NORTHWESTERN UNIVERSITY

High Tempo Knowledge Collaboration in Wikipedia's Coverage of Breaking News Events

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Program in Media, Technology, and Society, School of Communication

By

Brian C. Keegan

EVANSTON, ILLINOIS

December 2012

ABSTRACT

High Tempo Knowledge Collaboration in Wikipedia's Coverage of Breaking News Events

Brian C. Keegan

When major news breaks in our hyper-connected society, we increasingly turn to an encyclopedia for the latest information. Wikipedia's coverage of breaking news events attracts unique levels of attention; the articles with the most page views, edits, and contributors in any given month since 2003 are related to current events. Extant scholarship has made little effort to understand how online communities like Wikipedia are able to engage in high-tempo knowledge collaboration. Wikipedians editing these topics collaborate under conditions unlike those found on the vast majority of other articles: volatile information, highly-coupled work from dozens of simultaneous editors, and synthesizing new knowledge. This project analyzes how a large online community with diverse capacities for knowledge work is able to temporarily self-organize and rely on improvised responses, regenerated organizational forms, and knowledge encoded into artifacts to support high-tempo knowledge work before disassembling. These exchanges of knowledge and skill are occasions to diverse members of the community to come together and also supports the exchange of knowledge and skills to improve the collaborative capacity of the community. This project examines the historical and institutional contexts for encyclopedias incorporating new knowledge through history, characterizes the differences in the collaboration structures of articles about breaking and non-breaking articles, analyzes structural patterns in the sequences of edits made by editors, and develops a multi-level statistical model to understand the influence of users and artifacts on the self-organization fo these collaborations. Socio-technical systems like Wikipedia not only support novel forms of high tempo, distributed, and temporary organization, but they provide a roadmap for extending these successful collaboration processes to vital social needs like disaster response, journalism, and political reform.

Acknowledgements

A dissertation is the product of a hundred of small wins from dozens of mentors, colleagues, and friends.

My co-advisors Darren Gergle and Noshir Contractor gave me an exhilarating amount of freedom to develop my own research agenda over the past five years. Darren's affable candor counterbalanced my stream of half-baked ideas while Nosh's contagious enthusiasm for all things networks forever colored the way I see the world. Every student, scholar, and teacher should be so lucky to have had the resources they generously deployed to support me. They initiated me into the communities of computer-supported cooperative work and social network analysis respectively where so much magic now passes as research. My other committee members Daniel Gruber and Brian Uzzi were tremendously generous with their time and thoughtfulness and inestimably improved this project.

It has also been the greatest privilege to work with Muhammad Ahmad, Jaideep Srivastava, and Dmitri Williams as they fundamentally reorient the computational and social sciences toward towards each other. Patti Bao, Alan Clark, Kathleen Geraghty, Brent Hecht, and Lauren Scissors persevered through all the years in the broom closet together; Yun Huang, Brooke Foucault Welles, Ryan Whalen, and Mengxiao Zhu helped me understand the bestiary of SNA packages; Will Barley, Harsh Taneja, Jen and Chris Whitfield, and many other online and offline friends contributed immensely to my psychic well-being.

My deepest thanks go to my parents Debbie and Mike and my brother Philip who supported my every decision along the unlikely path that led me here. No article, book, or seminar prepared me one tenth as much for this undertaking as any one of our nightly dinner debates about current events.

Finally, to my most unfaltering critic, most motivational speaker, best friend and statistically significantly better half, my wife Ricarose: I tried emulating your passion, creativity, drive, and vision, but all I got was this dissertation, which I dedicate to you.

Table of Contents

Acknowledgements 3
Table of Contents 4
List of Figures 6
List of Tables 9
Chapter 1: Introduction 10
Problem Statement 11
Background 13
Approach 17
Chapter 2: Theoretical Framework 20
Temporary and high tempo organizations 20
Knowledge collaboration 26
Online community 29
Integrated framework 35
Chapter 3: The Convergence of Encyclopedism and Journalism 39
Boundary work and boundary objects 41
New knowledge and the encyclopedia 44
Emergence of news values 60
Promise of new mediums 68
Discussion 73
Chapter 4: An Encyclopedia with Current Events 77
Creation of Wikipedia 79
Wikipedia and the September 11 attacks 80
Current events on the homepage 91
Policies for news content 95
Emergence and irrelevance of WikiNews 102
Discussion 111
Chapter 5: Coauthorship Network Analysis 117
Background and approach 117
Data and methods 121
Results 129
Discussion 155
Chapter 6: Trajectories 157
Background 158
Approach 160
Research questions and hypotheses 166
Study 1: Structure and dynamics 167
Metrics 170

Results 172

Study 2: Repeat Collaborations 178

Results 183

Discussion 188

Chapter 7: Editor trajectories and social roles 190

Data and methods 191
Background 193
Bipartite coauthorship network 194
Article trajectory networks 216
Shifting attention 230
Editor trajectories 237
Discussion 250

Chapter 8: Statistical Network Analysis 255

Background 256
Approach 264
Data, Variables, and Methods 270
Results 272
Discussion 280

Chapter 9: Conclusions 284
References 290
Appendix A – Data acquisition and analysis code 311

List of Figures

Figure 3-1: Chambers' preface to the Cyclopedia. 51

Figure 3-2: Advertisement in the December 1921 issue of National Geographic for Encyclopedia Americana's second edition. 52

Figure 3-3: Advertisement in the June 1916 issue of Southwestern Law Review for Nelson's Loose-Leaf Encyclopedia. 56

Figure 3-4: Advertisement in September 1947 LIFE Magazine for Encyclopedia Britannica's "10 Eventful Years". 57

Figure 4-1: Screenshot of the Wikipedia article for the 9/11 attacks as of October 9, 2001. 84

Figure 4-2: Edit activity on 606 articles in "September 11 attacks" category. 87

Figure 4-3: Screenshot of the Wikipedia category "September 11 attacks" as of May 2012. 91

Figure 4-4: The "In the News" section on the Wikipedia homepage before (top) and after 2004 redesign 93

Figure 4-5: WikiNews editing warning for published articles 107

Figure 4-6: Comparison of growth between English Wikipedia and English Wikinews 108

Figure 5-1: Distribution of article creation lags for articles with first revisions occurring the same year as the incident itself. Most articles' first revisions occur less than 1 day after the incident which I classify as "breaking news articles." 126

