the actors (cite) and I can follow specific content. However, there are also things I cannot see which limit my overall perspective of information operations. For example, unlike more traditional forms of ethnography that may rely on in-situ participant observation and perhaps involvement within a culture, my engagement was with the digital traces that were generated through their online activities and not with the participants themselves. As a consequence, I do not have access to emic first-hand accounts of the actors' motivations or intentions; neither do I have the 'insider' perspective that may have helped me understand the presence or arrangement of a formal structure or hierarchy, or 'backchannel' communications through private channels (that may or may not exist), which could suggest explicit coordination among actors that are not visible through public digital traces. Nevertheless, by integrating mixed methods and conducting iterative research, I do strive to build 'thicker' descriptions and gain a holistic view of the phenomenon under study.

Ethical Considerations of this Research

Using social media log data that is produced as the 'side effect' (Breiter & Hepp 2018) of online activities raises ethical considerations. Although this research used public social media data (public as in it was publicly posted versus posted e.g., in private channels) this does not mean that its use in research is automatically ethical nor that "participants" explicitly consented to be part of a research project (boyd & Crawford, 2012). Indeed, the use of social data poses particular challenges for the practice of informed consent (boyd & Crawford, 2012). A key ethical criterion is concerned with the assessment of benefits and non-maleficence (Olteanu et al., 2019), in other words the research should be beneficial and not cause intentional harm. To address this, the data used in the research within this dissertation was collected in accordance with a protocol that was submitted to and assessed to be acceptable by the Human Subjects Review Board at the University of Washington. Furthermore, respecting the privacy of individuals is an important aspect of preventing harm. To preserve the privacy of Twitter account owners I anonymize the account names of individuals and groups who would have a reasonable expectation of privacy for their social media communications (even though the communications were posted publicly on a social media platform). However, I preserve some account names for the following types of account (which are determined through public Twitter profile information, follower counts, presence of blue check marks):

- Media outlets and journalists
- Elected officials
- Highly visible accounts (>100,000 followers)

In cases where I provide textual examples, for example of tweets, I alter the text and metadata of anonymized tweets to make it more difficult to search for the original tweet. In Study 1 account types were used to characterize the accounts that are involved in the conversation; in Study 2 accounts were anonymized by community. In Chapter 7 (Study 3) we identified a small number of prominent authors publishing articles on websites outside of Twitter. These authors are not anonymized because they are self-identified journalists that wrote and published articles on public-facing websites. As a result their patterns of activity are important for understanding the nature of content sharing in this ecosystem. In Study 4 the above outlined policy is applied. Specific details about the labelling of anonymized accounts are specific to a study and will be described within that chapter.

Chapter 5:

Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community (Study 1)

Author Preface

Chapter 5 presents *Assembling strategic narratives: Information operations as collaborative work within an online community*[23]**,** previously published work that examines **how information operations manifest on Twitter** and investigates to **what extent the emergent online crowd 'work' together as part of the information operation**. This research presents a case study of a specific episode of the Syrian Civil War: the reemergence of Omran Daqneesh, who colloquially became known as "Aleppo Boy" on Syrian state television, almost a year after his family's home was bombed in an airstrike conducted by the Syrian government or their allies, Russia. In the aftermath of the strike, Omran was rescued by the White Helmets and placed in the back of an ambulance—a moment that was captured by the Western mainstream media and an image that illustrated the suffering of the Syrian people. We pick up the story around the reemergence of Omran, analyzing the English-language Twitter conversation surrounding the appearance of the Aleppo Boy and his father in a televised interview on Syrian state television—viewing this as part of

[23] This study is previously published work. To cite material from this Chapter, please cite the original: Wilson, T, Zhou, K., & Starbird, K. (2018). Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community. *Proceedings of the ACM on Human-Computer Interaction*, *2*(Computer Supported Cooperative Work), pp.1-26.

an information operation that sought to undermine the current understanding of events (that Omran was injured in an airstrike on his family's home), generate uncertainty, and to reignite—and then disrupt—the ongoing sensemaking about the event.

We used a mixed-methods approach to examine the network structure of the conversation, conduct qualitative content analysis of the tweets, and take a closer look at the micro-interactions between Twitter accounts involved in the discourse. In this study we present these as varying 'levels' of analysis — from macro (network), through meso (content), to micro (interactions). Our analyses shifted between these levels, with insights at one level forming the basis of questions at another. Through this iterative process we were able to obtain a comprehensive understanding of the Twitter conversation regarding the Omran, including *who* was involved, *what* they were doing, and *how* they were doing it.

The analyses of the Aleppo Boy conversation on Twitter uncovered a network of clustered users that contributes to a contested and politicized information space surrounding Omran's story; the presence of undermining narratives that serve to disrupt the mainstream media's narrative and confuse the audience; and a suite of techniques used when promoting, defending, or undermining narratives.

This work was published and presented at CSCW 2018 and was one of the first CSCW studies of disinformation and online information operations (see also Arif et al., 2018). Studying the Omran conversation as being part of an information operation was a decision based upon exploratory research conducted over the summer of 2017, engagement with historical accounts of similar activities, "gray" literature in the form of intelligence reports, and nascent (at the time) publications about information operations on social media platforms released by Facebook (Weedon et al., 2017) and Twitter (Twitter Public Policy, 2018). A contribution of this research was the introduction of the concept of online information operations to the CSCW community and the examination of these activities as a form of online collaborative 'work'—extending current conceptualizations of online activism and online volunteerism and of the collaborations that take shape around and through them.

This research is a collaboration between the emCOMP lab and DataLab at the University of Washington and was supported by National Science Foundation Grants 1749815 and 1715078, and Office of Naval Research grant N00014-17-1-2980. Thanks go to the anonymous CSCW reviewers for their constructive feedback that helped refine this paper, Daniel Scarnecchia and the Harvard Humanitarian Initiative for their guidance on this project, Ridley Jones for assistance during data analysis, and the UW SoMe Lab for providing infrastructure support.

<<THE PUBLISHED PAPER BEGINS BELOW>>

Abstract

Social media are becoming sites of information operations—activities that seek to undermine information systems and manipulate civic discourse (Lucas & Nimmo, 2015; Stamos, 2017; Waltzman, 2017; Twitter Public Policy, 2018). Through a mixed methods approach, our research extends investigations of online activism to examine the "work" of online information operations conducted on Twitter. In particular, we analyze the English-language conversation surrounding the reemergence of Omran Daqneesh (the "Aleppo Boy") on Syrian state television, almost a year after his family's home was bombed in an airstrike conducted by the Syrian government. We uncover: a network of clustered users that contributes to a contested and politicized information space surrounding Omran's story; the presence of undermining narratives that serve to disrupt the mainstream media's narrative and confuse the audience; and the techniques used when promoting, defending, or undermining narratives. In the current climate of increasing polarization in online social spaces, this work contributes an improved understanding of information operations online and of the collaborations that take shape around and through them.

Introduction

There is increasing awareness and concern regarding political propaganda, disinformation and "fake news" online. The same affordances of social media that enable individuals to connect with others and share content also present a weakness that can be exploited to undermine communities and manipulate civic discourse (Weedon, Nuland & Stamos, 2017). Large social media companies have publicly confirmed that their platforms—and therefore their users—had been subject to information operations conducted by state and non-state actors during the 2016 US Presidential Campaign (Stamos, 2017; Twitter Public Policy, 2018). Information operations is a term used to describe the suite of methods that are used to influence others through the dissemination of propaganda and disinformation (Lucas & Pomeranzev, 2016; Waltzman, 2017), for example through the deliberate amplification of false news that contains intentional misstatements of fact. Information operations are longstanding methods that have been used as a tool to gain asymmetric influence for centuries (Lucas & Nimmo, 2015; Waltzman, 2017) and an effective alternative for, or complement to, kinetic ("boots on the ground") warfare (Darley, 2006; Lucas & Pomeranzev, 2016). The advent of social media, and the associated interconnected online social networks and abundance of user data, mean that information operations are now both easier to implement and (likely) more effective (Lucas & Nimmo, 2015; Waltzman, 2017).

This research aims to understand how the manipulative mechanisms of information operations manifest online, specifically on social media. We focus on a particular episode of the Syrian conflict, which began in 2011—specifically the disruptive narratives surrounding Omran Daqneesh (the "Aleppo Boy"), as it exemplifies a case of online information operations. Omran was photographed in the back of an ambulance, in an image that garnered global media attention, after his family's home in Aleppo, Syria was reportedly destroyed by an airstrike carried out by Syrian government forces or their allies. Our research looks at a time period in June 2017, about nine months after the viral photo, when Omran and his family reappeared in interviews broadcast on outlets loyal to Syrian President Assad. In those interviews, new and contested facts were presented that challenged the original narrative presented in the mainstream media. We examine the English-language Twitter conversation about the reemergence of Omran, and the competing narratives that formed within online discourse on the platform.

Omran's story takes place within the context of a prolonged political conflict, and our work focuses on social media use in that context. Social media use is a common feature of crisis events, including natural disasters (Starbird & Palen, 2010; Vieweg et al., 2010; Dailey & Starbird, 2017), protests, terrorist attacks, and war or unrest (Al-Ani, Mark & Semaan, 2010; Howard et al., 2011). The Syrian conflict is no exception and has been described as the most socially-mediated conflict in history (Zaram, 2013). One salient dimension of social media use in the crisis context involves people coming together to collaborate and work together towards a common cause. This includes "digital volunteerism" that helps to meet the needs of disaster-affected people (Starbird & Palen, 2011; Starbird & Palen, 2013; White, Palen & Anderson, 2014) and "online activism" where people use available platforms to garner support and express solidarity for a political cause (Starbird & Palen, 2012) or a population affected by war (Al-Ani, Mark & Semaan, 2010). Our research extends these investigations of online activism to examine the "work" of online information operations. We look to the online conversation—and competing narratives—around Omran's story to identify and understand information operations through the lens of computer-supported cooperative work (CSCW).

In other words, we want to understand who was participating in these conversations, what they were doing, and how they were doing it. Using an iterative, mixed methods approach, we conduct analysis at three levels:

1) at the macro level we use network representations to determine the actors involved in the Omran conversation and their roles;

2) at the meso level, we analyze how information is added to the conversation and assembled to form competing narratives; and

3) at the micro level we focus on the specific actions of Twitter users engaged within this information space.

This analysis reveals that the Omran conversation represents a digital battlefield of sorts, where two (or more) "sides" worked to shape the information space to support their political goals. They did this by propagating their preferred narratives (in some cases) and working to undermine the other side's narratives (in others). They utilized specific techniques, such as introducing uncertainty into the information space and discrediting information providers to promote, defend, or challenge narratives. Though there is evidence that government-funded media and accounts were active in these conversations, these efforts were not merely constituted by a single set of "agents" from a particular government or other organization but were integrated into the social fabric of online interactions and, in this case, online activism.

Adopting a CSCW perspective, we examine and unpack this collaborative (but not necessarily explicitly coordinated) activity as a form of information operations—pushing our understanding of this phenomenon to account for complex and sometimes organic effects that emerge from interactions within the online crowd. In the context of an internationalized conflict and climate of increasing polarization in online social spaces, this work contributes an improved understanding of information operations online, and of the collaborations that exist between groups of users, who consciously or not, are involved in producing and amplifying politically-charged narratives.

Related work

Information Operations and their Use on Social Media

Information operations refer to a collection of methods that seek to disseminate crafted messages (propaganda) within an opponent's information system (Waltzman, 2017). Information operations are longstanding methods that have persisted for centuries, albeit under various guises such as 'information warfare' (Lucas & Nimmo, 2015), 'weaponization of information' (Waltzman, 2017), and 'active measures' (Bittman, 1985). Referring to operations orchestrated by the Kremlin during the Soviet era, Bittman (Bittman, 1985) describes clandestine operations that seek to extend influence by focusing on

vulnerabilities and festering malaise. These rifts are exploited to propagate disinformation, forgeries, rumors, and clandestine broadcasting through print media, television, and radio. Speaking of more recent strategies of information operations that leverage online connectivity, Pomerantsev and Weiss (2014) explain that the purpose is not to persuade the audience and bring them around to a particular way of thinking, but to 'muddy the waters'—to confuse, distract, and distort information spaces. Such measures ultimately paralyze the decision-making capabilities of the public as increasing uncertainty leaves individuals unsure of what to believe and how to react (Pomerantsev & Weiss, 2014).

The interconnected structure of online social networks and the abundance of user data collected by social media platforms allow information operations to be implemented more easily and more effectively (Lucas & Nimmo, 2015; Waltzman, 2017)—collecting personal and log data and segmenting communities along demographic information, political ideology, and 'likes' is fundamental to the advertising revenue of social media platforms. Recent reports point to the suspected misuse of personal data by companies seeking to disseminate highly targeted and personalized messages to manipulate the public (Adams, 2018). Social media platforms such as Facebook admitted that information operations—which they define as "actions taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment, to achieve a strategic and/or geopolitical outcome" (Weedon, Nuland & Stamos, 2017) (p.4)—were posing a significant risk to the social media platform. This risk became reality when Facebook confirmed it had been misused, with a network of 470 inauthentic accounts spending $100,000 on advertising between June 2015 and May 2017 (Stamos, 2017). In January 2018, Twitter also notified 1.4 million users that they may have engaged with suspected "trolls" from the Russian government-funded Internet Research Agency during the 2016 US Presidential Election (Twitter Public Policy, 2018). The realization that these information spaces—the fabric of online social networks—are being exploited to manipulate the public, stifle discussion, and sow discord is disconcerting. It also underscores the need for more research to understand how information operations work.

Shaping Narratives as a Form of Information Operations

One aspect of information operations involves the intentional creation, propagation, and shaping of narratives for (geo)political purposes (Chivvis, 2017). Narratives allow humans to structure information, giving meaning to actions that facilitate sensemaking (Miskimmon, O'Loughlin & Roselle, 2014). Aside from broad categories of fiction or nonfiction, researchers have characterized various types of narratives: Miskimmon, O'Loughlin and Roselle (2014) describe strategic narratives that are employed to shape discussions within international relations, to manage expectations, and shape the future behavior of domestic and international actors; Corman (2013) identifies master narratives as broad social narratives

which are persistent over time and familiar to members of a particular culture; Rogers and Tyushka (2018) define counter-narratives, which symmetrically confront an existing narrative and offer an alternative, and anti-narratives that aim to spread negativity and cynically spoil the existing narrative without offering an alternative explanation. The conversation surrounding Omran includes multiple and often conflicting narratives, and in this research, we examine those narratives as they emerged on Twitter—conceptualizing them as specific types narratives, describing the communities they emerged from, and identifying the techniques used to shape and propagate them as one aspect of online information operations occurring on social media.

Information Operations as Collaborative Work

Information operations represent a type of collaborative work undertaken by groups of actors integrated within online social networks. Recent studies have uncovered concerted and coordinated efforts to conduct information operations online. In the Philippines, Ong and Cabanes (2018) found that PR professionals are employed as 'architects of networked disinformation' (p.6), using fake accounts and paid influencers to distort trending topics on social media and 'hack' the attention of the public. In China, bots supplement low-level government workers, who are incentivized with additional vacation time, to promote anti-Taiwanese and pro-Chinese government propaganda on social media. These efforts are so well orchestrated that it appears wholly automated (Monaco, 2017). And what has been termed a "troll factory" was discovered in St. Petersburg, Russia. As part of a sophisticated operation to interfere with the 2016 US Presidential Election, young Russians with a strong command of English and knowledge of American culture worked in teams of three to 'hack' the comments sections of news sites and social media (United States House of Representatives Permanent Select Committee on Intelligence Testimony, 2017; Troianovski, 2018). In these examples, the information operations illustrate a formalized coordinated action—a central node is orchestrating the accounts within the disruptive network. Participants are likely aware that they are part of a network, motivated by financial, political, social, or psychological reasons (Ong & Cabanes, 2018). In our research, the nature of the network is distinctly different. Although we cannot rule out coordination between some nodes in the network, the collaborative work of information operations has an emergent property as journalists, media organizations, government officials, and information activists converge online to conduct information operations at least somewhat organically—i.e., not exclusively orchestrated through a centralized and coordinated effort.

CSCW provides a lens through which to study this phenomenon, allowing us to explore collaborations that are emergent, as demonstrated in prior research of computer-supported cooperative activities, specifically "digital volunteerism" (Palen & Liu, 2007; Starbird & Palen, 2011; White, Palen & Anderson, 2014) and

"online activism" (Howard et al., 2011; Starbird & Palen, 2012; Savage & Monroy-Hernández, 2015; Flores-Saviaga, Keegan & Savage, 2018). Palen & Liu (2007) introduced the term "digital volunteerism" to describe computer-supported cooperative activities that social media facilitate, which allow a collective public to play an active role in disaster response, aiding their own recovery but also participating as digital volunteers to help others (White, Palen & Anderson, 2014). Studies regarding the role of social media in online activism occurred in the wake of the 2011 Arab Spring. Local, affected individuals took to social media to apply political pressure, shape political discourse, and to extend their reach—informing western audiences about events on the ground (Howard et al., 2011). Starbird & Palen uncovered the collective work conducted as individuals on Twitter considered what information they shared (amplified) through Twitter's retweet mechanism—leading to a collective, crowd-powered recommendation system (Starbird & Palen, 2012). In other contexts, Savage and Monroy-Hernández (2015) demonstrate how the online presence of a militia group leveraged social media to facilitate the mobilization of individuals for participation in collective efforts offline. And a longitudinal study of a politically active Reddit community described how the online community worked together to sustain itself—more active users work to mobilize others and deploy technical tools (e.g., bots) to sustain engagement and help develop the groups' shared identity (Flores-Saviaga, Keegan & Savage, 2018).

In each of these (prosocial) examples, emergent collaborations are formed as individuals converge online to engage in collaborative work to meet a collective goal. Our research extends this idea to explore how similar emergent collaborations form to conduct the "work" of online information operations. Situated within a prolonged and internationalized conflict, this work aims to improve our understanding of information operations online. Through integrated analyses of the who, what, and how, we examine the collaborations that exist within groups of users who collectively work to produce, support, and challenge various narratives that reflect the goals and strategies of information operations.

Background

Syrian Civil War

The Syrian civil war began in 2011 after anti-government protests calling for President Bashar al-Assad to step down turned violent, leading to armed conflict that escalated into a full-scale civil war (International Committee of the Red Cross, 2012). Although technically a civil war between government loyalists and rebel opposition groups, the situation is further complicated by the presence of Islamic State (IS) militants, and the involvement of other countries including Russia, Iran, and the United States (Gilsinan, 2015).

Russia and Iran support the Assad government, targeting Syrian opposition forces and IS militants (Associated Press, 2013), while the US and Gulf League states provide assistance to the Syrian opposition groups and Kurdish forces that are also fighting IS (Gordon, 2014). The result is a complex, multi-faceted, and internationalized conflict in which the warring parties have demonstrated a "*lack of adherence to the norms of international law*" (United Nations Human Rights Office of the High Commissioner. 2015).

Event Background

Within the context of this protracted and bitter conflict, Omran Daqneesh, the 'Aleppo Boy', gained widespread media attention in August 2016 after his family's home in Aleppo was reportedly destroyed by an airstrike carried out by Syrian government forces or their allies. After being rescued from the rubble by the local volunteer group Syrian Civil Defense (the White Helmets), Omran was photographed in the back of an ambulance—an image that helped garner attention for and sympathy with the plight of the besieged citizens of Aleppo.

Shortly afterwards, media aligned with Russian and Syrian government interests began to question Omran's story, and Syrian President Assad reportedly claimed the photograph was fake (Frisk, 2016). Initially, efforts at introducing such counter-narratives did not spread widely. However, that changed on May 26 2017, when the Kuwaiti telecommunications company Zain released a video advertisement to mark the start of Ramadan. The video featured a child actor playing the role of Omran as part of an anti-extremist message. Contrary to the version of events presented in the global 'mainstream' media the previous year, the video portrays Omran as the victim of a suicide bomber affiliated with a terrorist group rather than a Syrian-government airstrike.

After the release of the Zain advertisement, Omran and his father reappeared in a series of interviews that were featured in media outlets loyal to Syrian President Assad in Syria and other countries including Russia, Iran, and Lebanon. These interviews presented new and conflicting information. For example, Omran's father claimed that his son had been used as 'propaganda' and asserted that the family home was not destroyed in an airstrike but by a bombing conducted by anti-government 'rebel' groups—contrary to the narrative presented in the media in 2016.

This research centers around the release of the Zain advertisement at the end of May 2017 and the reemergence of Omran in June 2017 and is situated within a conversation that involves voices that are pro-Syrian government, anti-Syrian government, and from representatives of the mainstream media. Emerging from this dialog are a series of disruptive and often conflicting narratives. These are both challenged and

promoted by networks of interconnected accounts on 'both sides' of this conversation and disseminated to a wider audience.

Methodological approach

In this research we adopt a mixed methods approach, iteratively blending quantitative and qualitative analyses. We first generate tables and graphs of the data from a high level to identify patterns and anomalies, and then using these as 'entry points' for more in-depth investigation of the data using qualitative approaches, expanding upon methods developed for the investigation of online rumors in the context of crisis events (Maddock et al., 2015; Starbird et al., 2015). This approach reflects an adaptation of constructivist grounded theory methods (Charmaz, 2014) to large-scale online interactions.

Within the context of the Omran case study, we conduct our analysis at three levels: At the macro level we use network representations and descriptive statistics to determine who is involved in the conversation and their positionality within the network. Using a temporal chart of tweet volume over time we transition to the meso level, unpacking how information is assembled to form narratives as the conversation around Omran progresses. And at the micro level we identify specific actions that Twitter users perform when promoting or contesting narratives, or when challenging others within the information space. We provide additional detail about specific methods in each section of our findings. Though we present these in three distinct Findings sections, these analyses were conducted iteratively, often moving back and forth from one level to another, as insights from one informed new questions to investigate at another.

Data

Data was obtained from a larger collection related to the Syrian conflict, which consists of tweets posted between May 27 and September 9, 2017. We collected the data using the Twitter Streaming API, tracking various keyword terms related to the Syrian conflict in real-time, including geographic terms of affected areas. We then scoped this data to tweets that contained the term "`Omran`" (case-insensitive), were posted between May 28 2017 (00:00:00 UTC) and June 16 (23:59:00 UTC), and that Twitter identified as English (denoted by the language code "`EN`" in the tweet metadata). The rationale for collecting English tweets was based upon our interest in how information operations designed for western audiences are taking shape in online spaces. We acknowledge that conversations around these same topics in other languages would likely look different. Our dataset consisted of 2,966 original tweets, 21,940 retweets, authored by 17,284 distinct accounts.

Twitter Account Coding

To help describe the underlying network and the collaborative work being undertaken while still preserving the privacy of individual account owners, we categorized each of the accounts that authored a tweet featured (i.e., excerpted) in the text of this paper. Two authors of this paper conducted the classification using consensus coding, which involved discussion between researchers to resolve any disagreements. The categories and the categorization process were informed by both a long-term study into the information ecosystem surrounding the broader Syrian conflict, and through analysis of users' Twitter profiles and tweets specifically related to Omran's story. Through this process five account types were identified: Journalists—accounts that were identifiable as real people who produce content for one or more blogs, news sites, or media organizations; Media organizations—accounts associated with websites or other outlets that publish news and other content both in the "mainstream" (e.g. CNN) and "alternative" (e.g. 21WIRE) media; Government Officials—individuals who are verifiable members of state governments; Academics—individuals who are researchers at academic institutions; and Information Activists—accounts that authored high volumes of politically charged content and were active within this and other spaces. Many of the information activist accounts appeared to be genuine accounts (of people who are not affiliated with a government or political organization), however at least one is suspected of being an agent of a political group (BBC Trending, 2018). Other account types, such as celebrities, concerned citizens, and everyday users appear in the larger dataset, but not among the highly retweeted users featured in our graph or data excerpts.

Statement about Account Anonymization

To respect the privacy of individuals who may not understand their content is public and persistent, we have anonymized all account names in this paper with the exception of media organizations (e.g., CNN, RT, The Daily Telegraph). In lieu of account names we have used account types (derived from the account coding explained above) to characterize the accounts, providing insight into the types of accounts that are involved in this information ecosystem. Due to their influence and significance within the information space, accounts within the top-10 most retweeted users (Table 5.1) were assigned a persistent handle denoting their account type and position within the top-10 (e.g., Journalist1; InformationActivist3). For other non-media organization accounts, we provide the account type based upon their classification (e.g., Government Official; Information Activist), for context.

Paper Organization

Our findings are organized around the three levels of analysis in increasing order of granularity: First at the macro-level we describe the structure of the information space that takes shape through Omran-related tweets, following retweet patterns to reveal the most retweeted users and the broader patterns of information flow. Second, at the meso-level, we present a temporal analysis of the events and conflicting narratives within the Omran conversation, demonstrating the evolution of the dialogue and providing exemplar tweets which characterize the various phases of the conversation. And third, at the micro-level we explore three salient activities we observed in this conversation—the techniques of introducing uncertainty into the information space, and discrediting information providers.

Structure of the information space (macro level)

To understand who is involved in the contested and politicized information space surrounding Omran's story, we conducted and integrated two complementary analyses:

1) exploring the structure of the conversation through a retweet network graph; and
2) identifying and analyzing the most highly retweeted accounts.

Retweet Network Graph

We developed a shared network graph (Figure 5.1) to determine the information-sharing patterns across Twitter users. In the graph, each node represents an account, sized by the number of times they were retweeted—i.e., large nodes represent highly retweeted accounts. Nodes are connected via a directional edge from a retweeting account to a retweeted account. Each edge is weighted by the number of retweets between the two. The structure of the graph was determined using the ForceAtlas2 algorithm (Jacomy et al., 2014). The colors represent clusters or "communities" determined by the Louvain algorithm (Blondel et al., 2008). In the following sections we analyze the main clusters of the network and describe the positionality of the most retweeted users within the information space.

Blue Cluster: An Active, Cohesive Group of Pro-Syrian Government Journalists and Activists

The graph reveals a large central cluster, in blue, featuring many of the most highly retweeted accounts in the Omran dataset. These influential blue accounts occasionally retweet each other, but the cluster is primarily held together by a large number of small accounts (not highly retweeted) who send many Omran-related tweets, retweeting several of the influential blue accounts. This cluster was by far the most active in the Omran conversation, containing more than half of the participating accounts and producing 63%

(15,807 tweets) of the entire dataset. The content shared within this cluster was supportive of the Syrian government and President Assad and highly critical of "mainstream" media. Accounts in this cluster produced and amplified many alternative narratives to the original story of Omran and his family.

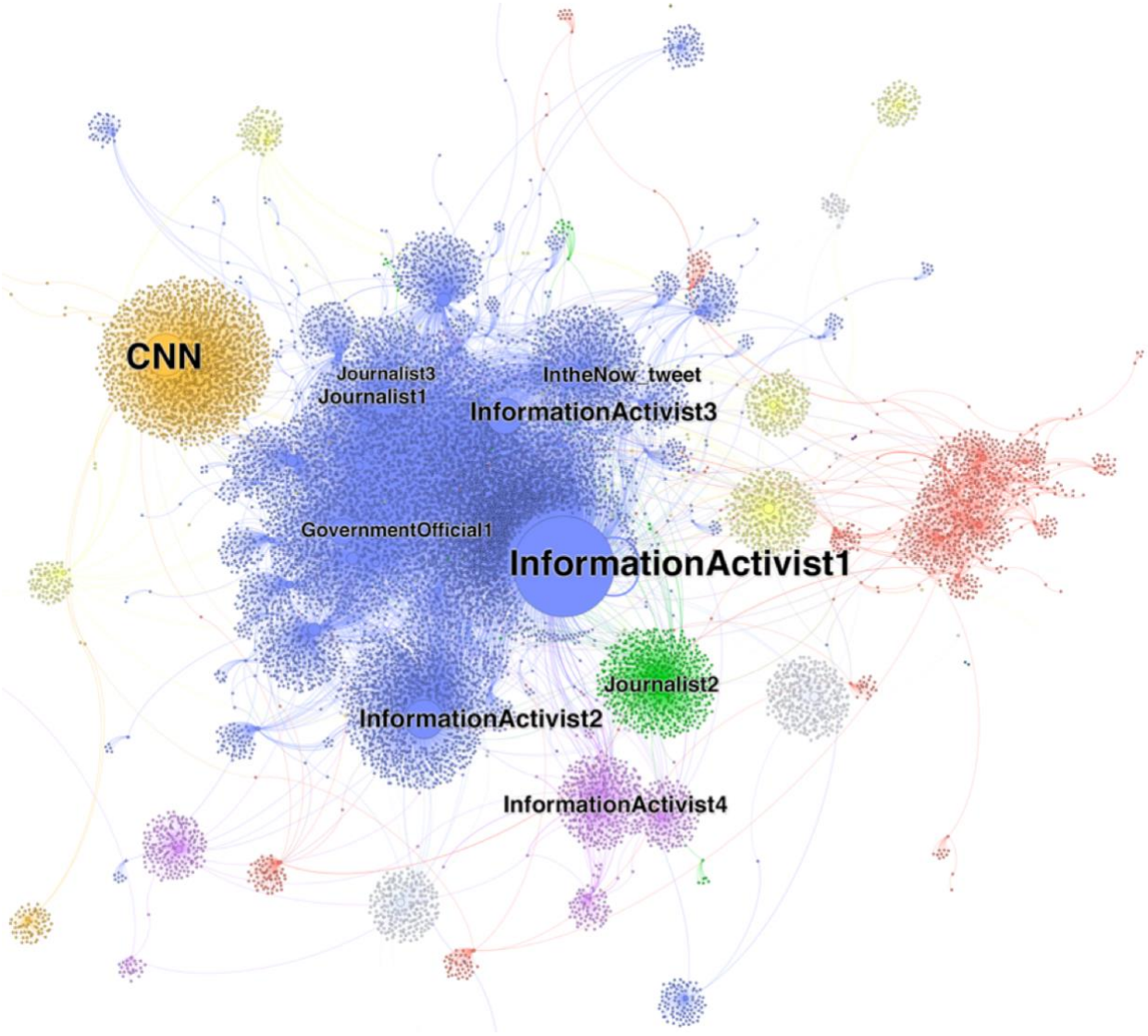

Figure 5.1. Network graph of the accounts involved in the Omran conversation.

Seven of the ten most-retweeted accounts (Table 5.1) appear in this cluster. Examining these accounts underscores both the diverse and multifaceted nature of the Syrian conflict and the informational operations within that conflict. It also highlights their convergence around a set of common themes.

Central to the Blue cluster are several diverse political actors and "information activists" who are highly retweeted within the Omran conversation, each of which expresses a decisively pro-Syrian government stance. *@InformationActivist1* is the most heavily retweeted user in our dataset. This politically-active

account, which features an attractive young woman in its profile image, consistently tweets pro-Syrian and pro-Russian government content, including commentary challenging politicians in the US and other NATO countries. In the Omran data, this account criticizes both the White Helmets and western media for producing propaganda. *@InformationActivist2* is another highly politically active account whose tweets often focus on issues in the Middle East. In the Omran conversation, this account expresses an anti-rebel (pro-Syrian government) stance with tweets centered around claims that Omran is well and living in Aleppo and describing a prior situation when rebel fighters kept civilians hostage in Aleppo. *@InformationActivist4,* a self-described citizen of a south Asian country and Twitter commentator, participates in various political conversations all around the globe, often as a critic of "mainstream" media and Islam. During the Omran conversation, he reports Omran's father saying he was "*coerced to do interviews and attack the Assad regime.*"

There were two other highly politically active accounts that focused almost exclusively on the Syrian conflict (from a pro-government stance). *@InformationActivist3,* which claims to be reporting from the "battlefields" of the conflict, live-tweets statements from Omran's father during the interview and is critical of the White Helmets who reportedly rescued Omran (suggesting they fabricated the whole event) and the opposition rebel forces (labeling them as terrorists). *@GovernmentOfficial1*, a serving member of the Syrian government, is highly critical of the "mainstream" media, referring to Omran's case as a "fake story" and using terms such as "presstitutes" and "NATO media".

Within the Blue cluster we also note the presence of two independent journalists who are supportive of the Syrian government: *@Journalist1*, and *@Journalist3*. *@Journalist1* is a European independent journalist, visitor to Syria as a guest of the Syrian government, and prominent voice in the Twitter conversation surrounding the Syrian conflict. *@Journalist1* is critical of NATO, foreign intervention in Syria, and of the mainstream media's reporting in the region. Her tweets seek to correct the version of events presented by the mainstream media, using excerpts from the interview with Omran's father to support a counter position that Omran is alive and well. *@Journalist3* is also a North American independent freelance writer who has spent time reporting from Syria, including interviewing Omran's father in June 2017. *@Journalist3* is also critical of mainstream media, calling Omran a child victim of propaganda (at the hands of the White Helmets and "mainstream" media).

Finally, we note the presence of Russian media within the blue cluster: *@RT_com* (RT or Russia Today) and *@IntheNow_tweet.* RT_com is not among the ten most retweeted accounts (11th with 423 retweets), but *@IntheNow_tweet,* an account connected to RT's digital platform, is. *@IntheNow_tweet* reports that

Omran supports Syrian government forces, and that the family were furious at being used for anti-Syrian government propaganda—an alternative, competing narrative to that presented by the mainstream media.

Orange, Yellow and Green Clusters: The "Mainstream" Media

In Orange, Yellow, and Green are three clusters that are primarily constituted by mainstream media accounts (*@CNN* in Orange; *@Journalist2* in Green; and *@Independent*, *@WashingtonPost*, *@NYTimesWorld*, and *@Telegraph* in Yellow). Together, these three clusters constitute 17% of the participating accounts and produced 12% of the content in the dataset. The mainstream media clusters are somewhat integrated into the other clusters, through a few common retweeters, but mostly have their own slightly-engaged (in terms of this topic) audiences who only participated in the Omran conversation through retweets of that media outlet. Each of the highly retweeted accounts in these clusters only posted a single Omran-related tweet. For example, *@CNN* tweeted a link to an article on their website that ran after the interview with Omran's family, describing his return to the media spotlight, but aligned with the original 2016 narrative of how his family's home was hit by Syrian-government airstrikes. After posting that article, *@CNN* is subject to a great deal of criticism and directed challenges by accounts from the Blue Cluster (which results in the proximity between those two clusters in Figure 5.1), but they do not respond.

Red Cluster: A Small Cluster of Anti-Syrian Government Activists

The third largest cluster in the graph is colored in Red, spread along the right side of the graph. This cluster includes 5.7% of participating accounts, which generated about 6% of the content in the Omran conversation. This cluster does not contain any of the top-10 most retweeted accounts, but it does contain some of the most highly-retweeted accounts during specific time periods within the Omran conversation. This cluster includes a few dedicated, politically active accounts similar to those that we see in the Blue, but on the opposite "side" of the conversation. Tweets from accounts in this cluster were highly critical of the Syrian government and primarily focused around defending the original story of what happened to Omran and his family. We explore some of the content from these accounts in the temporal analysis below.

Other Clusters

In purple are a few other peripheral influencers who have a mostly distinct audience (from other nodes in the graph) but are also connected, via common retweeters, to the blue cluster. Hundreds of other tiny clusters at the periphery of the graph are not featured in the analysis here. The largest of these are colored grey. Others have been excluded from the edges of the graph.

Most Retweeted Users

Table 5.1 lists the top-10 most retweeted accounts in the Omran data. Retweets of these accounts constituted more than half of the entire dataset. We also include a qualitative characterization of these accounts derived from long-term (10 month), ethnographic study of the information ecosystem surrounding the Omran story and broader Syrian conflict.

Later in this section we describe some of these accounts in the context of the cluster that they appear in the retweet network graph (Figure 5.1). The Cluster column in Table 5.1 refers to the account's position in that graph.

Screen name and description	No. followers	No. tweets (retweets)	Cluster
@InformationActivist1: Profile claims she is an "Independent Lebanese geopolitical commentator." Posts are consistently supportive of the Syrian government and its allies, including Russia and Iran	88,793	5 (3772)	Blue
@CNN: Official account of western "mainstream" media outlet. Related tweet states that Omran reappeared on Syrian TV containing two images of Omran and a link to an article (with a video of his father's interview).	35,947,914	1 (1522)	Yellow
@InformationActivist2: Profile claims that s/he is Syrian and "not a rebel", positioning her/him as a supporter of the Syrian government. Posts mix pop culture and political messaging.	43,186	4 (1403)	Blue
@InformationActivist3: Account focused on news related to Syria, including updates from battlefields. His posts are consistently strongly supportive of the Syrian government.	28,570	15 (1307)	Blue
@GovernmentOfficial1: Member of the Syrian parliament for Aleppo and Chairman of the Syrian Federation of Industry. Posts are strongly supportive of the Syrian government.	20,581	5 (912)	Blue
@IntheNow_tweet: Account of In The Now, an online video channel operated by Russian government-funded RT.	53,387	2 (838)	Blue

@Journalist1: Western European journalist who reports on the Syrian conflict from a pro-Syrian government stance. Omran-related posts were critical of the mainstream media and cast doubt on the original Omran story.	18,138	5 (692)	Blue
@Journalist2: British journalist and former writer and editor of several UK tabloids. On Syria, tweets are anti-Syrian government, e.g., questioning the UK's inaction of chemical weapons. Posts one tweet about Omran, relief that he was 'alive and well'.	5,824,512	1 (597)	Green
@InformationActivist4: Concerned about the rights of the Hindu population. Syria-related tweets have an anti-opposition stance, critical of 'ISIS rebels'.	46,367	2 (488)	Purple
@Journalist3: North American journalist who reports on the Syrian conflict from a broadly pro-Syrian government stance: "…ample experience in Syria and occupied Palestine." Interviews Omran's family during this study.	25,095	5 (436)	Blue

Table 5.1: Top-10 most retweeted accounts in the Omran dataset

Contested facts, conflicting narratives (meso level)

To understand the conflicting narratives within the Omran conversation, and the nature of participation in the conversation, we:

1) generated temporal charts to identify patterns and spikes in the conversation, helping us identify and distinguish between interesting phases in the Omran conversation, which became entry points for;
2) 2) an in-depth investigation of the data at the individual tweet level, including the various contested facts and narratives surrounding Omran. All times are in UTC.

Temporal Analyses

We generated temporal graphs of the volume (per hour) of English language tweets and retweets from May 28 2017 00:00 until June 16 2017 23:59 (Figure 5.2). By analyzing the contents of the

original tweets (not retweets) that occurred around the spikes in volume, we were able to identify (and subsequently describe) key events in the conversation (Figure 5.3).

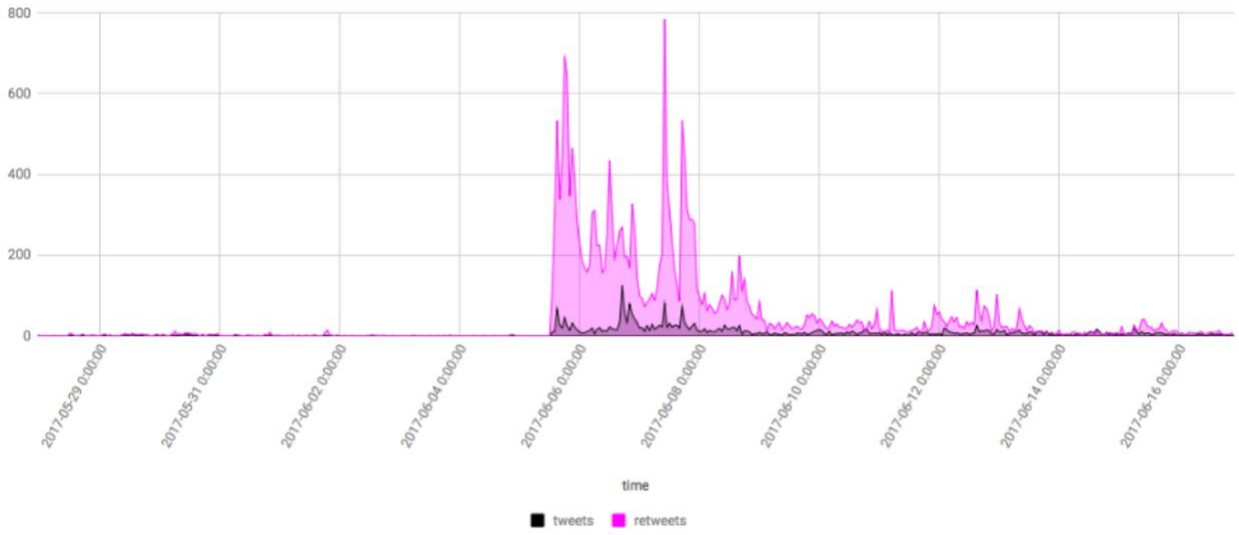

Figure 5.2: Temporal graph showing the volume of tweets and retweets during the Omran conversation (2017-05-28 - 2017-06-16)

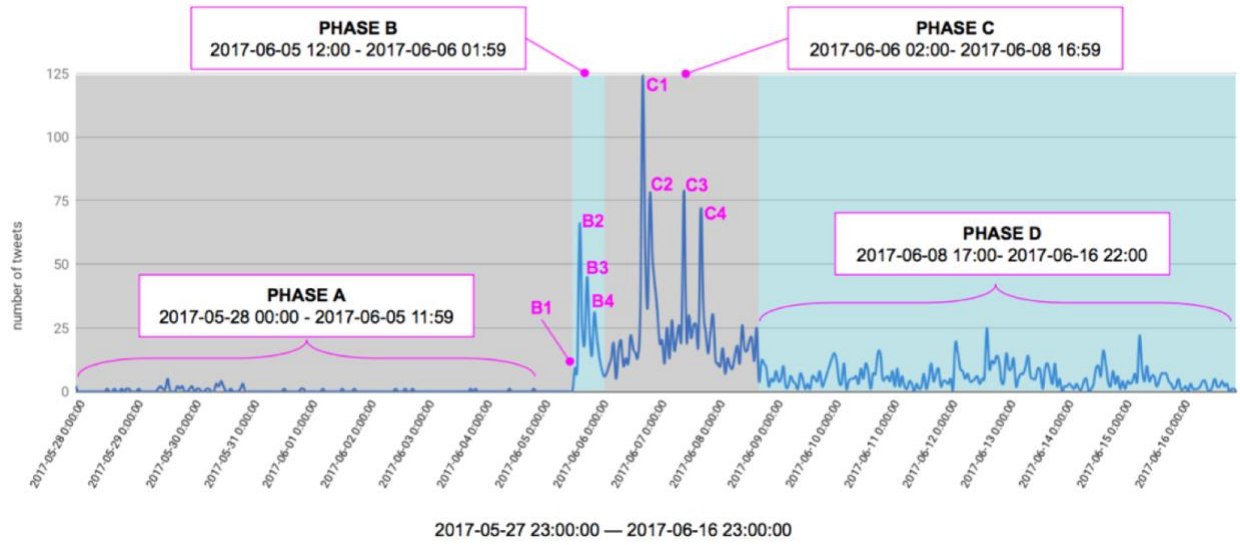

Figure 5.3: Temporal graph showing various phases in the Omran conversation

The volume of tweets can broadly be characterized as a low volume conversation (Phase A) followed by sudden spikes in activity, demonstrated by several peaks in the center of the graph (Phase B and C). After that there is a sustained low-volume conversation (Phase D). In the following sections we describe the conversation during the various phases (A-D), with a particular focus on the tweets sent around the peaks

in volume (B1-B4 & C1-C4). We contextualize the phases and peaks by illustrating the kind of content that was circulating on Twitter at that time.

Content Analysis

The presence of contested facts was identified through a process of open coding conducted by two researchers. In an iterative and collaborative process, individual tweets were examined in temporal order, with researchers independently noting new information that was being introduced into the information space. This procedure was carried out for the (temporal) first 150 tweets within the dataset and also for the 752 tweets that were retweeted once or more during the study period. This process resulted in the identification of dozens of pieces of 'new' information that were added to the conversation. A process of affinity-diagramming was used to remove redundancy and inductively assemble the narratives that were present in the conversation. We analyze each phase in more detail in the following sections.

PHASE A: Controversy Surrounding Release of Zain Advertisement.

In Phase A tweets are predominantly critical of the advertisement released by Kuwaiti telecommunications company Zain on May 26 2017, which implied that Omran was the victim of a suicide bomber—an alternative and in significant ways conflicting narrative to that presented in the mainstream media (Table 5.2). Tweets posted during this period criticize the Zain advertisement, calling it out as 'propaganda' or a 'distortion of facts'. Authors also use tweets to correct the version of events presented in the advertisement, clarifying that they believe that Syrian (or Russian) airstrikes were responsible for injuring Omran, not a suicide bomber affiliated with a terrorist organization (e.g. A_2). Much of this content during this period came from accounts in the Red (anti-Syrian government) cluster of the retweet network graph (Figure 5.1).