Figure 5-2: Schematic representation of edit history transformation and alignment for two breaking articles and two non-breaking articles. 127

Figure 5-3: Frequency of article types per event year. 128

Figure 5-4: Editor degree distributions, by year for breaking articles. 133

Figure 5-5: Article degree distributions, by year for breaking articles. 134

Figure 5-6: Edge weight distributions, by year for breaking articles. 135

Figure 5-7: Editor degree distributions, by year for non-breaking articles 136

Figure 5-8: Article degree distributions, by year for non-breaking articles. 137

Figure 5-9: Edge weight distributions, by year for non-breaking articles. 138

Figure 5-10: Article degree correlation, by year for breaking articles. 140

Figure 5-11: Article degree correlation, by year for non-breaking articles. 141

Figure 5-12: Editor overlap for breaking articles. 144

Figure 5-13: Editor overlap for non-breaking articles. 145

Figure 5-14: Number of editors in the snapshot networks for each time threshold for breaking articles (red), non-breaking articles (blue), and historical articles (green). 149

Figure 5-15: Average article degree in the snapshot networks for each time threshold. 150

Figure 5-16: Editor clustering in the snapshot networks for each time threshold for breaking articles (red), non-breaking articles (blue), and historical articles (green). 151

Figure 5-17: Article clustering in the snapshot networks for each time threshold. 152

Figure 5-18: Articles in the largest connected component (LCC) for the cumulative network for breaking articles (red), non-breaking articles (blue), and historical articles (green). 153

Figure 5-19: Editors in the LCC for the cumulative network for each article type and threshold time. 154

Figure 6-1: Trajectory construction example. 164

Figure 6-2: Article trajectory visualizations for three breaking and three non-breaking articles. 169

Figure 6-3: Largest component in repeat interaction multigraph. Editors (in grey, sized by degree) are linked with weighted edges if their interactions with another editor appear on two or more distinct articles. Bluer edges have greater fractions of non-breaking article interactions, redder edges have greater fractions of breaking article interactions. 182

Figure 7-1: Two-mode network visualization of the 72 articles (red squares) and 6,003 editors (blue circles). Black links indicate an editor contributed to that article after March 11, 2011. 196

Figure 7-2: Degree and edge weight distributions for the bipartite network in Figure 5 on log-log scales. 199

Figure 7-3: Distributions for average bytes changed per edit from bot users. 204

Figure 7-4: Distributions for average bytes changed per edit by non-bot users. 205

Figure 7-5: Distributions for unique non-bot users by article type, normalized by article age. 206

Figure 7-6: Distributions for unique bot users by article type, normalized by article age. 207

Figure 7-7: Normalized total revisions (strength/degree) plotted against degree, points colored by Gini coefficient. 213

Figure 7-8: Gini coefficients plotted against degree, points colored by total revisions. 214

Figure 7-9: Gini coefficients plotted against total revisions (strength), points colored by degree. 215

Figure 7-10: Article trajectory for “2011 Tōhoku earthquake and tsunami.” Nodes are editors, sized by in-degree (number of unique editors modifying their revisions) and colored by minimum ordinal revision history position (bluer nodes joined collaboration earlier, redder nodes later). Isolate nodes are consequence of thresholding tie strengths above 2 contributions to reveal core network. 218

Figure 7-11: Article trajectory for “Fukushima Daiichi nuclear disaster”. 221

Figure 7-12: Article trajectory for “Sendai”. 222

Figure 7-13: Article trajectory for “Fukushima Daiichi Nuclear Power Plant”. 223

Figure 7-14: Subgraph of the article trajectory for “2011 Tōhoku earthquake and tsunami” for nodes with in-degree greater than 2 with nodes labeled by editor name. 224

Figure 7-15: Multigraph for the corpus of revisions to all 74 articles. Nodes are colored by minimum position in any article from blue (joining collaborations early) to red (joining collaborations later). Edges are colored by the number of distinct articles editors. 227

Figure 7-16: Page views, number of unique editors, and number of revisions per day for the “2011 Tōhoku earthquake and tsunami” English article’s revision history. 228

Figure 7-17: Revisions (solid lines) and unique editors (dashed lines) per day for the 62 “old” articles predating the earthquake and the 22 “new” articles created afterwards. 229

Figure 7-18: The variance in time between edits to any Tōhoku article increases dramatically in parallel with the dilution of editor attention and activity among actively-edited articles. 232

Figure 7-19: A screenshot of an in-line casualty template. Only the text “12,087 deaths,[1][2]” would be transcluded (appear) in the body of the article when this template is called. 233

Figure 7-20: User Flodded's edit trajectory. Numbers in red boxes refer to distinct sub-graphs or phases of Flodded's editing career. 243

Figure 7-21: User L.Tak's edit trajectory. Numbers in red boxes refer to distinct sub-graphs or phases of LTek's editing career. 246

Figure 7-22: User Sandpiper's edit trajectory. Numbers in red boxes refer to distinct sub-graphs or phases of Sandpiper's editing career. 249

Figure 8-1: Visualization of p*/ERGM attribute interaction parameters that capture varying attributes of both the editors and the articles in Wikipedia. Dark-red circles are expert editors, medium-red circles are apprentice editors, and light-red circles are non-e 263

List of Tables

Table 5-1: Examples of breaking news articles from each year. 125

Table 5-2: Summary bipartite network statistics for breaking and non-breaking articles. For cells with two values, editor metrics are top and article metrics are bottom. “LCC” reflects the fraction of nodes in the largest connected component. 132

Table 6-1: Standardized beta regression coefficients (t statistics) for all articles.....176

Table 6-2: Regression coefficients (t statistics) for breaking articles. 177

Table 6-3: Regression coefficients (t statistics) for editors in breaking article (n = 173,292) and non-breaking article (n = 81,682) article trajectories. 185

Table 7-1: Editors with most connections. 200

Table 7-2: Editors with most revisions. 208

Table 7-3: Editors with greatest Gini coefficients. 211

Table 7-4: Namespace summary.....238

Table 8-1: “Main effect” p*/ERGM estimates (t-statistics). Estimates are net of parameters in Table 2. Cells are shaded green for positive & significant estimates and red for negative & significant estimates. 275

Table 8-2: Editor-article “interaction” p*/ERGM estimates (t-statistics) capturing the tendency for the row title to edit/be edited by the column title. Estimates are net of parameters in Table 1. Labels in parentheses next to attribute names are references for 279

Chapter 1: Introduction

When major news breaks in our hyper-connected society, dozens of channels stream this information to us. Traditional news media sources like CNN and Associated Press broadcast accounts from journalists on the scene while online social networking sites like Facebook and Twitter diffuse the information to friends, feeds, and followers. But what we collectively do next as we try to make sense of the people, location, and event would surprise both journalists and information scholars as late as 2005; an encyclopedia named Wikipedia became a crucial source of information about current events. This dissertation examines how this very unique kind of encyclopedia is able to be a central and authoritative source of information in the hours and days following breaking news events.

The proliferation of information and communication technologies has transformed the ways in which knowledge is created, shared, and used. Knowledge is stored in both people (from individuals to organizations) and documents (from papers to databases) and the networks that link them are being woven with ever-greater complexity. The increasingly collaborative, temporary, distributed, and mediated nature of this knowledge work is obsoleting organizations built around self-interest, stability, co-location, and interpersonal relations. Understanding the dynamics of knowledge collaboration in highly networked systems like Wikipedia can illuminate patterns at the leading edge of this new kind of knowledge work. However, there is a substantial gap between our understanding of how large-scale collaborations, like those that we find on Wikipedia, produce knowledge on conventional topics making up the vast majority of articles on Wikipedia versus a substantially smaller set of articles about current events that are the focus of significantly more user attention. The mechanisms for collaborations involving dozens of users contributing over a period of months do not extend to articles involving hundreds of editors contributing over a period of hours. This project fills that gap by analyzing collaborations about breaking news articles to understand how large-scale online collaborations engage in high-tempo knowledge work by

understanding the processes that enable effective collaboration and self-organization in response to shocks and surprises.

The coalescence of organizations following unexpected events vividly illustrates the origins and transformation of social structure and mediation of these processes through computer databases logging fine-grained behavior provides an opportunity to develop rich, *post hoc* accounts of these dynamics. To that end, this project makes two contributions. First, it demonstrates the similarities within and differences between high tempo knowledge work on breaking news collaborations and steady-state knowledge work on conventional articles. Second, it uses methods in network analysis to identify the roles played by users within and across these breaking news collaborations and the temporal evolution of this knowledge work.

Problem Statement

Wikipedia is a peer-produced online encyclopedia using the affordances of "wiki" technology to empower users to alter and revise the content of a webpage. Because Wikipedia is increasingly perceived as a credible, neutral, and accurate source of information (Messner & South, 2011), it is a popular reference website influencing web users' information seeking repertoires (Rieh & Hilligoss, 2007). Following unexpected and traumatic news events such as a major natural disaster, transportation accident, or mass shooting, familiar reference sources such as Wikipedia become loci for people seeking information to help them make sense of the event. In an internal 2009 memo, the Associated Press notes:

> "…a new pattern of [news] consumption was validated in the confusing minutes that followed the first reports of [Michael Jackson's death]: users shared; they searched and they clicked on Wikipedia. …The new routine of Twitter-to-Google-to-Wikipedia contrasts sharply with the behavior of users [after Princess Diana's death]. …the key to Wikipedia's success is that its pages are designed to catch traffic, provide key information and then send users on their way to deeper engagement on the subjects they're interested in. …for large portions of the Internet audience,

> [Wikipedia was] an authoritative source working to verify an important news development." (Seward, 2009)

While journalistic accounts and anecdotal evidence have suggested wikis are well-suited to supporting the temporary organizations which emerge to document breaking news events (Cohen, 2007), there has been little if any empirical work done to examine the ways in which Wikipedia articles about breaking news events differ from traditional articles' patterns of coauthorship and collaboration. Identifying and unpacking these differences could open the door to understanding the processes of how design technical systems and organize organizations to that can rapidly scale to meet the challenges of responding more quickly and effectively disasters and accidents.

Disasters and accidents are endemic to social life but so are the unique forms of social behavior and organization that emerge following them. Rich literatures on disaster sociology and crisis informatics examine how people self-organize to support information dissemination, response, and sensemaking during a crisis. The "improvisation of order out of chaos", equanimity of victims, emergence of serendipitous and egalitarian social ties, and redemptive moments of solidarity have characterized post-catastrophe communities for centuries but are also intrinsically ephemeral and recede as the most acute phase passes (Quarantelli & Dynes, 1977; Solnit, 2009). Social media spaces exhibit analogous processes as citizens and responders use mobile phones, text messaging, GPS, and microblogging to organize ad hoc responses, process and disseminate information, and provide social support (Shklovski, Palen, & Sutton, 2008).

Yet most Wikipedia users are personally unaffected by the immediate consequences of these events and this may restrict their access to information about these events and even inhibit their motivation to devote their limited volunteering time to topics so remote from their expertise and interests. Furthermore, centuries of precedent have likewise tightly proscribed the role that tertiary sources like encyclopedias

should play in including or adjudicating claims over new knowledge. Wikipedia's identity are grounded in this encyclopedic tradition and its policies repeatedly emphasize that the content of its articles should take a historical perspective and rely upon neutral and reliable secondary sources; prerequisites obviously absent in the coverage immediately following a crisis. In addition to these barriers, developing a collaborative account of a breaking news event on a site where "anyone can edit" would seem to inhibit rather than promote the generation of a reliable account. Editors' diverse motivations and skills, their lack of experience working together, no expectation of working together again, and their volition to contribute as much as they prefer should surely lead to major breakdowns in the process of collaborating together. Furthermore, the volatile information environment, lack of a central authority to assign tasks, make decisions, or enforce rules, and need to sustain attention to developments over long periods of time and across broad topical areas likewise should be a recipe for profound organizational dysfunction.

Yet practice proves otherwise. Since 2003, the top 25 Wikipedia articles with the most contributors every month consists nearly exclusively of articles pertinent to current events (Wikipedia, 2011). For example, the articles which attracted the most contributors in February 2011 included "2011 Egyptian revolution", "2011 Libyan uprising", "2010-2011 Middle East and North Africa protests", "Hosni Mubarak", "Born This Way (song)", "Super Bowl XLV", and "2011 Christchurch earthquake" (Wikipedia, 2011). Similarly, articles receiving the most unique edits (Wood, 2011) and page views (Mituzas, 2011) in any given week or month likewise demonstrate a substantial bias towards articles about current events. Wikipedia's coverage appears to thrive in spite of the serious challenges for organizing and coordinating responses to breaking news events on an open and large-scale collaboration system. Simply put, how is Wikipedia able to cover breaking news events in spite of itself?