Phase	RTs	Time	Text	Cluster	Account Type
A_1	51	5/29/17 01:13	Omran and other Syrian children are primarily victims of criminality by Assad, Iran and Russia, before any other criminal.	red	Information Activist
A_2	4	5/29/17 10:03	this is beautiful, but IT IS NOT ISIS WHO KILLED OMRAN. so get your research together @zain	red	Journalist
A_3	63	5/30/17 05:38	@zain ad for Ramadan pisses off #Syria's jihadists by portraying the star of their biggest propaganda coup, Omran, as a victim of ISIS	blue	Information Activist

Table 5.2. Exemplar tweets from Phase A

However, tweet A_3, posted from an account in the Blue cluster, goes against this trend, defending the alternative narrative presented in the Zain ad, suggesting that the Zain ad was frustrating to the anti-Syrian government forces (referred to by the tweeter as 'jihadists'). This tweet also implies that Omran was used for propaganda by the western, "mainstream" media. From both sides of this argument we see people calling the others' perspective out as propaganda or a distortion of the truth.

PHASE B: Pre-interview Speculation followed by Response to the Interview

B1 (Table 5.3): This short (~30 minute) period consists of a small, but important, increase in activity caused by speculation surrounding what Omran's father would say in the imminent interview. Almost all of the prominent tweets during this period come from accounts within the Blue (pro-Syrian government) cluster of the network graph (Figure 5.1). The consensus is that Omran's father will 'speak the truth', refuting the mainstream media's version of events and providing an alternative explanation—implying that these users,

(93% of which are from the Blue cluster), are anticipating (positively) a challenge to the mainstream media's narrative and promoting this idea on Twitter. *@GovernmentOfficial1* sends two tweets (B1_1, B1_2), referring to the mainstream media's narrative as a "fake story" and accusing them of abusing Omran for political purposes. The family, it is claimed, are free in Aleppo and will speak the truth.

Phase	RTs	Time	Text	Cluster	Account Type
B1_1	233	6/5/17 13:42	Remember Omran?! His family tells the true story of an Aleppo boy abused by western media presstitutes for political propaganda!	blue	@Government Official1
B1_2	398	6/5/17 13:50	Omran Daqneesh.. the fake story that shocked the world! Another White Helmets lie as his family in free Aleppo tells all..!	blue	@Government Official1
B1_3	380	6/5/17 13:50	my friend @[hidden] with Omran Daqneesh,Aleppo boy whom pictures spreaded by MSM 2 demonize Syrian gov.revealed his family were pro-gov	blue	Journalist
B1_4	5	6/5/17 13:51	According to pro-Assad media sources, The father of Omran child will be on the Syrian state TV to talk about the real story of his son Omran	red	Journalist

Table 5.3. Exemplar tweets from Phase B (Around B1)

Phase	RTs	Time	Text	Cluster	Account Type
B2_1	74	6/5/17 14:28	After 5 months under arrest, Omran is used by the regime media. We know he must say what Assad wants	red	Information Activist
B2_2	5	6/5/17 14:43	Omran's father refused to give interviews after the regime's bombing of his home. Chose to stay in Aleppo under the regime	red	Academic
B2_3	54	6/5/17 15:09	Remember Omran @CNN used to push its 'regime change' propaganda? He's well living under Assad gov	blue	Information Activist
B2_4	412	6/5/17 15:11	Omran raising the syrian flag .. in aleppo	blue	Information Activist
B2_5	188	6/5/17 15:13	Little Omran is safe and happy in government-controlled Aleppo.	blue	Journalist
B2_6	65	6/5/17 15:13	After being under house arrest, Omran is "giving interviews" with Syrian media. Threatening and coercing a child #Assad	red	Information Activist
B2_7	325	6/5/17 16:28	Omran alive & well in #Aleppo. Father talks of bribes offered by NATO terrorists. Airing Syria TV tonight	blue	@Journalist1
B3_1	312	6/5/17 18:46	Omran Daqneesh is fine and living in Aleppo. (For those who understand Arabic, watch the video)	blue	Information Activist
B4_1	51	6/5/17 21:48	#Omran's Father: I didn’t hear any noise caused by fighting jets, I don't know how things happened. #Aleppo #Syria	blue	Information Activist

Table 5.4: Exemplar tweets from Phase B (Around B2-B4)

B2-B4 (Table 5.4): After the speculative behavior, we see the first large spike in volume (B2) as the first interview is broadcast, followed by two downstream peaks (B3 and B4). During this period, the conversation shifts back and forth from anti-Syrian government accounts in the Red cluster to pro-Syrian government accounts in the Blue cluster, with collective claims and counter-claims about Omran's story. Accounts from the Red cluster defend the original Omran narrative (as presented by the mainstream media) and challenge the new events that are introduced.

The Daily Telegraph (UK daily newspaper), publishes an article on the reemergence of Omran, including photos of Omran and his father being interviewed on Syrian state television. *@Telegraph* offers quotes from the interview (in which the father provides alternative versions of events), but they also point out that the family could be participating in the interview for their own safety. Others (e.g. B2_1, B2_6) challenge these new revelations by discrediting the interview—introducing the idea that the family have spent several months under house arrest and that the interview is being conducted under duress.

Counter to this position, accounts from the Blue cluster post tweets (e.g. B2_3, B2_5, B3_1) describing Omran as "alive and well" or "free and happy" in (government-controlled) Aleppo. Tweet B2_4 (the most highly retweeted during this period) celebrates Omran's—and by extension his family's—pro-Syrian government political leanings. The accounts from the Blue cluster, and hence the pro-Syrian government voices, are much more highly retweeted during this time.

As excerpts from the interview begin to appear on Twitter, it becomes apparent that Omran's father did challenge the mainstream media's August 2016 portrayal of events—providing additional information that is itself inconsistent and therefore contested. B2_2 (Red cluster) describes how the father refused to give interviews following the regime's bombing of his home (supporting part of the original narrative that the family's house was bombed in an airstrike), but B4_1 quotes the father saying that he did not hear any fighter jets (above his house on the day of the attack), suggesting another, unspecified cause (not a bombing) for his family's home being destroyed. Additional information to support this narrative is not provided.

Three of the most retweeted accounts in our data are also active at this time: *@InformationActivist1* posts a link to the interview showing Omran "fine and living in Aleppo". *@InformationActivist3* and *@Journalist1* raise the issue of bribes (by the mainstream media and others during the initial telling of Omran's story in 2016). *@InformationActivist3* doesn't mention the source of the bribes but *@Journalist1* places the blame with "NATO terrorists"—direct criticism of their involvement in the Syrian conflict.

PHASE C: Release of Post-Interview Articles in the Mainstream and Alternative Media

Phase C sees the release of articles from both the mainstream and alternative media reporting on the reemergence of Omran in televised interviews (Table 5.5). Consistent with their earlier reporting, The Independent, CNN, BBC, Washington Post, and New York Times (Yellow cluster), support the narrative that Omran was a victim of an airstrike and is now appearing on Syrian television in what could be propaganda. Countering the original narrative, Russian news agency Sputnik (Blue cluster) releases an article that promises the "true story". Outlets that position themselves as "alternative" media—21stCenturyWire (21WIRE), Veterans Today (VT), and ActivistPost—also publish articles about Omran which offer an alternative version of events. Several of the most retweeted users are active at this time. *@Journalist2* sends his only tweet, relief that Omran is alive and well, engaging his followers in the Green cluster (C1_1).

A variety of challenges to the mainstream media's narrative originate from the Blue cluster. *@Journalist1* promotes her article on 21WIRE, questioning the western (mainstream) media's lack of interest in Omran's reemergence (C1_2); *@InformationActivist1* levels criticism at the White Helmets for using Omran for propaganda (C4_1, C4_2); @IntheNow_tweet promotes Omran's support of the government and the family's fury at being used for propaganda; and *@Journalist3*, who interviewed Omran's father, quotes him as saying he was a member of the Syrian Army.

Phase C also features criticism leveled at the White Helmets, the local humanitarian group (based in rebel-held territories) that are reported to have rescued Omran from the rubble in August 2016. Tweets suggest the White Helmets lied, the truth has been revealed, and that the group had used Omran for propaganda purposes.

Phase	RTs	Time	Text	Cluster	Account Type
C1_1	600	6/6/17 12:36	Alive, well & smiling again! Fantastic. ????	green	@Journalist2
C1_2	243	6/6/17 15:30	ALEPPO: The Return of Omran to No Fanfare from Western Media	blue	@Journalist1
C1_3	60	6/6/17 16:27	Suffering and propaganda: The true story of #OmranDaqneesh	blue	Media Organization
C2_1	40	6/6/17 19:29	When last we saw Omran, his home was destroyed in an airstrike. Now he's appearing on Syrian TV	blue	Media Organization
C2_2	462	6/6/17 21:00	Turns out the family of THAT Syrian ambulance boy, Omran Daqneesh, 'supports' government forces #Aleppo	blue	Media Organization
C3_1	46	6/7/17 09:01	#Omran father: "The Syrian Army always protects the country. I served in the Syrian army. The army is the people."	blue	@Journalist3
C4_1	2040	6/7/17 17:16	Father of little Omran Daqneesh from #Aleppo, describes how the White Helmets used Omran as a tool in their propaganda against #Syria	blue	@Information Activist1
C4_2	414	6/7/17 18:54	Remember Omran?His family is furious he was used for anti-government propaganda	blue	Media Organization

Table 5.5: Exemplar tweets from Phase C

PHASE D: Post-interview — Anti-media sentiment

Phase D is characterized by a sustained conversation from the Blue cluster that is predominately critical of the mainstream media and their Omran narrative. The general feeling is that the mainstream media lied to spread propaganda about Syria and Russia (Table 5.6).

Phase	RTs	Time	Text	Cluster	Account Type
D_01	611	6/8/17 16:10	CNN's Christiane Amanpour used Omran's story to spread lies about Russia. Will she now go to #Aleppo for the truth?	blue	Information Activist
D_02	0	6/8/17 16:21	@USEmbassySyria @CNN's @camanpour used Omran's story to spread lies about Russia. Will go to Aleppo for the truth?	blue	Information Activist
D_03	17	6/8/17 20:39	#MSM coverage of #Syria #Aleppo #Omran 'PR support to terrorist organizations'	blue	Media Organization
D_04	168	6/12/17 15:36	Child victims used for Syrian war propaganda (Op-Edge by Eva Bartlett)	blue	Media Organization
D_05	4	6/13/17 23:49	White Helmets filmed #Omran before providing first aid @camanpour #CNNisISIS #NATO	blue	Information Activist

Table 5.6: Exemplar tweets from Phase D

Tweets in Phase D include directed challenges (using the mention function) at journalists and news organizations, in particular at CNN and CNN journalist Christiane Amanpour. This stems from an RT article in which Russian Foreign Ministry spokeswoman Maria Zakharova challenges Christiane Amanpour to travel to Syria and speak to Omran and his family to get the real story[24]. *@InformationActivist1* amplifies this article and the challenge. In a barrage of tweets, one user (not highly retweeted information activist) sends 14 tweets in 13 minutes, each directly targeted at organizations such as the *US Embassy in Syria*, the *UN Human Rights Agency*, *BBC World News*, and the *Democratic National Convention* to disseminate that CNN, and Amanpour specifically, had used Omran to spread lies about Russia.

[24] This in response to a 2016 interview with Russian Foreign Minister Sergey Lavrov in which Amanpour used the picture of Omran to question Russia's support of Damascus in the Syrian conflict.

Throughout Phase D there is a common message that the truth is now out and that the "MSM"—the "mainstream" media—spread lies when reporting the story of Omran. Tweets such as D_05 attack the mainstream media for lying, insisting that Omran was not injured in an airstrike, but offering no alternative explanation. *@21WIRE*'s attack on the mainstream media suggests they '*provide PR support to terrorist organizations*' (D_03), while D_05 implies CNN is a terrorist organization by using the hashtag "*#CNNisISIS*".

Techniques of information operations (micro level)

At the micro level we sought to identify the specific actions that tweet authors engaged in as they participated in the Omran conversation i.e., how tweet authors were involved in information operations. Our approach was inspired by Charmaz's (2014) construction of grounded theory—coding for the actions of the participants and providing an 'insider's view' while avoiding biased analytical insights. This process began with a series of inductive open coding, with three researchers independently listing the actions they felt were present within two samples from our Omran data: 1) a random sample of 50 retweets; and 2) the first 50 tweets temporally from the start of the dataset. Through affinity-diagramming we came to consensus around the specific actions observed. We focus on two specific actions that are salient to the discussion here: introducing uncertainty into the information space, and discrediting information providers. We unpack these in the following sections.

Introducing Uncertainty into the Information Space

In long-standing work from social psychology on rumoring, particularly in the context of crisis events, researchers have found that a lack of information, anxiety, and uncertainty contribute to the spread of rumors (Shibutani, 1966; Bordia & Difonzo, 2004; DiFonzo & Bordia, 2007). During these times, people participate in collective sensemaking that—although it can help reduce uncertainty—also contributes to the development and spread of rumors (Shibutani, 1966; Bordia & Difonzo, 2004). In the Omran conversation, we observe some elements of sensemaking behavior, particularly milling, which includes interpreting, speculating, theorizing, debating, and challenging various explanations of events (Starbird et al., 2016). This activity can be viewed as a collective attempt to resolve uncertainty; however, we also see specific efforts to highlight existing uncertainty and to introduce new uncertainty into the conversation—for example by suggesting we don't know the truth. We found users actively talking about the 'truth'—that it is 'coming soon' or about to be 'revealed'. Around B1, speculative tweets cast doubt on the current understanding of events, suggesting that we don't (currently) know the truth but that soon the truth is coming:

Maybe the world will get to know the truth about Omran!
(2017-06-05 15:21, Blue cluster, Information Activist)

Recovered from WHAT? You'll hear the truth on TV tonight.
(2017-06-05 16:05, Blue cluster, Information Activist)

The real story of Omran in Aleppo with his family. Soon on Syria TV
(2017-06-05 16:16, Blue cluster, Journalist)

In addition to implying that the current narrative is false, these claims also imply that until now there has been an orchestrated cover up that prevented the revelation of the truth, which contributes to the doubt. Post-interview, there is collective satisfaction within the conversation that the truth was revealed by Omran's father:

Truth emerges. New photos of Omran, the boy who became a symbol of Aleppo's suffering
(2017-06-06 13:03, Other cluster, Journalist)

Finally, the truth about Omran
(2017-06-06 02:45, Blue cluster, Information Activist)

Omran: the true story!
(2017-06-08 23:23, Blue cluster, Journalist)

Tweets of the 'true story' emerge after the interviews and continue into Phase D of our data. In Phase D, the notion that the truth is now revealed is 'weaponized' and used to attack the mainstream media—the emergence of the 'truth' proves that the mainstream media had lied. This leads to calls of "fake news" that work to undermine the mainstream media.

Despite frequent references to the 'truth' in tweets, there is little convergence around a single, coherent version of the events. Instead, alternative versions of events are presented that challenge aspects of the mainstream media's narrative (e.g., there was no airstrike) but do not provide a viable alternative explanation. At other times new information is introduced (e.g., Omran and his family are free and living

in Aleppo) but the details surrounding the family's whereabouts since the reported airstrike in 2016 are not provided. Omran is accused of being a "child actor" in one tweet while being labeled a victim of NATO or ISIS propaganda in others. The use of this kind of action does not reflect sensemaking as previously theorized—as a collective-problem solving effort to converge around a shared understanding of what actually happened (Shibutani, 1966; Bordia & Difonzo, 2004)—but rather reflects efforts to cast doubt on the existing narrative, destabilize the information space, and make others question what, if anything, they can believe.

Discrediting Information Providers

Within the Omran conversation we observed numerous actions that can be characterized as attempts to discredit the credibility of institutions (e.g., NATO, UN), large media corporations (e.g., CNN, BBC), and individual journalists. This behavior was not limited to one side of the Omran conversation. While some tweets (originating from the Blue cluster) would call out CNN's "fake news", others (from the Red cluster) would seek to discredit the Syrian journalist that was interviewing Omran on Syrian state television. Most salient in this category is the opposition towards the mainstream media. This anti-media sentiment, concentrated in the Blue cluster, occurs throughout the conversation, but is particularly notable during the sustained anti-media conversation that occurs during Phase D.

Calling Out Fake News and Propaganda

The hashtag #MSM or abbreviation MSM is popular within this conversation. MSM is not simply a succinct reference to the mainstream media, but a label of disrespect, signifying disdain with the large media corporations and a distinct lack of trust in their reporting. In our data, the hashtag MSM was exclusively used in tweets attacking the credibility of the mainstream media, often co-occurring with the tweet's author pointing to what they viewed as "lies":

Omran's Father **Exposes the MSM's lies** which lead to air strikes

(2017-06-06 23:42, Blue cluster, Information Activist)

These attacks become more pointed as organizations and individual journalists are labelled as "fake news media" or "propaganda". This technique—accusing each side of fake news or propaganda—occurs in both the red and blue clusters, as demonstrated below. The yellow cluster (the target of such attacks) was not heavily involved (i.e., there is little, if any, response to these accusations):

#FakeNewsMedia staged event for **Anti-Assad #propaganda**

(2017-06-06 03:43, Blue cluster, Information Activist)

Pro-Damas journalist announces he's met Omran and his father. **Typical propaganda** (2017-06-05 15:08, Other cluster, Journalist)

Targeting Attacks

Targeted attacks, made possible by Twitter's mention (@twitter_user) and reply functions, were conducted at organizations and individuals associated with the media. In a technique that can be seen in other types of online activism, one user single-handedly sent 57 such tweets to news organizations, journalists, and other information providers such as US government agencies and the UK Foreign Secretary, which all sought to discredit the mainstream media, CNN, or Christiane Amanpour. In the following example CNN Journalist Christiane Amanpour's journalistic credentials are challenged:

@camanpour Fake Journalist Amanpour worked with White Helmets Propaganda studio using child actor Omran

(2017-06-13 21, Blue cluster, Information Activist)

The targeted attacks were also used to demand action from journalists and media organizations:

Hey, @BBCNews @euronews @SkyNews @Channel4News **Quickly broadcast the True Story of #Omran**

(2017-06-05 17:52, Blue cluster, Information Activist)

Targeted attacks were most prevalent from the Blue cluster, and we did not see any responses to these attacks from news agencies, journalists, or other information providers in our data. However, these kinds of attacks did also emanate from the Red cluster, with the Syrian journalist who conducted the interview being subjected to strong personal criticisms that aimed to undermine her credibility. The following tweet

was in reference to a selfie that the journalist had apparently posted (prior to our collection period) showing her with dead opposition fighters:

same smile with bodies of prisoners tortured to death and Omran. **Assad's journalist ready to commit atrocities**

(2017-06-05 21:40, Red cluster, Journalist)

Suggesting Unproven Affiliations

Another form of discrediting was to suggest affiliations with institutions or groups with which others may hold negative views, for example NATO, the CIA, and terrorist organizations. This type of attack was used against the mainstream media in particular. Highly retweeted user *@Journalist1* was the first to introduce the term "NATO terrorists" into the Omran conversation:

Omran alive & well in #Aleppo. Father talks of bribes offered by **NATO terrorists.** Airing Syria TV tonight

(@Journalist1 2017-06-05 16:28, Blue cluster)

An hour later, influencer *@GovernmentOfficial1* tried to discredit the mainstream media referring to them as "NATO media", implying that they nothing more than the propaganda arm of the military alliance:

Now that the ultimate Aleppo story of Omran turned out to be false as we expected, **will NATO media** apologize for this terrible mistake?! (@GovernmentOfficial1 6/5/2017 17:52, Blue cluster)

Paradoxically, rather than suggesting that the media are part of a structured military alliance, an alternative approach was to suggest the mainstream media collude with terrorist organizations and support their PR efforts:

MSM coverage of Omran is **PR support to terrorist orgs**

(2017-06-08 19:19, Blue cluster, Media Organization)

Other users took this further. A more succinct way to expunge a mainstream media's credibility is to simply state that mainstream media organizations are terrorist organizations, specifically ISIS:

CNN is ISIS. Skilled in image propaganda for their own purposes

(2017-06-03 18:12, No cluster, Information Activist)

Omran's Father Exposed How the US & White Helmets Lied to the World **#CNNisISIS** #CIAisISIS

(2017-06-08 03:48, Blue cluster, Information Activist)

The technique of associating individual journalists, the mainstream media, and other information providers with propaganda efforts, immoral behaviors, and terrorist groups, is an action we observed by users engaged in the Omran conversation. Tweets making these unsupported claims are generating negative sentiment toward the targeted individual or group and challenging their credentials and (perhaps) instilling uncertainty in the information they provide. Although it remains important to question the source of information, this sort of behavior goes beyond simply doubting the information provider. These appear to be attacks on a more personal and emotional level, specifically targeted at the credibility of the provider with the aim of diminishing trust and introducing further uncertainty.

Discussion

The multi-part story of Omran Daqneesh, characterized by contested facts and disruptive, conflicting narratives, has acute geopolitical significance in a multilateral military conflict. In his initial appearance, his photo, taken in the wake of a family tragedy, was used to garner attention in the West to the human impacts of the war, particularly those caused by the actions of the Syrian government and their primary ally, Russia. Omran therefore became a symbol of suffering at the hands of Syrian President Assad. His reappearance, about a year later, introduced new information—and importantly new alternative narratives—that worked to undermine the initial "mainstream" narrative about the causes of innocent citizens' suffering in Syria. In this work, we consider social media activity around the reappearance of

Omran from a perspective of online "information operations"—investigating the underlying structure of the conversation, the techniques used, on both "sides", to create, promote, defend, and challenge narratives that support geopolitical aims, and the emergent collaborations that facilitate the "work" of information operations.

Information Operations as Collaborative Work in an Online Crowd

From the macro-level, the retweet network graph reveals a highly polarized conversation, characterized by distinct communities or clusters of participants who primarily share messages from within their own groups. A closer look reveals the work of information operations taking shape within heterogeneous assemblages of actors. These assemblages consist of information activists, bloggers and journalists, non-profit organizations (NGOs), government officials, and government-funded media outlets. This work is geographically distributed, including people from within the affected areas (in Syria), western journalists (some of whom travel to the area), and online "volunteers" from around the world. These entities work together to produce, amplify, and spread a preferred set of narratives.

Interestingly, it was the pro-Syrian government voices of the Blue cluster that dominated the conversation as they made up more than half of the participating accounts and content; this included seven of the ten most-retweeted accounts. Within this cluster, accounts of journalists and "alternative" media outlets (including Russian-government funded media outlets) were integrated into a larger community of accounts, including concerned citizens, online activists and diverse political actors. The influential accounts occasionally retweeted each other and shared many common retweeters—whose sustained work to retweet the influential, pro-Syrian government accounts led to the dense and interconnected Blue cluster as represented in Figure 5.1.

It can be productive to view these information operations as a form of collaborative work within an online crowd. This perspective allows us to conceptualize this "work" as not simply coordinated—i.e., there is no central node controlling all of the accounts in these networks—but instead as an assemblage of diverse actors, driven by a variety of motivations, loosely collaborating in the production and propagation of strategic narratives. This view extends previous descriptions of online information operations as perpetrated by armies of automated accounts (or "bots") (Woolley and Howard, 2017) and factories full of paid trolls (Troianovski, 2018), and suggests, at least in this case, a complex and in some ways organic system with emergent properties, similar to other configurations of online volunteerism (Starbird & Palen, 2011; Starbird, 2013) and online activism (Savage & Monroy-Hernández, 2015).

In this graph, we can see that actors and organizations with specific geopolitical agendas, including government-funded media and "cloaked" (Farkas, Schou & Neumayer, 2018) agents of political organizations, are integrated into the online "communities" working to spread conflicting narratives related to Omran's story. This is especially true for the pro-Syrian government (Blue) cluster, where the voices of Russian-government funded media, Syrian government officials, and Western activist-journalists are retweeted by a set of common amplifier accounts that include both automated accounts and sincere online activists. This perspective suggests a strategy of cultivation, rather than purely coordination, and aligns with historical accounts of information operations perpetrated by specifically Russian government campaigns (Lucas & Pomeranzev, 2016; Waltzman, 2017).

The activities of the anti-Syrian government (Red) cluster and the mainstream media (Yellow, Orange and Green) clusters can also be seen through this same lens—i.e., as a form of information operations (broadly interpreted), whereby participants, knowingly or unknowingly, work to spread, support, and defend the western, mainstream media narrative. However, this activity was far less organized, in terms of the structure of collective action (represented in the retweet network graph) than the pro-Syrian government operations in this dataset.

Collaborative Operating: Collective Efforts in Information Operations

In this work we also examined what kind of information operations were present in the Omran conversation, and how. Through our in-depth analysis of actions at the account level, we noted two particularly salient and related complementary techniques that were employed by accounts in the pro-Syrian government blue cluster: assembling "facts" into alternative narratives that could be propagated into the information space; and challenging the credibility of other information sources (e.g. mainstream media organizations and journalists)—undermining integrity and introducing doubt that would serve to make the alternative narratives more acceptable (to some). We unpack these concepts in the following sections.

Assembling "Facts" into Undermining Narratives

As we analyzed the Omran conversation at the tweet level, we identified the presence of contested facts. These are new pieces of information that are introduced to the conversation and then assembled into new narratives, which could be used to challenge the existing narrative. Rogers and Tyushka (Rogers & Tyushka, 2018) outline two approaches to this type of narrative challenge. The first approach is by means of a positive competing counter narrative—a confrontation and replacement of the existing narrative. Another is through an anti-narrative, which, rather than attempting to replace the existing narrative, seeks to undermine it. Within the Omran conversation on Twitter there are persistent challenges to the mainstream

media's narrative, however there is no single counter-narrative that coherently presents and fully explains the story and how it came to be. Rather, we observe the presence of anti-narratives that seek to "replace the established order with disorder" (Rogers & Tyushka, 2018).

These undermining narratives do not offer robust, standalone versions of events that provide closure to the reader. Anti- or undermining narratives are formed by introducing multiple, inconsistent, and often conflicting "facts" or "events" into the information space, which in turn destabilizes the "mainstream" narrative by distorting the information space and confusing the audience (Pomerantsev and Weiss, 2014; Rogers & Tyushka, 2018). This, and other techniques such as suggesting we don't know the truth, align with a known information operations strategy termed "cognitive hacking" (Cybenko, Giani & Thompson, 2002)—which is defined as an attempt to alter the perception of reality through the intentional introduction of false information. Reflecting a tactic theorized to be part of the Russian disinformation apparatus, the intentional introduction of uncertainty serves to distort the information space leading to "muddled thinking" and undermining trust in information providers and the institutions of western democracies (Pomerantsev & Weiss, 2014; Lucas & Nimmo, 2015). Pomerantsev and Weiss (2014) argue that the goal of these kinds of information operations is to reduce trust in institutions, weaken societies, and demotivate political action such as resistance.

Challenging the Credibility of Information Sources

The concerted, collective, and sustained efforts to challenge the credibility of information sources can be seen in this same way—as a form of "information operations" (Pomerantsev & Weiss, 2014; Lucas & Nimmo, 2015; Rogers & Tyushka, 2018). Interestingly, we saw this technique in action on both "sides" of the Omran conversation, as accounts in both the Blue and Red clusters attempted to discredit voices within the opposing cluster by referring to the other as propaganda or suggesting unsavory affiliations between them and other groups. Within the Red cluster, efforts focused around discrediting the interview with Omran's father, describing it as propaganda, questioning the past actions of the pro-Syrian government interviewer, and suggesting that the interview was being conducted under duress. However, the most salient activity of this kind involved the Blue cluster's attempts to discredit individuals and organizations in the "mainstream" media. Within the pro-Syrian government Blue cluster, accounts demonstrated a sustained and collaborative effort to discredit the media—using hashtags such as #MSM, sending targeted tweets to directly confront media accounts and garner attention, and suggesting affiliations with controversial organizations such as NATO, the CIA, and ISIS (#CNNisISIS). Through these actions, the crowd in the Blue cluster both reflected and supported a diminished trust in the "mainstream" media, which, over time, would leave the door open for the introduction of alternative narratives originating from other sources

Limitations and future work

This study of information operations online focused on the Twitter conversation of one specific case study. Although the data is from a larger collection, the case study presented here is just a snapshot of that, and specifically a period of time when the counter-narrative was ascendant (due to Omran's reappearance). We acknowledge that the network may have looked much different (for example, the red cluster might have been bigger and higher volume) if we had similar data from the weeks and months immediately following the initial photograph of Omran in August 2016. Future work is needed to apply this lens to other information operations online, including more case studies in different contexts, and using data from other platforms to build a better understanding of information operations. In particular, the development of methods and models that help us to differentiate between coordinated and organic, emergent information operations is a critical area for future research.

Conclusion

In this research we investigate information operations on social media by studying the conversation and narratives surrounding a specific episode of the Syrian conflict—the story of Omran Daqneesh (the 'Aleppo Boy')—through a CSCW lens. We use an iterative mixed-methods approach to understand who was participating in these conversations, what they were doing, and how they were doing it. We uncover networks of clustered users that contribute to a contested and politicized information space surrounding Omran's story; the presence of undermining narratives that serve to disrupt the mainstream media's narrative and confuse the audience; and techniques such as introducing uncertainty into the information space and challenging the credibility of information providers as tactics for producing and propagating the undermining narratives. In addition to illustrating what information operations are and how they play out on social media, we contribute an improved understanding of online activism as a vehicle for information operations whereby groups of diverse users—including government agents, citizen journalists, and online activists—form emergent collaborations to engage in the cooperative work of conducting information operations. Future work is needed to examine why these online activist communities are functioning in this way, including whether they are specifically cultivated or co-opted for the purpose of information operations.

<<END OF PUBLICATION>>

Chapter 6:

Examining The White Helmets Conversation on Twitter (Study 2)

Introduction

The aim of Chapter 6 is to provide the reader with detail about the WH conversation on Twitter. As presented, the work is a derivative of the Methodological Appendix to *Cross-platform disinformation campaigns: lessons learned and next steps*, which was published in the inaugural edition of The Harvard Kennedy School (HKS) Misinformation Review[25]. I have restructured that published document here and added additional insights, including a domain-retweet network visualization, which sets up of Chapter 7 (the presence of an alternative media echo-system) and the multi-platform analysis that will be presented in Chapter 8 by illustrating links to other social media platforms including YouTube.

The analyses presented in this chapter represent the long-form, detailed investigation of the WH conversation on Twitter (and elsewhere) over the course of more than three years. This work was conducted by myself and multiple researchers, notably my academic advisor Associate Professor Kate Starbird, PhD

[25] This Chapter is a derivative of work previously published in The Harvard Kennedy School Misinformation Review. Please cite the original work as well as this dissertation:
Wilson, T. & Starbird, K. (2020). Cross-platform disinformation campaigns: lessons learned and next steps. The Harvard Kennedy School (HKS) Misinformation Review.

peer Ahmer Arif, and graduate and undergraduate students that contributed to the research efforts through Directed Research Groups[26].

In the following sections I lay out our findings. Broadly speaking, we identified (using the methods described below) two distinct communities involved in the WH conversation on Twitter: a community of accounts that were supportive of the group, and; a community of accounts that were highly critical of the group. These communities (that are further characterized below) became a basis for further investigation including a temporal analysis of Twitter activity over time, a qualitative examination of tweets, and insights into their internal structure.

In additional analyses that did not appear in the original publication, we leveraged the URLs within tweets to look beyond Twitter to the wider information ecosystem of websites linked to in the Twitter conversation. This analysis highlights the extensive links to both Russian government-controlled and alternative media from the community of accounts that were highly critical of the WH, which will be covered in further detail in Chapter 7. It also prepares the reader for the research presented in Chapter 8, in which I look at the role of other social media platforms in multi-platform information operations.

Data

This analysis presented in this chapter are based on a Twitter dataset of English-language tweets referencing the "`white helmets`" or "`whitehelmets`", collected in real-time using the Twitter streaming API between May 27, 2017 and June 15, 2018. The collection was case insensitive. Due to affordances of the Twitter API, it includes tweets and retweets where the term "`white helmets`" or "`whitehelmets`" appears in the *text* field of the tweet or in the case of 'quote' tweets (retweets with comment), within the *text* of the embedded quote tweet.

These data—and therefore the subsequent analyses—are limited to English-language tweets. This limitation is in part due to language constraints in our research team. However, it also encapsulates a specific and strategic element of the White Helmets discourse: the group has made significant and sustained efforts to communicate in English to western and global audiences, while similarly a significant portion of the anti-White Helmets campaign has occurred within English-language discourse, attempting to counter the White

[26] Directed Research Groups (DRG) are a research-for-credit program that offers graduate and undergraduate students opportunities to do hands-on research at the University of Washington: https://www.hcde.washington.edu/research/directed

Helmets' communications by targeting the same or similar audiences. This research therefore focuses on the aspects of the White Helmets discourse that were designed for—and taken up by—English-speaking audiences.

Methods

We adopt a mixed-methods approach, extending methodologies developed from the interdisciplinary field of crisis informatics (Palen & Anderson, 2016) and prior studies of misinformation (Maddock et al., 2015). Our research is iterative and involves a combination of quantitative methods, for example to uncover the underlying structure of the online discourse and information flows (e.g., the network visualization in Figure 6.1), and qualitative methods to develop a deeper understanding of the content and accounts that are involved in seeding and propagating competing strategic narratives about the White Helmets. More specific methods and techniques are described in-line with the findings presented within the chapter.

Our methodological approach is also iterative, meaning that we apply both qualitative and quantitative lenses and move back and forth across different analyses—of different types and at different scales—as we develop an understanding of the phenomena of interest. In what follows, we have attempted to describe these analyses in a way that is easiest for readers to understand, but this ordering does not reflect a linear progression across these analyses.

Overview of the WH conversation

The White Helmets Twitter dataset contains 913,028 English-language tweets from 218,302 distinct Twitter accounts. 77.6% of tweets (708,334) in the dataset are retweets, and 33.2% (303,104) of all tweets contain a URL linking to a web domain outside of Twitter. This information will be used to infer the structure of the Twitter conversation and the surrounding ecosystem.

Distinct accounts	total number of (re)tweets	tweets	retweets	% retweets	% url
218,302	913,028	204,694	708,334	77.58	33.20

Table 6.1: A table summarizing the White Helmets Twitter dataset

Figure 6.1 is a temporal plot of the volume of English-language tweets per day for the duration of the dataset. In general, we can see that the volume of tweets per day is between 1,000-7,000 until April 2018,

when the daily volume of tweets increases substantially as efforts to delegitimize the WH intensified to discredit their reports of a chemical weapons attack.. Another important area of the graph is the lack of data in February 2018. Between 7 February 2018 and 24 February 2018, our collection infrastructure suffered an outage, and we are therefore unable to account for activity at that time. Unfortunately, this data loss occurred at a somewhat active time in the White Helmets conversation, during an escalation in the conflict in eastern Ghouta and leading up to the Oscar Awards ceremony where the documentary about the White Helmets was due to receive an award.

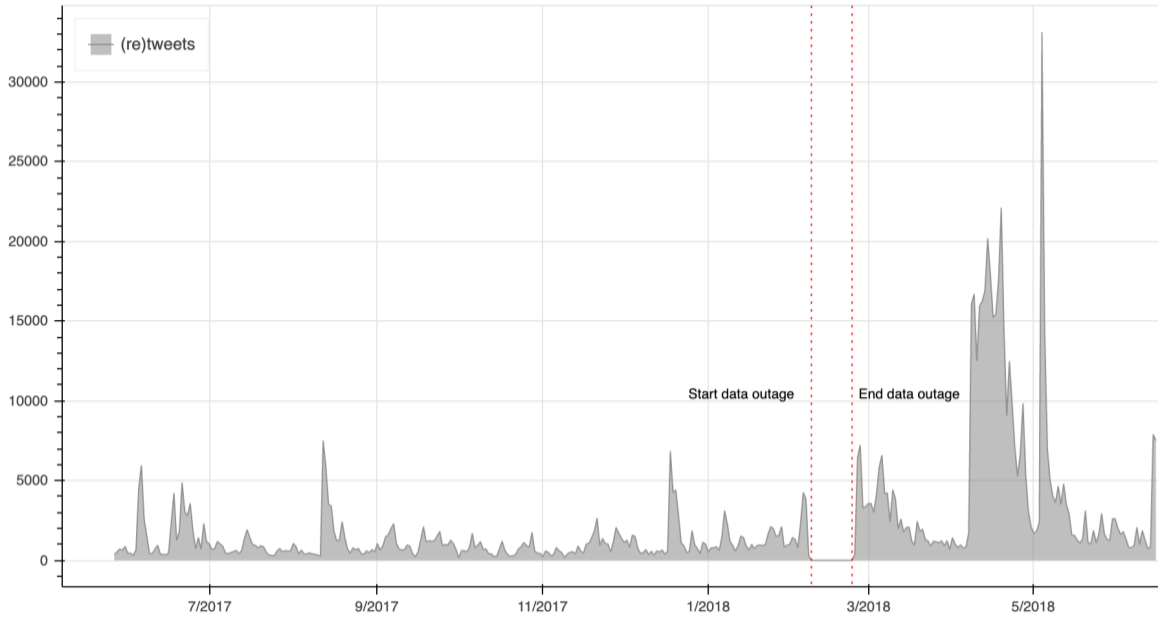

Figure 6.1: Temporal plot showing the number of (re)tweets per 24 hour period in the WH conversation on Twitter. Note the data outage in February 2018.

Retweet Network Visualization

To understand the underlying structure of the Twitter discourse surrounding the White Helmets, we generated a composite network visualization that infers relationships between accounts by identifying retweet trajectories. This visualization (Figure 6.2) shows both who the influential accounts are and how those accounts and domains are connected to each other through information-sharing patterns. In this graph, nodes (circles) represent accounts. Nodes are sized relative to the number of retweets that account received within the White Helmets Twitter dataset, providing some indication of the account's relative visibility within or influence on the conversation. Nodes are connected to other accounts via edges (lines) when one account retweets another. The edges become thicker relative to the number of times one account retweeted another.

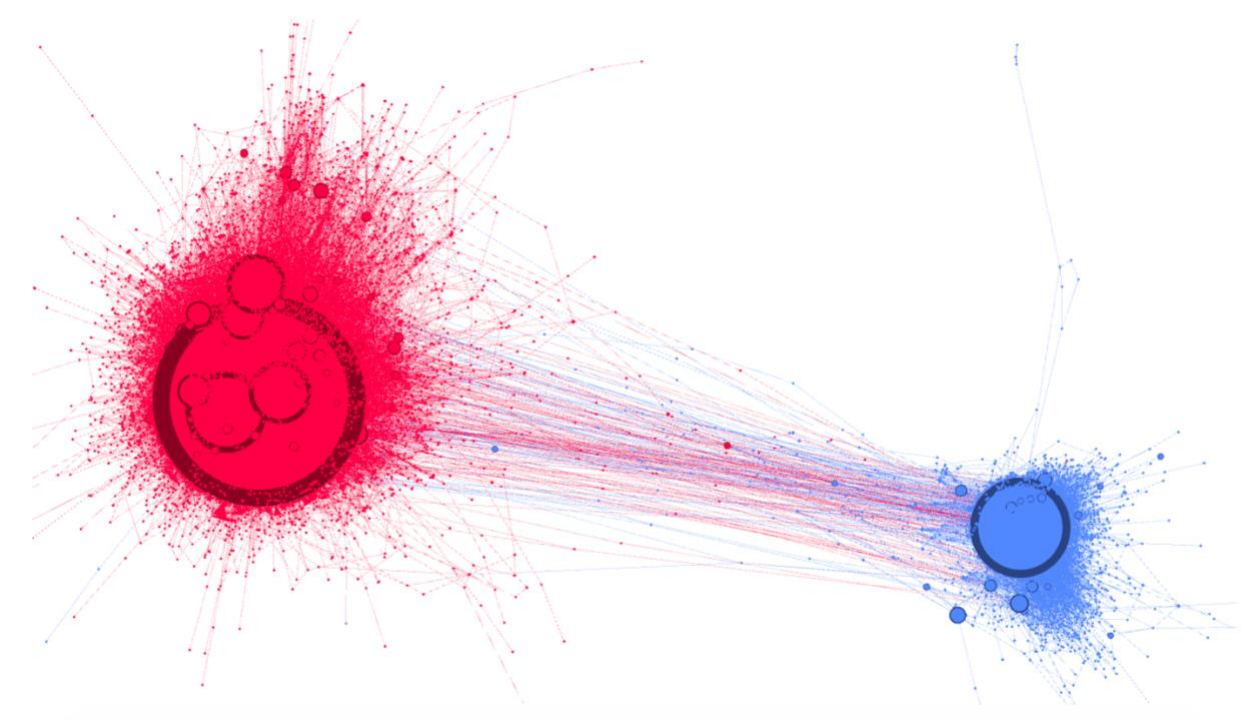

Figure 6.2: A retweet network visualization of the White Helmets conversation on Twitter illustrating the two distinct communities involved in the conversation: the anti-WH community (red, on the left), and the pro-WH community (blue, on the right). The nodes (circles) represent Twitter accounts which are sized relatively according to the number of times that account was retweeted. The edges (lines) between nodes represent a retweet between that pair of accounts.

We then applied the Louvain algorithm (Blondel et al., 2008) to detect communities of accounts. The Lovain algorithm is a greedy optimization method that optimizes modularity (the relative density of edges inside a community) compared to the modularity outside the community (Blondel et al., 2008). Since in our graph the edges represent retweets between a pair of accounts, the algorithm is effectively identifying accounts that retweet one another. This process revealed two main communities, which we colored Red and Blue (Figure 6.2). Of the 218,302 distinct Twitter accounts in the *white helmets dataset* as described above, 168,110 (77%) were classified as members of the red or blue communities (Table 6.2). The remaining (50,192 accounts) accounts were either:

- not present in the retweet network because the accounts did not retweet nor were retweeted by another account in the visualization (32,638 accounts), or;
- not within the two core communities (17,554 accounts retweeted others in the graph, but formed their own, smaller communities distinct from Red and Blue).

Qualitative analysis of the tweets in the smaller communities revealed that the conversation was about a motorcycle club (also called the White Helmets), and the headgear worn by a US football team. These accounts were removed from subsequent analysis as they were irrelevant to the conversation under examination here.

Accounts in the visualization were spatially arranged using *ForceAtlas2* (Jacomy et al., 2014), a force-directed layout algorithm that pulls accounts with edges (retweets between one account and another) closer together and pushes accounts without edges (i.e., that didn't retweet one another) further apart.

community	distinct accounts	total tweets	tweets	retweets	% retweets	% url	% replies	mean num tweets per account
blue, pro-WH	70,670	207,938	27,708	180,230	86.67	22.72	2.42	2.94
red, anti-WH	97,440	632,638	122,966	509,672	80.56	36.96	9.20	6.49

Table 6.2: Overview of the blue (pro-WH) and red (anti-WH) communities on Twitter

Table 6.2 provides a high-level overview of the red and blue communities, including the number of distinct accounts, the number of tweets and retweets, and the number of tweets and retweets that contain a URL. The red community is larger, in that it contains almost 27,000 more distinct accounts than the blue community. The accounts in the red community also play an outsized role when compared to that of the blue community: it contains 57.96% of the accounts, but it generated 75.26% of the (re)tweets in the WH conversation. On average, each account in the blue community generated 3 (re)tweets, while each account in the red community generated 6.5 (re)tweets. Also noteworthy is that 23% of the tweets from the blue community contain a URL, while for the red community this is 37%. This suggests that accounts in the red community include more content from external sources in their tweets.

In the following analyses, the focus is exclusively on accounts and activities within the Red and Blue communities. More information about the communities, including their stance toward the WH, will be presented in the following sections.

Temporal Signature by Community

Figure 6.3 contains the temporal signatures (the volume of (re)tweets[27] per day) for the Blue and Red communities, respectively. The graph reveals a period of data loss in February 2018, which as previously noted occurred due to an issue with our collection infrastructure.

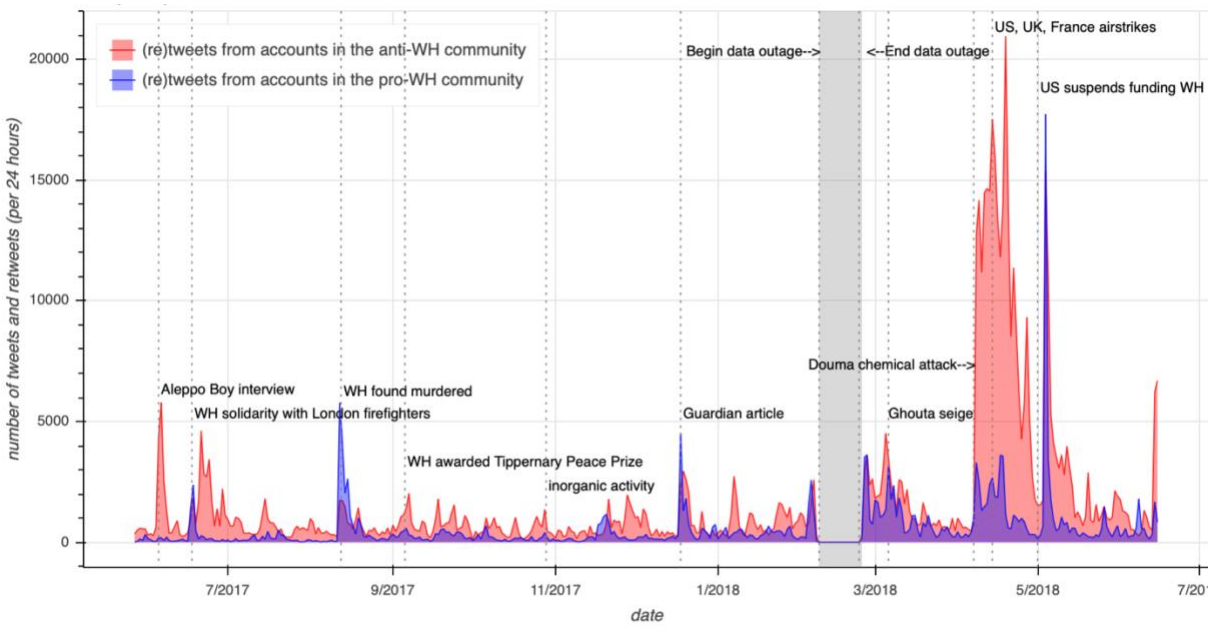

Figure 6.3: Temporal plot showing the number of (re)tweets per 24 hour period from the red and blue communities on Twitter. The plot is annotated with some key events regarding the WH and the Syrian conflict. Note the marked increase in activity—particularly from the red anti-WH community—after March 2018, and the data outage (grey area) in February 2018.

This plot is annotated with some key events related to the White Helmets that took place during this period, including:

- June 2017: The reemergence of Omran Daqneesh "The Aleppo Boy" with his father in a broadcast interview with a pro-Syrian TV journalist (see Chapter 5 of this dissertation);
- June 2017: WH send a message of solidarity to London firefighters after the Grenfell Tower fire;
- August 2017: The murder of seven members of the WH in their office;

[27] I use the term *(re)tweets* to mean non-retweets or retweets—in other words all tweets.

- September 2017: The WH are awarded the Tipperary International Peace Award[28] for humanitarian work.
- December 2017: Olivia Solon's article in the Guardian (UK newspaper) outlining the disinformation campaign against the White Helmets, titled "How Syria's White Helmets became victims of an online propaganda machine"

It also shows an increase in overall volume (across both communities) around that time and then again in April 2018. The first rise coincides with an escalation of violence as Syrian regime forces made advances in the rebel-held region of Ghouta. The second rise begins on 7 April 2018 with initial reports of a chemical gas attack in Douma, a major city in the Ghouta region, and continues into May and—for the Red community—through the end of our collection period. Comparing these graphs reveals the Blue community to consist of relatively low volume tweeting with occasional spikes, while the Red community's activities are more persistent over time.

This suggests that the bulk of activity in the Blue community is the result of sporadic attention from a more "mainstream" audience reacting to content shared by news media and social media influencers with large follower counts. Meanwhile, the activity in the Red is largely due to sustained and persistent activities of a group of activists, journalists, and 'alternative' or 'independent' media outlets whose influencers have relatively smaller follower counts (than mainstream media) but are more dedicated to tweeting about the White Helmets.

Focusing on the window of activity in April 2018 that began in the wake of the chemical attack in Douma, we can see the first tweets from our data that reference the attack appear in the Blue community, as members of the White Helmets were among the first to report the attack. But the Red community activates shortly afterwards (about four hours later), both to directly challenge the reports of a chemical attack and to share other types of content disparaging the White Helmets—both to undermine the legitimacy of their reporting and possibly also to distract from the news of the chemical weapons attack. This content in the Red community soon surpasses the content in the Blue and persists for weeks. 41% of all tweets posted by the Red community in the White Helmets Twitter dataset occurred in April 2018. Only 18% of the tweets from the Blue community were posted during that time. Eventually content from the Red community would converge on a set of narratives that disputes the nature and cause of the attack in Douma.

[28] The Tipperary Peace Award was created by residents of the Irish town to counter the association between their town and war, an association that was created by the song "*It's a long way to Tipperary*". The annual award is given for humanitarian work.

Another interesting feature in the temporal signature of the Red community is a large, narrow spike in late October 2017 that appears to be inorganic (due to how it rises so quickly and does not demonstrate the typical pattern of exponential decay). Looking more closely, we find that during a four-hour period on 28 October, 994 distinct accounts posted 1070 tweets. Almost all were retweets of just two tweets, part of a 'thread' (of a tweet and its replies) challenging the legitimacy of a recently published report that had concluded that the Assad regime was responsible for a chemical weapons attack in Khan Sheikhoun in April 2017. More than 50% of the accounts participating in this spike have now been suspended by Twitter—a much higher rate than average (this is further discussed below). Considering the inorganic temporal signature and the mass suspensions of participating accounts, it is likely that this spike is the result of coordinated, and likely automated, activity.

Content and Narrative Analysis by Community

This quantitative approach to community detection was supported by our qualitative analyses of hundreds of tweets, articles, and videos by multiple researchers over the course of several years. This long-form analysis enabled us to characterize the communities based upon their narrative stance toward the WH. We conducted qualitative content analysis to surface the salient narratives and determine the underlying stance of each community toward the White Helmets. This analysis included reading and manually analyzing hundreds of tweets—and dozens of articles and videos linked-to from those tweets—posted by accounts in the two communities. We identified exemplar tweets, organized those into groups representing salient narratives, and wrote memos describing those narratives.

This analysis revealed that the content from the blue community to be almost exclusively supportive of the WH, and the content from the red community to be almost exclusively critical of the WH. Below, we provide a sample of content from each community to demonstrate these distinctions and to surface some of the most salient narratives. To capture the discourse from different perspectives, for each community we present three samples:

(1) five random tweets from the most retweeted account in the community: @SyriaCivilDef (the official account of the White Helmets) from the Blue and @VanessaBeeley (an active and influential journalist) from the Red;
(2) a random sample of ten tweets selected from all tweets in the community; and
(3) the five most retweeted tweets in the community.

For each example tweet we include the number of followers of the authoring account and denote whether the account is active or suspended (as of December 2019). These samples are presented here to illustrate the types of tweets that we included in our content analysis, but they only represent a sample of the analysis.

The Blue (pro-WH) Community: Supportive of the White Helmets

In this section we present three samples of tweets from the Blue community with summary analysis accompanying each sample. Across the samples, we find that content from the Blue community is consistently supportive of the White Helmets.

Tweets from the @SyriaCivilDef (the White Helmets)

@SyriaCivilDef is the official, English-language account of the White Helmets. They are the most retweeted account and among the most active accounts (in terms of "white helmets" tweeting) in the Blue community. 14.3% of the total tweet volume in the Blue community are tweets and retweets of this account. They posted 444 "white helmets" total tweets during our collection period. 42.8% of those were original tweets (non-retweets). Below are five original tweets selected randomly from that set:

@SyriaCivilDef | Active | 106452 followers | 2017-12-30

5 civilian martyrs so far and many injuries since morning, the eastern neighborhoods of #Ghouta are subjected to air raids & heavy artillery shelling by the regime forces.
#WhiteHelmets #CivilDefense <*embedded video of fire across a horizon in the aftermath of bombing*>
[Retweeted 73 times]

@SyriaCivilDef | Active | 107084 followers | 2018-01-23

2 civilians were killed after artillery shelling targeted #Jisr_alshoughor city in #Idlib countryside, #Civil_Defense teams continue the rescue operation. #White_Helmets #With_Idlib 23.Jan <*embedded images of White Helmets volunteers recovering the bodies of victims*>
[Retweeted 45 times]

@SyriaCivilDef | Active | 117880 followers | 2018-03-06

White Helmets chief warns Eastern Ghouta worse than Aleppo <*link to article in thenational.ae titles "White Helmets chief warns Eastern Ghouta worse than Aleppo"*> via @anonymized @TheNationalUAE

[Retweeted 123 times]

@SyriaCivilDef | Active | 119451 followers | 2018-03-14

Having worked for 18 hours sifting through the massive destruction left by intensive Russian air raids on #Idlib city Tuesday, #WhiteHelmets teams ended their rescue operation. The confirmed civilian toll was 3 rescued alive and 9 dead. <*embedded video of White Helmets volunteers working to provide rescue and medical aid in the aftermath of air raids*>

[Retweeted 211 times]

@SyriaCivilDef | Active | 126532 followers | 2018-05-25

White Helmets firefighter teams managed to extinguish massive fires, that broke out after surface to surface missile carrying community bombs hit civilians farms south of town of #KhanShaykhun in #Idlib countryside, and were controlled without casualties. <*image of a bloodied victim*>

[Retweeted 121 times]

These example tweets demonstrate how the White Helmets used their English-language Twitter account to promote the humanitarian work that their organization was doing (firefighting, rescue, victim recovery) and to document civilian casualties of the civil war in Syria—including specifically calling out the impacts of the Syrian regime's shelling and Russian air raids. Many of @SyriaCivilDef's tweets include embedded images and videos aligned with the textual content in the tweet. In particular, 17.3% of original tweets sent

by @SyriaCivilDef contained an embedded video using Twitter's native embedding feature. Only 1.6% of their original tweets contained a URL link to a video on YouTube.

Random tweets from the Blue community

Below are ten original tweets (non-retweets) randomly selected from the Blue community. They demonstrate consistent support for the White Helmets across other accounts in that community.

@anonymizedBlue | Active | ~200 followers | November 2017

7th Nov: Upcoming Screening of the documentary "The White Helmets", with Q&A <*link to a webpage with information about the event*> #ESRCFestival @anonymizedBlue @anonymizedBlue

[0 retweets]

@anonymizedBlue | Active | ~500 followers | December 2017

"The way the Russian propaganda machine has targeted the #WhiteHelmets … case study in the prevailing information wars. It exposes just how rumours, conspiracy theories and half-truths bubble to the top of YouTube, Google and Twitter search al's" <*embedded link to article in the Guardian titled: "How Syria's White Helmets became victims of an online propaganda machine."*> #Syria

[0 retweets]

@anonymizedBlue | Active | ~2000 followers | December 2017

How Syria's White Helmets became victims of an online propaganda machine <*link to article in the Guardian titled: "How Syria's White Helmets became victims of an online propaganda machine."*> @anonymized @anonymized @anonymized @cbc

[2 retweets]

@anonymizedBlue | Suspended | ~6000 followers | January 2017

The White Helmets Several casualties among civilians across besieged #EasternGhouta, victims of aerial and ground bombardment campaign against the families. #CivilDefense teams are responding rapidly to distress calls and urgently assisting the wounded <*embedded content whose nature is unclear due to suspended status*>

[1 retweet]

> **mentioned tweet: @anonymizedRed (January 2018):** The same pit from the explosion, only later did they showed the funnel (from the explosion). #WhiteHelmets have done a lot of fake reporting recently. I have very little trust in this report. Lots of propaganda and #FakeNews.

@anonymizedBlue | Active | <100 followers | January 2018

@anonymizedRed @anonymizedBlue @anonymizedBlue Propaganda is the claim that the White Helmets aren't real rescuers and are a "legitimate target". Because they rescue people in rebel-held areas, and because they show the world these war crimes against civilians, so they are a threat to the regime who claims to only target “terrorists”.

[0 retweets]

> **mentioned tweet: @SyriaCivilDef (2018-02-01):** 3 civilians killed and many injured, including suffocation cases among children, women and #CivilDefense volunteers! 28 air raids, 70 surface-to-surface missiles, some loaded w/chlorine gas and 200+ artillery shells targeted cities in #Eastern_Ghouta today! #Doumasuffocating

@anonymizedBlue | Active | ~500 followers | February 2018

@SyriaCivilDef Absolute savagery. My heart goes out to everyone living in such horror. God bless the White Helmets.

[0 retweets]

@anonymizedBlue | Active | ~1000 followers | March 2018

Very serious, can anyone give me a list of Al Qaeda victims in the world by age? I hate them, but don't they deliberately avoid targeting children? Those figures might be a big wake up call #bonkersBeeley @VanessaBeeley #whitehelmets #Syria <*embedded image of a tweet from @VanessaBeeley claiming that content from the White Helmets does not include the context of the role of "Western backed terrorists"*>

[0 retweets]

@anonymizedBlue | Active | ~3000 followers | March 2018

.#Syria: Graduation for #WhiteHelmets trainees in #Idlib. These volunteers will make the world a better place. God bless them. #BraveHearts <*embedded photograph of 5 existing White Helmets volunteers with ~20 new recruits holding certificates*>

[0 retweets]

@anonymizedBlue | Active | <100 followers | March 2018

.@anonymizedBlue Today another Double-Tap attack on a School in Idlib. They bombed the school, the kids went to shelter, & they bombed the shelter. 16 Kids Dead, Youngest 3. White Helmets identified the plane as Russian Su34. RUSSIA IS MAIN AGGRESSOR IN SYRIA! @anonymized @anonymized

> **quoted tweet: @SyriaCivilDef (2018-03-21)** At 10:07 today 21 Mar 18 a Russian Su34 aircraft dropped a bomb on a school in Kafr Batikh (near Saraqueb, Idlib). The children evacuated to a nearby underground shelter which was bombed in a second strike minutes later. 16 children were killed, together with 3 women and 1 man. <embedded images of child victims>

[0 retweets]

@anonymizedBlue | Active | ~4000 followers | April 2018

@SyriaCivilDef @SyriaCivilDefe @SFUDialogue @anonymizedBlue @SFUint @UKinCanada @anonymized PeaceTalk with the White Helmets (@SyriaCivilDef). Darren Schemmer, E.D. of SFU International (@SFUint), introduces WH Volunteer <anonymized>. Simon Fraser University, Centre for Dialogue (@SFUDialogue). | #WHPeaceTalk #vanpoli #cdnpoli #WhiteHelmets #peacetalks2018 #vancouver <*embedded video of a panel conversation*>

[2 retweets]

@anonymizedBlue | Active | <100 followers | April 2018

Everyone needs to know this! There is a GENOCIDE happening! Children are being brutally killed by poisonous gas by PUTIN's best friend, Assad

WARNING: These images are horrifying.

#Syrian_crisis is a GENOCIDE. #Douma #EastGhouta

HELP: Please donate to The White Helmets @SyriaCivilDef

> **quoted tweet: @SyriaCivilDef (2018-04-07):** Entire families in shelters gassed to death in #Douma #EastGhouta hiding in their cellars, suffocated from the poisonous gas bringing the initial death toll to more than 40. @SyriaCivilDefe is still in the process of rescue and recovery. <embedded images of victims of the alleged chemical attack in Douma, Syria>

[3 retweets]

These random tweets from the Blue community demonstrate an underlying sentiment of support for the White Helmets. They also reflect some of the salient themes that emerged in that community.

Two of the randomly selected tweets are ‘quotes’ of tweets posted by the White Helmets official account (*@SyriaCivilDef*). These quote tweets function to pass along the White Helmet’s content, often with a

comment of affirmation or support. One quote tweet was directed (via an @mention) to the Deputy Prime Minister of Ireland. Another included a call for donations.

Another tweet in the sample is a supportive reply to a White Helmets' tweet featuring images of impacts from air raids and artillery attacks from regime forces on Syrian civilians. There is also a reply challenging a tweet (from the red community) that claimed the White Helmets propagate "fake news" and propaganda.

Three of the randomly selected tweets explicitly reference the campaign against the White Helmets, calling it "propaganda". Two of these tweets link to an article titled "*How Syria's White Helmets became victims of an online propaganda machine*" written by Olivia Solon and published in the Guardian on 18 December 2017. That article catalyzed a small burst of activity in the Blue community immediately following its publication (see Figure 7.3) and continued to shape some of the conversation there for weeks afterwards. This random sample also included advertisements for events featuring content about the White Helmets and a tweet celebrating the graduation of a new cohort of White Helmets volunteers.

Most-retweeted tweets from the Blue community

Below are the five most retweeted tweets in the Blue community, ordered by the total number of retweets they received in the White Helmets Twitter dataset. We include a short description of each to help contextualize the content and author.

@oliviasolon | Active | 32938 followers | posted: 2018-04-17

1. Cardiologist alleges White Helmets faked photos because EEG pads are misplaced: 12,000 retweets

2. Cardiologist admits he was mistaken & the EEG pads are placed correctly: 22 retweets <*embedded images of two screen-captured tweets, arranged side by side, illustrating the "mistake" and its correction*>

[Retweeted 4652 times]

Olivia Solon is a journalist for the Guardian who wrote an article in December 2017 about how the White Helmets were being targeted by a disinformation campaign (Solon, 2017). Her most highly retweeted tweet in our data did not feature her article, but came months afterwards, in the aftermath of the Douma chemical attack. It contains an embedded image, a screenshot of a popular tweet from the Red community, where a

cardiologist claimed that a photo of injured children taken by the White Helmets had been staged. The cardiologist explicitly mentioned (within his tweet) influencers in the Red community, who then retweeted his tweet, setting off cascades among their audiences. Later, after substantial directed criticism and refutation, the cardiologist corrected his original claim. Solon's tweet notes how the misinformation spread much further than the correction. Her tweet, making an example of this case and implicitly providing evidence to support the argument she made in her article published four months prior, received 4757 retweets.

@krassenstein | Suspended | 419991 followers | posted: 2018-05-04

Trump's State Department just froze funding for the famed "White Helmets," who served as emergency responders in the Syria.
-Bomb the country
-Refuse to allow refugees to enter ours
-Defund the primary group saving lives.
"Evil" is what you call that.
[Retweeted 3789 times]

This tweet criticizes President Trump for freezing funding to the White Helmets. It was posted by a social media influencer who produced content consistently critical of Donald Trump (across a wide range of topics). It is the only tweet that *@krassenstein* posted about the "White Helmets". This account had 419,991 followers at the time of this tweet, which is likely why this tweet is among the most retweeted tweets in the blue community (3789 retweets). Months later, this account was suspended for violating Twitter's terms of service related to manipulating the platform to grow its audience.

@Channel4News | Active | 1569575 followers | 2017-06-17

This is the message of solidarity to Britain's firefighters from Syria's White Helmets. <*embedded video of two White Helmets volunteers, standing in front of a destroyed building, displaying a series of signs sharing an English-language message of solidarity for firefighters responding to the Grenfell Tower fire*>
[Retweeted 2658 times]

@Channel4News is a broadcast media outlet in the United Kingdom. In this tweet, it features a video of members of the White Helmets holding up a series of messages (on white poster board) of solidarity for firefighters in Britain who were responding to the catastrophic Grenfell Tower fire. This is an example of how the White Helmets attempted to—and in this case managed to—garner international attention for their organization and its work. The *@Channel4News* account has over 1.5 million followers, which is likely why it was so highly retweeted here. In our data, this video (and a Channel4News article featuring it) was heavily criticized by accounts in the Red community, who claimed that the White Helmets were more focused on seeking attention than on doing humanitarian work.

@ajplus | Active | 864446 followers | 2018-02-25

The White Helmets were trying to save victims of an airstrike in eastern Ghouta, Syria.

They called for backup ... and then another bomb hit. <*embedded video of a news report from the Al Jazeera Plus outlet, featuring the work of the White Helmets*>

[Retweeted 2560 times]

This tweet, authored by an account associated with the Al Jazeera news network, describes a "double-tap" airstrike—where Syrian and Russian airplanes would bomb a location twice, with the second strike allegedly perpetrated with the specific goal of targeting first responders. It includes an embedded video with a news report on the same topic and was retweeted 2560 times. This tweet again reflects the White Helmets receiving attention in the international news media.

@mitchellreports | Active | 1618778 followers | 2018-02-26

Invaluable reporting on Assad's use of banned chemicals against civilians (any use is a war crime). Enabled by Russia. Anyone listening in DC?

> **quoted tweet: @60minutes (2018-02-25):** White Helmets had to load victims into a truck and drive them 30 miles to one of the nearest surviving hospitals. Doctors tried everything, but were unable to save many. Dr. Mamoun Morad told us the scene "was like Judgement Day, the Apocalypse." #60Minutes <embedded video of news report from the 60 Minutes television program>

[Retweeted 1508 times]

This tweet, posted by an NBC news anchor, uses the quote tweet functionality to pass along a post and video news report from the 60 Minutes television program documenting how the Assad regime—"enabled by Russia"—has used chemical weapons against civilians in Syria. The large number of retweets (1508) was perhaps due to the relatively large number of followers that this account has.

Reflecting the nature of social media influence, the most retweeted tweets in the blue community were originally authored by high-follower accounts of journalists and social media influencers. These examples demonstrate that the White Helmets were at least somewhat successful at garnering attention from international audiences through coverage by "mainstream" news media.

Summary of narratives from the Blue (pro-White Helmets) community

From the examination and qualitative analysis of tweets we surfaced about a dozen narratives that take shape and propagate within tweets in the blue community of the WH conversation on Twitter. The examples above illustrate many of the most salient, including:

- The White Helmets are a humanitarian response organization that provides firefighting, medical aid, rescue, response, and body recovery in rebel-held areas of Syria.
- Military actions taken by the Syrian/Assad regime are injuring and killing Syrian civilians.
- The Syrian/Assad regime and their allies specifically target humanitarian responders.
- The Syrian/Assad regime uses chemical weapons against Syrian civilians.
- Russian forces allied with the Syrian/Assad regime are participating in airstrikes that target civilians and humanitarian responders.
- The White Helmets are the target of a disinformation campaign.

The Red (anti-WH) Community: Critical of the White Helmets

In this section we present three samples of tweets from the Red community with summary analysis accompanying each sample. Across the samples, we find that content from the Red community is consistently critical of the WH.

Tweets from influential Red-community journalist @VanessaBeeley

The most retweeted account in the Red community—and in the entire dataset—is *@VanessaBeeley*, a highly active tweeter and journalist whose tweets and articles are consistently critical of the White Helmets.

@VanessaBeeley posted 2043 tweets, including 937 original tweets, and was retweeted 63,380 times in the WH Twitter dataset. Taken together, her content production (her tweets and retweets of her account) constitutes 10% of all activity in the Red community. Below are a sample of five of their original tweets. We include a description of each for context.

@VanessaBeeley | Active | 15941 followers | 2017-06-22

Exactly! Cant keep allegedly sacking people for murder, wont be any left.

> **quoted tweet: @anonymizedRed (2017-06-22):** How can it be a smear campaign when the #WhiteHelmets are constantly_apologizing for murdering people? @VanessaBeeley @anonymizedRed <*link unrecoverable because tweet has been deleted*>

[18 retweets]

This tweet (and the one below) are in reference to an incident in June 2017 where a White Helmets volunteer was captured on video participating in executing two soldiers of the Syrian regime. The volunteer was later expelled from the Syria Civil Defence organization for a "gross breach of SCD's Code of Conduct." Accounts in the Red community (and media aligned with anti-White Helmets sentiments) seized on this incident to paint the organization as aligned with terrorists.

In the above tweet, *@VanessaBeeley* quotes a tweet from another account in the Red community. The original tweet uses the incident as evidence to deny that anti-White Helmets discourse constitutes a "smear campaign". The original tweeter 'tags' (or 'mentions') *@VanessaBeeley* in their tweet, attempting to draw her attention to it through in-platform notifications. *@VanessaBeeley* does attend to it and retweets/quotes it with a comment of support.

> **mentioned tweet: @KreaseChan (2017-06-21):** The #WhiteHelmets at least have some level of credible accountability..rare in Syria where impunity rules. Another reason Assad targets them <*embedded image of a text excerpt from the Syria Civil Defense (SCD: the White Helmets organization) describing the expulsion of a volunteer for a "gross breach of SCD's Code of Conduct"*>

@VanessaBeeley | Active | 15905 followers | 2017-06-21

@KreaseChan So I guess he will next appear as cameraman for the #WhiteHelmets? Hand in Hand with #AlQaeda. <*embedded image of a White Helmets volunteer carrying a dead body*>
[7 retweets]

This tweet references the same incident as the one above. Here *@VanessaBeeley* is replying to a tweet from *@KreaseChan*, an account that belongs to the Campaigns Manager of Amnesty International UK. *@KreaseChan* has tweeted out an image of the statement by the Syria Civil Defense explaining how they have dismissed the volunteer in question for a "gross beach of SCD's Code of Conduct" and commended the White Helmets for exhibiting accountability. *@VanessaBeeley* replies to this tweet with implied incredulity, suggesting that the volunteer will find another role within the organization.

@VanessaBeeley | Active | 20103 followers | 2017-11-19
28/11 3pm speaking at @UNGeneva Press Club screening "They Dont Care About Us" #WhiteHelmets True Agenda. @BorisJohnson @anonymizedRed @anonymizedRed @anonymizedRed @anonymizedRed @anonymizedRed @anonymizedRed @RussianEmbassy @suspended <*embedded image for an event hosted by the Geneva Press Club featuring Vanessa Beeley and two other speakers*>
[101 retweets]

This tweet publicizes an event hosted by the Geneva Press Club where Vanessa Beeley was set to appear on a panel with two other speakers and share their (critical) views on the White Helmets. News that the Geneva Press Club would host this event — and these speakers — was met with concern. The Swiss section of Reporters Without Borders (Reporters Sans Frontieres, RSF), a member of the Geneva Press Club, wrote an open letter to the club requesting the event be cancelled, specifically noting that some of the speakers had appeared regularly in Russian state media. However, the event went forward as planned.

> **mentioned tweet: @VanessaBeeley (2017-12-21):** Understanding The Guardian's Latest 'Russia-White Helmets' Conspiracy Theory <*link to article on 21stcenturywire.com titled "Understanding The Guardian's Latest 'Russia-White Helmets' Conspiracy Theory"*> via @21WIRE
> **mentioned tweet: @anonyizedRed (2017-12-21):** Yes, curious isn't it, how conspiracy theorising is acceptable when its target is Russia?

@VanessaBeeley | Active | 21344 followers | 2017-12-21

`@anonymizedRed @21WIRE Imagine if Russia were funding the #WhiteHelmets. LOL Monbiot would be all over THAT story!` [1 retweet]

This is a back-and-forth exchange between *@VanessaBeeley* and another account in the Red community. Our random sample surfaced the third tweet in this thread, but we provide the previous two for context. The exchange took place in December 2017 in response to Solon's article in the Guardian claiming that the White Helmets were the targets of a disinformation campaign. In the original tweet, *@VanessaBeeley* posts a link to an article on the alternative news website 21stCenturyWire titled "*Understanding The Guardian's Latest 'Russia-White Helmets' Conspiracy Theory.*" Following that link to the website reveals the article to be a short blurb featuring an embedded YouTube video from the "Going Underground on RT" channel, a channel associated with Russia's state media apparatus.

Another account from the Red community replies to *@VanessaBeeley's* tweet with a statement implying that claims of a disinformation campaign against the White Helmets are due in part to anti-Russia sentiments and that reactions to the White Helmets would be different if they were supported by Russia. *@VanessaBeeley's* replies with a tweet supporting that view and exhibiting a humorous or mocking tone.

This example is one of hundreds showing the complementary use of Twitter, alternative news media websites, and YouTube—and the integration of content from Russia's state media apparatus within these interlinked webs of content.

@VanessaBeeley | Active | 37943 followers | 2018-06-01

`"There were a lot of people fm Nusra Front working with the #WhiteHelmets & other immigrants. There were many White Helmets from more than one country like Jordan or Saudi Arabia & there were some – we didn't even know what language they were speaking." Mahmoud Al Khatib #Ghouta <embedded image of a man sitting in a kitchen looking at the camera>`

[248 retweets]

This tweet features a quote from a person who is purported to be a former ambulance driver for the White Helmets. The quote suggests that the White Helmets were working with people from the Nusra Front, a jihadist organization (distinct from ISIS) that aimed to establish an Islamic state in Syria. The content here supports narratives that the White Helmets work with foreign "terrorist" groups.

Looking across these example tweets and more broadly at *@VanessaBeeley's* activity throughout the White Helmets Twitter dataset reveals her to be a central figure in the anti-White Helmets discourse on Twitter (and beyond). Vanessa Beeley introduces new content into the space through embedded images and videos, as well as links to external websites and YouTube videos. Vanessa Beeley also authors content that is published on multiple websites that she and other Red-community accounts link to in their tweets. They often interacts with other accounts, especially other accounts in the Red community—through replies, quotes, and mentions. Vanessa Beeley also specifically calls out other influential Red-community accounts in her tweets (through @mentions) to draw their attention to their content. They receive more retweets than any other account in the data (about 67.6 retweets per original tweet). Over the course of our collection period, @VanessaBeeley's follower count nearly doubles.

Random tweets from the Red community

Below are ten original tweets (non-retweets) randomly selected from the Red community, ordered temporally from earliest to latest. They show how accounts in that community consistently tweeted content that challenged the White Helmets. They also demonstrate how those criticisms ranged across themes and narratives (though they reflect an incomplete set of the narratives used to attack the White Helmets).

@anonymizedRed | Active | ~6000 followers | 2017-06-15

@anonymized @anonymized @anonymizedRed @anonymizedRed @Presidency_Sy @suspendedRed @suspendedRed @anonymizedRed @anonymizedRed Here's a short video explanation of who the #WhiteHelmets are: <*link to YouTube video titled "The Syria White Helmets Exposed as US UK Agents Embedded with Al Nusra and ISIS"*> #Syria

[2 retweets]

@anonymizedRed | Active | ~4000 followers | 2017-06-27

sy#Syria #NuNusra #FalseFlag

Just a reminder…

> **quoted tweet: @anonymizedRed (2017-05-03):** Keep an eye on what may be happening in #Ariha and #Saraqib with the #WhiteHelmets teams, @AJArabic #Aljaqeera film crews, GKW <*embedded images of two Facebook posts (textual content) purportedly describing how the White Helmets worked with television crews to stage the impacts of chemical weapons attacks*> #Syria #Idlib

[56 retweets]

> **mentioned tweet: @Thomas1774Paine (2017-08-24):** George Clooney and his wife donate $1 million to group targeting conservatives

@anonymizedRed | Suspended | 21 followers | 2017-08-25: @anonymizedRed Clooney is the lover of the White Helmets terrorists.

[0 retweets]

@anonymous701 | Active | 95904 followers| 2017-11-12

West Propoganda to White Wash Terrrorist “'White Helmets” <*link to article in SputnikNews titled “Heroes or Zeroes? Questions Raised About Upright Image of White Helmets”*>

[3 retweets]

@anonymizedRed | Active | <100 followers | 2018-01-31

WAKE UP. The truth about the White Helmets. This is not going to stop in a far off country! <*link to article on TheWallWillFall blog titled, “Examining the Truth about Syria and the White Helmets”*>

[0 retweets]

@anonymizedRed | Active | ~300 followers | 2018-04-03

Now that #Ghouta is over, the next target is #Idlib. And we can expect more videos and tweets from the #WhiteHelmets talking about the last hospital in Idlib, etc. More outrage from the Western world until the jihadists are all quietly bussed into Turkey. So it goes…

> **quoted tweet: @anonymizedBlue (2018-04-03):** A father and his children were killed today after warplanes bombed their home in al-Asadiya, #Idlib. #Syria <*embedded images of White Helmets volunteers working in rubble and carrying the body of a dead child*>

[0 retweets]

> **mentioned tweet: @Charles_Lister (2018-04-10):** Facing potential U.S strikes, #Assad & #Moscow accept an @OPCW fact-finding team to #Douma - it'd likely take several weeks for any results from a site already muddied by Russian army. Watch UNSC vote today in 2hrs 20mins - a #Russia veto is *very* likely & may catalyze action.

@anonymizedRed | Suspended | <100 followers | 2018-04-10

@Charles_Lister @OPCW #SyriaGasAttack #Syriahoax #Syria #Whitehelmets by Day, AlQaeda terrorists, torturers and executioners by Night. These are @RealDonaldTrump helpers!

[0 retweets]

@anonymizedRed | Active | ~2000 followers | 2018-04-13

I replied on this tweet... <*link to an archive.is page capturing a tweet thread where this account argues with an account from the other side of the conversation and is referred to as a "bot"*> of "Investigative Reporter" @OliviaSolon's article in #Guardian (which whitewashes all of the #WhiteHelmets

crimes in #Syria) and voilà - he thinks I’m a bot! Ok and you’re a prime example of a Jew with Paranoia - wanna get a Beer?🍺

> **quoted tweet:** @anonymizedBlue Perfect example of a bot responding to a hashtag.

[0 retweets]

@anonymizedRed | Suspended | 47457 followers | 2018-04-17

Roger Waters brought up the White Helmets after an audience member wanted him to condemn the Syrian government for its alleged complicity in a chemical weapons attack that led the US, UK, and France to bomb Syria. <*link to an article on MintPressNews titled “Pink Floyd’s Roger Waters Takes on Syrian White Helmets, Thought-Control”*>

[56 retweets]

@anonymizedRed | Active | ~1000 followers | 2018-05-18

@J_Donaldson_MP Our govt supports the 'White Helmets' -- a group that claims to be a humanitarian org. Please view this video of White Helmets cheering a public execution and tell me your reaction, along with what you plan to do about it. Thanks. <*link to a YouTube video titled “White Helmets cheering during extremist execution”*>

[0 retweets]

These randomly selected tweets demonstrate that the content in the Red community is consistently critical of the White Helmets. They are called terrorists and torturers, and repeatedly derided for a video that purportedly shows White Helmet’s volunteers celebrating the executions of Syrian soldiers. They are accused of being “crisis actors” and staging impacts—especially impacts from chemical attacks—to portray the Syrian government and their Russian allies in a negative light. And they are referred to as “Western propaganda”—positioned as favored and assisted by “Western” and “mainstream” media.

Accounts in the Red community rarely rely on "mainstream" news media, but instead leverage alternative media outlets and social media platforms to spread their content. Three of the Red-community example tweets contain links to articles on external websites, including MintPressNews and SputnikNews. These websites are nodes within an "alternative media ecosystem" that repeatedly shares content (copied and pasted articles) and functions to support anti-White Helmets messages (Starbird et al., 2018).

One of the example tweets describes how Roger Waters (the Pink Floyd musician) criticized the White Helmets on stage in a concert in April 2018 (a week after the chemical attack in Douma). That tweet linked to an article in MintPressNews that described the incident, saying that Waters called the group "fake" and claimed their main purpose was to spread propaganda. A YouTube video of Waters' making the comments, edited and hosted by a YouTube channel associated with Russian state media outlet RT (RT UK), is embedded within the MintPressNews article. That YouTube video is also linked to by 71 tweets within the White Helmets Twitter dataset. 54 of those were posted by accounts in the Red community; one was posted by an account from the Blue. Other video clips of the Waters video constitute a significant portion of all the YouTube video links posted in our data.

Another interesting aspect of the example tweets in the Red community is that three of the accounts have been suspended since they posted these tweets. This is a trend we examine further later in this chapter.

Most-retweeted tweets from the Red community

Below are the five most retweeted tweets in the Red community, ordered by the total number of retweets they received in the White Helmets Twitter dataset. To provide context about the author and the content, we include a short description of each.

@sahouraxo | Active | 81425 followers | 2017-06-05

Wow. Last year, the White Helmets said Assad bombed this boy. Turns out they lied. He is fine and his family supports the Syrian government. <*embedded image of "the Aleppo Boy" who had been rescued by the White Helmets a year prior, now a year older and healthy*>

[7605 retweets]

This tweet references the story of Omran Daqneesh, a boy who became known as the "Aleppo Boy" after he was photographed in an ambulance operated by the White Helmets, covered in blood, after an airstrike on his family's home in Aleppo that took eight lives (including one of his brothers). About a year later, he

and his father appeared in an interview on a Syrian government-run television station. The boy appeared healthy at that time and the father asserted that he was pro-regime and claimed that Omran had been used as a 'propaganda tool' by rebel forces. The account that shared this tweet is an active poster in a range of geopolitical topics, consistently critical of the US, UK, and other western governments.

@RussiaUN | Active | 28290 followers | 2018-04-19

#BREAKING 11-year-old #Syrian boy Hassan Diab who was showed in the video on the alleged chemical attack in #Douma witnesses that everything was staged by #WhiteHelmets. Watch the full interview by Russian war reporter @anonymized here: <*link to yadi.sk*> @VanessaBeeley <*embedded video with child explaining how he was unnecessarily treated for a chemical exposure*>

[2978 retweets]

This highly-retweeted tweet was posted by *@RussiaUN*, the official accounts of Russia's mission to the United Nations, in the aftermath of the purported chemical attack in Douma, Syria in April 2018. The tweet features an embedded video with an interview of a child who explains how he appeared in a video being treated for a chemical exposure, and how that treatment was unneeded because he had not experienced any chemical exposure. The tweet and video support a narrative that the White Helmets staged the effects of a chemical attack. *@RussiaUN* includes a mention of *@VanessaBeeley* to draw their attention to this content. *@VanessaBeeley* retweets this content about six hours later.

@RussianEmbassy | Active | 78712 followers | 2018-04-14

Breaking - people from #WhiteHelmets "chemical attack" video explain how it was staged.

Retweet to beat jihadist lies! <embedded video with interviews of two young men who had treated victims of the purported chemical attack in Douma and whose testimony suggests that the victims had not experienced a chemical attack>

[2889 retweets]

This tweet was posted by *@RussianEmbassy*, the official English-language account of the Russian Embassy in London. Like the tweet above, it was posted in the aftermath of the April 2018 chemical attack. It features a different video, embedded using Twitter's native video sharing feature. The video features interviews with two men who claim to have been medics treating victims after the explosion, and who say that White Helmets volunteers and film crews interrupted their treatments to give patients additional treatments for chemical exposure. The video was produced by Russia's state media outlet, RT, whose logo appears in the bottom-right corner.

@PrisonPlanet | Active | 853078 followers | 2018-05-04

`Trump is cutting off funding to the Islamist-linked White Helmets. Good news. <link to article on ZeroHedge titled "Trump Cuts of Funding For Syria's 'White Helmets'">`

[2845 retweets]

This tweet was authored by *@PrisonPlanet*, a verified account that belongs to Paul Joseph Watson, a media personality who has long served as an editor of Alex Jones' website InfoWars. Paul Joseph Watson has been characterized as a conspiracy theorist—for supporting numerous conspiracy theories, including that the 9-11 attacks were perpetrated by the US government. His tweet shares a link to an article on alternative news website ZeroHedge that celebrates US President Trump's move to cut funding to the White Helmets. This tweet demonstrates an intersection in appeal of anti-White Helmets narratives to both international audiences on the left of the political spectrum (featured elsewhere in these examples) and a section of the US and UK domestic audiences aligned with the "alt right" (in this example). @PrisonPlanet had 853,078 followers at the time of this tweet, which received 2845 retweets.

@EmbassyofRussia | Active | 15545 followers | 2018-04-13 18:24:03

`Russian Defence Ministry: We have evidence that #UK had a direct involvement in arranging the #DoumaProvocation in Eastern Ghouta. We know for certain that between 3 & 6 April the White Helmets were pressured from London to speed up the provocation they were preparing <embedded image of Russian military officer with overlaid text>`

[2313 retweets]

@EmbassyofRussia is the official Twitter account of the Embassy of Russian in South Africa. Their highly retweeted tweet, also posted in the aftermath of the Douma attack, calls the reporting of a chemical attack in Douma a "provocation" and claims that the United Kingdom had coordinated with the White Helmets to stage and publicize the event. The tweet features an image of a Russian military officer with overlaid text:

> "*Russian Defense Ministry has evidence that Britain had a direct involvement in arranging provocation in Douma. We know for certain that between 3 and 6 April the so-called White Helmets were seriously pressured from London to speed up the provocation that they were preparing*"
> Igor Konashenkov, Russian Defence Ministry Spokesperson

The account had 15,545 followers at the time of this tweet and the tweet received 2312 retweets.

Similarly to the Blue community, the most retweeted tweets from the Red community also originated in accounts with high numbers of followers. However, in contrast with the Blue community, these accounts did not include "mainstream" media or journalists. Instead, they consist of one high follower account dedicated to a wide range of political messaging in alignment with the political objectives of Hezbollah (@sahouraxo), an "alt-right" media influencer (@PrisonPlanet), and three accounts that are officially affiliated with the Russian government.

Summary of narratives from the Red community

The example tweets from these three samples reveal a subset of the salient narratives within the Red community. All are highly critical of the White Helmets. One prominent narrative is that the White Helmets staged evidence of the Douma chemical attack (and other previous chemical attacks). Conversations around this narrative include a kind of open source "sensemaking" where participants attempt to assemble evidence to support that theory. However, most of the narratives are not focused on directly denying specific claims or evidence provided by the White Helmets, but instead function more diffusely to delegitimize their work and their messages through a range of ad hominem attacks. Salient narratives from example tweets in the Red anti-WH community include:

- The White Helmets are a propaganda construct of the West.
- The White Helmets are assisted by "Western" and "mainstream" media.
- The White Helmets are not humanitarian actors.
- The White Helmets are terrorists.
- The White Helmets are "crisis actors" who stage impacts to make the Syrian government and their Russian allies look bad.

- The White Helmets stage chemical weapons attacks.
- The White Helmets are agents of the US and UK governments.
- The Douma chemical attack was a staged provocation by the White Helmets working with the UK
- The US government should not be supporting the White Helmets.

Suspended Accounts by Community

Twitter suspends accounts for breaking the platform's terms of service—for reasons including hate speech, harassment, spam, impersonation, and other inauthentic behavior. Unfortunately, the platform does not publish the reason that any specific account in the data has been suspended, but it is possible to use patterns of suspensions as a rough proxy for behavior that violated the terms of service.

In December 2019, we used the Twitter API to identify any accounts within either of our two communities that are currently listed as "suspended" by Twitter. Table 6.3 (below) contains descriptive statistics of the numbers of accounts that were suspended within each community as well as the amount of content and types of content (original tweet or retweet) that came from those suspended accounts.

Community	Num accounts	Num accounts suspended	% accounts suspended	% (re)tweets from suspended accounts	% original (non-retweets) from suspended accounts	% retweets from suspended accounts
Blue (pro-WH)	70670	4090	5.8	7.8	8.2	7.7
Red (anti-WH)	97440	13514	13.9	15.7	26.6	13.1

Table 6.3: Descriptive Statistics of Account Suspensions by community

This table shows that accounts in the Blue community are far less likely to be suspended than accounts in the Red community. In addition, more content from the Red community originated from suspended accounts—15.7% from the Red, compared to 7.8% of the Blue. Perhaps most interesting is the over-representation of original content (non-retweets) posted by suspended accounts. More than a quarter of all original "White Helmets" tweets posted by accounts in the Red community during our collection period were tweeted by accounts that have since been suspended. Much of this now-removed content was created by a small number of highly active accounts that repeatedly tweeted content criticizing the White Helmets.

k-core Analysis

The two communities also differ structurally. The Red community is larger overall (in number of accounts and number of tweets), but it also has a larger number of central accounts that are all connected—due to retweeting each other or having mutual retweeters. Table 6.4 presents a *k*-core decomposition of the two communities. A *k*-core is a subgraph of a given graph where all remaining nodes have a degree of at least *k*. In this case a degree is equal to the number of retweets an account gave to or received from another account remaining in the graph.

***k*-value**	**% of accounts from blue community**	**% of accounts from the red community**
1	100%	100%
2	25.2%	41.8%
5	5.5%	15.5%
10	2.1%	6.9%
25	0.6%	2.1%
50	0%	0.6%
100	0%	0%

Table 6.4: k-Core Decomposition of the Blue and Red communities

Table 6.4 reveals that the Red community retains a larger proportion of its graph at higher *k* values than the Blue community. At *k*=60, the Red community still has 653 accounts, including most the highly-retweeted accounts. This structure reflects some of the dynamics of the campaign within the Red community, where a number of dedicated accounts, including some journalists and other influencers as well as some activists, consistently tweet and retweet each other. The 'glue' that holds this network together is a large number of accounts that participate primarily through retweeting many different influential accounts in the space.

One explanation of the difference in structure is that the pro-White Helmets Blue community takes shape around a smaller number of influential accounts, centered around *@SyriaCivilDef*, but including an array of high-follower journalists and media accounts. Those accounts pull in larger audiences, but those audiences are less engaged in the "white helmets" discourse, with participation often limited to a single retweet. There are a few dedicated pro-White Helmets accounts in the Blue community, but not as many as there are dedicated anti-White Helmets accounts in the Red. In contrast, the Red community takes shape

around a larger number of accounts that consistently participate in tweeting (negatively) about the White Helmets.

Looking Beyond Twitter to the Wider Information Ecosystem

After identifying the communities that make up to two "sides" of this conversation—the blue pro-WH and red anti-WH communities—we 'extended' the retweet network representation to include the domains that are most often linked-to from each community (Figure 6.5). This was achieved by first unshortening the URLs within tweets and then resolving the expanded URLs to their destination domains (websites). These domains were added to the network representation provided earlier in this chapter (Figure 6.1), forming a circle around the communities at the center of the visualization. The visualization features 65 domains, each of which was linked-to by at least 500 tweets in the White Helmets Twitter Dataset. The size of the nodes represents the relative number of times that domain was linked-to in a tweet or retweet. The thickness of the edges between the red/blue communities and the domain nodes represents the number of times that a source account (in the respective community) tweeted or retweeted a link to that domain. A large domain node with a thick line from one of the communities would mean that the accounts in that cluster posted a relatively large number of URLs to that domain.

For each domain we calculated a *modularity ratio* — a metric that denotes the proportion of tweets to a domain from accounts in the red and blue communities. The *modularity ratio* is on a scale between 0 and 1, where, at the extreme, 0 denotes the domain was only tweeted by accounts in the blue community and 1 means the domain was only tweeted by accounts in the red community. The color of the domain nodes visually corresponds with the modularity ratio: For example, a domain node that appears red was predominately (re)tweeted by accounts in the red community; domain nodes that appear a shade of purple were (re)tweeted more proportionally from accounts in both the red and the blue communities (Figure 6.4).