Background

Wikipedia is an online encyclopedia that "anyone can edit." Unlike traditional encyclopedias or dictionaries that are authored by credentialed experts, edited by professional staffs, and then printed on

paper and distributed to libraries and bookstores, any user can log on to the website and change any part of almost any article on Wikipedia. There is no central editorial board that assigns users tasks, there is no process for determining the credentials of its contributors, contributors are not even required to register an account to make a change, and there is nothing to prevent malicious users from deleting the content of entire articles. Yet it is because of the radical openness of this project, not in spite of it, that has contributed to its success: users can develop articles reflecting their own interests, non-credentialed experts can contribute information, low barriers to entry promote wider participation, and malicious changes are reverted as quickly as they are made. This openness also means that the encyclopedia is never finished but is a perpetual work-in-progress.

Wikipedia's name is a portmanteau of the word "wiki" and "encyclopedia." The concept of a "wiki" was developed by Ward Cunningham in 1994 as a mechanism to allow users to edit a webpage using a webpage. Traditionally, webpages had to be written in a specialized code called Hyptertext Markup Language (HTML), and uploaded to a web server before users could access them. Users could use a web browser to render this code, but the content was static and there was no way to make changes to this code through the browser itself. Only users with the skill to write in HTML and access to the server to upload files could make changes. Cunningham's "wiki" translated HTML into a more readable plain text markup format, allowed the webpage to accept input from the user, and incorporated the changes submitted by the user into the page. The implications of these changes, while not immediately realized, were nevertheless profound as it challenged the prevailing model of the World Wide Web as a one-way broadcasting method and made it a two-way communication channel. The origins and history of Wikipedia are covered in more detail in Chapter 4.

As an encyclopedia, Wikipedia consists of articles about a wide variety of topics. However, articles are simply the public face of a much larger and more complex system of discussion pages, user pages, policies, noticeboards, templates, categories, projects, and portals that operate behind the scenes to

support the task of writing and maintaining this massive collaboration. An article is viewed as a static page, containing encyclopedic content about a topic, hyperlinks to related topics, and tables, figures, images, and other multimedia about the topic. At the top of each article is an "Edit" link that takes the user to another page where the source text and markup for the current version of the article can be viewed. On most Wikipedia pages, users can change this source code in the browser, make a comment explaining or justifying the changes made, and submit it back to the server where it immediately updates the article. The server keeps a time-stamped log of every change made by every user to every article since it was created and can be viewed in each article's "revision history." This revision history makes it relatively simple to track which users are responsible for which changes, and to immediately undo any controversial or unwanted changes by reverting to a previous version of the article which is saved on the server. Every Wikipedia article also has an associated discussion page wherein users can post comments and coordinate on how to co-author an article without these debates spilling over onto the main article.

The non-profit Wikimedia Foundation makes Wikipedias available in over 200 languages as well as analogous wiki projects for dictionaries, textbooks, newspapers, and other reference sources. However, the Foundation is primarily focused on ensuring the operation of the web servers and underlying MediaWiki software platform and does not endorse or review the contributions made by the thousands of volunteer editors co-authoring content on each of these projects. Furthermore, these volunteers receive no acknowledgement or credit and their contributions are automatically licensed under a permissive Creative Commons "copyleft" license that prevents them claiming ownership over or preventing others from changing to the content they have authored. The English language encyclopedia project (English Wikipedia) is by far the largest project in this ecosystem with more than 4.1 millions articles, 32,000 active editors, as well as 3.5 million changes and 8.3 billion page views per month.[1] Wikipedia's core

[1] http://stats.wikimedia.org/EN/SummaryEN.htm

identity is reflected by its "Five Pillars":[2] (1) Wikipedia is an encyclopedia, (2) Wikipedia is written from a neutral point of view, (3) Wikipedia is free content that anyone can edit, use, modify, and distribute, (4) editors should interact with each other in a respectful and civil matter, and (5) Wikipedia does not have firm rules. These rules form the backbone for a much larger set of policies and guidelines governing the types of content to be included, style and formatting standards, dispute resolution, and many other organizational matters and are developed and edited like the articles themselves.

The result of the fundamental components of motivated volunteer users, wiki interface and server, article and discussion page, and organization identity and policies has created a complex socio-technical system where each of these interacts and depends on the others. Users maintain the encyclopedia by self-organizing sub-communities dedicated to tracking recent changes, anti-vandalism, welcoming newcomers, and creating projects dedicated to editing particular topics such as military history, meteorology, or video games. The incremental nature of Wikipedia revisions and the system for logging and storing these accumulated changes means that the article itself encodes information about how it came to be. Wikipedia has attracted substantial scholarly interest to understand how the assembly of these socio-technical processes enables the sustained and coherent peer production of quality information.

However, this understanding of the relationship between the characteristics of editors, attributes of articles, practices of contributions, norms of the community, and features of the software systems are incomplete because they treat the development of all Wikipedia articles as identical. This project examines how breaking news articles about events like an airplane crashes unfolds and compares the collaborations around these articles with articles about similar incidents that occurred in the past and are therefore already well documented in other sources. These breaking articles—which are all but ignored by previous research in Wikipedia—attract more contributors (Wikipedia, 2011), more unique edits (Wood, 2011), and more page views (Mituzas, 2011) in a given month than other article types. Breaking news

[2] http://en.wikipedia.org/wiki/Wikipedia:Five_pillars

incidents are highly salient, and the behavior of users engaging in information seeking, the sources from which knowledge is being synthesized, and the contribution patterns associated with articles about these incidents operate under very different dynamics than those of other Wikipedia articles. Articles about breaking news events, as this project will demonstrate, have different editorial revision practices involving larger groups of editors collaborating in shorter amounts of time with substantially less information than compared to articles about conventional topics. Following a catastrophic incident, many users are motivated to seek and monitor information about an event. This upsets the ability for a small, dedicated cohort to engage in concentrated structuring and synthesis that prevails in most articles' collaborations (Kittur & Kraut 2008). Our understanding of how Wikipedia works is profoundly incomplete because the vast majority of studies are grounded in analyses of the entire corpus of articles co-authored under conditions of relative stability even though the articles which are among the most prominently featured and intensely edited are co-authored under conditions that bear faint resemblance to this assumption.

Approach

This project will examine high tempo online knowledge collaboration in the context of co-authorship relations and structures for Wikipedia articles about breaking news events. Wikipedia contains thousands of articles about natural disasters (i.e., earthquakes, hurricanes), technological accidents (i.e., industrial disasters, airline crashes), and violent conflict (i.e., wars, terrorist attacks, revolutions). Some of these events occurred before Wikipedia existed making their coverage historical in nature by drawing on authoritative accounts of relatively settled facts from official reports, books, and academic articles. Other events occurred while Wikipedia existed, but the articles were created so long after the event occurred they also are also able to draw on authoritative accounts of settled facts like the historical articles. The final class is breaking news articles that are written contemporaneously with the events they document. These articles, unlike either the historical or non-breaking articles, do not have the benefit of stable

accounts of the event because the accounts they draw upon are themselves changing as more information comes to light. Furthermore, because these events are in the news, these breaking news articles likely also attract substantially more attention than articles about historical disasters. This attention combined with the openness of the wiki model leads to substantially more editors making contributions, complicating the assumptions about collaboration involving stable knowledge and small co-authorship.

These genres of breaking news articles are not constitutive of all articles about current events. Articles about elections, sports, awards, or other scheduled events can be coordinated and structured ahead of time. Articles about politicians, artists, athletes, and other celebrities likewise can become topics of current interest, but the biographical content of their articles typically requires only minor revision to be updated rather than being written anew. The articles this project analyses will be distinguished by the fact that the events they reflect are sudden and surprising, which makes it impossible to frame or coordinate ahead of time. While this strong limitation reduces the set of current and breaking news event articles that qualify for analysis, it ensures the remaining set of articles were developed under analogous constraints.

This project employs a variety of theories and mixed methodological approaches to fully explore and understand Wikipedia's breaking news events. Chapter 2 will integrate siloed theories about high tempo organizing, knowledge collaboration, and online community into a framework for understanding Wikipedia's high-tempo online knowledge collaboration. Chapters 3 and 4 it will review the history of encyclopedias and Wikipedia to situate how encyclopedias incorporated new knowledge as technology changed and how Wikipedia emulated or diverged from these practices. Chapter 5 will compare the network structure of editors and articles involved in collaborations for breaking news articles to historical and non-breaking articles. Chapters 6 and 7 will propose a method for understanding the "trajectory" of editors' and articles' activities and analyze if breaking news articles and their co-authors have distinct patterns. Chapter 8 will develop a multi-level, multi-theoretical statistical model of the processes that structure these co-authorship networks. Finally, Chapter 9 will review these findings, discuss their

implications for designing socio-technical systems for high-tempo online knowledge collaboration, review the limitations of these methods, and propose a research agenda for on-going work in the domain.

This project's findings have substantial implications for understanding the role of online knowledge collaboration. Ensuring the stability of the community of contributors and motivating sustained contributions over time is paramount to the success of many online communities such as Wikipedia. However, participation in online communities does not always occur under conditions of stasis. The case of breaking news articles on Wikipedia highlights how online communities need to be flexible enough to support high-tempo collaborations involving many participants attempting to coordinate and process information about unexpected and complex events. This mirrors the post-crisis use of social media to facilitate coordinative processes, but also describes contexts with less immediate consequences, more poorly defined organizational factors, and less rich interpersonal social interactions.

Chapter 2: Theoretical Framework

Sociotechnical systems are complex and dynamic assemblies of human agents and their social organizations, technological artifacts and their infrastructures, and cultural values and their rules. These agents, artifacts, and values are embedded within and depend on each other and interactions among them cause them to change over time (Geels, 2004). However, the ways in which we have theorized about socio-technical systems and their roles as organizations as well as forms of organizing has omitted a convergence of critical themes related to high-tempo organizing, knowledge collaboration, and online community. While the permutations of pairs of these themes have been explored in a variety of literatures such as online knowledge collaboration or crisis informatics, organizations of every type are increasingly tasked with responding to contexts that integrate all three simultaneously. However, the convergence of all three has been under-theorized and consequently, under-analyzed. Each of these three themes is defined below, explore how prior literature has examined the intersection of these themes, and highlight instances of crucial topics omitted by failing to consider all three themes together. This review motivates the development of a theoretical framework for understanding how socio-technical systems to regenerate role structures and improvise domain expertise to enact the organizational prerequisites for *high-tempo online knowledge collaboration*.

Temporary and high tempo organizations

The shift away from inflexible bureaucratic hierarchies has re-oriented many forms of organizing towards project-based and temporary teams. But teams typically studied by organizational scholars rarely face the unpredictability, urgency, and re-configurability demanded by groups in high-tempo situations such as disaster response (Majchrzak, Jarvenpaa, & Hollingshead, 2007), emergency medicine (Faraj & Xiao, 2006), aircraft carrier flight decks (Weick & Roberts, 1993), or spot journalism (Berkowitz, 1992). These teams are often distinguished by their finite life span, project-oriented shared goals, geographic distribution, mediation through information and communication technologies, self-organization, tightly

coordinated activity, and particularly, their ephemerality. The terms “temporary” and “high tempo” will both be employed to characterize organizations like breaking news Wikipedia article coauthorship. Each term connotes distinct tasks, activities, and domains; temporary organizations need not be high tempo – a friendly poker game is low risk but is only constituted for a few hours (Zurcher, 1970) – nor are high tempo organizations necessarily temporary – flight deck operations on an aircraft carrier involve high risk activities among a team sustained over a period of months (Weick & Roberts, 1993). However, as is discussed below, the concepts are imbricated and difficult to disentangle in a variety of other domains. Theories about how organizations coordinate their work in high-tempo contexts have emphasized swift trust, improvisation of emergent response groups, negotiated role structures, and regeneration of prior organizational forms. Each of these approaches to understanding high-tempo and temporary organizations rests on varying assumptions about the stability of the team, clarity of roles, and constancy of resources that are problematic when extended to the context of Wikipedia’s articles about breaking news events.