Figure 6.4: The modularity ratio (measured on a scale of 0-1) was used to color the domain nodes in the domain retweet network visualization.

For clarity, the domain nodes are positioned so that they broadly align with the community from which predominantly link to that domain, so domains primarily tweeted by accounts in the blue community are

on the right side of the visualization and domains primarily tweeted by accounts in the red community are on the left side of the visualization.

The network visualization shows that accounts in the red community tweet links to more external domains (more domain nodes are red), and they do so in greater volume (the red nodes are among the largest). Alternative media and government-controlled news websites that will be discussed in Chapter 7 (e.g., ClarityofSignal, 21stcenturywire, MintPressNews, ZeroHedge, RT, and SputnikNews) are salient domain nodes in this visualization: The color of the nodes, the bold edges (lines) connecting them to the red community, and the high *modularity score* (0.99 or greater) demonstrates that more than 98% of all the tweets that contain a URL to those domains are from accounts in the red community.

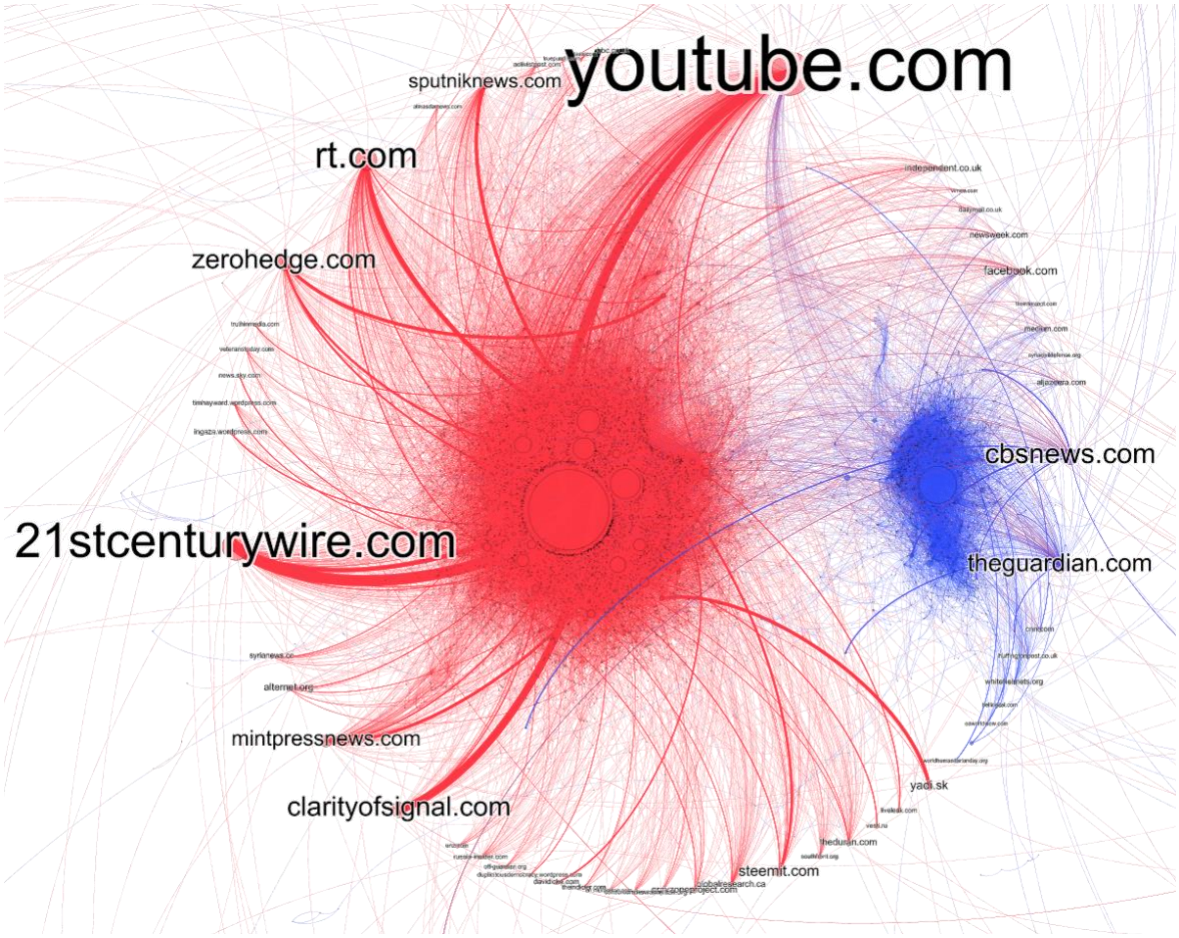

Figure 6.5: A domain retweet network visualization that looks beyond Twitter, showing the surrounding ecosystem of websites and social media platforms that are linked to in the Twitter conversation

Particularly striking is YouTube, which is by far the most cited domain and social media platform. As discussed further in Wilson and Starbird (2020), which looks in more detail at the cross-posting between YouTube and Twitter, the red anti-White Helmets community is more effective in leveraging YouTube as a resource for their Twitter campaign: They shared 43,340 total tweets, linking to 1604 distinct YouTube videos, hosted by 641 channels. YouTube videos cross-posted by accounts in Red were longer and had more views and likes (on YouTube) than videos cross-posted by Blue.

It is not particularly surprising that 94% of tweets containing a URL to the WH's website are from accounts in the pro-WH blue community. Several domains are predominantly linked-to by accounts in the blue community (illustrated by a modularity ratio closer to 0), including a variety of mainstream media websites such as the Huffington Post (81% of links are from accounts in the blue community), CNN (77%), The Guardian (75%), and CBS News (65%). Of all links to the social media site Facebook, 35% are from accounts in the blue community, while for the publishing platform Medium this is 58%.

Summary

In this Chapter I presented a long-form analysis of the WH conversation on Twitter, including the structure, temporal activity, and narrative analysis of the two communities involved in the discourse. The empirical observations presented here set up Study 3 (Chapter 7), which will examine the role of government-controlled and alternative media in fomenting, shaping, and propagating anti-WH narratives. The domain retweet network visualization analysis also reveals the salience of YouTube (the role of YouTube is examined in further detail in Wilson and Starbird (2020)) plus other social media platforms such as Facebook in the WH conversation on Twitter. In Chapter 8 I will investigate further the links between Twitter and other social media platforms as I examine multi-platform information operations.

Chapter 7:

The Role of Government-Controlled & Alternative Media in Shaping Disruptive Narratives (Study 3)

Author preface

Chapter 7 looks beyond the Twitter conversation to examine **how government-controlled and alternative media websites shape and propagate disruptive narratives** about the role and function of the White Helmets. With the permission of the first author, this chapter is based around an article published in the AAAI International Conference on Web logs and Social Media (ICWSM) titled *Ecosystem or Echo-System: Content Sharing Practices of Alternative Media Domains*[29] (Starbird et al., 2018). In this version I have included previously unpublished content that was outside the scope (and word length) of the published paper. The additional content is written in a sans-serif font to denote that it was not in the original published article.

In particular, I present the work that was done to surface the salient narratives about the White Helmets that featured on websites linked-to in the White Helmets conversation on Twitter. For two quarters (September 2017-March 2018), I co-managed a Directed Research Group (DRG), a research-for-credit program which

[29] This study is previously published work. To cite material from this Chapter, please cite the original work:
Starbird, K., Arif, A., Wilson, T., Van Koevering, K., Yefimova, K., & Scarnecchia, D. (2018). Ecosystem or echo-system? Exploring content sharing across alternative media domains. In Twelfth International AAAI Conference on Web and Social Media.

offers graduate and undergraduate students a wide range of opportunities to do hands-on research. I led the students conducting content analysis of the articles about the White Helmets and the subsequent qualitative work to identify the salient narratives circulating about the group on websites linked to on Twitter. I explain more about this process at the end of the Methods section and describe the salient narratives that we identified at the beginning of the Findings section. To distinguish this from the previously published work, the additional text appears in a sans serif font.

The remainder of the chapter will focus on the content-sharing practices of the websites involved in fomenting, shaping, and propagating competing strategic narratives about the White Helmets. This work uncovers a predominantly alternative media "echo-system" of websites that repeatedly share content about the White Helmets. This work also provides insight into the network of government-controlled and alternative media websites that function to shape anti-White Helmets narratives. The variety of websites within this network appeal to diverse audiences, and by publishing anti-WH articles alongside content that resonated with readers' existing beliefs, the websites within the echo-system were able to draw in audiences from diverse communities into a set of common narratives, critical of the White Helmets.

I made some minor stylistic edits to this published paper: I anonymized some account names in line with the anonymization policy of this dissertation; and I isolated some references to content that is problematized as misleading within the article. The inline citation for these articles appears as subscript and the references appear at the end of the chapter, rather than in the bibliography section at the end of the dissertation.

<<THE PUBLISHED PAPER BEGINS BELOW>>

Abstract

This research examines the competing narratives about the role and function of Syria Civil Defence, a volunteer humanitarian organization popularly known as the White Helmets, working in war-torn Syria. Using a mixed-method approach based on seed data collected from Twitter, and then extending out to the websites cited in that data, we examine content sharing practices across distinct media domains that functioned to construct, shape, and propagate these narratives. We articulate a predominantly alternative media "echo-system" of websites that repeatedly share content about the White Helmets. Among other findings, our work reveals a small set of websites and authors generating content that is spread across diverse sites, drawing audiences from distinct communities into a shared narrative. This analysis also reveals the integration of government-funded media and geopolitical think tanks as source content for anti-

White Helmets narratives. More broadly, the analysis demonstrates the role of alternative newswire-like services in providing content for alternative media websites. Though additional work is needed to understand these patterns over time and across topics, this paper provides insight into the dynamics of this multi-layered media ecosystem.

Introduction

In September 2016, a documentary featuring the Syrian Civil Defense, a volunteer response group in Syria who are also known as the White Helmets (WH), was released to wide acclaim. Later, the film would win an Oscar for Best Documentary. It was set within the context of the then five-year-old civil war in Syria, troubling claims about the brutality of Syrian President Bashar al-Assad who had been accused of bombing civilians and medical workers (Fouad et al., 2017), the rise of the Islamic State (IS) and other extremist groups in areas outside the government's control, and a massive exodus of refugees escaping the violence (BBC News, 2016). The documentary and subsequent sympathetic articles by mainstream media outlets worldwide brought global awareness to the plight of Syrian people, especially those resisting the Assad government.

This narrative, which promoted the role of the WH as a humanitarian response organization in Syria, functioned to grow solidarity between many Western audiences and the WH—as well as Syrian people from rebel-held areas who were seen as victims of the Assad regime. However, this narrative was not aligned with other views of the complex geopolitical landscape of the conflict. In particular, representatives and supporters of the Syrian government and its allies in that conflict (including Russia and Iran) resisted this sympathetic perspective. In response, a counter-narrative took shape and eventually spread to other online communities. Critics argued that the group, which is funded by Western governments, was a propaganda construct, supported by mainstream media, and employed as a tool of NATO interests in Syria (RT 2017). Some claimed that the WH aided and in some cases were themselves active in terrorist organizations (Beeley 2017). Supporters of the WH, in turn, accused their critics of orchestrating a propaganda campaign to spread a "conspiracy theory" about the group (e.g., Grey Ellis, 2017). These disruptive[30] narratives, which are still active, are spread through and shaped by various media—including mainstream news articles, alternative media articles, blog posts, social media posts and interactions, etc. The resulting information space is an evolving and multi-layered one.

[30] The original publication used the term "contested narratives", but for consistency within this dissertation I have updated that to "disruptive narratives" to describe the same phenomenon.

This paper explores the dynamics of a subsection of the media ecosystem that was active around Twitter conversations about the WH during a three-month period in the summer of 2017. Using a mixed-method approach based on seed data collected from Twitter, and then extending out to the websites cited in that data, we examine content sharing practices across distinct media domains—articulating an alternative media, and, to a lesser extent, a mainstream media "echo-system" of websites that repeatedly share content about this topic. Among other findings, our analysis reveals a small set of source domains (and authors) generating content that is spread across diverse domains, drawing audiences from distinct communities into a common narrative. This analysis also reveals the integration of government-funded media (RT, SputnikNews) and geopolitical think tanks (GlobalResearch) as source content for anti-WH narratives. More broadly, the analysis demonstrates the role of alternative "newswire" services in providing content for alternative media websites. Though more study is needed to understand these patterns over time and across topics, this paper provides insight into the dynamics of this multi-layered media ecosystem.

Background

The White Helmets and the Syrian Civil War

The ongoing conflict in Syria has taken more than 400,000 lives and displaced millions more. Armed conflict began during the 2011 Arab Spring, when anti-government protests calling for President Bashar al-Assad to step down escalated into full scale civil war between the Syrian government and those opposing Assad's rule (ICRC, 2012). It has since evolved into an internationalized, multi-sided conflict. Militants from the IS took advantage of the conflict to capture much of Eastern and Northern Syria (Whewell, 2013). Russia and Iran initially supported the Assad government with materiel and financial assistance, and later committed troops, targeting Syrian opposition forces and IS militants (Associated Press, 2013; Kramer & Barnard, 2015). In 2014, the United States and Gulf League states began bombing IS in Syria and providing assistance to Kurdish forces fighting IS and Syrian opposition groups (Gordon, 2014). The UK joined this coalition in 2015.

The WH are a group of trained volunteer rescuers that operate throughout Syria's opposition-controlled areas to assist civilians affected by the violence. According to the organization's website (http://syriacivildefense.org), they abide by the fundamental principles of the Red Cross and work in accordance with Article 61 of the Additional Protocol I, which defines the activities that constitute civil defence: protecting civilians from hostilities or disasters, aiding recovery in the aftermaths of such events, and providing conditions for the survival of the civilian population (Pictet, 1979). WH activities include

search and rescue, evacuating buildings, firefighting, medical aid, providing emergency shelter and supplies.

Online Propaganda & Disinformation in 2017

This research took place during a period (June 2017- January 2018) of global attention to the threat of misinformation, disinformation, and political propaganda, and the role of technology in facilitating their spread. In prior years, researchers optimistically noted the rise of citizen journalism (Gillmor, 2006), and also acknowledged a related "crisis in journalism" (Fuller, 2010) as economic, production and distribution models for "news" became disrupted. As traditional journalists and media outlets attempted to adapt to these new conditions, citizen journalists and media outlets outside the mainstream worked to establish their legitimacy—resulting in an information space where new voices were heard in new ways. Simultaneously, it became increasingly difficult for information consumers to assess the validity of the information they saw. The rise of partisan "fake news" sites and the subsequent appropriation of that term to challenge "mainstream" outlets (Qiu, 2017) corresponded with record-low levels of trust in media and information (Swift 2016; Barthell & Mitchel 2017).

Additionally, for decades critics have explicated systemic biases in "mainstream" media[31] —e.g., towards neo-liberal, pro-Western, and colonialist/imperialist ideologies (e.g., Herman & Chomsky, 1988). As a prominent example, relevant here, were claims within the New York Times (Narasimhan, 2005) that Saddam Hussein had weapons of mass destruction, a premise used to garner public support for the 2003 US invasion of Iraq. These examples and the arguments constructed around them likely contribute to diminished trust in mainstream media.

Information Warfare and Online Disinformation

While these claims suggest strategies tied to traditional means of news production, recent evidence suggests others—including state and non-state actors—are working to leverage online technologies to forward their geopolitical goals (e.g., Weedon, Nuland & Stamos, 2017). These actors are using a mix of automation and human curation to intentionally spread misleading information using online technologies such as social media (Woolley & Howard, 2017).

[31] The term "mainstream media" has historically been used as a pejorative, especially by those who identify with "alternative" perspectives or media, to criticize the agenda-setting power of mass media. How- ever, the "mainstream" and "alternative" terms have also become a com- mon way to distinguish between media types. We employ these terms here for efficiency, but also with acknowledgment of their political roots and the tension between them.

In particular, Russia has been accused of conducting an "information war" that extends long-standing tactics of disinformation to new Internet-enabled channels (Pomerantsev & Weiss, 2014). Though researchers and intelligence communities are still working to understand these strategies, evidence suggests that Russia and others are utilizing social media in conjunction with other channels to spread their messages (Weedon, Nuland, & Stamos, 2017; Paul & Matthews, 2016). The Russian government also utilizes its own media apparatus, including RT and SputnikNews, to forward their geopolitical aims. RT receives the vast majority of its funding from its government (Moscow Times, 2014), and its editor-in-chief has argued that RT uses that funding to support "information warfare ... against the Western world" (DFRLab, 2018).

Pomerantsev & Weiss (2014) define Russian disinformation as intended not to simply convince, but to confuse—to sew doubt and distrust across a society. The idea is that doubt can act to reduce agency. In other words, if we are not sure about what the truth is, we cannot choose the best action to take, and therefore will take no action. For this reason, disinformation campaigns do not need to rely on a single narrative or counter-narrative but can work by presenting diverse and even contradictory narratives.

Information Operations and Non-Governmental Organizations (NGOs)

The Russian government has previously taken issue with— and actively worked to undermine the mission of—NGOs which they see as a threat to its geopolitical interests (Ambrosio 2007). In the early 2000s, Western NGOs supported "pro-democracy" civil society groups in Russia and its neighboring states and played a role in facilitating a shift away from Russia-aligned governments and policies. Following the color revolutions that took place in the former Soviet states, the Russian government argued that these activities represented unfair interference. Vladimir Putin specifically called out and criticized "pseudo-NGOs" funded by foreign governments and corporations for their role in destabilizing other countries (Putin, 2012). In this study, we can see an extension of that criticism to a humanitarian response organization working—both through its efforts to assist affected people and to garner attention for their cause—against the geopolitical interests of Russia and its ally, the Syrian government.

Conducting Research on Information Operations

This paper is a small component of a larger research effort examining disruptive narratives involving the WH. Over the course of several months, our team spent hundreds of hours analyzing this data at multiple levels. This information space can be intensely disorienting. Our researchers repeatedly use this word to describe how the qualitative analyses affect us. The arguments and evidence presented in support of narratives on both sides are often compelling. Despite, or because of, deep engagement with this content,

our researchers are often left in a state of confusion about what and whom to believe. In this study, we do not speak directly to this question. Instead, we focus on describing the media ecosystem surrounding these conversations—especially the dynamics of content-sharing practices—with the goal of gaining insight into how these narratives and counter-narratives are produced and disseminated.

Methods

Data Collection and Processing

Our White Helmets dataset (WH dataset) consists of tweets posted between May 27 and September 9, 2017. We created this collection using the Twitter Streaming API, initially tracking various keyword terms related to the Syrian conflict including geographic terms of affected areas. Later, we scoped this data to tweets that contained “white helmet” or “whitehelmet”, resulting in 135,827 tweets.

To understand the role played by external websites in Twitter conversations about the WH, we examined the links embedded within these tweets. 52,903 tweets contained a URL link. To process this data, we expanded shortened links, removed HTML parameters, and filtered out duplicates. We also removed links (approximately 35% of the total) that resolved to social media domains (i.e., Twitter, Facebook, YouTube) and newsreaders (e.g., feed-proxy.google.com).

The resulting set of 3410 distinct URLs was used to extract articles in a structured and automated fashion via a tool built using Newspaper, a python library designed for full-text and article metadata retrieval. Due to how some web servers and content delivery net- works were configured, we were unable to automatically scrape content from 111 domains. Subsequent analysis suggested that some of these domains play a prominent role in this information space (e.g., GlobalResearch), so we manually captured content from the top-10 most tweeted of these missing domains by traversing the URL to the article and copy-pasting its content. However, 322 URLs from the other 101 domains were omitted from the analysis.

Detecting Article Similarity

The text of the scraped articles was passed through an algorithm for detecting similarity. This was done by computing the term frequency–inverse document frequency (tf-idf) statistic (Salton, Fox & Wu 1983) for each article and obtaining the cosine distance between the tf-idf vector for each pair of articles. The resulting matrix of similarity scores was used to identify duplicate articles across domains. We selected a threshold

of >=85% to identify two articles as containing shared content. This level of similarity generally captured identical articles without being overly sensitive to small changes in image captions and article bylines.
After collapsing links to similar articles within the same domain, there were 1680 distinct news articles. From there, we identified 558 articles that had significant (>=85%) overlap with another article within another domain in our set. Interestingly, nearly two-thirds (63%) of tweets with URLs cited one of these 558 articles that appeared on more than one domain.

Construction of "paths"

Next, we constructed "paths" for each article—tracing all URLs in our dataset where that article appeared. We identified 135 paths, which provide insight into how content was shared across domains in the media ecosystem. We used these 135 paths to construct a network graph (Figures 6.1-6.3) where two nodes (domains) are connected if they appear in the same path—e.g., if one domain hosted an article that had >=85% similarity with an article in the other domain. The edge weight represents the number of similar articles shared by the two domains. These edges do not encode directionality, but merely reflect similar content. Using manual analysis to determine the original source of each article, we labeled each node as primarily a source (publishes original content), an amplifier (republished content from others), or a hub (published original and borrowed content). Nodes are sized by the number of tweets in the WH data that link to that domain, therefore representing the salience of this domain—and its articles— in the Twitter conversation. We use the ForceAtlas2 algorithm to determine the visual layout of the graphs and the Louvain method to detect communities for Figure 6.1.

Interpretative, Mixed-Method Analysis

We conducted interpretive, mixed-method analysis of this data, expanding upon methods developed for the investigation of online rumors in the context of crisis events (Maddock et al., 2015). This approach iteratively blends quantitative and qualitative analyses—in this case generating a network graph to see larger patterns of content sharing across domains, and then using that representation as an entry point for a closer examination of both the practices of content sharing and the influential domains within this ecosystem. For the qualitative analysis of the domains and the "paths", we focused primarily on the content within each domain, including its home page, about page, and the content-sharing practices visible within the specific articles cited in the WH dataset.

In a parallel research effort, we conducted qualitative analysis to surface the salient narratives being propagated within articles about the White Helmets. This analysis included reading and manually analyzing tweets and articles linked-to from those tweets.

This research was conducted as part of a Directed Research Group (DRG), a research-for-credit program which offers graduate and undergraduate students a wide range of opportunities to do hands-on research. Members of the group included 5 undergraduate students from Computer Science and the Department of Human Centered Design and Engineering, 1 PhD student from the Information School, and 2 PhD students from the Department of Human Centered Design and Engineering who led the group. The students came from a diverse set of countries including China, Indonesia, Kazakhstan, Pakistan, the United Kingdom, and the United States.

The DRG was held once per week, meeting for 90 minutes each week during the Fall Quarter 2017 (10 meetings). The meeting time was used to surface interesting findings and discuss the next steps. Students also spent an additional 6-8 hours on research outside of the meeting time, which was used for in-depth reading of the articles, qualitative coding, and memo-writing.

An initial exercise involved each researcher reading the same 6 articles, which were randomly sampled from the set of 118 English-language articles that were linked-to from the Twitter dataset. As they were reading the article, each researcher was asked to identify specific events and themes that stood out to them. The purpose of this step was to familiarize each researcher with the content about the WH and provide an opportunity to surface interesting and salient themes.

In the next step, each researcher labelled a sample of 30 articles each along several dimensions: The narrative stance toward the White Helmets (whether the article conveyed a pro-WH or anti-WH position, or whether this was ambiguous); The certainty around the narrative stance (whether the author distanced themselves from the source of the information using attribution shields or expressed doubt on the plausibility of the information); Explicit references to other authors/sites/sources (did the article include references to other authors, websites, or sources). This stage classified the articles into pro-WH/anti-WH.

The final step involved an in-depth analysis of each of the articles to surface the salient narratives. This was an iterative process that involved each researcher highlighting specific events and themes that stood out to them in the articles. Each week a list of these would be compiled, shared among researchers, and discussed during the weekly DRG meeting. Over the period of several weeks and cycles through this process, the group was able to surface a set of salient narratives that were frequently seen within the sample of articles.

Note on Data and Privacy

In this paper, we identify a small number of prominent authors within the alternative media ecosystem. We considered anonymizing these names but chose to publish real names because these authors are self-identified journalists and their patterns of activity—both within our data set and before/after—are important for understanding the nature of content sharing in this ecosystem. Several articles from both "sides" of this conversation are cited in the references.

Findings

Salient Narratives within Articles about the White Helmets

Pro-White Helmets Narratives

The pro-WH narratives accept and reflect the framing of the White Helmets as a humanitarian response organization. In our data, the pro-WH articles were mainly sympathetic articles relating the murders of six White Helmets volunteers which happened in August 2017.

The White Helmets are a volunteer humanitarian response organization who participate in medical-related activities within rebel-held areas in Syria: Pro-WH narratives describe the White Helmets as volunteers and part of a rescue group whose role is that of first responder—rescuing victims, treating the injured, and tending to those who have lost their lives. The group is described as independent and impartial, coming to the aid of people on all sides of the fight, though stories covering this narrative often acknowledge that the vast majority of their activities occur in rebel-held opposition areas. This narrative is often connected to narratives that problematize the actions of the Syrian government and their allies in the conflict.

White Helmets are being targeted by Syrian and Russian forces: This narrative highlights the risks that the White Helmet volunteers assume by being members of the group—at times connecting these risks to actions by Syrian and Russian forces. In preceding months, the White Helmets' regional centers, their vehicles, and personnel were targeted by airstrikes, with increased frequency since Russian air support for the Syrian government commenced. It was reported in 2018 that around one in ten White Helmet volunteers had been killed since the beginning of the conflict (Di Giovanni, 2018). This narrative positions these targeted attacks as violations of humanitarian norms.

Anti-White Helmet Narratives

Our analysis suggests that the anti-WH narratives were the dominant narratives in the broader White Helmets Twitter conversation during the summer of 2017. These anti-White Helmet narratives are complex and somewhat intertwined. Taken together, they promote a perspective that the White Helmets are a front for "soft power" in Syria, serving the interests of western governments.

White Helmets are funded by and therefore a tool of foreign governments (US, UK): That the White Helmets are funded by western governments is not disputed—this fact is often acknowledged in pro- or neutral-White Helmet articles, but this narrative positions this funding as reflecting a quid pro quo relationship, with western governments utilizing the group as a geopolitical tool.

White Helmets are associated with extremist groups operating in Syria: This narrative seeks to undermine support and foment disgust for the White Helmets by connecting them to extremist groups operating in Syria—e.g., Al-Nusra, Al-Qeada, or ISIS. These range from suggestions that they are "tied to" or "affiliated with" these groups to direct claims that they are a terrorist organization. In places, this narrative calls attention to perceived violations by the White Helmets of the "impartiality" principle of humanitarian response. Videos showing White Helmets moving bodies that have been executed and in some cases beheaded by rebel groups are used support this narrative. In our data, these claims appear in 21stCenturyWire, RT, ClarityOfSignal, GlobalResearch.ca, and ActivistPost (among others).

White Helmets use propaganda—publicizing their activities to garner positive attention from the Western media: In 2016, the White Helmets were featured in a Netflix documentary that was nominated for and won an Oscar award. Supporters of the Syrian government framed this documentary as a propaganda film in service of US-UK foreign policy. This narrative integrates arguments that White Helmets focus more on promotion of their activities than on actually helping people. These arguments are not entirely without basis (i.e. they contain a "kernel of truth" (Bittman, 1985)), as the White Helmets have clearly made an effort to publicize their work and situation to garner international support, both in sympathetic solidarity and in financial resources.

White Helmets stage events to use as propaganda: This narrative extends the propaganda narrative to claim that the White Helmets fake events to promote their cause and/or make the Syrian government look bad. These claims align with "false flag" and "crisis actors" narratives from other crisis events (Starbird, 2017). Articles promoting this narrative appear in 21stCenturyWire, RT, Sputnik News, Sott.net, and UNZ.

Corporate and/or "mainstream" media are active conduits of White Helmets "propaganda": A related narrative—one that appears as subtext across much of the anti-White Helmets content—

asserts that “corporate” media are purposefully creating and spreading a false narrative of the White Helmets to support western geopolitical interests. This narrative is salient across the White Helmets dataset, appearing in tweets and in articles within many websites including 21stCenturyWire, MintPressNews, ActivistPost, DCClothesLine, Salon, and Sott.net. For many of these websites, this de-legitimization of mainstream and corporate news can be seen as a foundational premise that reflects their reason for being and helps to promote their mission.

Content-Sharing Domain Network Graph

Figure 7.1 shows the complete content-sharing domain network graph for the entire WH dataset (nodes sized by tweet volume and colored by Louvain-detected community). From a high level, this graph has several key features: two large clusters—Cluster A in pink on the left and Cluster B in blue on the right, with some connective tissue between them; and a small distinct community (Cluster C, in yellow) loosely connected with Cluster B. There are also a large number of small, distinct clusters that are unconnected to the other clusters (in grey). One of these (Cluster D, in red) is interesting because it contains a highly tweeted, but disconnected, domain (see Table 7.1).

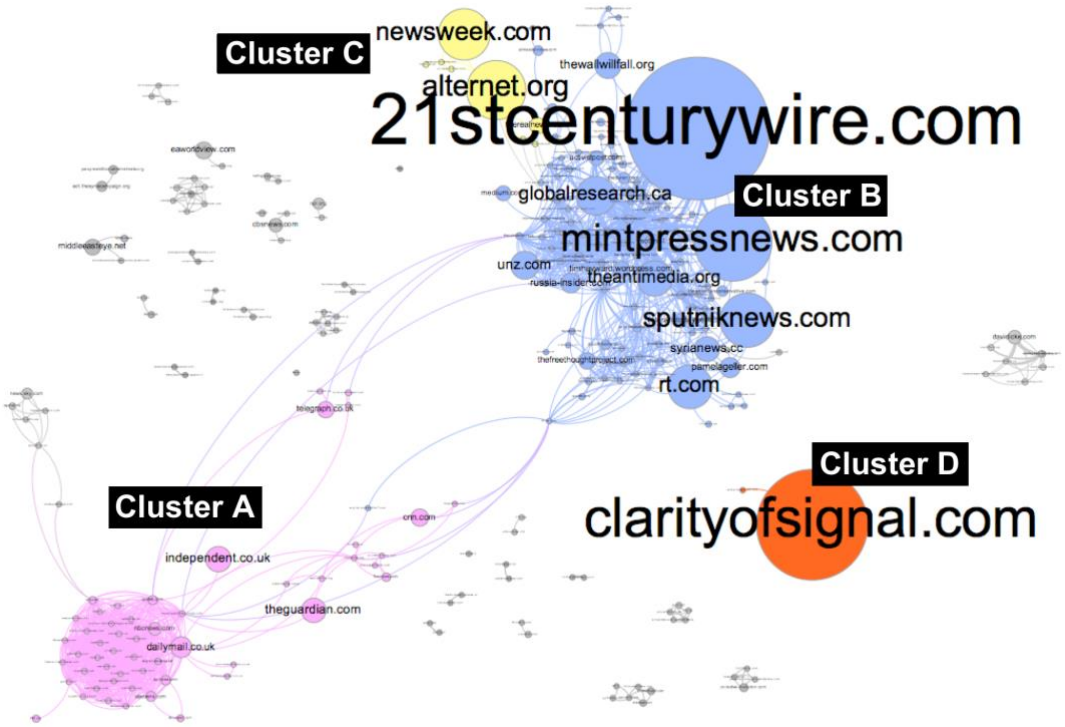

Figure 7.1: WH Content Sharing Domain Network Graph

Cluster A: An Associated Press News Cluster

Cluster A (Figure 7.2) is a relatively large component of 40 nodes, consisting of several western "mainstream" media sites (i.e., wsj.com, dailymail.co.uk, apnews.com), news outlets from the Arab world (aljazeera.com, arabnews.com, english.alarabiya.com), and other local and alternative media outlets from around the world. There are also a few news aggregators in this cluster, including castwb.com. Most of the edges in this cluster have weight of one, representing >=85% overlap of a single WH-related article within both web domains.

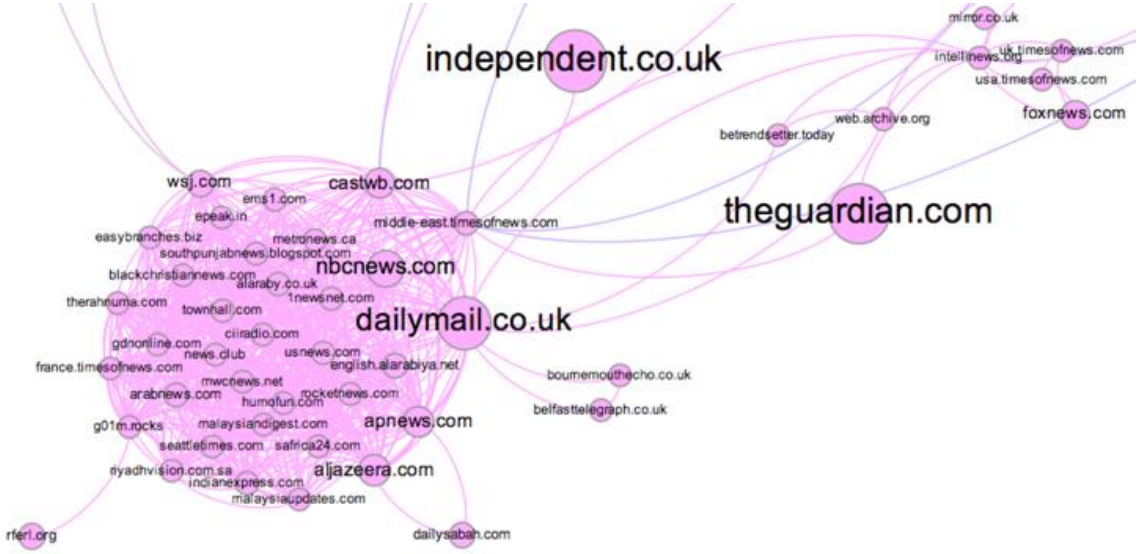

Figure 7.2: Close-up of Cluster A

Almost all of the connections in this cluster were related to a single article published in August 2017 (Mroue, 2017), describing the murder of seven WH volunteers at their office in Idlib, Syria. This article was sympathetic to the WH, presenting them as "first-responders who have been known to risk their lives to save people from the civil war."

This article constituted the largest path in our data—its content appeared in 44 different domains. Its original source, almost always cited in the downstream articles, was the Associated Press (AP). The AP operates as an international, non-profit news cooperative and allows its partner news outlets to reuse its content. These partners pay to use the AP's content in their own newspaper or website. This content-sharing model allows media outlets to provide coverage of diverse topics across the globe. Another similar agency, Agence France-Presse (AFP), appears within a small isolated cluster elsewhere in the graph. This model, which is a long-standing one that pre-exists the Internet (Fenby, 1986), results in content sharing across news outlets in our data, including many that are considered "main-stream". However, the relative scarcity of paths other

than this AP path suggests that intensive content sharing about the WH was not observed in this set of websites.

A small number of nodes serve to connect—over one to three degrees—Cluster A (the AP cluster) to Cluster B (on the right of the graph). Within this connective tissue are a few "mainstream" media domains including the Telegraph, Independent, CNN, and the Guardian. These domains' content was re-published by news aggregators (i.e., intellinews.org, f3nws.com) that connected those websites to other mainstream and alternative media domains. Most edges in this section have an edge weight of one—a single article, of which the mainstream media domain was the original source. The articles featured in this "connective tissue" area were generally supportive of the WH—promoting narratives that featured the WH as courageous volunteers who were risking their lives to rescue and provide medical assistance to Syrians who were injured by Syrian government and Russian military operations.

Cluster B: The Alternative Media Ecosystem

Most of the volume represented in the complete content- sharing network graph (Figure 7.1)—in terms of tweets, articles, and distinct domains—resides in Cluster B (Figure 7.3). The articles cited within these domains were highly critical of the WH. This cluster contains 110 nodes or web domains. 10,821 tweets in the WH data included a URL that linked to Cluster B, compared to only 2526 in Cluster A. Unlike Cluster A, which exhibits a consistent, nearly symmetrical structure, Cluster B is more heterogeneous in terms of both node size and edge weight. Edges vary in strength from one to seven articles. Thicker edges represent more consistent content sharing patterns over time.

Structural analysis on the domain graph in combination with content analysis of the articles within the content-sharing paths and the domains that hosted them reveal a few salient categories of domains: a small set of prominent alternative media "hub" domains that produce source con- tent for the rest of the graph and occasionally re-publish each other's articles (21stCenturyWire, MintPressNews, GlobalResearch); two Russian government-funded outlets.

(RT, SputnikNews) that provide source content and occasionally amplify articles from the prominent hub domains; and a diverse set of alternative news aggregators that consistently amplify content from the peripheral sources.

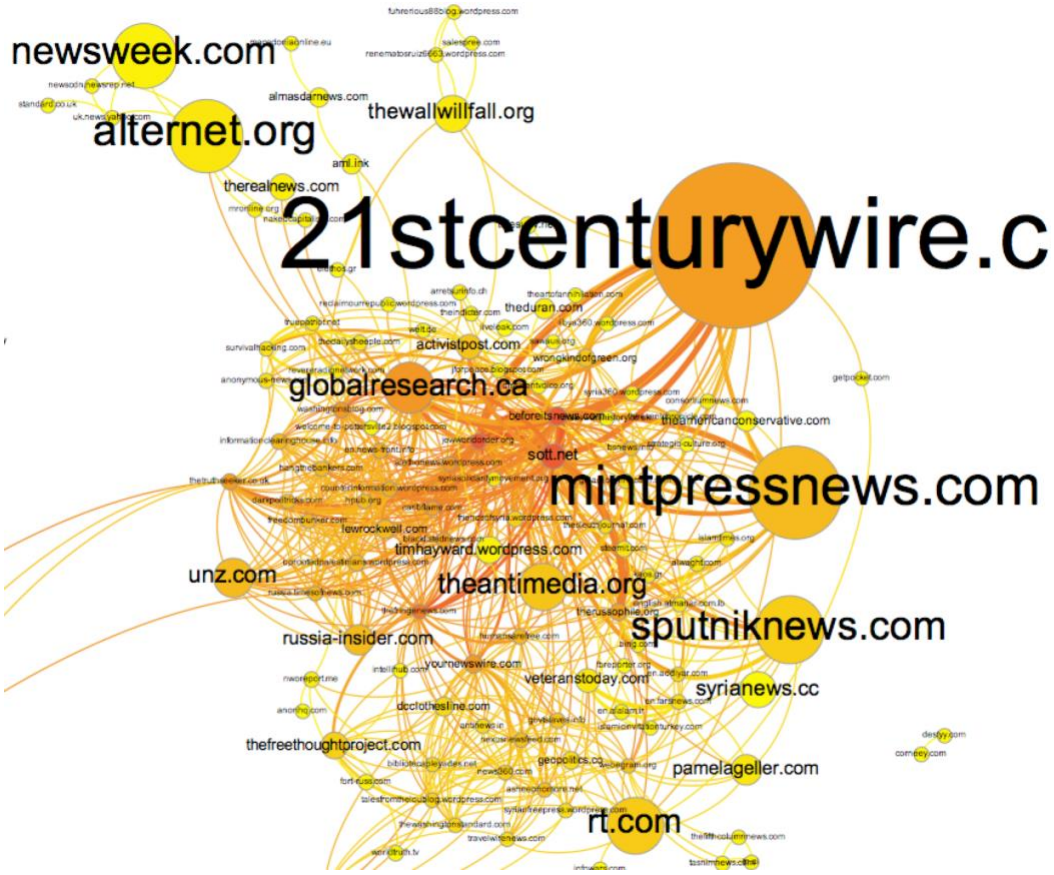

Figure 7.3: Close-up, expanded view of Clusters B and C; Colored by degree, Yellow=lower; Red=higher

As evidenced by their size in the graph, Cluster B includes many of the most highly tweeted domains in the data. Table 7.1 lists Top 10 domains within the WH collection in terms of tweet volume. Seven are located in Cluster B, and each of these was cited for multiple articles that were critical of the WH. Table 6.1 also provides the number of WH tweets that link to each domain and the "degree" of each domain in the graph—e.g., the number of other domains in the graph that are cited in the WH tweets for an article that has high similarity to one of the articles cited from this domain. We also note whether the domain is a source, a hub, or primarily an amplifier of content in this ecosystem.

Domain	Tweets	Degree	Network Role
21stcenturywire.com	3119	30	Central Hub
clarityofsignal.com	2391	1	Isolate
mintpressnews.com	1630	22	Central Hub
alternet.org	1219	6	Peripheral Hub
sputniknews.com	1110	16	Central Source
newsweek.com	1046	2	Peripheral Source
rt.com	879	17	Central Source
globalresearch.ca	707	33	Central Hub
theantimedia.org	682	19	Amplifier
unz.com	512	22	Central Source

Table 7.1: Top 10 Most Tweeted Domains in WH Dataset

Central Hubs in the Content Re-sharing Ecosystem

One key finding is that three of the most-tweeted domains (21stCenturyWire, MintPressNews, and GlobalResearch) generated the majority of source content for the re-sharing practices reflected in Cluster B. Interestingly, these domains were not exclusively source domains, but also borrowed content from each other, and published original content from some of the same authors. One author, Vanessa Beeley—a British journalist and leading critic of the WH—had original articles on each of the three domains, content which later appeared on one or more of the others. These three domains are central hubs of content re-sharing in the anti-WH conversation.

21stCenturyWire is by far the most cited domain in the WH collection—3119 tweets link to 26 distinct articles within this web domain. 13 of these articles appear elsewhere in the ecosystem, republished in all or large parts on other domains. Most of these articles were written by Beeley. 21stCenturyWire was founded by Patrick Henningsen, whose Guardian byline includes affiliations to RT and alternative news site Infowars. The website positions itself as grassroots, independent media that provides “news for the waking generation”. In the WH data, 21stCenturyWire is often the source of content that spreads across other prominent domains in the cluster (i.e., MintPressNews and theAntiMedia) and across several less trafficked domains (i.e., BeforeItsNews, YourNewsWire, and JewWorldOrder). 21stCenturyWire also republishes content that originally appears elsewhere in the graph, including articles from GlobalResearch, MintPressNews, and RT.

MintPressNews (MPN) is the third most-tweeted domain in the WH data—1630 tweets, 12 distinct articles. This domain is connected to 22 different domains in the graph, including many of the same domains as 21stCenturyWire. MPN describes itself as an "independent watchdog journalism organization" that features original reporting "through the lens of social justice and human rights." Its home office is located in Minnesota (USA), but the website covers both national topics and foreign affairs. In our data, their content is strongly pro-Syrian government and critical of the WH. Apart from their original content, MPN has multiple news and syndication partners whose content they frequently re-publish. Their original articles appear on other prominent domains (21stCenturyWire and GlobalResearch) as well as common amplifiers (i.e., theAntiMedia.org, YourNewsWire, BeforeItsNews, Sott.net, and JewWorldOrder). They also re-share content from other domains, including 21stCenturyWire, ActivistPost, and TheAmericanConservative.

GlobalResearch is an influential hub within the content sharing network that appears 8th on our most tweeted list, being tweeted 707 times for 17 different articles. Seven of these had high similarity with other articles in the WH data. GlobalResearch is operated by the Centre for Re- search on Globalization, "a non-profit independent research and media organization" that describes itself as a think tank on economic and geopolitical issues. The center is operated by Michel Chossudovsky, Professor Emeritus at the University of Ottawa. Chossudovsky has previously published claims of conspiracies related to world events—including that the September 11, 2001 attacks were not perpetrated by Islamic terrorists (Chossudovsky 2015). In our data, GlobalResearch is both a source and an amplifier. Underscoring its role in supporting this information ecosystem, of the top-10 most-tweeted domains, GlobalResearch has the highest degree, sharing articles whose con- tent overlaps with 33 different domains. Though it has a strong, multi-article connection to 21stCenturyWire and shares several overlapping amplifiers with 21stCenturyWire and MintPressNews, its content also reaches a subset of domains outside of that subnetwork (i.e., LewRockwell.com, WashingtonsBlog, and Freedom-Bunker).

Government-Funded News Outlets

Two government-funded news outlets (SputnikNews and RT) are also within the Top 10 most-tweeted domains and Cluster B. RT is a Russian government-funded media out- let that provides content to international audiences. Founded in 2005 with the stated purpose of improving Russia's image abroad, it has been accused of spreading disinformation and its US-based affiliate has been forced to register as a foreign agent (Stubbs & Gibson, 2017). Its current tagline is "Question More," and its content often encourages readers to question western and mainstream narratives of world events. RT.com was tweeted 879 times for 16 different articles which all take a critical perspective of the WH. Within our content-sharing paths, RT is primarily a source domain. Its content is re-shared entirely or in large excerpts across

17 other domains. Interestingly though, all of its edges have the weight of one (article). The graph shows a large number of domains borrowing a single RT WH article (not always the same one), rather than consistently re-publishing their content. In addition to content-sharing that we can see through the similarity graph, several articles within other domains embed videos from RT in their content.

SputnikNews, founded in 2013 as the replacement for the "Voice of Russia", is another Russian government-funded media outlet that features radio, television, and on-line content. Like RT, they have also been accused, primarily by western governments and media, of spreading disinformation and political propaganda that is favorable to the current Russian government (Dearden, 2017). They are slightly more highly tweeted than RT in our WH data— 1110 tweets for 15 articles, all critical of the WH. They are also primarily a source domain in this set. Their content is re-shared across 16 domains. Their most common amplifiers are Sott.net, theRussophile.org, and en.Addiyar.com (a Lebanese news outlet with a pro-Syrian government leaning). Each of those websites re-shared multiple articles from SputnikNews in our data.