Temporary organizations involve a set of “diversely skilled people working together on a complex task over a limited period of time” (Goodman & Goodman 1976). The occasion for organizing temporary systems is the unique, one-off, and exceptional tasks demanded of these systems which enduring organizations are incapable of addressing (Goodman and Goodman 1976, Meyerson et al. 1996). These temporary groups often work on tasks with high degrees of complexity yet operate without the formal structures, routines, and norms that facilitate coordination and control in traditional, enduring organizations (Thompson 1967). Temporary teams are distinguished from other types of teams based upon the shared perception that a definite deadline or finite time limit exists (B. Bell & Kozlowski, 2002; Saunders & Ahuja, 2006). The conception of time as cyclical (repeating) or linear (start and finish) can differentially motivate participants by emphasizing the scarcity of time. The majority of participants in high tempo collaborations may interpret high tempo knowledge collaboration as a form of organizing which is unique and one-off. However, participants who have contributed to many collaborations have

seen multiple instances of these collaborations unfold and potentially conceive of these collaborations in more cyclical frames following defined stages with repeating roles, demands, and routines to be managed (Lundin & Söderholm, 1995). The tasks in such organizations are complex, interdependent, immanent, non-routine, and consequential and the ways these organizations manage these dependencies can better inform work done in increasingly prevalent project-based teams characterized by few formal organizational structures, temporary duration, and high tempo work (Bechky, 2006).

Temporary organizations are particularly likely to emerge in high risk contexts where no existing structures can handle a suddenly significant but non-routine issue (Meyerson, Weick, & Kramer, 1996). Theories of organizational design emphasize how organizations facing uncertainty shift coordination from formal organizational structures like rules, schedules, and division of labor and move to interpersonal and informal coordination mechanisms (March and Simon 1958, Thompson 1967). Studies of coordination have emphasized that teams have higher levels of performance when tasks with high uncertainty are matched with less formal modes of coordination (Van de Ven, Delbecq, & Koenig Jr, 1976). Temporary organizations need to manage extremely urgent work demanding interrelated action, high levels of interdependence requiring knowledge integration, and information processing in volatile environments which leads them to be governed through networks of relations emphasizing reciprocity, socialization, and reputation rather than lines of authority (S. Brown & Eisenhardt, 1997; Jones, Hesterly, & Borgatti, 1997). These organizations depend on collective knowledge and diverse skills of its constituents; yet individual members may lack familiarity, shared experience, reciprocal disclosure, and other forms of trust to support organizing and collective action (Meyerson et al., 1996). The preconditions to facilitate efficient coordination such as known membership, expertise specialization and credibility, shared goals, and interaction routines are limited or non-existent (Majchrzak et al., 2007). Examples of this type of high-tempo work include journalists managing the uncertainty and complexity of covering breaking news

events by categorizing then adapting routinized practices and expertise within the news organization (Berkowitz, 1992; Tuchman, 1973).

Because temporary organizations often operate under constraints of high uncertainty and interdependence (Jones & Lichtenstein, 2008), the mechanisms by which temporary organizations resolve vulnerabilities, uncertainty, and risk is crucial because the circumstances of these organizations cannot engage in traditional forms of confidence building like enduring organizations. Theories of "swift trust" emphasize the collective perception and relating between participants' shared trustworthiness to manage vulnerability, uncertainty, risk, and expectations emerges among people drawn from small and vulnerable populations, reciprocally recognizing and expecting contributions from each other's quantized roles or specialties, and emphasizing speed and categorical heuristics rather than evidence and accuracy for information processing (Meyerson et al., 1996). However, Wikipedia collaborations are drawn from large and incidental populations of users, many of whome have never worked together before. Because these collaborations are online, it is difficult to assess others' qualifications and there is little accountability for one's actions. Furthermore, Wikipedia collaborations are focused on the production of an immaterial knowledge artifact, not coordinating the logistics of humanitarian aid.

While high reliability perspectives emphasize how actors within an organization internalize predispositions and practices which enable them to support high levels of reliability and resistance to failure (Weick & Roberts, 1993; Weick & Sutcliffe, 2007), much of the literate emphasizes role systems (Bechky 2006, Weick & Roberts 1993). Faraj and Xiao (2006) argue high tempo collaborations rely on protocols and repertoires of proven interactions to manage task and role ambiguity, plug-and-play teams with interchangeable roles, and communities of practice to manage participation and organizing within specialties. However, in organizations like emergency rooms or aircraft carrier flight decks, roles are often clearly credentialed, which allows individuals to adopt a swift and depersonalized trust based on arbitrary category membership heuristics alone (Meyerson et al., 1996). The roles in these systems are

often stable and endure beyond single temporary team memberships (Bechky, 2006; Weick, 1993; Weick & Roberts, 1993).

However, organizations like disaster response teams lack the role clarity or group stability found in flight deck crews or emergency medical teams that enable traditional forms of coordination (Majchrzak et al., 2007). These *emergent response groups* are unique because group members lack clearly credentialed roles and have diverse motivations, mixed perspectives, varied resources to contribute, and substantial volition to come and go as they please. These factors contribute to unstable task definitions and the pursuit of multiple and potentially conflicting goals. In the context of an emergent response group, participants orient to what is known about the situation as well as the history of actions leading to group cognition centered on relationships between people and tasks rather than people and expertise. Emergent response groups lack pre-existing organizational structures such as group membership, tasks, role or expertise that can be specified *ex ante* (Faraj & Xiao, 2006; Weick & Roberts, 1993). In contexts characterized by novelty, unpredictability, and ever-changing combinations of tasks, actors, and resources, participants attend to progression toward or deviance from a goal rather that executing routinized sequences of steps (Faraj & Xiao, 2006). Members of emergent response groups re-tailor expertise and repurpose resources, emphasize trust through action rather than credentialed expertise or roles, and rely on narratives and artifacts to document and track knowledge and activity (Majchrzak et al., 2007). These approaches again assume group members are physically co-located and their work is fundamentally material, neither of which apply to Wikipedia collaborations. However, this approach emphasizing improvisation of resources and social bonds is appealing for modeling Wikipedia's "anyone can edit" ethos.