Central Source Domains

In addition to RT and SputnikNews, there are two other central source domains—UNZ.com and ActivistPost.com. UNZ is an alternative media outlet founded by Ron Unz, a former (conservative) political candidate in California. The outlet's national security editor, Philip Giraldi, was the author of an article arguing that the WH are "a fraud". This article, hosted on the UNZ website, was tweeted 512 times within our data. It was also re-published on 21 other do- mains that appear in the WH data. UNZ therefore performed as a central source domain, though solely through sharing of this single article.

Another central source domain is ActivistPost, an alternative independent media outlet whose tagline is "propaganda for peace, love, and liberty." Their WH-related content is consistently critical, echoing many of the common narratives, claiming that they are a propaganda construct of mainstream media and western government interests. In terms of tweet count and compared to the more visible hubs, ActivistPost is relatively small—the domain was only tweeted 213 times. However, their role in the content sharing is significant. They are the source domain for eight different "paths" in the graph—e.g., eight of their original articles were re-published in all or part by other domains in the graph. In total, their content appeared in 18 domains, including central hub domains GlobalResearch and MintPressNews. All of their articles were authored by Brandon Turbeville and include a Creative Commons license that enables the free distribution of the work.

Alternative News Aggregators

Another core component of Cluster B is a large number of Alternative News Aggregators that repeatedly share content that originally appears elsewhere in the graph. The most prominent of these aggregator domains, in terms of tweet volume, is theAntiMedia.org. This website positions itself as the "homepage for the independent media movement", claiming to be a "non-partisan, anti-establishment news publisher and crowd-curated media aggregator." TheAntiMedia functions in part as a news aggregator, pulling in articles from other alternative and independent media outlets and mixing those with its own original articles. In the WH data, theAntiMedia was tweeted 682 times for one original and three borrowed articles (from MintPressNews and 21stCenturyWire). Most of these tweets link to a single article, re-shared from MintPressNews.

A large number of other domains in the graph function exclusively as amplifiers. In Figure 6.3, domains are colored by their degree (number of edges), from yellow (few edges) to red (many edges). Many of the most connected web domains (in red) are primarily content borrowers that repeatedly republish content from other websites in the graph. Sott.net, theRussophile.org, JewWorldOrder.org, and BeforeItsNews.com, are the domains with the highest degree in the graph, which also have thick, multi-article edges with the three central hub sites. All are exclusively amplifiers in this conversation—serially reposting content that first appeared elsewhere. Two other domains, YourNewsWire and FringeNews, are slightly less connected, but serve similar roles in a subnetwork in the lower-left-center of Cluster B. Many of these exclusively amplifier domains receive far fewer tweets for this content than other domains in our graph.

Another core component of the graph are the small (in terms of tweet volume) domains that are connected via thin edges to a relatively small number of domains. These domains typically appear in the graph for re-publishing one or two WH-related articles. Their relative positioning, near certain hubs and not others, may reflect a particular type of ideological targeting. For example, in the lower-left of the graph, near RT, TheFreeThoughtProject (a small source domain) and YourNewsWire, are domains like ASheepNoMore, GovtSlaves, and HumansAreFree which promote content questioning many mainstream narratives and suggesting large-scale geopolitical conspiracies. And in the upper left, near GlobalResearch and UNZ are a collection of libertarian-leaning domains (LewRockwell, FreedomBunker, HangTheBankers). Most of these political and/or ideology-centered domains appear in the graph for a single article re-shared from one of the source or hub domains. The domains do not necessarily amplify everything in the ecosystem, but may pick and choose content to reshare, as their focus is not necessarily on the WH, but on a specific worldview that these anti-WH narratives reflect.

Cluster C: A Peripheral Hub: Reframing Mainstream Content for the Alternative Ecosystem

Cluster C is a small (in terms of number of domains), distinct community that is loosely connected to Cluster B. Two of the Top 10 domains are in this peripheral cluster: Newsweek and Alternet. Newsweek, a "mainstream" media outlet, was cited for six articles in the WH data. However, we only found evidence of one of these articles being re-shared on other domains. Rather, Newsweek appears in Cluster C, and is peripherally connected to Cluster B, through the Alternet domain—due to one highly-tweeted article that described how a WH volunteer was caught on video (and subsequently fired for) disposing of the mutilated bodies of Syrian soldiers. Alternet, an alternative news site that both aggregates content and posts its own articles, re-published this article with attribution. Alternet's version, however, uses a different title, re-framing the original content to suggest that this incident was part of an ongoing pattern of misbehavior by WH volunteers, which aligns with other Alternet content critical of the WH. In this case, Alternet functioned as a peripheral hub, borrowing source content from "mainstream" media and reframing it to fit the predominant narrative of Cluster B.

Content Remixing Practices and Echo Effects

Though there are hundreds of distinct URLs in our tweet data, a significant percentage of the linked-to content is authored by a small number of prolific authors whose content is often re-shared and re-mixed elsewhere in the ecosystem. Beeley is the author of record for at least a dozen articles that appear in "paths" within the WH data—shared across several domains in Cluster B. In our tweet data, she was cited for original content in at three source domains: 21stCenturyWire, MintPressNews, and TheWallWillFall. She also repeatedly appears as a secondary source in articles by other authors through quotes, excerpts, and embedded videos of interviews. Similarly, Turbeville, who primarily publishes in ActivistPost, authored eight articles that were source articles for multiple content-sharing paths across the ecosystem.

The following example traces a single, short path that includes both Beeley and Turbeville and illustrates several of the diverse content remixing practices and echo effects that manifest in this ecosystem. On May 2 2017, ActivistPost published an article by Turbeville titled "Photos from Syria Show White Helmets and Nusra/Qaeda Are The Same Organization" (Turbeville 2017). This article used photos and videos that Beeley captured while in Syria and posted on her Facebook account. Later in the article there is a textual excerpt, citing Beeley, that describes the content in one of the videos. At the foot of the article there are seven links to other articles about the WH: four authored by Turbeville and published on ActivistPost, and three authored by Beeley published on 21stCenturyWire.

This same article (including Beeley's photos, videos, and excerpt) is published on MintPressNews on the same day, citing Turbeville and ActivistPost, but removing the links to related content on 21stCenturyWire and ActivistPost. Thirteen days later, on May 15 2017, Beeley published the same ActivistPost article on TheWallWillFall, her personal blog. In this version the article is titled "WHITE HELMETS: Living next door to Al Qaeda in Aleppo" and Beeley is listed as the author. However, below an additional image that did not appear in the original version, Turbeville at ActivistPost is cited as the author, followed by the original article in its entirety—including the photos, videos and the quote from Beeley. There are now 13 links to other related White Helmet-related articles—ten of them on 21stCenturyWire, plus TruthDig, WrongKindofGreen, and Wikipedia. These circular citations and remix practices create another kind of echo effect within this system.

Discussion: Alternative Media Echo-System

In this research, we explored content sharing practices across media domains, using URL links in tweets to capture domains that were active in an online conversation, and an article similarity metric to determine domains that shared articles with high similarity. The conversation we focused on—views of the WH in relation to the ongoing civil war in Syria—is a highly contested one with geopolitical significance. Using content similarity, we generated a network graph of shared content, and utilized that network graph to conduct a mixed-method, interpretative analysis of the structure and dynamics of content sharing across active domains.

Our analysis uncovered sharing practices among both mainstream and alternative media domains, as well as a few aggregator domains that bridged the two. Articles that originated (or echoed) within mainstream media (Cluster A) were largely supportive of the WH, reflecting some of the critique (from those in Cluster B) that the WH are favored by Western, mainstream media. Our graph shows a couple of clear examples of content sharing of and by mainstream media. In particular, Cluster A represents (primarily) a single "path" of an article originally posted by the AP and re-published by dozens of domains, including global and local mainstream media outlets and news aggregators. This activity is not insignificant, and clearly demonstrates that 1) mainstream media were participating in the WH conversation, primarily through the production and diffusion of pro-WH narratives; and 2) content-sharing is a component of mainstream news distribution.

However, for the WH conversation happening on Twitter during the summer of 2017, the vast majority of content production and amplification occurred on and through the alternative media domains represented

in Cluster B. The content shared within these domains was strongly critical of the WH, promoting several related narratives that framed them as a propaganda construct and accused them of aiding, working with, or being terrorists. Using tweets as seed data, we were able to unwind trajectories of content sharing across domains, articulating an alternative media ecosystem—or "echo-system"—of 130 distinct domains that provided source content for, or re-published existing content from, another domain in Cluster B. Analysis of this content-similarity structure, the domains that played significant roles within it, and the practices of content-sharing and remixing across domains, provides several interesting insights and leads to additional questions about how and why this echo-system has these properties.

Explicit Critique of Mainstream Media

One widespread theme within the domains that constituted the alternative media ecosystem in Cluster B is criticism of mainstream media and skepticism or outright rejection of its narratives. Messaging across many domains suggests that the mainstream media is lying to members of the public who should come to this website to get the truth. While some of these domains are extremely conspiratorial in nature (a synergistic worldview to the anti-media arguments), others are more focused on questioning the motives of western governments, in some cases specifically around conflicts in the Middle East and positioning mainstream media as tools of those governments in those conflicts.

Support of Russian Government

Perhaps not surprising, considering the position of Russia as an ally of the Syrian government (which views the WH as assisting rebel forces), many of the domains in this ecosystem are explicitly supportive of the Russian government. Beyond RT and Sputnik, there are a few other sites that focus on Russia-related topics from a point of view favorable to the current regime: Russia-Insider.com, Russophile.org (Russia News Now), and Fort-Russ.com. Many of the other domains in Cluster B feature content supportive of Russian geopolitical positions (abroad) and specifically resistant to accusations that Russia had an impact on recent elections in the US and elsewhere.

Shared Content across Ideologically Diverse Sites

But perhaps more interesting—or more impactful—than the commonalities across domains are the differences between them. Superficially, many of the domains in the alternative media echo-system articulated here appear to promote different ideologies. Consider a selection of domains that appear in this echo-system—i.e., MintPressNews, JewWorldOrder, LewRockwell, FreedomBunker, UprootedPalistinians, TruePatriot, TheDailySheeple, TheFringeNews, Anonymous-News,

MakeWarHistory, ActivistPost, and TheRussophile. This list includes websites with strong political themes reflecting distinct (and in some cases, seemingly conflicting) ideologies—including antiimperialist left, libertarian, conservative and alt-right; as well as other more niche ideological leanings, including explicit anti-Semitism. These websites are publishing the same content, but inside very different wrappers. The content itself is not necessarily tailored for each community (though each domain may select the articles most likely to resonate with its audience), but it is packaged up for them, appearing within a domain that features other material that may appeal to a reader's existing ideology. The effects of this kind of sharing may be to draw people from diverse, niche, political and ideological communities into a set of common narratives. We may think of these niche communities as being isolated and distinct, but here they are connected (in terms of common content) with other quite different communities.

Future Work to Understand the Drivers and Impacts of the Information Echo-System

Although in our Twitter seed data we found hundreds of domains and dozens of articles, we discovered that a small number of authors are responsible for a large proportion of highly-cited articles. We uncovered instances of circular attributions when authors cite themselves from other sources. We also demonstrated how similar content is repeated across in some cases vastly different domains. While such practices could reflect a more coordinated strategy, our evidence suggests that this complex ecosystem both has organic properties and is strongly influenced by a small set of politically, ideologically, and financially motivated actors and organizations.

Prior work has found that exposure to repetition of misinformation (which is not necessarily intentional) leads to a fluency effect—as people become familiar with claims they are more likely to judge them as true (Nyhan and Reifler, 2012). By disseminating unverified or falsified stories to audiences through various channels, and from multiple sources, people may begin to assume they are true, regardless of the credibility of the individual sources (Paul & Matthews, 2016). Although this research does not provide evidence of a coordinated strategy, the distribution of content across these seemingly distinct domains resembles a kind of intentional "astroturfing" campaign (Ratkiewicz et al., 2011) meant to exploit these cognitive biases. Future work is needed both to better understand the mechanisms underlying these patterns of content sharing and their effects on online audiences.

Limitations

One limitation of this work is the use of tweet data to "seed" the investigation of the surrounding ecosystem. This method had the advantage of allowing us to measure the impact of these domains on the online conversation, but the disadvantage of having our view of the content-sharing shaped by the contours of a single social media platform. The vast majority of tweets with URLs in our dataset link to domains that were critical of the WH and appear in Cluster B in the graph—the area that we have termed the alternative media echo-system. Additionally, and perhaps consequently, the majority of content-sharing "paths" across domains are in this area of the graph as well. It is likely that this tweet-seeded method resulted in a better view of content-sharing practices across alternative media than across mainstream media. Other less significant limitations include the loss of some URLs (and the articles/domains they pointed to) that we were unable to resolve, and the exclusion of article content that was not publicly available when we completed our automatic and manual scraping (in November 2017).

Acknowledgments

This research was supported by U.S. NSF grant #1715078 and ONR grants N00014-17-1-2980 and N00014-18-1-2012. We wish to thank students who provided significant assistance to this project, including Gordon, Ostin, and Vera.

<<END OF PUBLICATION>>

Summary and Takeaways

In this chapter I set out to understand how government-controlled and alternative media websites shape and propagate competing narratives about the WH. In the first part of the work, I presented the salient pro- and anti- WH narratives that were present in articles published on mainstream, alternative, and government-controlled media. The narratives identified here are important in the following studies, as the analysis will focus on the role of social media platforms in the information operation surrounding the WH.

The findings in this study build upon the findings in the Omran case study (Study 1, Chapter 5) where new "facts" and "evidence" were introduced to the Twitter conversation, and these in turn formed into disruptive narratives, and also the narrative analysis conducted in Study 2 (Chapter 6). In Studies 1 and 2 I did not follow URLs within tweets (beyond identifying the salient websites linked to in tweets and retweets in Study 2). However, in this study, by following URLs in tweets about the WH and analyzing the articles on

a variety of websites, it is revealed that some of this is seeded in the hybrid assemblage of government-controlled and alternative media websites. In particular, Russian government-controlled media RT and Sputnik served as source content domains for a wider network of alternative media websites that would republish their articles. This raises the questions surrounding the independence of these websites: alternative media outlets generally position themselves as independent (Kinix, 2011) but here they are publishing content that is the product of government-controlled media, which are working in alignment with the government that they represent.

The alternative media outlets used to host anti-WH content were established, and actively producing content about other subjects also. In other words, these were not websites set up for the specific purpose of publishing stories about the WH: they were websites with established followings that served a "bridging" function (Freelon, Marwick, & Kreiss, 2020) between state media (or other websites in the alternative media "echo-system") and audiences from distinct communities. This illustrates how information operations can target, and work to onboard individuals and groups with different worldviews into shared narratives surrounding the WH.

References to Problematic or Misleading Content

For presentation in this dissertation, I have chosen to separate academic references from references to content that were problematized in the chapter to reduce the capacity of this campaign to impact citation metrics and search results for academic literature. Those references (which appeared as subscript inline citations) are listed below and not in the references section at the end of the dissertation.

Beeley, V. 2017. EXCLUSIVE: The REAL Syria Civil Defence Exposes Fake 'White Helmets' as Terrorist-Linked Imposters. 21st Century Wire, September 23, 2016

Chossudovsky, M. 2015. Saudi Arabia's Alleged Involvement in the 9/11 Attacks and the 28 Pages: "Red Herring", Propaganda Ploy. Global Research.

RT. 2017. UK's Shadowy £1bn Conflict Fund Being Kept Secret from MPs. RT, March 7, 2017

Turbeville, B. 2017. Photos From Syria Show White Helmets And Nusra/Qaeda Are The Same Organization. ActivistPost. May 2, 2017

Chapter 8:

Multi-platform Information Operations: Mobilizing Narratives & Building Resilience through 'Big' & 'Alt' Tech (Study 4)

Abstract

Despite increasing awareness and research about online strategic information operations, there remain gaps in our understanding, including how information operations leverage the wider information ecosystem and take shape on and across multiple social media platforms. In this paper we use mixed methods, including digital ethnography, to look beyond a single social media platform to the broader information ecosystem. We aim to understand how multiple social media platforms are used, in parallel and complementary ways, to achieve the strategic goals of online information operations. We focus on a specific case study: the contested online conversation surrounding Syria Civil Defense (the White Helmets), a group of first responders that assists civilians affected by the civil war within the country. Our findings reveal a network of social media platforms from which content is produced, stored, and integrated into the Twitter conversation. We highlight specific activities that sustain the disruptive narratives and attempt to influence the media agenda. And we note that underpinning these efforts is the work of *resilience-building*: the use of alternative (non-mainstream) platforms to counter perceived threats of 'censorship' by large, established social media platforms. We end by discussing the implications on social media platform policy.

Introduction

While there is increasing awareness and research about online strategic information operations—broadly defined actions by state and non-state actors to distort the information ecosystem to achieve strategic goals, through methods such as the dissemination of disinformation, amplification of specific accounts or content, or the manipulation of the information space (Weedon et al., 2017)—we do not fully understand the multi-platform dynamics of these operations. While information operations attempt to leverage the wider information ecosystem (Morgan & DiResta, 2018), focusing on a single platform oversimplifies the information ecosystem, providing a limited view of the dynamics of information flow and overlooks how variations in social media platforms contribute to this flow (Hughes et al., 2016).

The aim of this research is to look beyond a single social media platform to the broader information ecosystem to understand how multiple social media platforms are used, in parallel and complementary ways, to achieve the strategic goals of online information operations. In particular we interrogate the role that different social media platforms play. In other words, how are multiple social media platforms—as sociotechnical infrastructure comprising technology, users, norms, and business practices (Berkman Klein Center, 2020)—used? We also focus on the collaborative activities of actors on and across social media platforms to unpack the kinds of participatory "work" that users engage in while involved in multi-platform information operations.

The context of this research is the Syrian Civil War, and in particular the online conversation surrounding Syria Civil Defense (the White Helmets), a humanitarian response organization that serves as first responders, assisting civilians affected by the military actions within the country. The White Helmets (WH) predominantly operate in rebel-held (government-opposed) areas of Syria, and through their work have highlighted atrocities carried out against civilians by the Syrian government and their allies, including Russia. As a consequence, the group has been subjected to a long-term information operation that seeks to delegitimize them and weaken their support, particularly in the West (Solon, 2017; DiGiovani, 2018; Levinger, 2018; Starbird et al., 2018; Starbird, Arif, & Wilson, 2019; Wilson & Starbird, 2020).

Our point of departure is a Twitter dataset of English-language tweets about the WH, collected during 2017-2018 and almost four years of longitudinal observation and analysis of the online discourse surrounding the WH (e.g., Wilson, Zhou, & Starbird, 2018; Starbird et al., 2018; Wilson & Starbird, 2020). In this paper, we use mixed-methods research, including investigative digital ethnography (Friedman, 2020), to extend this "seed" data to look beyond Twitter to other social media platforms from which content was sourced,

to understand how information operations function within the broader ecosystem (Morgan & DiResta 2018). While this approach means we provide a "Twitter-centric" perspective of the information ecosystem, it does offer insight into how the discourse is shaped through different configurations of social media platforms, and how content flows "downstream" into the Twitter conversation—the breaking news platform of choice of major politicians or journalists (Lieberman 2020).

The empirical findings presented in this research provide insight into the established (and largely successful) multi-platform information operation conducted against the WH. We build a picture of the surrounding ecosystem of social media platforms from which content is produced, stored, and integrated into the Twitter conversation. We extend current understanding of the collaborative aspects of online disinformation (Wilson, Zhou, and Starbird, 2018; Starbird, Arif, and Wilson, 2019) revealing the "work" involved in conducting multi-platform information operations. One aspect of this is the work to sustain the disruptive narratives and attempt to influence the agenda through *content production*, *content synthesis*, and *content mobilization* to gain the attention of specific audiences at specific times. But underpinning these efforts is the work of *resilience-building*: the use of "*alt-tech"*—alternative (non-mainstream) platforms—to counter perceived threat of censorship from "big-tech", the large, established social media platforms. We end by outlining the implications of this work, particularly with regard to social media platform policy.

Background

The Syrian Civil War

The context of this research is the Syrian Civil War, an ongoing conflict that began in 2012 after Arab Spring protests against the rule of Syrian President Assad turned violent. The conflict descended into a multi-sided, internationalized conflict that involves global powers such as Russia and Iran, which are allied with the Syrian government; the US, UK and other global and regional allies, which support the anti-Syrian government 'rebels'; and ISIS, a now-weakened group that was attempting to establish a caliphate across the region (Gilsinan, 2015; Zaram, 2013). The ongoing conflict has taken many lives and displaced millions more: accurate figures for the number of people killed are disputed, but the Syrian Observatory for Human Rights (2020) put the figure at almost 600,000. The United Nations Refugee Agency in 2018 reported that 5.6 million Syrians were refugees in other countries while a further 6.6 million were internally displaced (United Nations High Commissioner for Refugees, 2018).

Alongside the kinetic war ‘on the ground’, the Syrian Civil War also manifests itself online as a battle of public opinion. Syria and its allies, including Russia and Iran, take one “side”, while on the other “side” is the US and its allies, including the UK. The research contained herein examines the online manifestation of the Syrian Civil War, and how this takes shape on and across social media. A particular focus of the online discourse is the White Helmets: a volunteer rescue group which receives financial backing from countries such as the US and UK and is applauded by western governments for their volunteer response activities in the war-torn country but is viewed by Syria and its allies as a ‘soft power’ in the region, operating on behalf of western governments, or as aligned with Islamic terrorist organizations.

Syria Civil Defense — The White Helmets

Amongst this complicated geopolitical terrain operate Syria Civil Defense, also known as the White Helmets (WH) due to their distinctive headgear. The WH first emerged as loose collaborations between local volunteers fulfilling the role of first responders, rescuing civilians and retrieving bodies in the aftermath of military attacks and airstrikes conducted by the Syrian government or their allies, Russia, in “rebel”-held areas that opposed the Syrian government. After some of the volunteers reached out for financial help from foreign humanitarian groups, James Le Mesurier, a former British army officer, established Mayday Rescue as a means to channel aid to the group, money that came from the US, UK, Canada, Denmark, Germany, and the Netherlands (Chulov, 2020).

The group actively promoted their work to garner attention for the plight of the Syrian people caught up in the war: they captured rescues using helmet cameras (Solon, 2017) and were the focus of a Netflix documentary that won an Oscar in 2017 (Larkin & Lewis, 2017). Their video footage documented atrocities committed by Syria and their allies, Russia, that the WH observed when responding to attacks. Such evidence directly challenged Russia’s claims that it respected international law and was upholding global stability through its role in the Syrian conflict (Levinger, 2018). In response, and in a bid to discredit the group and undermine their reporting of alleged atrocities, the WH became the targets of a sustained online disinformation campaign designed to influence public perceptions of the group (Solon, 2017) and in some cases justify physical attacks against them (Grey-Ellis, 2017).

For the most part, this information operation was effective in inundating the information space with negative content about the WH and nurturing a community that was willing and able to further advance online activities against the group. The success of the online operation against the White Helmets was noted

by prominent Russian General Gerasimov in early 2019[32] (Kramer, 2019), and as the group's reputation became more tarnished (e.g., by associating them with known terrorist groups), it became easier to justify physical attacks against them (Grey-Ellis, 2017). As of October 2018, it was reported the once 3000 strong group had lost more than 250 members to the violence (Di Giovanni, 2018). In November 2019, James Le Mesurier (founder of Mayday Rescue, the organization providing financial support to the WH), was found dead in Turkey with injuries consistent with falling from a height (Yee, 2019; Loveluck, 2019). The following investigation found that his death was caused by suicide and this has been attributed, at least in part, to the persistent efforts to discredit him through a disinformation campaign (Chulov, 2020).

Methods

This research is the culmination of almost four years of research of online activities surrounding the Syrian Civil War and the White Helmets in what can be described as an *investigative digital ethnography* (Friedberg, 2020)—a long-form, detailed investigation of a specific case study. While ethnographic insofar as the research was situated within a specific space where there exist distinct rituals and cultural production, rather than engaging with the subjects directly or becoming an active participant in the culture, we engaged longitudinally with the digital traces that were left behind as artifacts of online interactions[33] (Friedberg, 2020). We leveraged these traces through a mixed methods approach, using both quantitative and qualitative methods to query and analyze the data in an iterative way. We generated queries, plots, and network visualizations to understand the structure and higher-level patterns of the data, plus conducted qualitative analyses of thousands of tweets, and hundreds of articles, videos, and other social media posts. Through this research we were able to characterize communities of Twitter accounts based upon their stance toward the WH (Wilson & Starbird, 2020), and noted—among both supporters and detractors of the WH—the heterogeneous make-up of the communities as a confluence of online information activists, Western journalists, politicians, government-controlled media, and agents of governments that loosely collaborate through computer-supported cooperative work (Wilson, Zhou, & Starbird, 2018; Starbird, Arif, & Wilson, 2019). This work takes place with little evidence of top-down orchestration (Starbird, Arif, & Wilson,

[32] General Gerasimov realized the asymmetrical possibilities of leveraging technologies and information networks to reduce the fighting potential of the enemy earlier in the decade in what was termed the Gerasimov Doctrine (Kramer, 2019).

[33] At times we were brought into these discourses, such as when we shared research output, for example a network visualization, and were the recipient of tweets of support and condemnation from both "sides". With the exception of a small number of journalists, we chose to not directly engage with accounts on either side of this discourse as we focused on our data-driven investigation.

2019), in contrast to the e.g., the highly-coordinated efforts of the Internet Research Agency circa 2016-17 (Arif et al., 2018).

We began with the same Twitter dataset, and the research presented in this paper is supported by the longitudinal observation and analysis that came before it. In this research we adopt a mixed methods approach, extending upon methodologies of crisis informatics (e.g., Palen & Anderson, 2016; Palen et al., 2009), studies of misinformation (e.g., Maddock et al., 2015; Starbird et al., 2015), and document-driven digital trace ethnography (e.g., Geiger & Ribes 2011; Ribes & Geiger 2010) to examine the multi-platform dynamics by following both specific content and the actors involved in its production and dissemination. We combined both an actor and information-focused approach: identifying the individual accounts and communities that are involved and following these across the information ecosystem; and also tracking specific media and its appearance on and across multiple social media platforms[34] through URL (link) sharing (Friedberg, 2020). By applying quantitative, qualitative, and visual techniques to this data we sought to build a comprehensive understanding of this phenomenon, including revealing the activities—or 'work'—of disparate actors that collaborate to sustain an information operation over the long term. In the following subsections we describe in more detail the specific methods we used to prepare and analyze our Twitter data, including how we extended this dataset to look beyond Twitter to other social media platforms by leveraging URLs as digital traces.

Twitter Data Collection

This research is centered around a Twitter dataset of English-language tweets referencing the "white helmets" or "whitehelmets" (case insensitive), collected in real-time using the Twitter streaming API between May 27, 2017 and June 15, 2018. The data includes tweets and retweets where the term "white helmets" or "whitehelmets" appears in the text field of the tweet, or in the case of 'quote' tweets (retweets

[34] We define social media as "forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content" (Merriam-Webster, 2020). The use of the term 'platform' is summarized by Gillespie (2010): "The more conceptual use of 'platform' leans in all of the term's connotations: computational, something to build upon and innovate from; political, a place from which to speak and be heard; figurative, in that the opportunity is an abstract promise as much as a practical one; and architectural, in that [they are] designed as an open-armed, egalitarian facilitation of expression, not an elitist gatekeeper with normative and technical restrictions" (p.352). In this way, the use of the term platform can be viewed as strategic— consistent with rhetoric surrounding the democratizing potential of the internet but at the same time serving to downplay the role of the social media companies, highlighting that they are not producing information but are just hosting it. This shifts the responsibility of the content to the user and therefore limits the liability of the company, which can deny their role as a publisher. This is supported through Section 230 of the Communications Decency Act (1997), which offers the social media platforms some legal protections from what their users may say and do on their infrastructure (47 USC 230, 1997)

with comment), within the text of the comment tweet or the embedded quote tweet. The dataset contains 913,028 English-language tweets from 218,302 distinct Twitter accounts. Unfortunately, between February 7-24 2018, our collection infrastructure suffered an outage and we are therefore unable to account for activity at that time (annotated by the grey area on the temporal plot in Figure 8.3).

These data—and subsequent analyses—are limited to English-language tweets, and we focus on the aspects of the WH discourse that was designed for and taken up by English-speaking audiences. The aim here is to encapsulate a specific and strategic element of the WH discourse: the group has made significant and sustained efforts to communicate in English to western and global audiences, while similarly a significant portion of the anti-WH campaign has occurred within English-language discourse, attempting to counter the WHs' communications by targeting the same or similar audiences. A limitation of this approach is the acknowledgement that this represents just part of the discourse around the WH, and that the conversation may look different in other languages.

Characterizing the Communities in the WH Discourse on Twitter

As detailed in Wilson and Starbird (2020) we generated a retweet network visualization by identifying retweet trajectories to detect communities of accounts. We used the Louvain algorithm, a greedy optimization method that optimizes the relative density of edges inside a community (modularity) compared to that outside the community (Blondel et al., 2008). Since edges represent retweets between a pair of accounts, the algorithm effectively identified two communities[35] of accounts that retweet one another (Figure 8.1). This quantitative approach to community detection was supported by our qualitative analyses of hundreds of tweets, articles, and videos over the course of almost 4 years, enabling us to characterize the communities based upon their narrative stance toward the WH. We describe the communities here because they are significant as we describe the methods applied to the data herein.

[35] Of the 218,302 distinct Twitter accounts in the White Helmets dataset, 168,110 (77%) were classified as members of these communities (Table 8.1). The remaining 50,192 accounts were not included because the accounts did not retweet nor were retweeted by another account (32,638 accounts), or; the accounts formed their own, smaller, distinct communities (17,554 accounts). Qualitative analysis of the tweets in these smaller communities revealed that the conversation was not about the White Helmets in Syria, but rather a motorcycle group in the UK and the headgear of a US football team. These unrelated accounts and their tweets were excluded from the subsequent analyses.

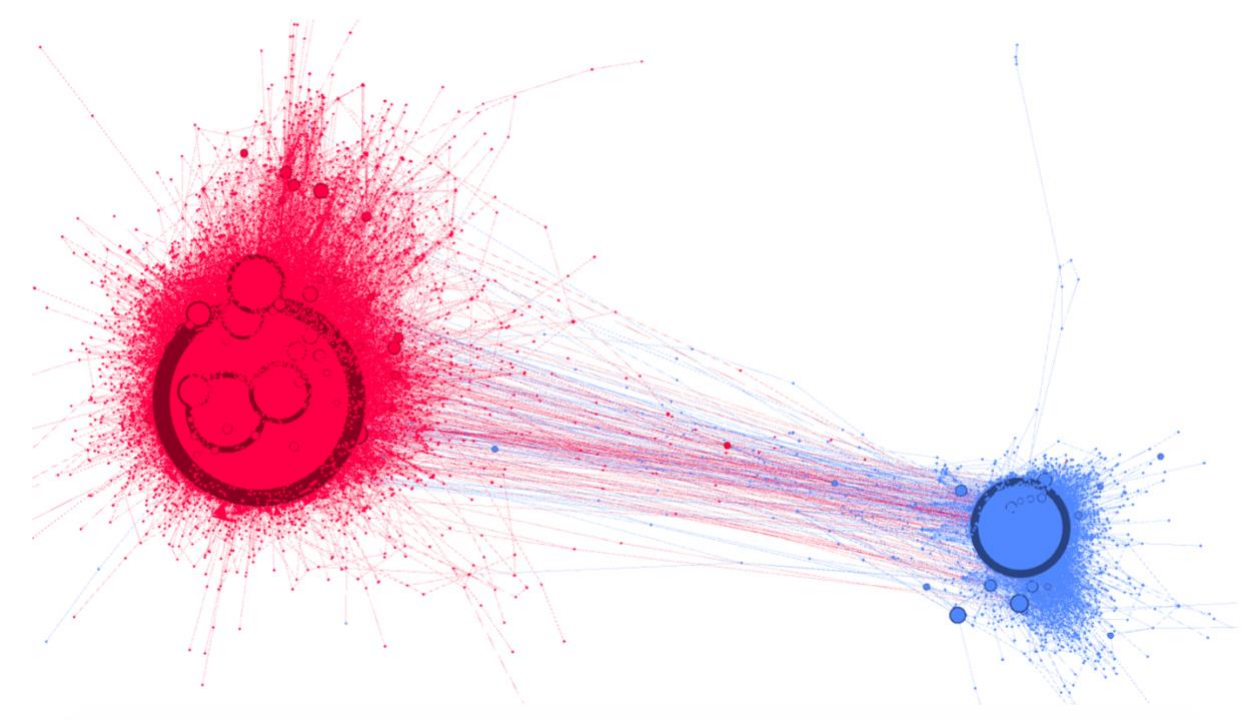

Figure 8.1: A retweet network visualization of the White Helmets conversation on Twitter illustrating the two distinct communities involved in the conversation: the anti-WH community (red, on the left), and the pro-WH community (blue, on the right). The nodes (circles) represent Twitter accounts which are sized relatively according to the number of times that account was retweeted. The edges (lines) between nodes represent a retweet between that pair of accounts. This is a reproduction of the network visualization published in Wilson and Starbird (2020)

The pro-WH (blue) and anti-WH (red) communities on Twitter

The first community—the pro-WH community (blue)—was almost exclusively supportive of the White Helmets. The *blue pro-WH* community contributed to and promoted the broad narratives that the White Helmets are a humanitarian response organization that provides firefighting, medical aid, rescue, response, and body recovery in rebel-held areas of Syria; that military actions taken by the Syrian/Assad regime are injuring and killing Syrian civilians, including through the use of chemical weapons; and the White Helmets are the target of a disinformation campaign.

In contrast, content from the second (*red, anti-WH*) community was almost exclusively critical of the White Helmets. Within that network multiple, often conflicting, narratives surfaced, all of which functioned to delegitimize the White Helmets through a range of ad hominem attacks, including claims that: the White Helmets are a propaganda construct of the West; the White Helmets are assisted by "Western" and "mainstream" media; the White Helmets are not humanitarian actors; the White Helmets are terrorists; the White Helmets are "crisis actors" who stage impacts to make the Syrian government and their Russian

allies look bad; the White Helmets stage chemical weapons attacks; the White Helmets are agents of the US and UK governments.

Looking Beyond Twitter to Understand Cross-Platform Activities

We next looked beyond Twitter to other social media platforms from which content was sourced[36] — in other words the platforms that the two communities on Twitter linked-to within their (re)tweets. This can provide insight into the ecology of the wider information space (Poell, 2014, Foot, Schneider, & Cornfield, 2006; Marres & Rogers, 2005; Segerberg & Bennett, 2011) and how information operations function within this ecosystem (Morgan & DiResta 2018). We unshortened URLs in tweets and resolved to their destination domains. By ranking the domains by the number of (re)tweets[37] that linked to them, we identified the five social media platforms that were linked-to in at least 1000 (re)tweets posted by accounts in pro- or anti-WH communities on Twitter: YouTube, Steemit, Facebook, Medium, and Liveleak.

To scope the research, we determined the 25 most (re)tweeted distinct URLs from each platform and followed those URLs to gather additional data about the content being shared. We focused on the most (re)tweeted as a proxy for what was considered the most important by users (Poell & Borra, 2012; Cha et al., 2010; boyd et al., 2010). For the 25 most (re)tweeted posts on each platform we collected the publicly available metadata (e.g., title, author/channel, published date, interactions (likes, shares), tags); and analyzed (i.e., through watching/reading) the content (video/post/article). We built a small database of the public metadata, notes that summarized the contents of the post/video, the narrative stance of the content toward the WH, and a breakdown of (re)tweets from the pro- and anti-WH communities on Twitter. Collectively, these data enabled us to build a qualitative understanding of the 125 posts and their contribution to the WH conversation on Twitter.

Examining Twitter Content Integration Through Retweet Cascades

The integration of content to the Twitter conversation was examined through retweet cascades—tweet-retweet combinations visualized temporally. We first identified "introductory" tweets—a tweet, reply, or quote tweet (non-retweet)—that contained a URL, which represented digital links to other regions of the

[36] This approach places Twitter temporally downstream insofar as content shared to Twitter has already been produced and hosted elsewhere.

[37] We focused on both tweets and retweets (herein referred to as "(re)tweets") so we could analyze both the introduction of content to Twitter (which happens through non-retweet tweets, replies, or quotes) and dissemination of content (which occurs through retweets).

information ecosystem and the point in time that the content was introduced into the Twitter conversation. Introductory tweets were assigned a *retweet sequence* of 0. The second step involved finding all retweets of each introductory tweet (if they existed) and ordering these temporally. The first (temporal) retweet of an introductory tweet was assigned the *retweet sequence 1*, the second as 2, etc. In the third step, the cascades were plotted to axes, with date-time on the *x*-axis and the retweet sequence on the *y*-axis. The points on the graph were shaded by community membership (red/blue), assigned a shape (filled circle: tweet, quote tweet; filled diamond: reply; empty circle: retweet), and relatively sized according to the number of followers of the (re)tweeting account (Figure 8.2).

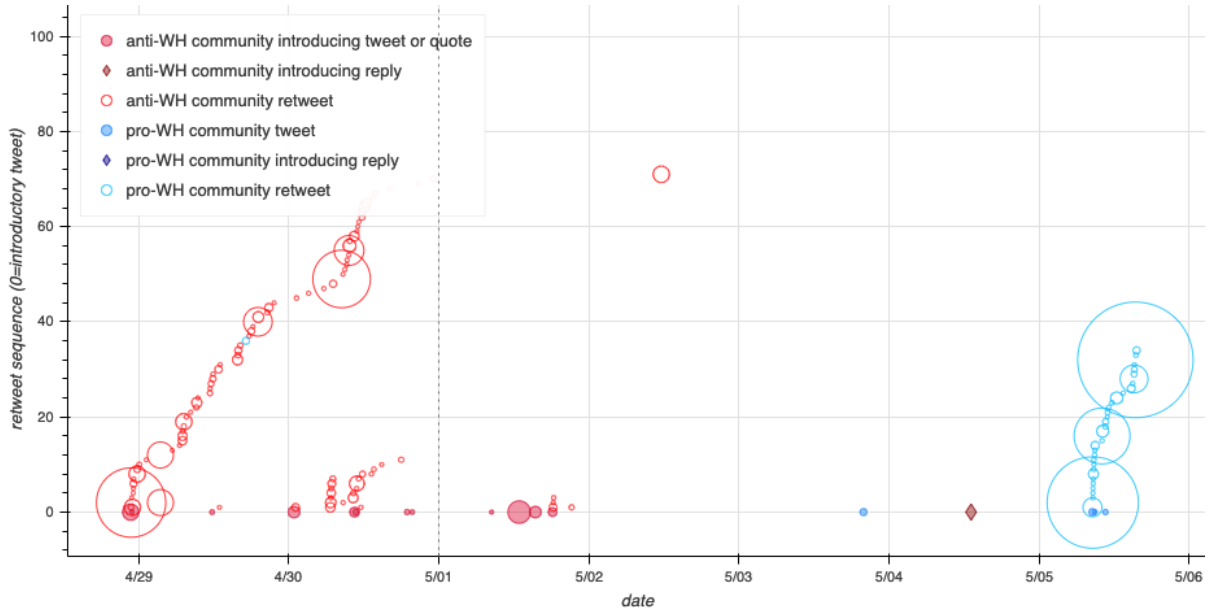

Figure 8.2: A retweet cascade plot used for examining content integration to Twitter. In this example, the retweet cascades show content introductions from Medium (different distinct URLs) by accounts in the anti-WH red community on the left, and the pro-WH blue community to the right. The pro-WH content is introduced as a tweet (darker blue circle at y=0, and then retweeted by some accounts with larger followings. It receives more than 30 retweets over the course of two hours. Various anti-WH content is introduced, several of which are not retweeted, including the reply just to the left of the pro-WH cascade, which is not retweeted.

The retweet cascade plots enabled interactive examination of the integration of content into the Twitter conversation, for example they reveal which community on Twitter introduced (through introductory tweets) or disseminated (through retweets) specific pieces of content. Retweet cascade plots were used as an investigative tool in the analyses presented below to help understand when and by whom content was introduced to the Twitter conversation and which accounts subsequently retweeted it. Due to space constraints, we do not include these plots each time they were used.

Findings

We first describe the structure of the WH conversation on Twitter then examine each of the five social media platforms that played a role in shaping the WH discourse. We look at key platform features, provide an overview of the content, describe the introduction of content to Twitter, and summarize how the platform is 'used' within this information ecosystem.

The WH Conversation on Twitter

As previously reported (anonymized for review), the anti-WH community (red) on Twitter is larger, containing 57.96% of all accounts in the WH discourse, and more active, generating 75.26% of (re)tweets in the conversation (Table 8.1). On average an account in the red community generated, over the course of our data collection 6.5 WH-related (re)tweets, versus the 3 by accounts in the blue community.

community	distinct accounts	total tweets	tweets	retweets	% retweets	% url	% replies	mean num tweets per
blue, pro-WH	70,670	207,938	27,708	180,230	86.67	22.72	2.42	2.94
red, anti-WH	97,440	632,638	122,966	509,672	80.56	36.96	9.20	6.49

Table 8.1: Overview of the blue (pro-WH) and red (anti-WH) communities on Twitter

The red anti-WH community is also persistently more active: (re)tweet volumes from the blue community seldom surpass volumes of the red. There are some exceptions to this, and we see increased activity from the pro-WH community in response to several key events: the murder of seven members of the WH in August 2017; an article in the Guardian (UK newspaper) that brought attention to the disinformation campaign being conducted against the WH in December 2017; and news that the US government had ceased its funding for the group in May 2018.

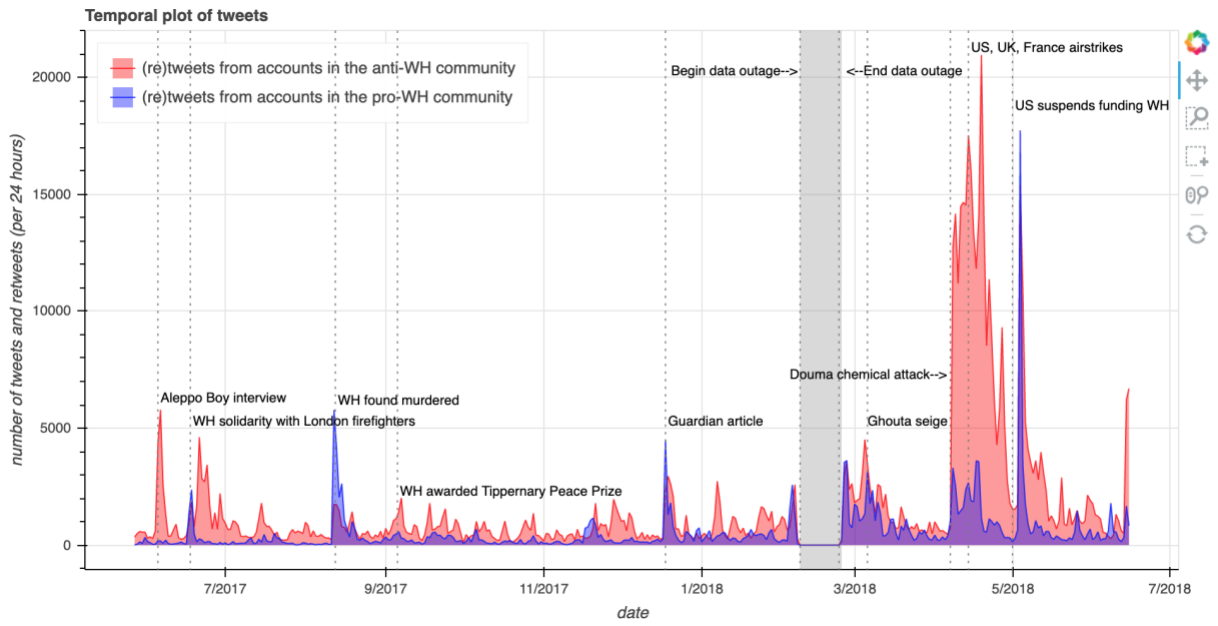

Figure 8.3: Temporal plot showing the number of (re)tweets per 24-hour period from the red and blue communities on Twitter. The plot is annotated with some key events regarding the WH and the Syrian conflict. Note the marked increase in activity—particularly from the red anti-WH community—after March 2018, and the data outage (grey area) in February 2018.

We see marked increase in activity on Twitter in April 2018, which is more pronounced and sustained in the anti-WH community. This period coincides with some major events in the Syrian conflict, including the military offensive by the Syrian Army to capture for rebel-held Eastern Ghouta (Rif Dimashq offensive) that began in February 2018; heavy airstrikes of Douma that included the use of chemical munitions on April 7; missile strikes on multiple Syrian government targets by the US, UK, and France on April 14; and the May 4 announcement by the US government that it would cease funding the WH. Immediately after the Douma chemical attack on April 7 there is a sharp increase in (re)tweets, volume that is sustained through the other events. Due to the salience of YouTube URLs on Twitter during this period, we focus in more detail on this period in the YouTube section of our findings below.

A further interesting distinction between accounts in the red and blue communities is their use of the reply function on Twitter. We observed three kinds of reply:

- *in-network replies*: an account in a community replying to an account in the same community (e.g., account in anti-WH community replies to another account in the anti-WH community);
- *cross-network replies*: an account in one community replying to account in the other community (e.g., account in anti-WH community replies to account in the pro-WH community); and

- *out-network replies*: an account in the anti- or pro-WH community replying to an account outside of these communities (i.e., an account not in our retweet network due to limited or no participation in the WH discourse on Twitter).

In general, accounts in the anti-WH community demonstrated a greater propensity to use the reply function: 9.20% of the total tweets sent by the red community were replies to other tweets (compared to 2.42% for the pro-WH community). Both the red and blue communities direct the majority of their replies in-network (3.61% and 1.22% of all tweets, respectively), but a much higher proportion of tweets from the red community are out-network replies (bottom right cell, Table 8.2)—with 3.25% of anti-WH community tweets replying to accounts outside of the WH discourse on Twitter. The out-network replies are generally directed at prominent politicians (e.g., *@realDonaldTrump*; *@jermeycorbyn*) and large media organizations (e.g., *@guardian*; *@cnni*; *@bbc*, *@derspiegel*). This behavior—which illustrates attempts to gain the attention of these prominent figures and organizations—is something we will resurface throughout findings.

reply <u>from</u> account in:	**reply <u>to</u> account in:**		
	blue (pro-WH) community	**red (anti-WH) community**	**other (not in red/blue communities)**
blue (pro-WH) community	in-network replies 1.22% (2542 replies)	cross-network replies 0.69% (1444 replies)	out-network replies 0.51% (1056 replies)
red (anti-WH) community	cross-network replies 2.33% (14729 replies)	in-network replies 3.61% (22885 replies)	**out-network replies** **3.25%** **(20592 replies)**

Table 8.2: The percentage (and count) of reply tweets that were in-network, cross-network, and out-network by each of the communities on Twitter.

Top-5 Most Linked-to Social Media Platforms

Patterns in URL sharing (Table 8.3) demonstrate that accounts in the anti-WH community link to more externally-sourced content in their (re)tweets. Prior work has examined some elements of this wider information ecosystem, including networks of government-controlled and alternative media websites that function to foment and distribute anti-WH narratives (Starbird et al., 2018) and the use of YouTube by the anti-WH community on Twitter (Wilson & Starbird, 2020).

social media platform	distinct urls linking to platform	num (re)tweets from red community	% of red community (re)tweets containing URL	num (re)tweets from blue community	% of blue community (re)tweets containing URL
youtube.com	1822	43,275	18.5	2,221	4.7
steemit.com	95	5,559	2.4	13	0.0
facebook.com	1660	3,435	1.5	1,850	3.9
medium.com	54	747	0.3	1,013	2.1
liveleak.com	64	1,564	0.7	2	0.0

Table 8.3: The number of URLs linking to content on each platform (re)tweeted by the red (anti-WH) and blue (pro-WH) communities on Twitter

	Breakdown of 25 most (re)tweeted				Tweets (introduction to Twitter through URL)		Retweets (content dissemination)	
social media platform	pro-WH content	anti-WH content	unique content producers	content removed	number of tweets	number unique accounts	number of retweets	number unique accounts
youtube.com	0	25	14	1	3548	1085	13945	8004
steemit.com	0	25	5	0	589	108	4491	2887
facebook.com	5	13	9	7	190	48	2213	1693
medium.com	15	8	11	2	292	134	1405	1215
liveleak.com	0	23	2	1	470	37	959	637

Table 8.4: Overview of the 25 most (re)tweeted social media posts from each of the platforms, including the narrative stance toward the WH, number of unique content producers, and details of the introductions to Twitter.

The focus herein will be on the links from Twitter to YouTube, Steemit, Facebook, Medium, and Liveleak, which were identified as the most linked-to social media platforms (Table 8.3). Accounts in the red community sourced content from a wider variety of platforms, including those less familiar such as Steemit and Liveleak. Table 8.4 provides an overview of the 25 most (re)tweeted social media posts (could be posts, articles, videos, depending on the platform) from each of these platforms, including a breakdown of pro- or anti-WH posts, the number of unique content producers, and how many posts had since been removed

(as of December 2020). To the right of the table, we present the number of introductory tweets and number of retweets of the 25 posts from each platform.

We next present our findings from Twitter: YouTube, Steemit, Facebook, Medium, and LiveLeak. For each we provide an overview of the platform and its features, a description of the types of WH content within our sample of data, and analysis of how the content was introduced into the WH conversation on Twitter.

YouTube

YouTube is the worlds' largest video-sharing service and a dominant platform within the digital media environment (Burgess & Green, 2017). Accounts are free and offer users unlimited space to upload and store video content to their *channel*. Social share buttons enable sharing to a wide variety of other platforms, in addition to a specific video URL that can be shared. YouTube videos can be labelled with hashtags, which group similar content and facilitate the surfacing of content through search functionality powered by Google.

YouTube has rules against sensitive content (nudity and sexual content), spam and deceptive practices, and violent or dangerous content (YouTube, n.d.). Since YouTube was established its policies have not explicitly prohibited the publishing of videos featuring conspiracies or misleading information (Matsakis, 2018). However, in March 2018 YouTube announced *information cues*—links to Wikipedia below content that is contested or conspiratorial in nature (Matsakis 2018), or to denote channels that are government-funded (Horwitz 2018).

WH Content on YouTube

As previously reported in Wilson & Starbird (2020), YouTube is the most linked-to domain in the WH discourse on Twitter, used extensively to host content containing anti-WH messaging. In this sample, 24[38] of the most-(re)tweeted YouTube URLs include a variety of the anti-WH narratives in circulation, for example claiming terrorist affiliations; accusing the group of staging rescues; and seizing upon their funding by western governments as proof they are "agents of change" or a "soft power" in the region. YouTube content was hosted on 14 distinct channels, a heterogeneous mix of independent journalists, influential individuals in the Syrian context, government-controlled media (e.g., *RT*), and alternative media (e.g., *One America News Network*; *UK Column*).

[38] As of December 2020, 1 video was unavailable, however during a prior round of data collection and content analysis (when the video was still available) we noted that this video was of "hand held footage purportedly showing the WH cheering during an execution". Assumedly this content went against YouTube policy surrounding violent or dangerous content and was subsequently removed from the platform.

Several of the YouTube videos were 'old' i.e., they predate our Twitter data (i.e., pre-May 2017), including several compilation videos that combine "evidence" of various accusations levelled against the WH. Such compilations are flexible insofar as they can be "dug up" from the YouTube "archives" (Thorson et al., 2013) and reintegrated to the mainstream discourse to serve a variety of purposes. One example is a video from 2016 in which Boris Johnson, then Foreign Secretary of the UK Government, is watching a demonstration exercise by the WH, praising the group, and announcing additional funding for the WH. While the video demonstrates support for the WH and their activities from a western government, it was "dug up" (Thorson et al., 2013) and mobilized by anti-WH voices as evidence that the WH were serving the interests of the UK government in Syria.

On the other hand, 12 out of the 24 available videos were 'new', uploaded to YouTube in April and May 2018 yet received enough (re)tweets to make it into the top-25 most (re)tweeted. This is due to the spike in Twitter activity—and links to YouTube—toward the end of our collection period on Twitter (Figure 8.3). This video content—and the subsequent efforts to introduce it to the WH conversation on Twitter—are in response to events of the Syrian conflict at this time, the most significant of which was the Douma chemical attack that took place on April 7 2018 and of which the WH provided evidence (BBC, 2018; Loveluck & Cunningham, 2018)

These newer videos reiterate many of the disruptive narratives surrounding the WH—their purported links to terrorism and their funding by the US and UK governments. For example, the most (re)tweeted YouTube video (2199 (re)tweets) is a cellphone recording of Roger Waters (musician) at a concert in early April 2018, in which Stone describes the WH as *"fake...jihadists...and terrorists"*. But there are also concerted efforts to propagate claims that the WH conduct 'false flag' events staged by 'crisis actors', and that they have been doing so for some time. The purpose is to discredit the WH and, by association, the evidence that the group provided of a Syrian/Russian chemical attack in Douma. To support the 'crisis actors' claim of false flag events, the anti-WH community appropriated the WH's recording of themselves involved in the "Mannequin Challenge", a 2016 social media trend that involved people remaining frozen in action (like mannequins) while they were filmed. This clip was reframed (and remixed as a clip in other videos) as evidence that the WH stage rescues, picked up by OANN, RT, and the Russian government to support otherwise unfounded claims that the WH stage attacks and rescues. The tweets introducing this content to Twitter carry increasingly extreme interpretations, including claims that the WH drug children (with anesthesia) before using them to stage fake events; and that they burn barrels of plastic and wait for them to explode to create scenes (and injuries) that are consistent with a chemical attack.

Introducing YouTube Video Content to Twitter (April 1 -May 15 2018)

As illustrated by the temporal plot in Figure 8.4, YouTube is the most linked-to social media platform in the WH discourse between April 1-May 15 2018: There were 26,365 (re)tweets posted to Twitter: 8,484 (36%) of these contained a link to one of the 25 most (re)tweeted videos on YouTube. In other words, during this 6-week period more than a third of (re)tweets in the WH discourse on Twitter contained a URL to one of these 25 videos. The majority (8436) came from accounts in the red anti-WH community, versus 48 from the blue pro-WH community.

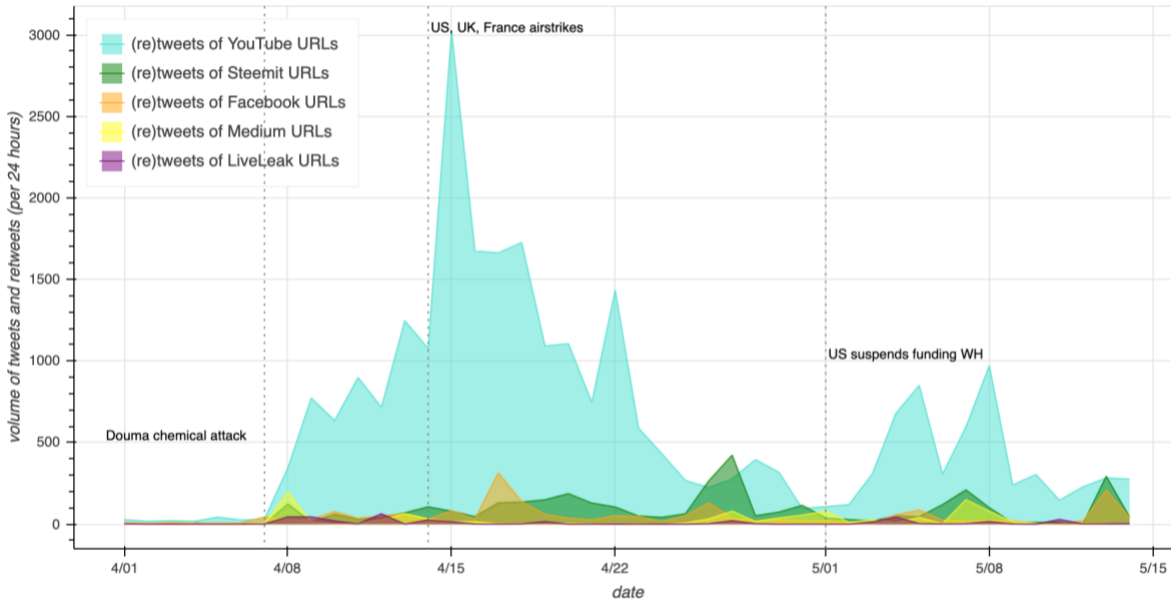

Figure 8.4: A temporal plot showing the number of (re)tweets, per day containing URLs to Facebook, LiveLeak, Medium, Steemit, and YouTube between April 1 2018 and May 15 2018 inclusive.

The largest peak on the graph occurs around April 15, coinciding with US, UK, and French airstrikes in Syria in response to the Douma chemical attack. Preceding this there is a substantial increase in (re)tweets containing URLs to YouTube, beginning on April 7 when the WH reported on the chemical attack in Douma. The work to undermine the credibility of the WH intensifies in the wake of the Douma chemical attack in a bid to tarnish their claims that a chemical attack took place. This sustained activity lasts almost three weeks, slowly tapering off toward the end of April, and then picking up again slightly as the US government announced it would cease funding the WH (Atwood, 2018)[39].

[39] The funding was reinstated in June 2018 when the US government pledged $6.6 million to the WH and the UN's International, Impartial and Independent Mechanism, a group that assists with the investigation and prosecution of war crimes (Nauert, 2018).

Within the 8,484 (re)tweets, 1,742 are introductory tweets (non-retweets) that function to introduce YouTube content to the Twitter conversation. More than a third of these introductory tweets (604) are posted as replies to other users.602 replies are sent by 151 distinct users in the anti-WH community[40]. A small group of five Twitter accounts are responsible for more than half (356) of these replies, illustrating a dedication, among a small number of accounts, to repeatedly introduce this content to Twitter.

These replies are sent to 396 distinct users on Twitter. 172 in-network to accounts such as *@RT_com* (14), *@PrisonPlanet* (10), *@RussianEmbassy* (9), and *@Partisangirl* (7); 73 cross-network including to *@BBCNews* (16), *@BorisJohnson* (10), and *@guardian* (4); and 151 to out-network accounts such as *@realDonaldTrump* (14) and *@jeremycorbyn* (5).

In-network replies often brought new information or "evidence" to the attention of influential accounts in a bid to disseminate it further. For example, providing evidence that there was no chemical attack to RT, the government-controlled Russian broadcaster:

Tweet by @RT_com: 'No role for WHO' in confirming or denying suspected chemical attack in Syria's Douma

> **Reply to above tweet:** @RT_com There was no chemical attack in #Douma. 13 health workers, including 8 doctors and 3 nurses, tell in detail… <<*link to YouTube video of Douma Chemical Attack False Flag Operation EXPOSED!*>>
> April 18 2018

In-network replies were also sent to influential accounts in the anti-WH community on Twitter. These replies attempt to garner the attention of these users to leverage their influence within the community, i.e., to retweet the content to their large number of followers who are likely to be receptive:

@Partisangirl: <<tweet is no longer available>>

> @Partisangirl BREAKING!!! VITAL INFORMATION ON #SYRIA FROM @VanessaBeeley REGARDING #ChemicalAttacks #Douma <<*link to YouTube video of UK Column News featuring Vanessa Beeley*>>
> April 13 2018

[40] Two distinct accounts in the pro-WH community also each send a reply to themselves during this period.

Cross-network replies generally sought to correct or dismiss reporting, offering alternative versions of events as in the following examples to The Guardian and Boris Johnson (then UK Foreign Secretary).

@guardian: UK denounces Moscow claims country was behind 'staged' Syrian gas attack

> @guardian There's a lot of info to suggest, even to a novice, that these are staged events <<*link to YouTube video of Boris Johnson visiting a WH training session and pledging further monetary support for the group*>>

@BorisJohnson: Welcome the news of UK military strikes against major chemical weapons facilities in Syria alongside our US and French allies. The world is united in its disgust for any use of chemical weapons, but especially against civilians

> @BorisJohnson ⚠ There was no chemical attack at Douma It was a staged provocation by the White Helmets <<*link to YouTube video Douma Chemical Attack False Flag Operation EXPOSED!*>>

Out-network replies sought to delegitimize the WH in the eyes of prominent political figures such as Donald Trump, in effect lobbying them to not take retaliatory action against the Syrian government.

@realDonaldTrump: Many dead, including women and children, in mindless CHEMICAL attack in Syria. Area of atrocity is in lockdown and encircled by Syrian Army, making it completely inaccessible to outside world. President Putin, Russia and Iran are responsible for backing Animal Assad. Big price...

> @realDonaldTrump You've been duped by the deepstate and their white helmet terrorist propagandists <<*link to CorbettReport video on YouTube titled The White Helmets Are A Propaganda Construct*>>
> <<date will be added>>

A single account was responsible for the majority of replies sent during this period, resurfacing the 2016 video of Boris Johnson pledging funding for the WH (discussed above) and persistently posting it as a reply to other accounts 296 times between April 14 and May 12. The start of this effort coincided with US, UK, and French airstrikes against military targets in Syria, to highlight that the WH were funded by western governments and strategically frame this to mean that they were a military tool of these governments.

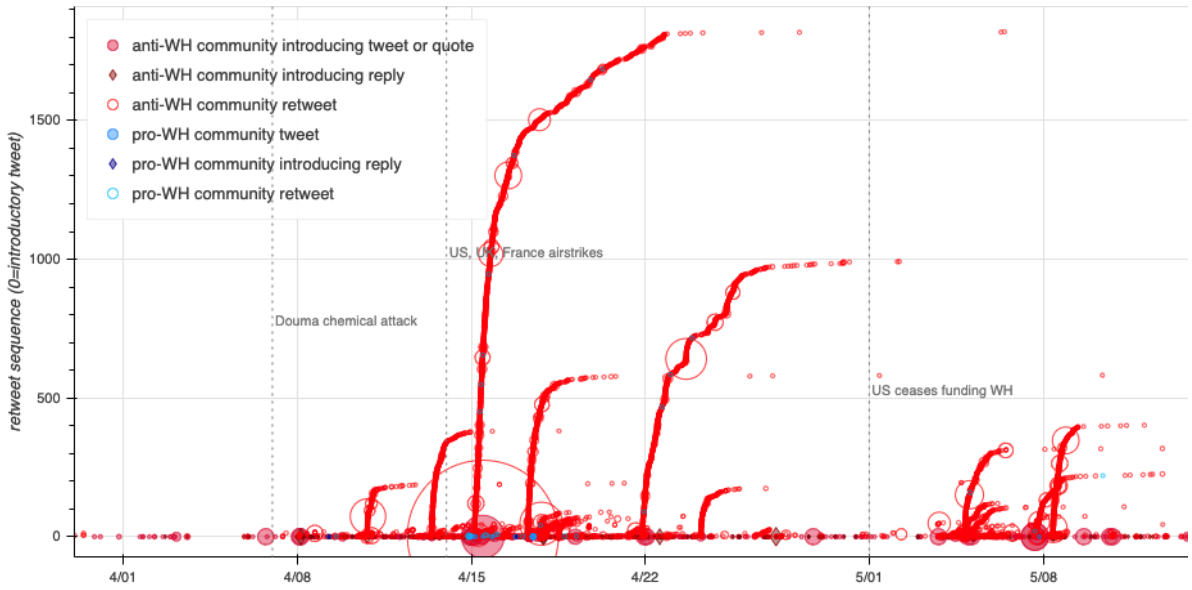

Figure 8.5: Retweet cascades of YouTube videos in the top-25 most retweeted, April 1 to May 15 2018.

Aside from replies, other accounts also appeared 'energized' during this period and were actively and persistently introducing YouTube content to Twitter. Figure 8.5 illustrates the retweet cascades of YouTube content during this period. The vast majority of (re)tweets are from the anti-WH community, generally from accounts with smaller followings, but the content is retweeted by more influential accounts (with more followers). The largest cascade here is the Roger Waters video, receiving 1822 retweets, the majority of those over the course of a week. Overall, the vast resources of YouTube meant it was used as a source of both new and old content. While new content was being produced and introduced to Twitter in "real-time" during this period, content produced many months earlier was also being appropriated and recirculated in new context (Thorson et al., 2013).

Steemit

Steemit is a blockchain-based social media platform that supports community building and social interaction with cryptocurrency rewards (Steem, 2017). Steemit is decentralized, existing on a distributed peer-to-peer network, which eliminates control by a single entity (unlike e.g., Facebook). Steemit's terms

of service posits "freedom of speech" as a foundational principle of civil liberties and calls out censorship as a means of limiting public discussion. It is free to join Steemit[41], author posts, and upvote others' posts. Users can be rewarded in Steem (cryptocurrency) for contributions to the community, including writing an article that is upvoted (valued) by other members, or upvoting content that later becomes popular (Steemit. n.d.). Steemit includes social share buttons to Facebook, Twitter, Reddit, and LinkedIn, plus a URL-sharing. An author must include at least one relevant tag on their post; the use of an irrelevant tag can lead to downvoting. Posts can be edited by the authoring user for 7 days, after which they are immutable. However, since all changes are logged on the public blockchain, even if deleted, traces of the content history will persist.

Mirroring Content on the Blockchain

Steemit is almost exclusively (re)tweeted by accounts in the red anti-WH community on Twitter, and all content within this sample promotes anti-WH narratives. There are five distinct content creators, although 19 of the articles are authored by a single user. Steemit is generally used to host compilations of content, with links to other Steemit posts, URLs to other external websites, screenshots of relevant tweets, and embedded videos. These posts offer a kind of clearing house of anti-WH content, with titles such as "massive compilation of material…", "huge cache of…", and "huge information drop".

Our analysis revealed the use of Steemit as a means of content preservation whereby a copy of content published elsewhere was also published on Steemit. This wrote the content to the blockchain and meant that in the event of its removal from another platform (e.g., due to a policy violation), it would still exist on the decentralized peer-to-peer network. Within our sample we found copies of two articles published on Medium and one video from YouTube that were 'mirrored' on Steemit. We also noted other activities on Steemit that would support efforts to preserve content: When embedding YouTube content from an alternative (non-mainstream) media source in a Steemit post, the video was often copied to D.Tube[42] and appeared alongside screenshots and commentary describing the video. Similarly, tweets were included as screenshot images alongside a text link to Twitter. Screenshots (as images) offer persistence, which would exist even if the source content was subsequently removed.

[41] The Steemit FAQ reveals there is a cost to joining the Steem blockchain, but this cost is absorbed by Steemit for new signups. For this reason, users applying to join must supply their email address and phone number, which will be verified to prevent a user having multiple accounts. Some authors argue that the associated cost with participation can help with issues of authenticity online (Guidi, 2020).

[42] D.Tube is short for Decentralized Tube, a decentralized open source video-sharing platform that also utilizes the Steem blockchain

Steemit-Twitter content integration

The majority of Steemit content is integrated to Twitter by a small number of highly active users, in particular the now-suspended *@WhiteHelmetsEXP* (382 tweets). These accounts undertake a concerted effort to attract the attention of large media organizations (e.g., *@BBCWorld* 24 replies; *@FoxNews* 24 replies) and prominent politicians (e.g., *@realDonaldTrump* 19 replies; *@NYCMayor* 18 replies) by posting URLs to Steemit content via replies on Twitter. For example, in the reply is to Fox News (conservative cable broadcaster) the reply attempts to bring the WH "exposé" to one of their primetime presenters, Sean Hannity (also an ally of then US President of the Donald Trump) and his audience:

Tweet by @FoxNews: MONDAY: Watch @seanhannity at a new time, 9p ET, when he'll interview Steve Bannon! #Hannityat9
Sep 24, 2017 2:19 PM

> **Reply:** @FoxNews @seanhannity White Helmets Fully Exposed as Hollywoods Favorite Terrorists-Over 200 Revealing Facebook Images <<*link to Steemit post*>>
> September 24 2017. 10:12 PM

We also observed accounts with large followings (re)tweeting Steemit content. For example, an anonymous account with ~100,000 followers that was—and remains in December 2020—highly active in the anti-WH discourse, posts 19 tweets that contain 10 of the 25 Steemit posts in the sample and retweets other Steemit content an additional 18 times that results in downstream retweets by other accounts in the red community.

Facebook

Facebook has grown from a social networking site connecting college friends to an influential company with billions of users (Gartenberg, 2019). Facebook offers a wide variety of functionality for its users, but important here is that it allows users to connect with others and to author content in the forms of posts (images and text), videos, and live videos. Within Facebook, this content can be shared 'privately' (with 'friends' within one's social network) or publicly. Sharing to other platforms (e.g., Twitter) requires a user to ensure the post is shareable (i.e., the permissions are not limited to friends only) and copy the URL to a tweet. Facebook policies were somewhat in flux during 2017-2018, and content moderators were overwhelmed by the volume of posts (Hopkins, 2017). After a Guardian investigation into the content moderation policies, in 2018, Facebook released its Community Standards (Facebook, 2020), which outlined that although the platform values expression, they also recognize the potential for abuse. In

particular, the policies focused on authenticity, safety (being a safe 'place'), privacy, and extremist content. Although no explicit policy on mis- or dis- information was in place, Facebook said it wanted to talk about what they felt was important, even if objectionable, but would consider the 'newsworthiness' and weigh the potential harms (Facebook, 2020).

WH content on Facebook

The sample of Facebook posts linked-to in the Twitter data is a mix of pro-WH and anti-WH content that take the form of Facebook posts (image and text) or Facebook Live (videos) produced by nine unique content producers. Seven of the 25 posts are no longer accessible, the highest proportion across the five platforms; the redirect page for these posts suggests that four of the posts were removed by the user or they changed the permissions and three are "no longer available". Six of the seven removed posts received the majority of (re)tweets from accounts in the anti-WH community on Twitter, suggesting they contained anti-WH messaging.

Pro-WH Facebook posts were authored by the official Facebook account of the WH (4 posts of the group involved in local civil service) and the Tipperary Peace Convention (1 post announcing that the WH had been awarded the 2018 Tipperary Peace Prize[43]). Anti-WH content comes from a variety of sources including the *Russian Embassy in the USA*, Russian state-backed broadcaster RT, and a journalist with Syrian state TV al-Sama TV. The most active content producer in this sample is a Bolivian celebrity and philanthropist who posts five Facebook Live videos, on location in Syria, in which she is visiting various buildings that are purportedly occupied by the WH with claims that they are connected (physically, by underground passages) to terrorist groups affiliated with a known terrorist group Al-Nusra. In another video, this individual interviews and discusses her findings with Pearson Sharpe, a journalist with One America News Network (OANN), a far-right, pro-Trump, conservative news network recently involved in spreading covid and election-related misinformation (Frieman, 2020; CNN Business, 2020).

Facebook-Twitter Content Integration

Accounts in the pro-WH community on Twitter involved in content introduction and retweeting of Facebook content have, in general, relatively large followings. As illustrated below, some pro-WH content seemed to benefit—in terms of retweeting—from coordinated boosting by activists and possibly even automated accounts. Three of the pro-WH posts from the official WH Facebook account were introduced

[43] The Tipperary International Peace Award was created by residents of the Irish town to counter the association between their town and war, which was created by the song "It's a long way to Tipperary". The annual award is given for humanitarian work.

to Twitter through cross-posting by the respective *@SyriaCivilDefense* Twitter account. They were then retweeted by accounts with large followings, for example @Free_Media_Hub, an activist account with 225K followers (and a second account called @Free_Media_Hub2), and *@9b3OR2qdITMbDcd* a Japanese account with (at the time) a large following of 1.18M. Interestingly, in December 2020 *@9b3OR2qdITMbDcd* still exists on Twitter but the number of followers is down to 242K, suggesting that the majority of its followers were part of coordinated inauthentic behavior and were subsequently removed by Twitter. One account in the pro-WH community posts a reply to themselves, in which they translate an earlier tweet of a Facebook Live video of a WH volunteer. This is retweeted by some other accounts with modest followings, before a retweet by *@9b30R2gdITMbDcd*. This tweet ultimately receives 43 retweets. In general, there are fewer content introductions by the pro-WH community, and only a single reply. However, when content is introduced to Twitter, the retweet cascade quickly builds as accounts with large followings work together to promote it.

On the anti-WH side, the timing of FB content appears to coincide with current events on the ground, for example posts in response to the reappearance of Omran Daqneesh (The 'Aleppo Boy') and when the WH were awarded the Tipperary International Peace Award. Despite being positive news for the WH, introduction of the Tipperary International Peace Award Facebook posts to Twitter is done by the anti-WH community who frame the award and support for the WH as a "stain on Ireland". The Bolivian celebrity (with ~85K followers) introduces her own Facebook content to Twitter (5 introductory tweets). The anti-WH community uses a combination of in-network, cross-network, and out-network replies to draw attention to the content. For example, one user sends 45 replies containing a URL to an anti-WH Facebook post to different Twitter posts over a 20-minute period on April 13 2018.

Medium

Medium is an online publishing platform that encourages users to express their ideas in the form of written posts. Medium posts can include embedded images, videos and content from other platforms such as Twitter. All posts include social share buttons to Facebook, Twitter, LinkedIn, or a shareable URL. Each post can have up to five tags that serve as a means of grouping content; clicking tags reveals similarly tagged posts. Medium provides users a comprehensive set of rules in its policy documents. Explicitly disallowed are hate speech, harassment, violations of privacy, deceptive or inauthentic content, graphic content, and pornography. Medium also singles out the promotion of harmful conspiracies, which includes pseudo-scientific claims, conspiracy theories with an associated history of harassment, hate, or violence, and intentional distortions, in particular systematic false claims about historic events and facts. Reported

violations are investigated by the company on a case-by-case basis according to a risk analysis (Medium, 2019).

WH Content on Medium and its Introduction to Twitter

The sample of Medium posts presents a mix of both pro-WH (15) and anti-WH (8) articles and is the only platform (in this sample) that is predominately (re)tweeted by the blue pro-WH community on Twitter. The Medium posts are generally longer articles, including a blog post of an academic paper[44]. One post has been removed by Medium for being "*in violation of the Medium rules*", and one article has since been deleted by the author.

The pro-WH content can broadly be categorized as either research-based articles that describe the disinformation campaign against the WH or as posts that celebrate the local volunteer efforts of the WH as they organize events for children and clean schools. Two of the articles in support of the WH come in the form of obituaries to the seven WH members that were murdered in Idlib in August 2017. Several of these articles are authored by an account called *The Syria Campaign*, a group that describes itself as supporting the WH through advocacy and fundraising. On the anti-WH side, seven (out of the 8) articles are authored by a single author. These are also longer posts that call into question the role and function of the WH, both alleging ties to terrorist groups in the region and at the same time calling the group out as part of a western "war machine". These articles are also highly critical of the western "mainstream media", particularly journalists, such as Olivia Solon at the Guardian who wrote an article about the disinformation campaign against the WH (Solon, 2017). In general, there is a criticism of what are termed "establishment" narratives and how those that challenge these narratives are labelled as Russian trolls or "useful idiots". In an example of cross-posting beyond Twitter, two of *Caitlin Johnstone's* posts also appear in our sample of Steemit posts. In general, due to the Medium interface and comprehensive community guidelines, we generally see Medium used for longer posts and opinion pieces.

Medium content introductions to Twitter are done by pro- and anti- WH communities. There are 292 introductory tweets containing a Medium URL from our sample, 201 from pro-WH accounts and 91 from anti-WH accounts. The majority of these introductions are tweets, not replies that attempt to gain the attention of large media or politicians.

[44] We were also contributing authors to this published article (Starbird et al., 2018).

LiveLeak

LiveLeak ("*redefining the media*") is a video-sharing platform founded in 2006, emerging from a place of skepticism with the news media (Cook, 2014) and a desire to host real footage of real-world events such as war. In 2019, the platform attracted between 16-20 million visitors per month (Herrman, 2019). LiveLeak describes itself as *"as 'free' as possible"* (LiveLeak, n.d.), though warns users that graphic media may be removed "*on a case by case basis*"; it also prohibits use of its services by terrorist or banned organizations (LiveLeak, n.d.). However, the platform has a relatively relaxed policy on what can and cannot be uploaded, and videos often contain violence (Cook, 2014). LiveLeak hosted footage of Saddam Hussein's execution and the beheading video of American journalist James Foley (Stryker, 2014). Its decision to ban future ISIS beheading videos and refusal to host the 2019 Christchurch shooting led to complaints from some users (Herrman, 2019). In 2014, LiveLeak and Ruptly (the video news agency of RT, the Russian government-controlled media corporation) announced a video distribution partnership (Ruptly, 2014).

LiveLeak features a search function, and videos can be tagged to facilitate searches (although clicking tags does not initiate a search for related content but must be manually typed into the search bar). The platform does not have social share buttons: instead, a "share link" must be copy-pasted elsewhere, which adds additional effort when sharing content. On Twitter LiveLeak videos appear as a text URL that when clicked redirects to the content on LiveLeak.

WH Content on LiveLeak and its Introduction to Twitter

23 of the LiveLeak URLs were to video content, one to the LiveLeak homepage, and one to a user's page (the most active content creator in this sample). The 23 videos were produced by two distinct users (producing 1 and 22 videos). As of December 1 2020, only one of the videos in the sample is unavailable, due to "*a possible violation of our terms of service*". This is consistent with a platform that exhibits relaxed policies and user base that accepts violent or gory content (Herrman, 2019).

Aside from two notable exceptions (a video produced by a European news agency, and a compilation titled "*Tapestry of Terror*" that we discuss in detail in the next section), the production quality of videos in this sample is generally low: raw footage, captured from cell phones on the ground. All LiveLeak videos in this sample promote anti-WH messages, mainly presenting the narrative that the WH are aligned with terrorist groups such as Al-Nusra and Al-Qaeda in Syria. Three videos purport to show the WH involved in executions and beheadings. In effect, LiveLeak appears to be the venue for hosting content that would otherwise fall foul of community guidelines on YouTube due to depictions of violence.

Introduction of LiveLeak content to Twitter is conducted by a small set of 37 determined users that post 470 tweets. 373 of the tweets are by a single user. In the majority of cases, the content is introduced as replies to cross-network or out-network accounts of large media organizations or politicians, including *@guardian* (77), *@cnni* (28), and *@realDonaldTrump* (8). This once again speaks to the persistent effort, by a subset of users, to lobby politicians or set the media agenda from via Twitter.

Multi-platform Concurrent Introduction of an Evidence Collage

The *Tapestry of Terror* (ToT) is a 65 minute long video compilation that presents multiple disruptive narratives about the WH, but for the most part tries to reframe the WH as aligned with regional terrorist organizations. The production is a remix of video clips, images, and screenshots of social media and websites (some of which appear in the 125 posts from across 5 social media platforms analyzed in this paper). These media resources are compiled into this production, tied together with narration that signposts the various anti-WH narratives. ToT also implicates the western mainstream media as complicit in the "conspiracy" of the WH. The Tapestry of Terror video can be described as an *evidence collage*—files that aggregate evidence (Kraft & Donovan, 2020). In our data, ToT appears on YouTube, LiveLeak and Steemit (Table 8.5)

Tapestry of Terror video		**Tweets (introduction to Twitter through URL)**			**Retweets (content dissemination)**		
platform	**upload/ post date**	**num tweets**	**num unique accounts**	**avg tweet per user**	**num retweets**	**num unique accounts**	**avg retweet per user**
YouTube	2017-03-25	207	26	8	153	141	1.1
LiveLeak	2017-03-26	374	10	37.4	442	345	1.3
Steemit	2017-12-04	41	5	8.2	26	23	1.1

Table 8.5: ToT appeared on 3 social media platforms in this study: YouTube, LiveLeak, and Steemit. This table breaks down the tweets and retweets of each version.

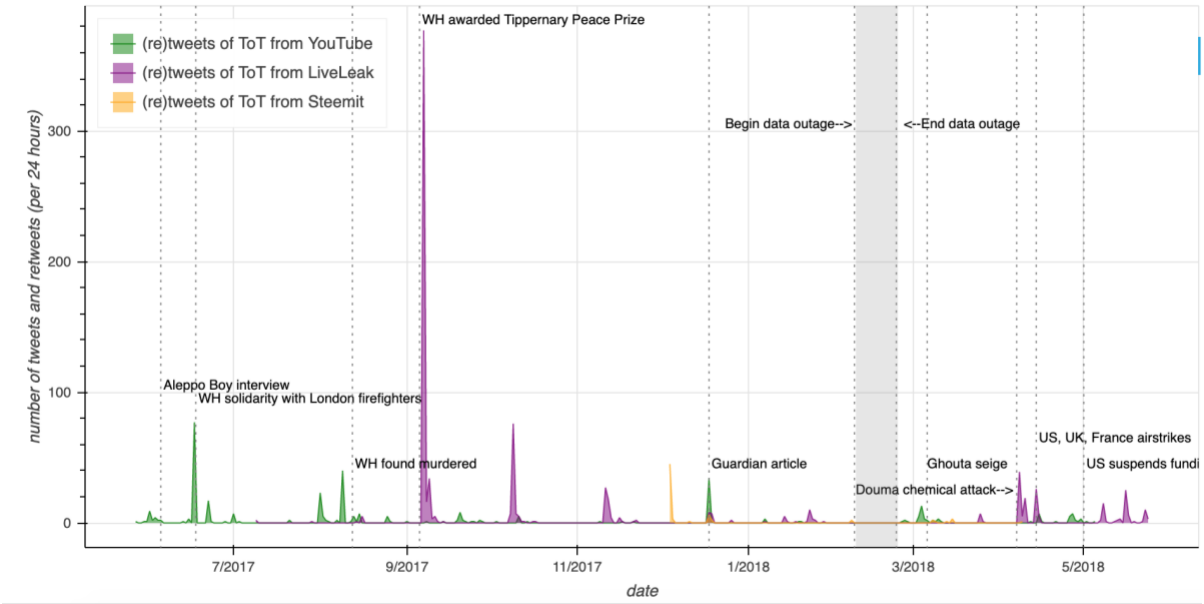

Figure 8.6: A temporal plot showing the number of (re)tweets, per day, for each YouTube, LiveLeak, and Steemit versions of *Tapestry of Terror*. This illustrates when the video was recirculated on Twitter.

ToT was uploaded to YouTube and LiveLeak at the end of March 2017.The YouTube video description includes a call to action for others to download, reupload, and share and to prevent "*the truth be[ing] hidden and censored*". The Steemit version of ToT was uploaded in December 2017 and appears as part of a longer post, itself an evidence collage, including screenshots from the video and embedded frames of both the YouTube and LiveLeak versions of ToT. The LiveLeak one is labelled as the "*mirrored/shared version*".

The YouTube version of ToT is the first introduced to the Twitter conversation (Figure 8.6), but the LiveLeak version led to spikes in the volume of tweets. The Steemit version, uploaded several months later, only features in 67 (re)tweets, so is not visible in the plot. The LiveLeak version of ToT is introduced to the Twitter conversation 374 times by 10 distinct accounts. Concurrently, the YouTube version is introduced 205 times by 26 distinct accounts. In other words, the work of ToT content curation and integration to Twitter is fulfilled by a small number of dedicated users who persistently work (for between 2 and 10 months) to introduce this content to Twitter.

Examining this work in more detail, the effort to introduce the LiveLeak version of ToT to Twitter was predominantly divided between two Twitter accounts: a now suspended account that was dedicated to attacking the White Helmets and another active account in this conversation that seek to bring attention to the video by replying to the tweets of major politicians (e.g., *@realDonaldTrump*, *@jeremycorbyn*) and

large media organizations (e.g., *@Guardian*, *@DerSPIEGEL*, *@cnni*, *@IrishTimes*). Oftentimes, these replies were in response to tweets about unrelated matters:

@Guardian: Russian women win partial victory in Aeroflot gender discrimination case <<*link to this article on the Guardian*>>
September 7, 2017 4:39 AM

> ***@WhiteHelmetsEXP***: @guardian Tapestry of Terror (Highly Graphic) - White Helmets Exposed As FSA Terrorists Linked With ISIS <<*shortened link to ToT on LiveLeak*>>
> September 7, 12:59 PM

@RealDonaldTrump: We want our companies to hire & grow in AMERICA, to raise wages for AMERICAN workers, & to help rebuild our AMERICAN cities & towns! #USA <<USA flag emoji>> <<*image of Trump speaking outside a factory with message about raising wages for American workers*>>
September 6 2017, 6:26 PM

> ***@WhiteHelmetsEXP***: @guardian Tapestry of Terror (Highly Graphic) - White Helmets Exposed As FSA Terrorists Linked With ISIS <<*shortened link to ToT on LiveLeak*>>
> September 7, 11:52 AM

In other cases, the replies were targeted and specific, for example when CNN journalist Jake Tapper recommended people to follow the account of Bana al-Abed, a Syrian girl from Aleppo who (with the help of her mother) documented the siege of the city via Twitter. In the reply to this tweet, this is called out as "fake news" with Tapper's credentials as a journalist challenged:

@jaketapper: Follow @AlabedBana <<*link to Twitter account*>>
October 6 2016, 8:01 AM

> @jaketapper @AlabedBana So disappointing if I thought u were a REAL journalist. It's #FakeNews. WATCH ↓↓↓↓↓↓↓↓↓↓ <<*shortened link to ToT on LiveLeak*>>

In total, 93% of all non-retweets containing the ToT LiveLeak URL are replies. A similar pattern of behavior exists around the YouTube version of ToT—80% of tweets are replies, predominately to major news outlets and politicians, although the replies are sent by a different group of Twitter users.

Further examination of the temporal activity reveals an interesting pattern—ToT is introduced to Twitter in response to events that present the WH in a positive light. For instance, in September 2017 the WH were awarded the Tipperary Peace Prize and in response we see the largest spike of (re)tweets of the LiveLeak version of ToT. Similar (but smaller) spikes occur in December 2017, in response to the Guardian article that reported on the disinformation campaign being waged against the WH (Solon, 2017); and January 2018 when the Netflix documentary about the WH was awarded an Oscar (Larkin & Lewis, 2017). In such cases the re-introduction of ToT was meant to distract from the positive media coverage of the WH and replace it with the alternative disruptive narratives presented within the video. These examples highlight YouTube's role as a repository of content that can be referenced over time (Thorson et al., 2013), and the reintroduction of ToT serves to sustain the anti-WH narratives, in particular by reintroducing them into the Twitter discourse at strategic times to distract from otherwise positive reporting of the group.

Discussion

The empirical findings presented in this research provide insight into the established (and largely successful) cross-platform information operation conducted against the WH in Syria. Our work contributes to research of multi-platform information operations by extending our initial Twitter dataset, using the URLs in tweets to build a picture of the surrounding ecosystem of social media platforms from which content is produced, stored, and integrated into the Twitter conversation. While imperfect insofar as it provides a "Twitter-centric" perspective of the wider information ecosystem, it does offer insight into how the two sides of the WH conversation make use of different configurations of social media platforms, and the ways in which the information subsequently flows into the Twitter conversation, where it can potentially be noticed by major politicians or journalists (Lieberman, 2020).

Our research leveraged the digital traces left by online interactions, following both the content and the actors involved in its production and dissemination. We adopted a CSCW perspective, meaning that we focused on the collaborative activities of these actors on and across multiple platforms, enabling us to surface the "work" involved in conducting multi-platform information operations including sustaining the narratives and attempting to set the agenda, oftentimes to distract or counter positive reporting about the

WH; and resilience-building through the use of alt-tech to counter perceived threats of censorship. We elaborate on each of these forms of work in the following sections.

Sustaining Narratives and Working to Set the Agenda.

The sustaining of narratives—keeping them relevant and maintaining attention of a mainstream audience—is achieved through three forms of work: *content production, content synthesis,* and *content mobilization*.

On all platforms in this study, we saw a small number of dedicated accounts involved in ***content production***—producing original content that supported one or more of the WH narratives in circulation by covering a new story. While the content in our sample contained both pro- and anti- WH narratives, anti-WH content was far more visible in the trace data. Content came in a variety of forms (posts, images, articles, videos), and varied in production quality. *Content production* involved authoring content; uploading it to one or more platforms (i.e., storing it); and labelling it by adding titles, tags, and descriptions (facilitating its retrieval in the future). Content production generated *communal goods* (Bimber, Flanagin, & Stohl, 2005)— resources for the community that could be searched for, selected, and synthesized into new content.

Content synthesis is the process of synthesizing existing content resources into *evidence collages* (Kraft & Donovan, 2020)—compilations or aggregates of evidence in the form of 'new' (remixed) content. For example, we observed the appropriation of the WH's "*Mannequin Challenge*" video as "evidence" that they staged rescues; the assembling of multiple URLs into Steemit posts that could serve as a clearing house of anti-WH content; and the compilation of video and image "evidence" into feature-length videos such as ToT. A particular strength of evidence collages is their flexibility—because they compile a variety of evidence that can be synthesized to support or challenge a range of narratives they are particularly useful *communal goods* that can be leveraged by other users.

Content mobilization is the activity of searching for and selecting content and trying to amplify it, or gain the attention of specific audiences, at specific times, in particular by moving content across platforms. The Twitter-centric nature of this case study generally meant that we could see content mobilization taking place by Twitter accounts, which would curate content from the other platforms and opportunistically introduce it to the Twitter conversation where it could potentially be broadcast and received by a mainstream audience, including politicians, news organizations, and journalists. A specific form of *content mobilization*, particularly salient in the anti-WH community, was to surface (search for, select) content and then publicly mobilize it *at* specific Twitter users through use of the *reply* function. This form of

mobilization was often targeted at politicians and journalists to get their (and their audience's) attention. These deliberate, and often persistent efforts to introduce content from the fringes to the mainstream are known as *trading up the chain* (Marwick & Lewis, 2017; Kraft & Donovan, 2020), which aims to build up the credibility of the message and appeal to a wider audience (Donovan 2019, Ryan, 2012).