Other scholars criticize approaches emphasizing temporary organizations' management of ephemerality and instability through "swift trust" alone. Coordination and self-organization in temporary teams also proceeds by participants regenerating, adapting, and improvising roles and routines used in previous

projects and collaborations (Bakker, 2010; Bechky, 2006; Bechky & Okhuysen, 2011). Minimal structures such as general rules, guidelines, norms, and attitudes can facilitate organizational improvisation generating alignments necessary for coordinated action (Orlikowski & Hoffman, 1997). Bechky (2006) advanced a framework in which temporary organizations are organized around enduring, structured role systems whose nuances are negotiated, reproduced, and reinforced across collaborations within industries characterized by temporary organizing. This "negotiated order" perspective emphasizes how social structure and order is reconstructed through negotiation and interaction in temporary organizations. Sources of permanence such as role structures and expectations are imported into temporary organizations and enable self-organization and speedy engagement (Ratcheva & Simpson, 2011). Entrants to a position find expectations through socialization and interaction, encounter and deploy resources with which to negotiate expectations, and enact the position in response to particular situations. Role expectations guide relationships and the execution of tasks, but this role structure simultaneously provided continuity and stability which the temporary projects lacked while conformity to role expectations was exerted through social pressure. The dynamics by which participants thus learn the general role structures in a context such as collaborating on a breaking news article needs to reflect the way in which this structure is provisional, influenced, and maintained by their behavior on each article (Bechky, 2006). This approach aligns with the emergent response group as it also emphasizes processes of improvisation and negotiation

In the context of Wikipedia, this would imply that coauthorship of breaking news articles on Wikipedia are not "one-off" but rather involve editors who have repeatedly work together or even specialize in editing content about breaking news articles. Wikipedia's ability to coordinate high tempo work may reflect the regeneration of organizational forms such as editors inhabiting roles, enacting routines, and employing resources successfully used in previous collaborations. Following a catastrophe such as an earthquake or commercial airliner crash, the facts must be reconstructed, negotiated, and integrated into

the account even as information continues to unfold. Co-authors of an article are self-organized from Wikipedia users' diverse skills, expertise, and motivations. Some editors have previously collaborated on a breaking news event, while other editors may make several contributions and never edit Wikipedia again. The responsibilities for integrating and updating content, reverting vandalism, formatting citations, and mediating disputes are likewise diffused among all editors.

High tempo collaborations on Wikipedia articles about breaking and current news events, unlike the editing patterns typical to most articles on Wikipedia, exhibit clear features of temporary and high tempo organizing. However the lack of clear roles or strong ties to bind participants together in the context of online rather than co-located contexts and knowledge-focused rather than relief-focused efforts undermines crucial but unstated assumptions in many of these approaches. However, breaking news article collaborations do operate under similar constraints as emergent response groups, which suggests they may emulate their coordination practices. In particular, Wikipedia's ability to manage the complexities of breaking news collaborations may derive from its ability to improvise responses and rapidly iterate, retailor their expertise to adapt to new role structures, and regenerate organizational resources and structures from prior breaking news article collaborations.

Knowledge collaboration

Another rich body of literature examines organizations not as temporally bound or driven by coordinating urgent and interrelated demands, as in the previous section, but focused on managing and supporting the creation and exchange of knowledge among members of the organization. This emphasis on knowledge collaboration is clearly fundamental to studying the mechanisms by which volunteers in an combine information and translate their expertise into a knowledge artifact like an encyclopedia. These approaches emphasize the need for information technologies like transactive memory systems and cultural norms for socializing newcomers like communities of practice. While this literature clearly informs how collaboration happens on Wikipedia, it makes assumptions the stability of this knowledge work and

embeddedness of participants within traditional organizations. Both of these assumptions are clearly problematic in the context of Wikipedia's high tempo and collaborations about breaking news events.

The shift to knowledge-based perspectives of organizing emphasizes the know-how of its members matters is as essential for organizational performance as material resources and institutional contexts. Definitions of knowledge vary and emphasize it as a state of knowing and understanding, objects to be stored and manipulated, a process of applying expertise, a condition of access to information, or the potential to influence actions. Knowledge may be explicitly articulated, individually held or socially developed, consciously or subconsciously held, and explicitly codified or tacitly shared (Alavi & Leidner, 2001). A variety of frameworks describe processes of knowledge collaboration and management, but they share four major processes: creating knowledge, storing and retrieving knowledge, transferring knowledge, and applying knowledge (Davenport, Jarvenpaa, & Beers, 1996). Successful knowledge-based organizations develop practices that make knowledge visible and accessible, develop cultural practice for seeking, exchanging, and combining knowledge rather than hoarding it, and create socio-technical infrastructures to link and surface the knowledge embedded within people, documents, and other artifacts (Davenport, 1997). Wikipedia complicates these models because its knowledge transfer and organizational learning occur in contexts of fluid membership and radical orientation to tasks rather than interpersonal relationships or institutional affinity.

Theorists employ transactive memory systems to explain knowledge coordination as individuals learn who knows what and use this knowledge to decide who does what (Wegner, 1987). Transactive memory helps teams manage cognitive load about the task itself and the coordination of the elements of the task by relying on one another to furnish the detail only knowing the location rather than the content of what is being stored. Crucially, the performance of these teams relies on individuals knowing where external expertise and information such as skilled people as well as documents and files are located. This suggests team members must know who has what skill and knowledge to facilitate the emergence of expertise

coordination. The success of collaborations employing transactive memory systems depends on the ability of participants to deploy their expertise on appropriate tasks, share common mental models, willingness to contribute expertise, the resources and capacity to engage in this knowledge work, assessing the credibility of other participants' expertise, and sharing similar values and repertoires for action. However, the assumptions of transactive memory systems as applied to Wikipedia articles about breaking news become problematic as these articles involve large numbers of co-authors who are unlikely to have all worked together before and are unlikely to work together in the future.

"Communities of practices" relax the strong assumptions found in transactive memory system approaches and are also useful for understanding coordination mechanisms in online communities like Wikipedia. CoPs emphasize co-located and tightly knit groups practicing together long enough to develop mutual and shared understandings through sustained interactions and coordination and promoting knowledge transfer through legitimate peripheral participation (LPP) (J. Brown & Duguid, 2001; Bryant, Forte, & Bruckman, 2005). In this framework, new participants see the opportunities to not only fulfill niche roles that others have not filled, but also contribute to a salient and completely new knowledge artifact (Antin & Cheshire, 2010; Bryant et al., 2005). However, the assumptions of CoPs clearly break down in the context of temporary and high tempo collaborations because tenure is no guarantee of relevant knowledge (newcomers may know more than old-timers) and learning occurs through bricolage rather than apprenticeship (newcomers acquire requisite tacit knowledge without peripheral participation).