There was a distinct temporal element to this work: The anti-WH community on Twitter would work to reintroduce existing content into the Twitter discourse when the time was 'right', in what Chadwick (2017 p.64) termed *opportunity structures*. In other words, while beneficial to publish something at the moment it occurs, there can also be an advantage in reintegrating dormant "fragments of stories" at specific times when they will have maximized impact. In this case study, the 'right' time most often coincided with otherwise positive reporting of the WH. At those times (e.g., amid reports that seven WH were murdered; when the WH were awarded the Tipperary Prize), we observed reactive and distracting dissemination practices conducted by what could be considered *episodic volunteers* (Starbird & Palen 2011; Voida et al., 2012)—subgroups of accounts working in alignment with the goals of the community, leveraging *opportunity structures* to take advantage of real-world events. Sometimes this was achieved through the use of influential accounts that stepped up to quickly propagate new content to shift the narrative, while other times a single or small group of accounts worked intensely—with the lightweight help of an audience who assisted in disseminating the content—over a period of hours or days to persistently reintroduce content that would challenge or distract from the positive reporting of the WH.

Resilience-Building Using Alternative Infrastructure

Social media platforms are sociotechnical, comprising technological features, users, and content moderation policies, which collectively form a communication infrastructure that can support online movements (Donovan, 2018). Since each social media platform is a unique combination of these sociotechnical aspects, they can be particularly suitable to serve a specific role or purpose within the larger information ecosystem.

A distinct advantage of our document-driven approach (i.e., following the content through digital traces) was that it enabled us to build up a picture of the distributed ecosystem of infrastructure through which information flowed. By reconstructing the digital traces, we were able to look beyond a single platform (Twitter) to the wider information ecosystem. This allowed us to see the work of resilience-building—working toward stability against the perceived threats of "censorship" from "big tech" (a reference to the large technology companies that own the platforms such as Google, Facebook, Twitter, and Medium) (Donovan, Lewis, & Friedberg, 2019).

The work of resilience-building was more visible on the anti-WH side of the conversation, in part due to their over-representation in our data (more active on Twitter and (re)tweeted more URLs), but also because there was explicit evidence of this strategy (and the necessity of it) in posts, articles, and videos in our sample. Resilience-building was broadly concerned with preserving content through the use of alternative platforms or *alt-tech*—a term used to describe how the "alt-right" shifted to the use of alternative (non-mainstream) platforms because of perceived threats to their stabilization as a result of efforts to deplatform and censor the movement (Donovan, Lewis, & Friedberg, 2019). Like the "alt-right", the anti-WH voices in this research shared disdain for the mainstream media and the evolving content moderation policies of the established social media platforms.

In striving for resilience, the anti-WH voices sought out alternative platforms to circumnavigate policy and carve out direct distribution channels. Working toward resilience involved the strategic placement of content on a specific platform or by mirroring across multiple platforms. An artifact of this activity is visible in our data—the anti-WH community on Twitter shares more content from a greater variety of platforms that it uses as a means of storing and organizing their content. The alternative platforms of choice were LiveLeak, with its relaxed policies, editorial partnership with Russian government-controlled media (Russian media were prominent members of the anti-WH discourse), and user base that has an affinity for violent content; and Steemit, the blockchain-based platform built on an immutable decentralized peer-to-peer network. These platforms provided the anti-WH voices with infrastructure that was able to protect the availability and accessibility of their content in the long-term: technologically, both platforms inherently enabled the hosting of content (with Steemit offering the additional advantage of decentralized storage not overseen by a single entity); and from a policy perspective, both platforms claimed to oppose censorship (an implicit suggestion that other "mainstream" platforms such as Twitter, Facebook, and YouTube are engaging in censoring content). It is important to note however, that censorship describes the suppression of free speech backed up by law (i.e., conducted by a government), and here that term is being conflated with *content moderation*, which are the community guidelines set out by social media companies as private entities that lay out what is and is not appropriate (Walter, 2020). In our data, LiveLeak and Steemit were used exclusively to store anti-WH content, and, in the vast majority of cases, were linked to in (re)tweets posted by accounts in the anti-WH community on Twitter. As this community worked 'outside' of the mainstream—and its established norms and technological solutions—preserving content was important in protecting the longevity of the operation.

But despite technologically-enabled content storage and retrieval and relatively forgiving policies, an infrastructural limitation of these alternative platforms is their lack of mainstream appeal. Steemit is among the largest blockchain-based social media platforms (Guidi, 2020), and LiveLeak enjoyed 16-20M users per month in 2019 (Herrman, 2019), but this is a fraction of Twitter's 330M monthly active user base that includes major politicians and journalists. And while YouTube has 2B users, it is technologically suited to content storage and retrieval, built around recommendation algorithms that seek to keep users engaged on the platform, rather than providing a large, mainstream audience with a broadcast breaking news feed of what is *happening now*. So, while these platforms may be suited to storage and retrieval, we observed the persistent work of *content mobilization* as attempts are made to integrate anti-WH content to Twitter, to reach a broader audience and sustain the anti-WH narratives in the long-term.

Chapter 9:

Discussion

In the preceding chapters I presented four studies that each examine how information operations take shape on and across social media platforms and a network of news-producing mainstream, alternative, and government-controlled media websites. As they are presented within this dissertation, the studies represent an increasing level of complexity, both in terms of the temporal scope and the extent of the inquiry: from 3 weeks to 13 months; and from a single platform (Twitter; Study 1, Study 2) to a network of news websites (Study 3), and then social media platforms (Study 4) that were linked to in the Twitter conversation through URLs posted in tweets. Collectively these studies work to build up the broader view of the information ecosystem from the perspective of Twitter, which was used as the 'seed' data in each of the studies.

Study 1 (Chapter 5) shows how information operations manifest on a single social media platform, Twitter, demonstrating how distinct online communities of users contribute to a contested and politicized conversation by assembling undermining narratives that serve to disrupt the mainstream media's reporting and confuse the audience. Study 2 (Chapter 6) presents a long-form analysis of the WH conversation on Twitter, including the structure, temporal activity, and narrative analysis of the two communities involved in the discourse. In addition, a domain retweet network visualization analysis reveals the salience of content from YouTube and other social media platforms that is introduced into the White Helmets conversation on Twitter. Study 3 (Chapter 7) reveals an "echo-system" of websites that repeatedly share content about the White Helmets. This network is supported by the integration of government-controlled media that often serve as sources for this content. This content is then 'bridged' to other communities with the effect of

drawing distinct audiences into a common set of White Helmets-related narratives. And Study 4 (Chapter 8) looks beyond Twitter to the broader information ecosystem to understand how multiple social media platforms are used, in parallel and complementary ways, to achieve the strategic goals of online information operations. Across these studies I was also interested in the participatory nature of information operations, seeking to identify the collaborative 'work' of the online crowd that enabled information operations to take shape on and across these websites and platforms.

The remainder of this chapter is structured as follows: First I will discuss the need to reconceptualize information operations to acknowledge the role of the online crowd. Across the studies it became clear that information operations were not something conducted exclusively by orchestrated networks of bots, trolls, and agents of foreign governments. While those elements may exist, so too do authentic (as in real, sincerely participating) actors that loosely collaborate (as opposed to being explicitly coordinated) in the production, synthesis, and mobilization of disruptive narratives. Indeed, I view this as an important aspect of how an information operation can be sustained in the long term.

Second, I will integrate findings across the studies to consider how this information operation takes shape on and across websites and social media platforms—I will do so through the lens of the *hybrid media system*, a theory that provides a way of conceptually understanding the contemporary coexistence of older (traditional, broadcast) and newer (social) media, plus the technology, behaviors, and organization (infrastructure) that underpin these forms (Chadwick, 2013; Chadwick et al., 2016; Chadwick, 2017). A central characteristic of the hybrid media system is the *hybridity*, which in this context encourages us not to look at 'older' and 'newer' media as a dichotomy but rather to bring awareness to the blurring of the boundaries and integration of these forms (Chadwick, 2017)[45].

Third, I will probe further a common feature across the studies, which was the tension between the "mainstream" and "alternative" forms of (social) media. I will work to expand the conceptual 'older'-'newer' of the hybrid media system to also consider this 'mainstream'-'alternative' dimension. This will include a discussion surrounding the integration of government-controlled media within alternative

[45] Although my analysis did not include data from traditional broadcast media sources (e.g. i.e. I did not collect data from print newspapers; broadcast television), these 'older' media also have a 'newer' online presence: An article published in the Guardian (print newspaper) also appears on its website from where it can be shared on social media platforms; clips of CNN posted on Twitter first appeared on its cable TV channel; and content on RT's YouTube channel is also broadcast on one of RT's channels around the world. Such examples go to illustrate the hybridity that Chadwick describes. This is all to say that despite my reliance on digital trace data, I still gain visibility into the convergence of these different forms of media.

media, which brings into question the independence of these outlets. Furthermore, I will look at the migration from “big-tech” to “alt-tech” as a means of sustaining information operations.

Fourth, I will conclude the chapter considering the implications of my research for social media platforms working to tackle information operations and for researchers involved in studying disinformation and information operations.

Extending the Conceptualization of Information Operations beyond that of Coordinated and Inauthentic Behavior

As previously noted, information operations are not a new phenomenon. Although in the US the term information operations emerged from the US military around the turn of the 21st century, the phenomenon it describes is much broader, encompassing information-related capabilities (tools, techniques, or activity used to create operationally desirable conditions) used to influence, disrupt, corrupt, or usurp the decision-making of adversaries and potential adversaries during military operations (United States Army Combined Arms Center, 2018). These operations included techniques such as computer network operations; psychological operations; military deception; and operational security (The Joint Chiefs, 2006). The term information operations transitioned into the civilian vocabulary after it was adopted by social media companies such as Facebook and Twitter that initially used it to describe how their platforms had been exploited—and users manipulated—through the coordinated efforts of the Russian Internet Research Agency (RU-IRA). The RU-IRA operated the “troll army” that sought to on the one hand influence the outcome of the 2016 US Presidential Election while on the other sow discord and exacerbate existing rifts in American society through the manipulation of social media. When used in this way, information operations are less aligned with military doctrine and are more akin to Soviet *active measures*—“*covert or deceptive operations* [with the] *goal to influence opinions or actions of individuals, governments or publics* [through methods such as] *disinformation and forgeries* [and] *political influence operations*” (US Department of State, 1989). As previously noted, *active measures* were a particularly salient practice of the Soviet Union during the Cold War (Bittman 1972, 1985; Rees, 1984, US Department of State, 1989).

The initial (post 2016 US Presidential Election) conceptualizations of information operations characterized them as state-backed and orchestrated, such as the information operation conducted by the IRA, which involved explicitly coordinated groups of actors, in this case members of the Russian intelligence services (DiResta et al., 2019; United States House of Representatives Permanent Select Committee on Intelligence, 2018). While undoubtedly a large undertaking in terms of resources and effort, the activities of the IRA

were not particularly sophisticated: a significant number of IRA Twitter accounts signed up from the same IP address, and there were also shared email addresses used during account signup (DiResta et al., 2019). In response, under the umbrella term of information operations, Facebook and Twitter developed policies to prohibit—and methods to detect and enforce—what they labelled *coordinated inauthentic behavior* (Facebook, 2020; Twitter, 2020). For example, on Twitter these fall under platform manipulation policy, which include:

> "*inauthentic engagements, that attempt to make accounts or content appear more popular or active than they are*" and "*coordinated activity, that attempts to artificially influence conversations through the use of multiple accounts, fake accounts, automation and/or scripting*" (Twitter, 2020).

While on Facebook, coordinated inauthentic behavior is defined as :

> "*the use of multiple Facebook or Instagram assets, working in concert to engage in Inauthentic Behavior (the use of Facebook or Instagram assets (accounts, pages, groups, or events), to mislead people or Facebook about the identity, purpose, or origin of the entity that they represent…[or] the purpose of an audience or community…[or]...the source or origin of content), where the use of fake accounts is central to the operation*" (Facebook, 2020)

However, as illustrated in each of the studies of this dissertation, information operations are more nuanced than (explicitly) coordinated acts conducted by inauthentic accounts operating from the same location—indeed, as seen here, information operations involve authentic (as in real, sincerely participating) actors that loosely collaborate (as opposed to being explicitly coordinated) in the production, synthesis, and mobilization of disruptive narratives. In other words, the explicit coordination represents only part of the collaborative work that goes into an information operation, and to some extent oversimplifies what is a more nuanced and complex phenomenon. I elaborate on two aspects of this in the following subsections: the blurred distinction between active measures (information operations) and activism and conceptualizing online information operations as a form of collaborative work.

The Blurred Distinction between Active Measures and Activism

Participation of authentic (as in real, genuinely participating) people is evident across the studies. For example, there are online "information activists[46]" who devote time and resources to the anti-WH agenda, and Western "journalists," that have risen to prominence through their anti-White Helmets content production (Starbird, Arif, & Wilson, 2019). In Study 4 there were several examples of accounts involved in persistent efforts to bring content to the attention of prominent journalists and politicians. While this behavior can appear "bot-like" at first glance, a detailed look at the accounts and their activities revealed that real people are in control, some of which are still posting content about the WH in December 2020.

These activities could be viewed as a form of online activism as people use available platforms to garner support and express solidarity for a political cause (Starbird & Palen, 2012; Nielsen, 2013; Donovan, 2018). For example, during the 2011 Arab Spring, local, affected individuals took to social media to extend their reach—applying political pressure, shaping political discourse, and informing western audiences about events on the ground (Howard et al., 2011); online activist networks coordinated and facilitated the physical 2010 G20 summit protests in Toronto (Poell, 2014); a 'networked social movement' leveraged communication infrastructure that transitioned the 2010 Occupy Wall Street protests from distributed encampments into a coordinated movement (Donovan, 2018); and the extensive use of social media by the Black Lives Matter (BLM) movement to bring awareness to police killings of unarmed Black citizens (Freelon, McIlwain, & Clark, 2016). However, in his book charting a century of disinformation and political warfare, Thomas Rid (2020) on multiple occasions highlights the blurred distinctions between activism and active measures. Rid describes how it has become (through history) increasingly difficult to distinguish between a genuine activist and those involved in active measures (in other words an inauthentic activist). Rid also notes that this distinction is difficult for both the activist (authentic or inauthentic) and the target of their behavior (Rid, 2020). Considering this in the context of contemporary information operations, within the activities described in this dissertation there too is the blurring of this distinction between information operations and genuine online activism as actors and organizations with specific geo-political agendas, such as government-controlled media, are integrated within online communities that are working to shift the narrative. This leaves genuine online activists extremely vulnerable to the information operations as they exist in a position when they can be leveraged by malign actors, possibly as "unwitting agents" (Starbird, Arif, & Wilson, 2019; Bittman, 1985).

[46] *Information activist* is used to refer to accounts that do not have another affiliation in this list but were involved in authoring relatively higher volumes of content within this (and often other) conversations.

Positing Online Information Operations as a form of Collaborative Work

Across the studies I revealed the work of information operations taking shape within heterogeneous communities of actors: information activists, bloggers and journalists, non-profit organizations, government officials, and mainstream and government-controlled media outlets. This work is geographically distributed, including people from within the affected areas (in Syria), western journalists (some of whom travel to the area), mainstream media organizations in the US and UK (among others), and what may be termed online 'volunteers' from around the world. These entities work together to produce, amplify, and spread their preferred set of narratives, depending on which 'side' of the conversation they are. In the Omran case study it was the pro-Syrian government voices that dominated the conversation, while across the WH studies (2-4), it was the anti-WH (that were also pro-Russian, pro-Syrian government) side that made up the majority of participating accounts and content on Twitter and elsewhere.

As I analyzed the data for my research in Study 1 (Chapter 5), I and my colleagues began to view these information operations as a form of collaborative work within an online crowd (Wilson, Zhou, & Starbird, 2018). Furthermore, and based on the empirical observations, I looked beyond the conceptualizations of this work as purely coordinated by a central node or nodes in a network: Although elements of coordination exist—there are accounts that appear to be agents of foreign governments or other groups (Starbird, Arif, & Wilson, 2019) and the analysis presented in Study 2 (Chapter 6) reveals that more than 13,000 accounts that were critical of the WH were suspended (Wilson & Starbird, 2020)—I (and my co-authors, see: Starbird, Arif, & Wilson, 2019) argue that these assemblages of diverse actors, driven by a variety of motivations, loosely collaborate in the 'work' of information operations. This view extends previous descriptions of online information operations as perpetrated by armies of automated accounts (or "bots") (Woolley & Howard, 2017) or factories full of paid trolls (Troianovski, 2018), suggesting a more complex system with emergent properties. This 'loose' collaborative work took various forms but was visible as a shared set of practices among accounts. One example of this was the use of *@mentions* to call attention to specific content, which when used en-masse resulted in a cascade of reactions downstream as the mentioned accounts amplified the message in a bid to drown out the pro-WH narratives (Starbird, Arif, & Wilson, 2019). Another was the use of replies (on Twitter) by groups of users to bring anti-WH content to the attention of politicians and mainstream media organizations (and their audiences) as in Study 4. However, as previously noted, there are likely elements of coordination that are not visible through the methods used in this research including the direct communications between accounts (e.g., through Twitter direct messages) or other means.

Examining Information Operations as taking shape through a Hybrid Media System

In the empirical chapters, I demonstrated how Twitter (Study 1, Study 2), a network of government-controlled media and alternative media websites (Study 3), and other social media platforms are used in the information operation surrounding the WH (Study 4). In this section I aim to bring these aspects together and synthesize across the studies to discuss how this sociotechnical infrastructure— comprising technology, users, and policies (Berkman Klein Center, 2020; Donovan, 2018)—is leveraged in information operations. I will do so through the lens of *the hybrid media system*, a theory that provides a way of conceptually understanding the contemporary coexistence of older (traditional, broadcast) and newer (social) media, plus the technology, behaviors, and organization (infrastructure) that underpin these forms (Chadwick et al., 2016; Chadwick, 2017). In particular I will discuss hybrid news production; the use of the hybrid media system in timely interventions to shift the narrative, and; how hybridity can mean losing control of the message. I will then consider a persistent theme across the studies, namely the tension between the 'mainstream' and the 'alternative', which I posit as an additional dimension of hybridity.

Historically, (in the 'older', broadcast media sense), news and information production was something done by a small band of elites such as large media companies, politicians, communications professionals, and journalists. However, the advance of Web 2.0 technologies meant that individuals were also afforded the ability to participate in productive information use and content creation ("*produsage"* (Bruns, 2008)). Although the same band of elites referenced above can still produce the news in the same way, they have less control over how it plays out in the information cycle because non-elite *podusers* are also able to advance or contest their own news (Chadwick et al., 2016). Chadwick et al. (2016) posits that the hybrid media system is "*composed of multiple, loosely-coupled individuals, groups, sites, and temporal instances of interaction involving diverse yet highly interdependent news creators that plug and unplug themselves from the news-making process, often in real time*". Furthermore, these actors are able to "*create, tap, or steer information flows in ways which suit their goals…across and between a range of older and newer media settings*" (Chadwick, 2017: xi). Across the studies in this dissertation, we get a glimpse into the different parts of the hybrid media system, and the interactions within, as the different 'sides' of the discourse in a sense compete in pursuit of their goals.

Hybrid News Production

In Study 3 (Chapter 7) the production of articles about the WH takes shape on and across networks of mainstream, government-controlled, and alternative media websites. Russian government-controlled media

served as an information resource of the anti-WH campaign: in what was termed the alternative media "echo-system" (Starbird et al., 2018) there was hybridity and integration between government-controlled and alternative media that was evident through the content-sharing practices of the alternative media websites (this relationship, and the use of the term "alternative" is something I revisit in further detail below). The Russian government's media apparatus also shaped the anti-WH campaign in other, more subtle ways. For instance, through cross-posting between non-professional bloggers/journalists and government-controlled media, and through source content amplification of voices that aligned with the anti-WH narratives (Wilson & Starbird, 2020). In a self-reinforcing way, the amplification of voices that aligned with the anti-WH narratives helped to raise these figures to prominence, enabling them to establish credibility and extend their reach on alternative news websites and multiple social media platforms. These activities revealed a hybrid form of non-professional blogger/journalist and government-controlled media contributor.

The hybrid media system is visible on the pro-WH side of the discourse too, but the infrastructure underlying it is of a different composition: integrated and aligned with the reciprocal support of the Western mainstream media. Articles published in the mainstream media were supportive of the WH and their cause, celebrating their achievements and mourning their losses. The pro-WH side supported the prevailing narratives about the group (for example by tweeting and retweeting articles published mainstream media outlets such as the Guardian, CNN, etc.) and were able to rely on support (in the form of content distribution) from Western governments, politicians, and the mainstream media presence on social media platforms such as Twitter. While this content was often uncomfortable—insofar as it contained images, video, and descriptions of death, displacement, and war—it also appeared designed to appeal to and reach wider audiences (i.e., it was less extreme in nature).

Timely Interventions through Fluid Opportunity Structures

In Study 4 (Chapter 8) the production and mobilization of content on and across multiple social media platforms was revealed. In particular, WH content, which was produced and hosted on a variety of social media platforms, was (re)integrated into the Twitter conversation. Some content was integrated and reintegrated multiple times, perhaps many months after the content was first produced. Chadwick et al. (2016) refer to this as "*cross-media iteration and recursion"* that, when coupled with low-threshold practices of "produsage" (Bruns 2008; Pyrhönen & Bauvois, 2020), create "*fluid opportunity structures"* (Chadwick et al., 2016; Chadwick (2017)—the potential to intervene in the flow of information. Fluid opportunity structures allow for *produsers* to make "timely interventions" (Chadwick, 2017 p.64) in the information cycle. In other words, while it can be beneficial to publish something the moment it occurs,

there can also be an advantage in reintegrating otherwise dormant "fragments of stories" at very specific times for maximized impact. The temporal analysis of Twitter data presented in Studies 2 and 4 revealed the leveraging of opportunity structures as the anti-WH side of the conversation worked to strategically mobilize content in response to mainstream positive reporting about the WH. The effect of this, on Twitter at least, was to distract from or drown out positive news about the WH or negative news about the Assad government and replace it with disruptive narratives about the group. Access to "fragments of stories" was facilitated by infrastructure such as YouTube, which offers robust resources as a repository of video content that, coupled with a robust tagging and search system, can be "dug up" whenever needed (Thorson et al. 2013).

A Pitfall of Hybridity: Losing Control of the Message

The WH themselves (via their official Syria Civil Defence social media accounts) leverage the hybrid media system to promote their work and define their role as a humanitarian response organization in Syria. Hybridity can empower groups such as NGOs by offering them an interconnected environment from which they can reach large and diverse audiences (Powers, 2014). For example, the WH were highly successful in promoting their cause through a Netflix documentary (itself a hybrid form of media integrating television with virtual video rental and personal, on demand, time-shifting technology (Alvarez-Monzoncillo, 2011)). This Oscar-winning documentary served to educate viewers about the group and their work and bring Western attention to the plight of the Syrian people from rebel-held areas who were seen as victims of the Assad regime. However, the same hybridity that can empower groups can also lead to a loss of ability to shape their own message at the point of circulation and consumption (Powers, 2014). An example of this is the WH's participation in the social media trend known as the "Mannequin Challenge", which turned out in many respects to be detrimental to their cause. The 2016 challenge involved people remaining frozen in action (like mannequins) while they were filmed, with the resulting videos posted to social media. For the purposes of the challenge, the WH were filmed staging a rescue, during which they froze in action as the camera panned around them holding a stretcher. Although seeking to leverage newer forms of media to promote their cause (by creating a clip as part of a social media trend in the hopes it would go viral and gain additional attention), control of the message was lost. The clip was appropriated and reframed by the anti-WH community as video "evidence" to support the "fake rescues" and "crisis actors" narratives. This short clip was remixed and repurposed, appearing in various forms across hybrid media including Russian and Syrian government-controlled television, and in multiple YouTube videos hosted by diverse channels, such as those of politicians, alternative media journalists, and government-controlled broadcasters. The WH lost control of the message and ultimately, in this data at least, it was the anti-WH version of events that were more salient on Twitter, YouTube, and RT (via RT and its subsidiary's YouTube channels).

Furthermore, in an example of *opportunity structures* used in timely intervention, almost 18 months after it was filmed by the WH, the Mannequin Challenge clip was resurfaced. It was remixed into evidence collages—compilations or aggregates of evidence in the form of 'new' (remixed) content (Krafft & Donovan, 2020), hosted by various channels on YouTube (and LiveLeak), and reintegrated into the Twitter conversation in April 2018 as efforts to delegitimize the WH intensified to discredit their reports of a chemical weapons attack.

Extending the Conceptualization of the Hybrid Media System to consider the Mainstream-Alternative Dimension

A common trend across the studies in this dissertation were the challenges between what was often referred to as the 'mainstream' and 'alternative'. This appeared both in the context of news (mainstream media versus alternative media) and also social media platforms ("big-tech" versus "alt-tech"). In this section I unpack this dimension of 'mainstream'-'alternative' further.

Alternative of the mainstream, but not independent

The use of the label "alternative media" has historically been used as a means of differentiating from the "mainstream (corporate) media", which is framed as being focused on profit maximization, hierarchical, and largely monolithic (Atton, 2002). In this dichotomy, the alternative media offered a counterbalance: they were viewed as independent (of the corporate mainstream), participatory, and autonomous (Atton, 2002). Indeed, the term "mainstream media" developed into a pejorative against the agenda-setting power and corporate interests of the mainstream media (Kenix, 2011), while the alternative media were viewed as non-commercial and attempting to challenge the dominant power structures within society (Atkinson & Dougherty, 2006; Kinix, 2011).

In Study 1 I observed the use of the acronym "MSM", not simply as a succinct reference to the mainstream media, but as a label of disrespect, signifying disdain with the large media corporations and a distinct lack of trust in their reporting. #MSM was exclusively used in tweets attacking the credibility of the mainstream media, often co-occurring with the tweet's author pointing to what they viewed as "lies". This trend continued into Study 3, with articles that were critical of the WH also carrying criticism of mainstream media and skepticism or outright rejection of its narratives. From the hybrid media system perspective, the anti-WH side of the discourse uses their configuration of the hybrid media system to work against the western mainstream media as they produce and propagate competing, and frequently conflicting, disruptive narratives about the WH. Pyrhönen & Bauvois (2020) refer to this as "reinforming", describing it as

"*proactive, politically inspired, ideological, online community-driven side of produsage in the hybrid media system*" (p. 708). Content on the anti-WH side of the discourse often features criticism of the mainstream media that appears to serve a dual purpose: First, it provides a means with which to introduce the disruptive narratives; but second, and it forms part of an implicit strategy to cumulatively denigrate and undermine trust in the institution of the media—which has been a longstanding aim of disinformation campaigns (Bittman, 1985; Rid, 2020).

However, while the 'alternative media' websites featured in the echo-system of Chapter 7 are distinct from the mainstream media, they appear far from independent or autonomous. Rather they are aligned with government-controlled broadcasters such as RT, Sputnik, and FARS (Iran), which is revealed through the content-sharing practices of the alternative media "echo-system" (Chapter 7; Starbird et al., 2018). Additional integration is seen as journalists and contributors to alternative news sites such as 21WIRE also appear on RT; and these segments of RT broadcasting are then uploaded to the YouTube channels of the alternative media and their contributors. In summary, it appears that there is a specific subtype of alternative media that is perhaps better described as a hybrid form of government-supported-alternative media. This integrated form of media publishes content that merges fact with fiction, is spreadable in an alternative news format, and challenges the established authority of the mainstream media (Pyrhönen & Bauvois, 2020). While certainly situated outside and independent of the Western mainstream media, these alternative media do not appear autonomous and powered by citizen journalism. Rather they appear aligned with government-controlled media, serving that content to diverse audiences, at times without clear attribution, and to advance geopolitical goals.

Alternative tech as a form of stability

The mainstream-alternative tension (and hybridity) is also visible with regards to social media platforms. In this context, 'mainstream' refers to the large social media platforms (Facebook, YouTube, Twitter) or so called "big tech" (Freelon, Marwick, & Kreiss, 2020). In a similar way to which the "mainstream media" or "MSM" developed into a pejorative, so too has the term "big tech", with the belief, particularly in conservative circles, that these platforms are biased (Freelon, Marwick, & Kreiss, 2020). 'Alternative' in this context is referring to *alternative technology*, or "*alt-tech"*, platforms that "*provide a parallel online space for individuals, ideas, and causes that are outside the boundaries of speech permitted on mainstream social media platforms*" (Zuckerman & Rajendra-Nicolucci, 2021).

The term *alt*-tech was originally used to describe the alt-right's shift to the use of alternative infrastructure in response to perceived threats to de-platform and censor[47] the movement, in particular in the wake of the August 2017 Unite the Right Rally that took place in Charlottesville, VA. Alt-tech platforms are smaller than the mainstream equivalents (in terms of the number of users and therefore reach), but they are more permissive in terms of their content moderation policies (Freelon, Marwick, & Kreiss, 2020), often promising "free speech" without interference from "big-tech" (Zuckerman & Rajendra-Nicolucci, 2021). As the alt-right began to seek other platforms, we saw the emergence of alt-tech platforms such as Gab, a Twitter equivalent. More broadly, the shift provided new organizing and recruitment sites for the alt-right (Donovan, Lewis, & Friedberg, 2019).

The use of alt-tech was also revealed during the multi-platform analysis (Study 4, Chapter 8), particularly on the anti-WH side of the discourse. Within this case study, the use of alt-tech appeared to be a strategy of resilience-building—circumnavigating platform policy (to preserve content) and developing new channels of communication. For the anti-WH voices, the alternative platforms utilized were LiveLeak, which in 2014 signed a video distribution partnership with Russian government-controlled media outlet Ruptly (Ruptly, 2014), and Steemit, an immutable blockchain-based platform built on a decentralized peer-to-peer network. These platforms provided the anti-WH voices with infrastructure that was able to protect the availability and accessibility of their content in the long-term: technologically, both platforms inherently enabled the hosting of content (with Steemit offering the additional advantage of decentralized storage not overseen by a single entity); and from a policy perspective, both platforms claimed to oppose censorship—an implicit suggestion that other "mainstream" platforms such as Twitter, Facebook, and YouTube are engaging in censoring content.

Alternative but still relying on the mainstream

Despite the anti-WH community's disdain for the mainstream media and the evolving content moderation policies of the established social media platforms, they still sought access to the mainstream and their audiences—they needed them to further propagate their disruptive narratives to the broader public. The work to tap into these audiences was laid bare in Study 4 (Chapter 8) as we saw efforts to intervene in the flow of mainstream news by attempting to gain the attention of the western mainstream media (alongside journalists and prominent politicians) in order to reach their audiences. In particular there were efforts to

[47] Censorship describes the suppression of free speech backed up by law (i.e. conducted by a government with the power to enforce it through punishment). In the context of social media platforms, censorship is conflated with *content moderation*, which are the community guidelines set out by social media companies (as private entities) that lay out what is and is not appropriate while using their services (Walter, 2020).

"trade up the chain"—to have a story endorsed or repeated by a prominent figure or organization (Marwick & Lewis, 2017; Krafft & Donovan, 2020); and persistent efforts to use the reply function on Twitter to bring content (from alternative media and platforms) to the attention of mainstream media organizations, their journalists, and prominent politicians, in a bid to lend credibility to their messaging and reach these larger audiences. In this way, the anti-WH voices were generating hybridity in between the mainstream and the alternative as they worked to integrate content across these forms.

Implications for Social Media Platform Policy

Extending our conceptualization of information operations to account for the non-explicit (loose) collaborations among heterogeneous sets of actors has implications for platform policy and the detection of the activities. The activities of accounts observed across the studies in this dissertation, although sometimes persistent, do not appear automated or explicitly coordinated as was the case with other documented information operations such as that conducted by the Internet Research Agency (DiResta et al. 2019; Gadde & Roth, 2018). As such, it is not appropriate to characterized this as *coordinated inauthentic behavior*, a common policy used for remediations by the large social media platforms (Twitter, 2020; Facebook, 2020. Further complicating matters is that, in general, the social media platforms do not have broad policies regarding mis- and dis- information: Section 230 of the Communications Decency Act (CDA) of 1997 offers them some legal protections from what their users may say and do on their infrastructure (Electronic Frontier Foundation, n.d.; 47 U.S. Code § 230), while leaders, such as Facebook's Mark Zuckerberg, have previously stated they have no desire to be the "arbiters of truth" (McCarthy, 2020).

During 2020 the social media platforms such as Twitter and Facebook proactively introduced policies to facilitate content moderation in specific contexts, including the coronavirus pandemic and US presidential election-related disinformation and misinformation. By design, these policies are limited in scope to these specific contexts that can be justified by the associated demonstrable offline harms (i.e., the hindering of global efforts to fight a global health crisis; voter suppression or efforts to challenge the election results). But the development of these expanded policies is also limited to the "mainstream" platforms: "alt-tech" platforms have no such policies, and the 2020 US election cycle led to the rise in prominence of platforms such as Parler, a pro-Trump platform with minimalist community guidelines that was involved in amplifying post-election disinformation that culminated in the breaching of the US capitol in January 2021 (Romm & Lerman, 2021).

On the one hand, without a collective effort—meaning consistent policies across platforms, such as those in place for child sexual abuse material, terrorism-related content, and more recently coordinated

inauthentic behavior (Douek, 2020)—it remains difficult to moderate against information operations because, as demonstrated in this research, information operations are participatory, accounts are not necessarily explicitly coordinated, and content is also produced, synthesized, and hosted on alt-tech platforms and mobilized (introduced) to the mainstream platforms from there. Unified agreements across platforms could tackle this in two ways: from a content perspective it could make it easier to remove misleading content at its source to prevent it moving across the information ecosystem; or from an account perspective it could lead to the identification and labelling of problematic accounts across platforms in an effort to monitor or limit their activities—although this would clearly raise important data-sharing and privacy implications.

On the other hand, increased cooperation between platforms to develop unified policy could lead to what have been described as *content cartels*—"*arrangements between platforms to work together to remove content or actors from their services without adequate oversight*" (Douek, 2020). A unified policy is likely to favor the larger platforms that have more users, meaning that the policy for many will likely be shaped by a few. Furthermore, such efforts, particularly if enacted with a lack of transparency (in terms of the development and subsequent application of the policy), will reinforce the view that platform content or account moderation is equal to "censorship" and drive content (and its production) elsewhere—to the "alt-tech" platforms where policies are weaker, non-existent, or not enforced. In effect, such action would just push the content to the unmoderated fringes of the internet and, unless the mainstream platforms begin to denylist whole platforms, the content can—and will—make it back to the mainstream through the dedicated work of the online crowd.

Chapter 10:

Conclusion & Contributions

The empirical studies presented in this dissertation provide insight into the information operation conducted against the WH in Syria during 2017-2018. While it should be noted that the empirical evidence presented cannot speak to if and how online WH-related content determined behavior toward the group offline, the information operation was effective in inundating the information space with negative content about the WH and nurturing a community that was willing and able to further advance online activities against the group. This success was acknowledged in 2019 by Russian General Gerasimov, who had earlier in the decade spoken of the asymmetrical possibilities of leveraging technologies and information networks to reduce the fighting potential of an enemy (Kramer, 2019). Arguably, as the White Helmets' reputation became more tarnished (e.g., by associating them with known terrorist groups), it became easier to justify physical attacks against them (Grey-Ellis, 2017). In August 2017, seven members of the group were murdered by unknown gunmen in their offices in the town of Sarmin in Idlib province (Mroue, 2017); Later the same year, a report in The Economist claimed that approximately one in six WH had been killed, predominately by "double-tap" airstrikes that take place at the same site as a prior attack specifically targeting the first responders as they search for bodies (Gaziantep, 2016); As of October 2018, it was reported that the White Helmets had lost close to ten percent of its members to the violence (Di Giovanni, 2018); and in November 2019, James Le Mesurier OBE[48], a former British Army Officer that founded Mayday Rescue, the organization that provided financial support to the WH, was found dead in Turkey with injuries consistent with falling from a height (Yee, 2019; Loveluck, 2019). The subsequent

[48] OBE is the acronym of the "Order of the British Empire", which rewards contributions to the arts and sciences, work with charitable and welfare organizations, and public service outside the civil service.

investigation found that his death was caused by suicide, which was attributed, at least in part, to the persistent disinformation campaign that sought to discredit him (Chulov, 2020).

Summary of studies in this dissertation

At the beginning of this dissertation, I set out with the aim to build a comprehensive understanding of the structure and dynamics of multi-platform online information operations, including the integration of government-controlled and alternative media, and the roles of different social media platforms (i.e., how they are used in complementary ways). My inquiry was guided by three main research questions:

1. How do information operations manifest on Twitter? To what extent does the emergent online crowd 'work' together as part of the information operation, and what kinds of work are they involved in?

2. How do government-controlled and alternative media websites shape and propagate strategic narratives? In particular, what are the content-sharing practices of distinct media domains involved in fomenting, shaping, and propagating competing strategic narratives?

3. How are multiple social media platforms used in information operations? What roles do different platforms play? And what kinds of participatory "work" do social media users engage in while involved in multi-platform information operations?

In each case I was interested in the collaborative work of the online crowd that enabled information operations to take shape. In particular I sought to look beyond the conceptualization of information operations as activities conducted by bots and trolls, instead interested in the participatory nature of information operations.

The starting point of each inquiry in this dissertation was a Twitter dataset of English-language tweets and the associated metadata about the Syrian Civil War collected between May 2017 and June 2018. Each of the studies used a subset of this data, which varied in temporal scope and keyword selection. For each study I used a mixed-method approach, based upon techniques devised in the field of crisis informatics (Palen & Anderson, 2016), combining both quantitative and qualitative research methods to pose questions of the data. Oftentimes this involved first analyzing data from a high level to identify patterns or anomalies of interest, and then using these as an entry point for a further low-level in-depth qualitative investigation.

This process was highly iterative, and I would move back and forth between different levels of analysis, whereby answers to one line of inquiry would generate new questions to ask of the data. In particular, I was guided by approaches such as trace ethnography (Geiger & Ribes, 2011), and, as the research progressed, investigative ethnography (Friedberg, 2020), which I used to interrogate digital traces as artifacts of online interactions. I used these traces to reconstruct the online discourse and the wider information ecosystem, for example by understanding the structure of the online conversation through a retweet network visualization, examining temporal patterns in online behavior in relation to current events in the Syrian Civil War, and also following URLs in tweets to map out the wider network of websites and social media platforms that host WH-related content.

As they are presented within the dissertation, the studies represent an increasing level of complexity, both in terms of the temporal scope and the extent of the inquiry: from 3 weeks to 13 months; and from a single platform (Twitter; Study 1, Study 2) to a network of websites (Study 3) and social media platforms (Study 4) linked to in the Twitter conversation. This illustrates how I worked to build up the broader view of the information ecosystem, using Twitter as a starting point (or what I referred to as my 'seed' data). I provide an overview of this progression and summarize the main findings of each of these studies below:

- ***Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community*** **(Study 1, Chapter 5)**: I focused on how an information operation manifests on a single platform, Twitter, during a specific ~3-week case study concerning the re-emergence of Omran Daqneesh (The "Aleppo Boy"). Omran appeared on Syrian state television with his father almost a year after he was rescued from the rubble of his family's home by the WH and photographed in the back—a moment that, in the Western mainstream media, came to represent the suffering of the Syrian people caught up in violence. In particular, I focused on the types of accounts, the content, and the behaviors that were exhibited in the online conversation surrounding Omran. The investigation revealed a digital battlefield of sorts, where two (or more) "sides" worked to shape the information space to support their political goals, destabilizing the current mainstream narrative about the causes of innocent citizens' suffering in Syria. I examined these activities as a form of information operations—efforts to distort the information space through both the propagation of preferred narratives (in some cases) and working to undermine the other side's narratives (in others). Furthermore, I argued that information operations were a form of collaborative work within an online community, conducted by an assemblage of diverse actors, driven by a variety of motivations, loosely collaborating on Twitter. One form of 'work' involved the production and propagation of multiple, inconsistent, and often conflicting narratives that

destabilized the "mainstream" narrative by distorting the information space and confusing the audience (Pomerantsev and Weiss, 2014; Rogers and Tyushka, 2018); this was coupled with efforts to challenge the credibility of established information sources, for example through the sustained and collaborative effort to discredit the mainstream media and their reporting, which in turn could generate uncertainty around their reporting on the topic. At a high level, Study 1 posits information operations as a concern for CSCW researchers and also points to the implicit collaborative work that takes place beyond prior conceptualizations of information operations as solely the product of bots and trolls.

- ***The White Helmets Conversation on Twitter*** **(Study 2, Chapter 6):** Sought to provide the reader with more detail about the WH conversation on Twitter. The analyses presented in this chapter represent the long-form, detailed investigation of the WH conversation on Twitter (and elsewhere) over the course of more than three years. Through this work, we identified two distinct communities involved in the WH conversation on Twitter: a community of accounts that were supportive of the group, and; a community of accounts that were highly critical of the group. These communities became a basis for further investigation including a temporal analysis of Twitter activity over time, a qualitative examination of tweets, and insights into their internal structure. This study sets up the work of Study 3 (Chapter 7), which examined the role of government-controlled and alternative media in fomenting, shaping, and propagating anti-WH narratives; and Study 4 (Chapter 8) which further probes the links between Twitter and other social media platforms in the examination of multi-platform information operations.

- ***The role of Government-controlled and Alternative Media*** **(Study 3, Chapter 7):** [my coauthors and] I looked beyond Twitter to the network of websites that are linked-to in tweets about the WH over a period of 3 months. By following URLs within tweets, we found a network of alternative and (to a lesser extent) mainstream media websites that repeatedly publish content (articles) about the White Helmets. Through an examination of the articles and the content sharing practices across these websites, we revealed a small set of websites and authors generating content that is spread across diverse sites, which we termed the "echo-system". The echo-system consisted of 130 distinct websites that provided source content for, or re-published existing content from, another website within the echo-system. By 'repackaging' (tailoring) articles to appeal to specific groups, the echo-system could draw audiences from distinct communities into shared but disruptive narratives about the role and function of the WH. These efforts were supported by the integration of government-controlled media and geopolitical think tanks that provided source content for anti-WH narratives.

This raised questions about the independence of the alternative media websites, which often frame themselves as autonomous and independent.

- ***Multi-platform Information Operations*** **(Study 4, Chapter 8):** I extended the WH Twitter dataset represented more than a year of tweets and their associated metadata, following URLs within (re)tweets to build a picture of the network of social media platforms from which content about the WH is produced, stored, and integrated into the Twitter conversation. Borrowing from Donovan (Berkman Klein Center, 2020), I viewed social media platforms as a form of sociotechnical infrastructure, comprising technology, users, and norms. I posited that since each social media platform is a unique combination of these sociotechnical aspects, they can be particularly suitable to serve a specific role or purpose within the larger information ecosystem. On Twitter I identified two communities of accounts: one supportive of the WH (pro-WH) the other highly critical of the group (anti-WH). Looking to the social media platforms that are linked-to in each side's (re)tweets, I found that while the pro-WH community links to mainstream platforms such as Facebook and Medium, the anti-WH side used a wider variety of platforms as they work to build resilience by using alternative (non-mainstream) platforms to counter the perceived threat of "censorship" from "big-tech"—a pejorative term for the large technology companies that own the platforms such as Google, Facebook, and Twitter (Donovan, Lewis, and Friedberg, 2019). I also noted that both sides were working to sustain their set of preferred narratives. This was particularly salient on the anti-WH side of the conversation because despite being highly critical of, and to some extent working against the mainstream media to propagate disruptive narratives about the WH, there were efforts to set the media agenda by producing or synthesizing existing content and then mobilizing it by introducing it to Twitter audiences, in particular those of media outlets, journalists, and prominent politicians, which was done through targeted replies to these accounts.

Knowledge Contributions

The broad empirical contribution of this dissertation is the revelation and exploration of online information operations—activities by state and non-state actors to distort public opinion to achieve strategic goals, through the dissemination of disinformation, amplification of specific accounts or content, or the manipulation of the information ecosystem—how these take shape on and across multiple websites and social media platforms, and the collaborative 'work' required to sustain them over time. This work is situated within HCI and CSCW, but the empirical findings and theoretical knowledge contributions speak to a broader audience, including researchers involved in studies of disinformation and online manipulation,

and those tasked with the development of social media platform policy. In the following subsections I will articulate the knowledge contributions made by this work.

Reconceptualized the View of Online Information Operations

Through this dissertation I make a theoretical contribution by reconceptualizing information operations—broadening the scope beyond that of coordinated inauthentic behavior. My research highlights the role of a motivated contingent of online activists including academics, bloggers, self-declared journalists, and politicians that engage in collaborative 'work' to create and disseminate a set of preferred narratives (e.g., Wilson, Zhou, & Starbird, 2018; Starbird, Arif, & Wilson, 2019; Wilson & Starbird, In Review). This is empirical evidence that online information operations are not solely the product of explicitly orchestrated bots and trolls (although those elements likely exist), but rather that they integrate into and leverage existing online communities to meet their strategic goals (Wilson, Zhou, & Starbird, 2018; Starbird, Arif, & Wilson, 2019; Wilson & Starbird, In Review). This has implications for platform policy because policies that seek out inauthentic and/or coordinated behavior are not a sufficient challenge to information operations.

Established Information Operations as a form of Collaborative 'Work'

My research has contributed to establishing online information operations as a form of collaborative work, highlighting the role of the collaborative (but not necessarily explicitly coordinated) work of the online crowd. In Wilson, Zhou, and Starbird (2018) we introduced the concept of online information operations to the CSCW community and posited this as an extension of current conceptualizations of online activism and online volunteerism, and of the collaborations that take shape around and through them. The aim of this was to initiate research interest about information operations and disinformation campaigns within the CSCW community—a phenomenon that the CSCW community should be concerned about but also has the tools and techniques to research further (Starbird, Arif, & Wilson, 2019).

Increased Awareness of Disinformation and Information Operations to the HCI and CSCW Communities

This work contributes to a growing body of research about information operations and online disinformation campaigns within the HCI, and more specifically CSCW communities. At the time of writing the first study in this dissertation (Chapter 5) in late 2017 and early 2018, there was very little academic research about information operations, active measures, or disinformation campaigns in HCI and CSCW. As such, part of my work as a junior PhD student involved translating what I was observing through my research for the HCI/CSCW community. This involved turning to historical accounts of disinformation

and active measures (e.g. Bittman 1972; Bittman, 1985), intelligence reports (e.g. Rees, 1984; United States Department of State, 1989), and whitepapers (e.g. Pomerantsev & Weiss, 2014; Morgan & DiResta, 2018) to provide the vocabulary for what I was observing, and then synthesizing this with existing literature in HCI and CSCW communities in order to establish information operations and disinformation (campaigns) as relevant to CSCW researchers (Wilson, Zhou, & Starbird, 2018; Starbird, Arif, & Wilson, 2019).

Considered the Mainstream-Alternative Dimension of a Hybrid Media System

I used the theoretical framework of the *hybrid media system* to explore how the different aspects of the information ecosystem were leveraged in this information operation. In doing so I demonstrated the utility of this framework for studying phenomena such as information operations and disinformation, particularly for those interested in the interactions between different parts of the system (i.e., the interactions between newer and older forms of media and actors ability to shape information flows through low-threshold produsage. Furthermore, the hybrid media system perspective helped to reveal the salient tension between the 'mainstream' and the 'alternative', which I suggest as an additional dimension to the framework.

Presented a Mixed Method Approach to Studies of Online Information Operations and Disinformation

In addition, the work presented in this dissertation contributes to a growing body of research that uses mixed methods approaches to examine phenomena such as disinformation, information operations, and online manipulation (e.g. Acker & Donovan, 2019; Arif et al., 2018; Krafft & Donovan, 2020; Sommariva et al., 2018)—using quantitative methods as a means of gaining entry to the data, but combining this with qualitative analysis of content and account activity to reconstruct events and build thicker understandings. Such an approach allows researchers to analyze large amounts of data (e.g., using quantitative methods to look for patterns and anomalies as entry points for further investigation) while also encouraging in-depth analysis at various 'levels' to uncover the nuanced and sophisticated activities. The studies presented here also demonstrate the importance of looking at the information ecosystem more broadly (i.e., not constraining research to a single platform) because information operations are not confined to the limits of any one particular platform, rather they take shape on and across platforms and websites, and through hybrid forms of media.

References

Acker, A., & Donovan, J. (2019). Data craft: a theory/methods package for critical internet studies. Information, Communication & Society, 22(11), 1590-1609.

Adams, T. (2018, March 24). Facebook's week of shame: the Cambridge Analytica fallout. The Guardian. Retrieved from https://www.theguardian.com/technology/2018/mar/24/face

Al-Ani, B., Mark, G., & Semaan, B. (2010). Blogging in a region of conflict. In Proceedings of the 28th international conference on Human factors in computing systems - CHI '10 (p. 1069). New York, New York, USA: ACM Press. https://doi.org/10.1145/1753326.1753485

Allport, G. W., & Postman, L. (1947). The psychology of rumor.

Alvarez-Monzoncillo, J.M. (2011). Watching the Internet: The Future of TV?. Media XXI.

Ambrosio, T. (2007). Insulating Russia from a Colour Revolution: How The Kremlin Resists Regional Democratic Trends. Democratisation, 14(2), 232-252.

Arif, A., Robinson, J. J., Stanek, S. A., Fichet, E. S., Townsend, P., Worku, Z., & Starbird, K. (2017). A closer look at the self-correcting crowd: Examining corrections in online rumors. In Proceedings of the 2017 ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 155-168).

Arif, A., Stewart, L. G., & Starbird, K. (2018). Acting the part: Examining information operations within# BlackLivesMatter discourse. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-27.

Arif, A., Stewart, L. G., & Starbird, K. (2018). Acting the part: Examining information operations within# BlackLivesMatter discourse. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-27. http://doi.org/10.1145/3274289

Armistead, L. ed. (2004) Information operations: Warfare and the hard reality of soft power. Potomac Books, Inc

Associated Press. (2013, September 6). Russia to Keep Helping Syria If It's Attacked. Associated Press Financial Wire

Atkinson, J., & Dougherty, D. S. (2006). Alternative media and social justice movements: The development of a resistance performance paradigm of audience analysis. Western Journal of Communication, 70(1), 64-88.

Atton, C. (2002). Alternative Media. London: Sage

Atwood, K. (2018). U.S. freezes funding for Syria's "White Helmets. CBS News. Retrieved January 14, 2021 from: https://www.cbsnews.com/news/u-s-freezes-funding-for-syrias-white-helmets/

Barthel, M. & Mitchell, A., (2017). Americans' attitudes about the news media deeply divided along partisan lines. Pew Research Center, 10.

BBC News. (2016, March 11). Syria: The Story of the Conflict. BBC

BBC Trending. (2018). Syria war: The online activists pushing conspiracy theories. BBC News.

Benkler, Y., (2006). The wealth of networks: How social production transforms markets and freedom. Yale University Press.

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A. & Etling, B. (2015) Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate. Political communication, 32(4), pp.594-624.

Bennett, W. (2003). Communicating global activism. Information, Communication & Society, 6(2), 143-168.

Berkman Klein Center. (2020). Joan Donovan on domestic misinformation. The Breakdown. Retrieved January 11, 2021 from: https://medium.com/berkman-klein-center/the-breakdown-joan-donovan-on-domestic-misinformation-da4f60296420

Bimber, B., Flanagin, A. J., & Stohl, C. (2005). Reconceptualizing collective action in the contemporary media environment. Communication theory, 15(4), 365-388.

Bittman, L. (1972). The deception game: Czechoslovak intelligence in soviet political warfare. Syracuse University Press.

Bittman, L. (1985) The KGB and Soviet disinformation: an insider's view. Washington: Pergamon-Brassey's.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

Bordia, P., & Difonzo, N. (2004). Problem Solving in Social Interactions on the Internet: Rumor As Social Cognition. Social Psychology Quarterly, 67(1), 33–49. https://doi.org/10.1177/019027250406700105

boyd, D., & Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Information, Communication, and Society. 15, 662–679. doi: 10.1080/1369118X.2012.678878

boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In 2010 43rd Hawaii international conference on system sciences (pp. 1-10). IEEE.

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. Journal of the American society for information science and technology, 59(9), 1493-1512.

Breiter, A. & Hepp, A. (2018). The complexity of datafication: putting digital traces in context. In Communicative figurations (pp. 387-405). Palgrave Macmillan, Cham.

Bruns, A. (2008). Blogs, Wikipedia, Second Life, and beyond: From production to produsage (Vol. 45). Peter Lang.

Burgess, J. & Green, J. (2018). YouTube: Online video and participatory culture. John Wiley & Sons.

Castells, M. (2013). Communication Power. Oxford University Press.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring user influence in twitter: The million follower fallacy. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 4, No. 1).

Chadwick, A. (2013). The Hybrid Media System: Politics and Power. Oxford: Oxford University Press.

Chadwick, A. (2017). The Hybrid Media System: Politics and Power. Oxford University Press.

Chadwick, A., Dennis, J., Bruns, A., Gunn, E., Skogerbo, E., Larsson A.O., & Christensen, C. (2016). Politics in the Age of Hybrid Media: Power, Systems, and Media Logics: 7-22. Harvard

Chadwick, A. (2011). The Hybrid Media System. In European Consortium for Political Research General Conference, Reykjavik, Iceland (Vol. 25).

Chaney, A.J., Stewart, B.M. & Engelhardt, B.E. (2018). How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility. In Proceedings of the 12th ACM Conference on Recommender Systems (pp. 224-232).

Charmaz, K. (2014). Constructing Grounded Theory. Sage.

Chivvis, C. (2017). Understanding Russian “Hybrid Warfare”: and What Can Be Done About It. RAND Corporation. Retrieved January 11, 2021 from https://doi.org/10.7249/CT468

Chulov, M. (2020, October 27). How Syria’s Disinformation Wars Destroyed the Co-founder of the White Helmets. The Guardian. Retrieved January 11, 2021 from

https://www.theguardian.com/news/2020/oct/27/syria-disinformation-war-white-helmets-mayday-rescue-james-le-mesurier

CNN Business. (2020, November 23). Newsmax and OANN are telling lies about the election as more people tune in. CNN Business. Retrieved January 11, 2021 from https://www.cnn.com/videos/business/2020/11/23/newsmax-oan-trump-conspiracy-theories-ratings-orig-vf.cnnbusiness

Comfort, L.K., Ko, K. and Zagorecki, A. (2004). Coordination in rapidly evolving disaster response systems: The role of information. American behavioral scientist, 48(3), pp.295-313.

Cook, J. (2014). Profile Of Hayden Hewitt, Founder Of LiveLeak - Business Insider. Business Insider. Retrieved January 11, 2021 from https://www.businessinsider.com/profile-of-hayden-hewitt-founder-of-liveleak-2014-10

Corman, S. R. (2013). The Difference between Story and Narrative. Retrieved April 15, 2018, from http://csc.asu.edu/2013/03/21/the-difference-between-story-and-narrative/

Corpus Ong, J., & Vincent A Cabanes, J. (2018). Architects of Networked Disinformation. Retrieved October 15, 2020 from http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf

Cybenko, G., Giani, A., & Thompson, P. (2002). Cognitive hacking: A battle for the mind. Computer, 35(8), 50–56.

Dailey, D., & Starbird, K. (2017). Social Media Seamsters. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 (pp. 1277–1289). New York, New York, USA: ACM Press. https://doi.org/10.1145/2998181.2998290

Darley, W. M. (2006). Clausewitz's Theory of War and Information Operations. ARMY COMBINED ARMS CENTER FORT LEAVENWORTH.

Dearden, L. (2017, February 11). NATO Accuses Sputnik News of Distributing Misinformation as Part of 'Kremlin Propaganda Machine'. The Independent

DiFonzo, N., & Bordia, P. (2007). Rumor psychology: Social and organizational approaches (Vol. 1). American Psychological Association Washington, DC.

di Giovanni, J., (2018). Why Assad and Russia Target the White Helmets. The New York Review of Books.

Digital Forensic Research Lab (DFRLab). (2018) Question That: RT's Military Mission. Assessing Russia Today's role as an "information weapon". Medium. January 8, 2018

DiNucci, D. (1999). Fragmented future. Print, 53(4), pp.32-33.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. and Johnson, B. (2019). The tactics & tropes of the Internet Research Agency. Retrieved January 11, 2021 from https://digitalcommons.unl.edu/senatedocs/2

Donovan, J., Lewis, B. & Friedberg, B. (2019). Parallel ports: Sociotechnical change from the alt-right to alt-tech. Social Media + Society January-March 2018: 1–12 http://doi.org/10.14361/9783839446706-004

Donovan, J. (2018). After the# Keyword: Eliciting, sustaining, and coordinating participation across the occupy movement. Social Media + Society, 4(1), p.2056305117750720.

Donovan, J. (2019). How memes got weaponized: A short history | Berkman Klein Center. MIT Technology Review. Retrieved January 11, 2021 from https://cyber.harvard.edu/story/2019-10/how-memes-got-weaponized-short-history

Douek, E. (2020). The Rise of Content Cartels. Knight First Amendment Institute. Retrieved January 11, 2021 from https://knightcolumbia.org/content/the-rise-of-content-cartels

Dynes, R.R. (1970). Organized Behavior in Disaster. Lexington, MA: Heath Lexington Books.

Electronic Frontier Foundation. (n.d.) Section 230 of the Communications Decency Act | Electronic Frontier Foundation. Retrieved January 11, 2021 from https://www.eff.org/issues/cda230

Facebook. (2020). Community Standards | Facebook. Retrieved January 11, 2021 from https://www.facebook.com/communitystandards/inauthentic_behavior/

Facebook. (2020). Community Standards | Facebook. Retrieved January 11, 2021 from https://www.facebook.com/communitystandards/inauthentic_behavior/

Farkas, J., Schou, J., & Neumayer, C. (2018). Cloaked Facebook pages: Exploring fake Islamist propaganda in social media. New Media & Society, 20(5), 1850–1867.

Fenby, J. (1986). The International News Services. New York: Schocken Books.

Flores-Saviaga, C., Keegan, B. C., & Savage, S. (2018). Mobilizing the Trump Train: Understanding Collective Action in a Political Trolling Community. Stanford, CA. Retrieved from https://arxiv.org/pdf/1806.00429.pdf

Foot, K. A., Schneider, S. M., & Cornfield, M. (2006). Web campaigning (p. 263). Cambridge, MA: MIT press. http://doi.org/10.7551/mitpress/7186.001.0001

Fouad, F.M., Sparrow, A., Tarakji, A., Alameddine, M., El- Jardali, F., Coutts, A.P., El Arnaout, N., Karroum, L.B., Jawad, M., Roborgh, S. & Abbara, A. (2017). Health Workers and the Weaponisation of Health Care in Syria: A Preliminary Inquiry for The Lancet–American University of Beirut Commission on Syria (p. 2518). The Lancet.

Freelon, D., (2014). On the interpretation of digital trace data in communication and social computing research. Journal of Broadcasting & Electronic Media, 58(1), pp.59-75.

Freelon, D., Marwick, A., & Kreiss, D. (2020). False equivalencies: Online activism from left to right. Science, 369(6508), 1197-1201.

Freelon, D., McIlwain, C. D., & Clark, M. (2016). Beyond the hashtags:# Ferguson,# Blacklivesmatter, and the online struggle for offline justice. Center for Media & Social Impact, American University.

Freiman, J. (2020). YouTube suspends One America News Network channel for violating COVID-19 misinformation policy. CBS News. Retrieved January 11, 2021 from https://www.cbsnews.com/news/oan-suspended-youtube-covid-19-misinformation/

Friedberg, B. (2020). Investigative Digital Ethnography: Methods for Environmental Modeling. Retrieved January 11, 2021 from https://mediamanipulation.org/research/investigative-digital-ethnography-methods-environmental-modeling

Frisk, A. (2016, October 20). This iconic photo of Syrian boy pulled from rubble in Aleppo is fake, according to President Bashar Assad. Global News CA. Retrieved April 15, 2018, from https://globalnews.ca/news/3015545/this-iconic-photo-of-syrian-boy-pulled-from-rubble-in-aleppo-is-fake-according-to-president-bashar-assad/

Fritz, C.E., & Mathewson, J.H. (1957). “Convergence Behavior in Disasters: A Problem in Social Control.” Special Report. National Academy of Sciences- National Research Council 476.

Fuller, J. (2010). What is Happening to News: The Information Explosion and the Crisis in Journalism. University of Chicago Press.

Gadde, V. and Roth, Y. (2018). Enabling further research of information operations on Twitter. Twitter Blog. Retrieved January 11, 2021 from https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html

Gartenberg, C. (2019). What is Facebook? Just ask Mark Zuckerberg - The Verge. The Verge. Retrieved January 12, 2021 from https://www.theverge.com/2019/3/8/18255269/facebook-mark-zuckerberg-definition-social-media-network-sharing-privacy

Gasson, S. & Waters, J. (2013). Using a grounded theory approach to study online collaboration behaviors. European Journal of Information Systems, 22(1), pp.95-118.

Gaziantep. (2016, October 15). The rise of Syria's White Helmets. The Economist. https://www.economist.com/middle-east-and-africa/2016/10/15/the-rise-of-syrias-white-helmets

Geertz, C. (2008). Local knowledge: Further essays in interpretive anthropology. Basic books.

Geiger, R.S. & Halfaker, A. (2017). Operationalizing conflict and cooperation between automated software agents in wikipedia: A replication and expansion of 'even good bots fight'. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW), pp.1-33.

Geiger, R.S. & Ribes, D. (2010). The work of sustaining order in wikipedia: the banning of a vandal. In Proceedings of the 2010 ACM conference on Computer supported cooperative work (pp. 117-126). ACM.

Geiger, R.S. & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. In 2011 44th Hawaii International Conference on System Sciences (pp. 1-10). IEEE.

Gerasimov, V. (2013). "The value of science is in the foresight: New challenges demand rethinking the forms and methods of carrying out combat operations," Military-Industrial Kurier, 27 February 2013, Robert Coalson, translated., Military Review 96, no. 1 (2016), 23.

Gillespie, T. (2010). The politics of 'platforms'. New media & society, 12(3), 347-364.

Gillmor, D. (2006). We the Media: Grassroots Journalism By the People, For the People. O'Reilly Media, Inc.

Gilsinan, K. (2015). The Confused Person's Guide to the Syrian Civil War. The Atlantic. Retrieved April 15, 2018 from https://www.theatlantic.com/international/archive/2015/10/syrian-civil-war-guide-isis/410746/

Gordon, M.R. (2014, September 22). U.S. Strikes Follow Plea by Syrian Opposition Leader on Behalf of Kurds. The New York Times

Grey Ellis, E. (2017, April 30). Inside the Conspiracy which Turned Syria's First Responders into Terrorists. Wired. Retrieved February 1, 2021 from https://www.wired.com/2017/04/white-helmets-conspiracy-theory/

Guidi, B. (2020). When blockchain meets online social networks. Pervasive and Mobile Computing, 62, 101131. http://doi.org/10.1016/j.pmcj.2020.101131

Habermas, J., Lennox, S. and Lennox, F., (1974). The public sphere: An encyclopedia article (1964). New German Critique, (3), pp.49-55.

Halverson, J., Corman, S. and Goodall, H.L. (2011) Master narratives of Islamist extremism. Springer.

Halverson, J. (2011) Why Story is Not Narrative. http://csc.asu.edu/2011/12/08/why-story-is-not-narrative/

Herman, Edward S, & Noam Chomsky. (1988). Manufacturing Consent: The Political Economy of the Mass Media. New York: Pantheon Books.

Hermida, A., Lewis, S.C. & Zamith, R. (2014). Sourcing the Arab Spring: A case study of Andy Carvin's sources on Twitter during the Tunisian and Egyptian revolutions. Journal of Computer-Mediated Communication, 19(3), pp.479-499.

Herrman, J. (2014, March 24). The Internet's Endless Appetite for Death Video - The New York Times. Retrieved January 11, 2021 from https://www.nytimes.com/2019/03/24/style/really-bad-stuff.html

Holiday, R. (2013). Trust me, I'm Lying: Confessions of a Media Manipulator. Penguin.

Hollis, D.B. (2007). Why states need an international law for information operations. Lewis & Clark L. Rev., 11, p.1023. https://law.lclark.edu/live/files/9551-lcb114art7hollis.pdf

Hopkins, N. (2017, May 21). Revealed: Facebook's internal rulebook on sex, terrorism and violence | News | The Guardian. The Guardian. Retrieved January 12, 2021 from https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence

Horrigan, J. (2007). 'A Typology of Information and Communication Technology Users', Pe Internet and American Life, 6 May, URL (consulted April 2009): http://www.pewinternet.org/Reports/2007/A-Typology-of-Information-and-Communication-Technology-Users.aspx

Horwitz, J. (2018). YouTube is turning to Wikipedia to help it fight conspiracy theories — Quartz. Quartz. Retrieved January 11, 2021 from https://qz.com/1228635/youtube-is-turning-to-wikipedia-to-help-it-fight-conspiracy-theories/

Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Maziad, M. (2011). Opening closed regimes: What was the role of social media during the Arab Spring?. SSRN 2595096.

Huang, Y. L., Starbird, K., Orand, M., Stanek, S. A., & Pedersen, H. T. (2015). Connected through crisis: Emotional proximity and the spread of misinformation online. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (pp. 969-980).

Hughes, A., Palen,L., Sutton, J., Liu, S., & Vieweg, S. (2008). "Site-Seeing" in Disaster: An Examination of On-Line Social Convergence Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM 2008).

Hughes, A. & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. Proceedings of the 2009 Information Systems for Crisis Response and Management Conference (ISCRAM 2009), Gothenburg, Sweden.

Hughes, A. L., Starbird, K., Leavitt, A., Keegan, B., & Semaan, B. (2016). Information movement across social media platforms during crisis events. ACM Computer Human Interaction (CHI) 2016 Extended Abstracts, 1, 5.

International Committee of the Red Cross. (2012). Syria: ICRC and Syrian Arab Red Crescent maintain aid effort amid increased fighting. Retrieved from https://www.icrc.org/eng/resources/documents/update/2012/syria-update-2012-07-17.htm

International Committee of the Red Cross (ICRC). (2012). "Syria: ICRC and Syrian Arab Red Crescent Maintain Aid Effort amid Increased Fighting," Operational Update (Geneva: International Committee of the Red Cross, July 17, 2012)..

Jack, C. (2017). Lexicon of lies: Terms for problematic information. Data & Society, 3.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS One, 9(6), e98679.

Jenne, E. K., & Popovic, M. (2017). Managing Internationalized Civil Wars. In Oxford Research Encyclopedia of Politics.

Joint Chiefs of Staff. (2006). Information Operations https://www.hsdl.org/?view&did=461648

Jowett, G., & O'Donnell, V. (2015). Propaganda & persuasion (Sixth ed.). Thousand Oaks, Calif.: SAGE.

Karahalios, K., Fitzpatrick, G., & Monroy-Hernández, A. (2017). Editor's Note/Chairs' Welcome. Proceedings of ACM Human-Computer Interaction 1, CSCW, Article 16 (November 2017). DOI:https://doi.org/10.1145/3134651

Kendra, J., & Wachtendorf, T. (2003). Creativity in Emergency Response to the World Trade Center Disaster. Preliminary Paper. Presented at the 9th Annual Conference of The International Emergency Management Society. http://udspace.udel.edu/handle/19716/733.

Kenix, L. J. (2011). Alternative and Mainstream Media: The Converging Spectrum. London & New York: Bloomsbury Academic.

Krafft, P. M., & Donovan, J. (2020). Disinformation by design: The use of evidence collages and platform filtering in a media manipulation campaign. Political Communication, 37(2), 194-214. http://doi.org/10.1080/10584609.2019.1686094

Kramer, A.E., & Barnard A. (2015, October 6). Russian Soldiers Join Syria Fight. The New York Times

Kramer, A.E. (2019, March 2). Russian General Pitches 'Information' Operations as a Form of War. New York Times. Retrieved February 1, 2021 from https://www.nytimes.com/2019/03/02/world/europe/russia-hybrid-war-gerasimov.html

Larkin, A. & Lewis, S. (2017, February 26). Netflix documentary "The White Helmets" wins Oscar. CNN. Retrieved January 11, 2021 from https://www.cnn.com/2017/02/26/us/white-helmets-oscar/index.html

Levinger, M. (2018). MASTER NARRATIVES OF DISINFORMATION CAMPAIGNS. Journal of International Affairs

Lewin, K. (1943). Forces behind food habits and methods of change. Bulletin of the national Research Council, 108(1043), 35-65.

Lieberman, M. (2020). A growing group of journalists has cut back on Twitter, or abandoned it entirely - Poynter. Poynter. Retrieved January 11, 2021 from https://www.poynter.org/reporting-editing/2020/a-growing-group-of-journalists-has-cut-back-on-twitter-or-abandoned-it-entirely

LiveLeak. (n.d.) LiveLeak.com - FAQ. Retrieved January 11, 2021 from https://www.liveleak.com/faq

LiveLeak. (n.d.) LiveLeak.com - Rules. Retrieved January 11, 2021 from https://www.liveleak.com/rules

Loveluck, L. (2019, November 11). James Le Mesurier, backer of Syria's White Helmets, found dead in Istanbul. Washington Post. Retrieved January 11, 2021 from https://www.washingtonpost.com/world/james-le-mesurier-backer-of-syrias-white-helmets-found-dead-in-istanbul/2019/11/11/e75da222-04ab-11ea-a5e2-fccc16fa3576_story.html

Lucas, E., & Nimmo, B. (2015) CEPA INFOWAR PAPER No . 1 Information Warfare : What Is It and How to Win It? (CEPA Infowar No. 1). Washington D.C. Retrieved June 15 2020 from http://cepa.org/files/?id_plik=1896

Lucas, E., & Pomeranzev, P. (2016). Winning the Information War: Techniques and Counter-strategies to Russian Propaganda in Central and Eastern Europe. Center for European Policy Analysis.

Luhn, A. (2015, October 13). Russia's Reality Trolls and the MH17 War of Misinformation. Retrieved February 10, 2021 from: https://foreignpolicy.com/2015/10/13/russias-reality-trolls-and-the-mh17-war-of-misinformation-buk-missile/

Maddock, J., Starbird, K., & Mason, R. M. (2015). Using historical Twitter data for research: Ethical challenges of tweet deletions. In CSCW 2015 workshop on ethics for studying sociotechnical systems in a Big Data World. ACM.

Maddock, J., Starbird, K., Al-Hassani, H.J., Sandoval, D.E., Orand, M. & Mason, R.M. (2015). Characterizing online rumoring behavior using multi-dimensional signatures. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 228-241).

Marres, N., & Rogers, R. (2005). Recipe for Tracing the Fate of Issues and their Publics on the Web.

Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online. Data & Society Research Institute.

Matsakis, L. (2018). YouTube Will Link Directly to Wikipedia to Fight Conspiracy Theories | WIRED. Wired. Retrieved January 11, 2021 from https://www.wired.com/story/youtube-will-link-directly-to-wikipedia-to-fight-conspiracies/

McCarthy, T. (2020, May 28). Zuckerberg says Facebook won't be "arbiters of truth" after Trump threat. The Guardian. Retrieved January 12, 2021 from https://www.theguardian.com/technology/2020/may/28/zuckerberg-facebook-police-online-speech-trump

McCaughey, M., & Ayers, M. D. (Eds.). (2003). Cyberactivism: Online activism in theory and practice. Psychology Press.

McLuhan, M. (2003). Understanding me: Lectures & interviews. Cambridge, MA: MIT Press.

Medium. (2019). Medium Rules | by Medium | Medium Policy. Retrieved January 11, 2021 from https://policy.medium.com/medium-rules-30e5502c4eb4

Meikle, G. (2002). Future active: Media activism and the Internet. Psychology Press.

Merriam-Webster “social media”. (2020). Definition from Merriam-Webster.com. Retrieved January 10, 2021 from https://www.merriam-webster.com/dictionary/social media

Miskimmon, A., O'loughlin, B. and Roselle, L. (2014). Strategic narratives: Communication power and the new world order. Routledge.

Monaco, N. J. (2017). Computational Propaganda in Taiwan: Where Digital Democracy Meets Automated Autocracy, (2), 34. Retrieved from http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Taiwan-2.pdf

Morgan, J., & DiResta, R. (2018). Information Operations are a Cybersecurity Problem: Toward a New Strategic Paradigm to Combat Disinformation. Just Security, 10. Retrieved January 11, 2021 from https://www.justsecurity.org/59152/information-operations-cybersecurity-problem-strategic-paradigm-combat-disinformation/

Moscow Times. (2014, September 23). Looking West: Russia Beefs Up spending on global media giants. Moscow Times

Mroue, B. (2017, August 12). 7 White Helmets Medics Killed in Syria’s Idlib. Associated Press

Narasimhan, R. (2005). Looking Beyond Flawed Journalism: How National Interests, Patriotism, and Cultural Values Shaped the Coverage of the Iraq War.Harvard International Journal of Press/Politics 10.1:45-62.

Nardi, B.A. and O'Day, V. (1999). Information ecologies: Using technology with heart. MITPress

Nauert, H. (2018). Funding for the Syrian Civil Defense and UN International Impartial and Independent Mechanism. US State Department Press Statement.

Nielsen, R.K., (2013). Mundane internet tools, the risk of exclusion, and reflexive movements—Occupy Wall Street and political uses of digital networked technologies. The Sociological Quarterly, 54(2), pp.173-177.

Nyhan, B. and Reifler, J.A., (2012). Misinformation and Fact-Checking: Research findings from Social Science. New America Foundation.

Olson, J. S., & Kellogg, W. A. (Eds.). (2014). Ways of Knowing in HCI (Vol. 2). New York, NY, USA:: Springer.

Olteanu, A., Castillo, C., Diaz, F. & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Frontiers in Big Data, 2, p.13.

Ou-Yang, L. (2017). Newspaper [Python library]. Retrieved December 2017 from https://github.com/codelucas/newspaper

P. M. Krafft & Donovan, J. (2020). Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign. Political Communication. http://doi.org/10.1080/10584609.2019.1686094

Palen, L., & Liu, S. B. (2007). Citizen communications in crisis. In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07 (p. 727-736). New York, New York, USA: ACM Press. https://doi.org/10.1145/1240624.1240736

Palen, L., Anderson, J., Bica, M., Castillo, C., Crowley, J., Pérez, P.D., Finn, M., Grace, R., Hughes, A., Imran, M. & Kogan, M. (2020). Crisis Informatics: Human-Centered Research on Tech & Crises.

Palen, L., Vieweg, S., Liu, S.B.& Hughes, A.L. (2009). Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event. Social Science Computer Review, 27(4), pp.467-480.

Palen, L. & Anderson, K.M. (2016). Crisis informatics—New data for extraordinary times. Science, 353(6296), pp.224-225.

Paul, C & Matthews, M. (2016). The Russian "Firehose of False- hood" Propaganda Model: Why It Might Work and Options to Counter It. Santa Monica, CA: RAND Corporation, 2016

Paul, C. (2019). Is It Time to Abandon the Term Information Operations? https://thestrategybridge.org/the-bridge/2019/3/11/is-it-time-to-abandon-the-term-information-operations

Pennycook, G., Cannon, T.D. and Rand, D.G. (2018). Prior exposure increases perceived accuracy of fake news. Journal of experimental psychology: general.

Pictet, J. (1979). The Fundamental Principles of the Red Cross: Commentary International Review of the Red Cross, 210. Retrieved February 1, 2021 from https://www.icrc.org/eng/resources/documents/misc/fundamental-principles-commentary-010179.htm

Pilloud, C., Sandoz, Y., Swinarski, C. and Zimmermann, B. eds., (1987). Commentary on the additional protocols: of 8 June 1977 to the Geneva Conventions of 12 August 1949. Martinus Nijhoff Publishers.

Poell, T., & Borra, E. (2012). Twitter, YouTube, and Flickr as platforms of alternative journalism: The social media account of the 2010 Toronto G20 protests. Journalism, 13(6), 695-713.

Poell, T. (2014). Social media and the transformation of activist communication: exploring the social media ecology of the 2010 Toronto G20 protests. Information, Communication & Society, 17(6), 716-731.

Pomerantsev, P., & Weiss, M. (2014). The Menace of Unreality: How the Kremlin Weaponizes Information, Culture and Money. New York: Institute of Modern Russia.

Posetti, J. and Matthews, A. (2018). A short guide to the history of 'fake news' and disinformation. International Center for Journalists, https://www. icfj. org/sites/default/files/2018-07/A% 20Short, 20.

Powers, M. (2014) The Structural Organization of NGO Publicity Work: Explaining Divergent Publicity Strategies at Humanitarian and Human Rights Organizations. International Journal of Communication, 8, pp. 90-107.

Putin, Vladimir. (2012, February 22). Russia and the Changing World, Valdai

Pyrhönen, N. & Bauvois, G. (2020). Conspiracies beyond Fake News. Produsing Reinformation on Presidential Elections in the Transnational Hybrid Media System. Sociological Inquiry, 90(4), pp.705-731.

Qiu, L. (2017, February 24). Fact Check: Trump Blasts "Fake News" and Re- peats Inaccurate Claims at CPAC. The New York Times

Ratkiewicz, J., Conover, M., Meiss, M. R., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Detecting and Tracking Political Abuse in Social Media. In Proceedings of the International AAAI Conference on Web and Social Media, 11, 297-304.

Rees, D. (1984) Soviet Active Measure: The Propaganda War. Institute for the Study of Conflict, London, UK.

Reuter, C., Marx, A., & Pipek, V. (2011). Social Software as an Infrastructure for Crisis Management—A Case Study About Current Practice and Potential Usage. Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM 2011)

Rid, T. (2020) Active Measures: The Secret History of Disinformation and Political Warfare. Farrar, Straus and Giroux.

Rogers, J., & Tyushka, A. (2018). Hacking Into the West: Russia's Anti-hegemonic Drive and the Strategic Narrative Offensive, 35–60. Retrieved from https://www.stratcomcoe.org/james-rogers-andriy-tyushka-hacking-west-russias-anti-hegemonic-drive-and-strategic-narrative

Romm, T. & Lerman, R. (2021). Amazon suspends Parler, after Google and Apple also take action. Washington Post. Retrieved January 12, 2021 from https://www.washingtonpost.com/technology/2021/01/09/amazon-parler-suspension/

Rudin, M. (2016, April 25). Conspiracy Files: Who shot down MH17?Retrieved February 10, 2021 from: https://www.bbc.com/news/magazine-35706048

Ruptly. (2014). Ruptly Video News Agency and LiveLeak.com announce content partnership -- Ruptly | PRLog. PR Log. Retrieved January 11, 2021 from https://www.prlog.org/12299623-ruptly-video-news-agency-and-liveleakcom-announce-content-partnership.html

Rychiak, R. & Pacepa, I.M. (2013). Disinformation: Former Spy Chief Reveals Secret Strategies for Undermining Freedom, Attacking Religion, and Promoting Terrorism. Washington, Dc: WND books.

Salton, G., Fox, E.A. and Wu, H. (1983). Extended Boolean information retrieval. Communications of the ACM, 26(11), pp.1022-1036.

Savage, S., & Monroy-Hernández, A. (2015). Participatory Militias. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15 (pp. 724–733). New York, New York, USA: ACM Press. https://doi.org/10.1145/2675133.2675295

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. The Communication Review, 14(3), 197-215.

Shibutani, T. (1966). Improvised news: A sociological study of rumor. Ardent Media

Shoemaker, P. J. (1991). Gatekeeping.

Shoemaker, P. J., Eichholz, M., Kim, E., & Wrigley, B. (2001). Individual and routine forces in gatekeeping. Journalism & mass communication quarterly, 78(2), 233-246.

Shoemaker, P. J., & Vos, T. (2009). Gatekeeping theory. Routledge.

Soden, R. & Palen, L. (2018). Informating crisis: Expanding critical perspectives in crisis informatics. Proceedings of the ACM on human-computer interaction, 2(CSCW), pp.1-22.

Solon, O. (2017, December 18). How Syria's White Helmets became victims of an online propaganda machine. The Guardian. Retrieved January 11, 2021 from https://www.theguardian.com/world/2017/dec/18/syria-white-helmets-conspiracy-theories

Sommariva, S., Vamos, C., Mantzarlis, A., Đào, L. U. L., & Martinez Tyson, D. (2018). Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study. American journal of health education, 49(4), 246-255.

Spiro, E. S., Fitzhugh, S., Sutton, J., Pierski, N., Greczek, M., & Butts, C. T. (2012). Rumoring during extreme events: A case study of Deepwater Horizon 2010. In Proceedings of the 4th annual ACM web science conference (pp. 275-283).

Stamos, A. (2017). An Update On Information Operations On Facebook. Retrieved April 15, 2018, from https://newsroom.fb.com/news/2017/09/information-operations-update/

Starbird, K., & Palen, L. (2010). Pass it on?: Retweeting in mass emergency. Proceedings of the 7th International ISCRAM Conference, (December 2004), 1–10. https://doi.org/10.1111/j.1556-4029.2009.01231.x

Starbird, K., & Palen, L. (2011). Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis. In Proceedings of the Conference on Human Factors in Computing Systems (CHI), 1071–1080. https://doi.org/10.1145/1978942.1979102

Starbird, K., & Palen, L. (2012). (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising. In Proceedings of the acm 2012 conference on computer supported cooperative work (pp. 7–16). ACM.

Starbird, K., & Palen, L. (2013). Working and sustaining the virtual Disaster Desk. In Proceedings of the 2013 conference on Computer supported cooperative work (pp. 491–502). ACM.

Starbird, K., Arif, A., Wilson, T., Van Koevering, K., Yefimova, K. and Scarnecchia, D. (2018). Ecosystem or echo-system? exploring content sharing across alternative media domains. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 12, No. 1).

Starbird, K., Arif, A. & Wilson, T. (2019). Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), pp.1-26.

Starbird, K., Spiro, E., Edwards, I., Zhou, K., Maddock, J., & Narasimhan, S. (2016). Could This Be True?: I Think So! Expressed Uncertainty in Online Rumoring. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 360–371. https://doi.org/10.1145/2858036.2858551

Starbird, K., Spiro, E. S., Arif, A., Chou, F., Narisimhan, S., Maddock, J. & Robinson, J. (2015). Expressed Uncertainty and Denials as Signals of Online Rumoring. Proceedings of Collective Intelligence.

Starbird, K. (2013). Delivering patients to sacré coeur. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13 (p. 801). New York, New York, USA: ACM Press. https://doi.org/10.1145/2470654.2470769

Starbird, K. (2017). Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media (pp. 230-239).

Starbird, K. & Palen, L. (2012). (How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In Proceedings of the acm 2012 conference on computer supported cooperative work (pp. 7-16).

Steem. (2017). An incentivized, blockchain-based, public content platform. Retrieved January 11, 2021 from https://steem.com/SteemWhitePaper.pdf

Steemit. (n.d.) Faq — Steemit. Retrieved January 11, 2021 from https://steemit.com/faq.html#How_does_Steemit_work

Stryker, C. (2014, October 10). Murder, Mayhem and the Evolution of Website LiveLeak. Newsweek. Retrieved January 11, 2021 from https://www.newsweek.com/2014/10/10/murder-mayhem-and-evolution-website-liveleak-273963.html

Stubbs, J., & Gibson, G. (2017, November 13). Russia's RT America Registers as 'Foreign Agent' in U.S. Reuters.

Swift, A. (2016). Americans' Trust in Mass Media Sinks to New Low. Gallup. From http://news.gallup.com/poll/195542/americans- trust-mass-media-sinks-new-low.aspx

Syrian Observatory for Human Rights. (2020). Syrian Revolution NINE years on: 586,100 persons killed and millions of Syrians displaced and injured. Retrieved February 1, 2021 from https://www.syriahr.com/en/157193/

Thorson, K., Driscoll, K., Ekdale, B., Edgerly, S., Thompson, L. G., Schrock, A., & Wells, C. (2013). YouTube, Twitter and the Occupy movement: Connecting content and circulation practices. Information, Communication & Society, 16(3), 421-451.

Troianovski, A. (2018, February 17). A former Russian troll speaks: 'It was like being in Orwell's world'. The Washington Post https://www.washingtonpost.com/news/worldviews/wp/2018/02/17/a-former-russian-troll-speaks-it-was-like-being-in-orwells-world/

Tsvetkova, M., García-Gavilanes, R., Floridi, L. & Yasseri, T. (2017). Even good bots fight: The case of Wikipedia. PloS one, 12(2), p.e0171774.

Tufekci, Z. (2014). “Big questions for social media big data: Representativeness, validity and other methodological pitfalls,” in International AAAI Conference on Web and Social Media (Ann Arbor, MI).

Twitter. (2020). Platform manipulation and spam policy. Retrieved January 11, 2021 from https://help.twitter.com/en/rules-and-policies/platform-manipulation

Twitter Public Policy. (2018). Update on Twitter’s Review of the 2016 U.S. Election. Retrieved April 15, 2018, from https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html

United Nations High Commissioner for Refugees. (2018). Syria emergency. Retrieved January 12, 2021 from https://www.unhcr.org/en-us/syria-emergency.html

United Nations Human Rights Office of the High Commissioner. (2015). Human Rights Council holds interactive dialogue with the Commission of Inquiry on Syria. United Nations Human Rights Office of the High Commissioner. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=15708

United States. (1997). 47 U.S.C. 230 - Protection for private blocking and screening of offensive material - Content Details - USCODE-2011-title47-chap5-subchapII-partI-sec230. Retrieved January 11, 2021 from https://www.govinfo.gov/app/details/USCODE-2011-title47/USCODE-2011-title47-chap5-subchapII-partI-sec230

United States Army Combined Arms Center. (2018). The Conduct of Information Operations. Army Techniques Publication 3-13.1. Retrieved February 10, 2021 from https://fas.org/irp/doddir/army/atp3-13-1.pdf

United States Department of State. (1989). U.S. Department of State, Soviet Influence Activities: A Report on Active Measures and Propaganda, 1987–1988 (Washington, DC: U.S. Department of State, August 1989)

United States District Court for the District of Columbia. 2018. Case 1:18-cr-00032-DLF - USA v. IRA et al. (Feb. 2018).

United States House of Representatives Permanent Select Committee on Intelligence. (2018). Exposing Russia’s Effort to Sow Discord Online: The Internet Research Agency and Advertisements. Retrieved January 10, 2021 from: https://intelligence.house.gov/social-media-content/

United States House of Representatives Permanent Select Committee on Intelligence Testimony. (2017). Testimony of Sean J. Edgett Acting General Counsel, Twitter, Inc. November. Washington D.C. Retrieved from https://intelligence.house.gov/uploadedfiles/prepared_testimony_of_sean_j._edgett_from_twitter.pdf

Van Dijck, J. and Poell, T. (2013) Understanding social media logic. Media and communication, 1(1), pp.2-14.

Varnelis, K. (2012) Networked publics. The MIT Press.

Vegh, S., Ayers, M. D., & McCaughey, M. (2003). Classifying forms of online activism. Cyberactivism: Online activism in theory and practice, 71.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (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 (pp. 1079-1088).

Vieweg, S., Palen, L., Liu, S. B., Hughes, A. L., & Sutton, J. (2008). Collective intelligence in disaster: An examination of the phenomenon in the aftermath of the 2007 Virginia Tech shootings. In Proceedings of the Information Systems for Crisis Response And Management Conference (ISCRAM).

Voida, A., Harmon, E. and Al-Ani, B. (2012). Bridging between organizations and the public: volunteer coordinators' uneasy relationship with social computing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1967-1976).

Walter, J. (2020). Content Moderation Is Not Synonymous With Censorship. Public Knowledge. Retrieved January 11, 2021 from https://www.publicknowledge.org/blog/content-moderation-is-not-synonymous-with-censorship/

Waltzman, R. (2017). The Weaponization of Information: The Need for Cognitive Security, 10.

Wardle, C. and Derakhshan, H., 2017. Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making. Council of Europe report, DGI (2017), 9.

Weedon, J., Nuland, W., & Stamos, A. (2017). Information Operations and Facebook. Version, 1, 27

Whewell, T. (2013, November 13). Inside Syrian City Run by Al- Qaeda. BBC News

White, J. I., Palen, L., & Anderson, K. M. (2014). Digital Mobilization in Disaster Response: The Work & Self-Organization of On-line Pet Advocates in Response to Hurricane Sandy. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 866–876). ACM.

Wilson, T., Zhou, K. & Starbird, K. (2018). Assembling Strategic Narratives: Information Operations as Collaborative Work within an Online Community. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), pp.1-26.

Wilson, T. & Starbird, K. (2020). Cross-platform Disinformation Campaigns: Lessons Learned and Next Steps. The Harvard Kennedy School (HKS) Misinformation Review. https://doi.org/10.37016/mr-2020-002

Wilson, T. & Starbird, K. (In Review). Multi-platform Information Operations: Mobilizing Narratives and Building Resilience through 'Big' & 'Alt' tech. Submitted to Proceedings of the ACM on Human-Computer Interaction Computer-Supported Cooperative Work

Witschge, T., Anderson, C.W., Domingo, D. and Hermida, A. eds. (2016). The SAGE Handbook of Digital Journalism. Sage.

Woolley, S. C., & Howard, P. N. (2017). Computational propaganda, 36. Retrieved from http://comprop.oii.ox.ac.uk/publishing/working-papers/computational-propaganda-worldwide-executive-summary/

Woolley, S.C. & Howard, P.N. (2016). Political Communication, Computational Propaganda, and Autonomous Agents: Introduction. International Journal of Communication 10

Yee, V. (2019). James Le Mesurier, Backer of Syrian White Helmets, Is Found Dead in Turkey. The New York Times. Retrieved January 11, 2021 from https://www.nytimes.com/2019/11/11/world/middleeast/james-le-mesurier-white-helmets-dead.html

Yin, R.K. (2015). Qualitative Research from Start to Finish. Guilford Publications.

YouTube. (n.d.) YouTube Community Guidelines & Policies - How YouTube Works. Retrieved January 11, 2021 from https://www.youtube.com/howyoutubeworks/policies/community-guidelines/

Zaram, K. (2013). Syria's Civil War Plays Out On Social Media. Associated Press. Retrieved January 15, 2021 from https://apnews.com/article/9049ee92b1804bb88cbf6d75d0d61910

Zeng, L., Starbird, K., & Spiro, E. S. (2016). Rumors at the Speed of Light? Modeling the Rate of Rumor Transmission During Crisis. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1969-1978). IEEE.

Zuckerman, E., & Rajendra-Nicolucci, C. 2021. Deplatforming Our Way to the Alt-Tech Ecosystem. Knight First Amendment Institute at Columbia University. Retrieved January 15 2021 from https://knightcolumbia.org/content/deplatforming-our-way-to-the-alt-tech-ecosystem