Instead, knowledge collaboration in emergent response groups may be better understood as "knowledge collectivities" or "electronic networks of practice" rather than communities of practice (Lindkvist, 2005; McLure-Wasko & Faraj, 2005). A "collectivities" approach emphasizes in-situ knowledge and the agency of members to search for relevant knowledge, instead of codifying a communal knowledge base that everyone shares. As a result, individuals' knowledge and competencies are integrated based on their connectivity rather than similarity, learning proceeds via problem solving not socialization, work through

agentic disposition to act rather than enculturation and LPP, knowledge development from goal-directed trial and error rather than congruence with prevailing paradigms. Individuals in these knowledge collectivities rely on informal "network memories" to activate and exchange knowledge in spontaneous manners which rely on the connectedness of individuals rather than tenure within a community (Lindkvist, 2005).

Electronic networks of practice (ENoP) are characterized by large, self-organized, weakly tied, geographically distributed groups of individuals who practice an open activity but interact through information and communication technologies (McLure-Wasko & Faraj, 2005). These networks are characterized by voluntary affiliation, open participation, and mutual engagement. Despite the incentives for free loading in open collaborations and ENoPs, Wasko and Faraj (2005) adapted the framework of Nahapiet and Ghoshal (1998) to suggest the prevalence and type of knowledge contributions in ENoPs can be explained by individuals' patterning of relationships (structural capital), knowledge and expertise (cognitive capital), and commitment to the collectivity (relational capital). Locating coordination practices within structural relationships among participants in online knowledge collaborations suggests individuals possess variable capabilities to engage in these practices given their structural position.

Transactive memory, communities of practice, knowledge collectivities, and ENoPs are approaches that emphasize the participants' relationships and structural positions as key features enabling knowledge collaboration. Themes from this literature emphasize embedding knowledge within the ties that bind people together as well as encoding it in artifacts. While many of these approaches also make assumptions about co-location and task predictability that are problematic in the context of Wikipedia articles about breaking news events, collectivities and ENoP approaches employing the logic of networks.

Online community

A constant critique through the prior two sections emphasized the lack of theorization about how organizational processes map to online contexts. In particular, the mediation of organizational processes through information and communication technologies reduces the overhead and operating costs and opens the possibility of massive collaborations involving hundreds of thousands of people (Benkler, 2006; Shirky, 2008). The creation of online communities of users volunteering their effort and expertise towards developing open collaboration projects like encyclopedias or operating systems has been the subject of substantial amounts of scholarly work, but these too often omit the crucial dimensions laid out above. Much of this literature makes assumptions about the stability of community members' activity rather than examining instances of temporary communities engaged in high-tempo work.

As distinct from the virtual teams literature emphasizing collaboration within traditional institutions, online communities are collectives of distributed individuals who share common interests and attend to both individual and collective welfare (Butler, Sproull, Kiesler, & Kraut, 2002). Large-scale online communities are characterized by their self-organization, sustained social and information exchange, and complex organizational forms (Faraj & Johnson, 2010). Peer production projects such as Wikipedia add a further theoretical dimension as they allow participants to engage in collective action and collaborate around producing artifacts like encyclopedia articles in addition to providing social support and interaction. Online communities like Wikipedia do not bring all their members' attention and effort to bear on each artifact. Instead, the co-authorship of each article involves a group of editors who may be focused on editing only one article, editing related articles, or simply making minor, incremental changes on a large number of articles. Previous studies of coordination in Wikipedia suggested implicit and informal coordination mechanisms support article development when the article is young and intensively worked by a dense cohort of authors. As the article ages and more editors participate, coordination shifts to explicit and more formal mechanisms such as discussion for further improvement in article quality (A Kittur & RE Kraut, 2008; Kittur, Lee, & Kraut, 2009). This suggests that concentrated activity by a small

number of editors early in the life of an article, followed by contributions from a larger group of editors once the structure has been set, is a model arrangement for successful article creation.

Within an online community, individuals offer knowledge to others while adding, recombining, modifying or integrating existing knowledge (Faraj, Jarvenpaa, & Majchrzak, 2011). This form of collaboration is a crucial component of the sustainability of online communities because it signals and recognizes important information about the contributor and improves the community's worth (Faraj & Sproull, 2000; Murray & O'Mahony, 2007; Von Krogh & von Hippel, 2006). These contributions often occur without the structural mechanisms found in traditional organizations like stable membership, repeated interaction, goal-sharing, and mutual interdependence. Instead, coordination processes emphasize relationships between agentic actors, individuals' motivations, structural position, availability of alternative ties to actors, as well as resources and skills, individuals will exchange these assets given appropriate opportunities (Brass, Galaskiewicz, Greve, & Tsai, 2004; Faraj & Xiao, 2006; McLure-Wasko & Faraj, 2005). Online communities' flexible membership, rapid changes in attention, diverse levels of commitment, and varying commitments to norms or prior consensus that complicate traditional conceptions of how to manage these tensions. Understanding the dynamics of knowledge collaboration in online communities because they illuminate more general phenomena of organizational knowledge collaboration (Faraj et al., 2011; Von Hippel & Von Krogh, 2003).

Online communities like Wikipedia are also distinct from the other forms of organizing discussed above because they are able to pool the contributions and work performed by members of these communities in ways that neither high-tempo and temporary organizations nor knowledge-intensive organizations typically are able to support. Benkler identifies the "modularity" and "granularity" as properties of the projects and work done in them that support forms of coordination allowing the integration of diverse and dispersed contributions. Specifically, modularity is a "property of a project that describes the extent to which it can be broken down into smaller components that can be independently produced before they are

assembled into a whole." Highly modular tasks can be developed and contributed independently by autonomous agents making a single or thousands of contributions. Granularity is the size and effort necessary to make contributions to these modules. Contributions that require less time can attract more participants who may have more diverse knowledge, skills, and access to resources and create a virtuous circle. Benkler highlights Wikipedia as an exemplar which tolerates high diverse in style and level of development without losing its utility like a more coherent reference work such as a textbook demands (Benkler 2006).

As is the case with many online communities, the majority of contributions to Wikipedia come from a fraction of the entire user base (Kittur, Chi, Pendleton, Suh, & Mytkowicz, 2007; Panciera, Halfaker, & Terveen, 2009). Despite this disparity in effort, online communities like Wikipedia are able to escape traps such as the tragedy of the commons and social loafing owing to members' uses and gratifications (Lampe, Wash, Velasquez, & Ozkaya, 2010), diverse motivations (Rafaeli & Ariel, 2008), modularity and granularity of knowledge work in peer production (Benkler, 2006), design features of the community itself (Ren, Kraut, & Kiesler, 2007), and diverse social roles (Welser et al., 2011). These approaches generally emphasize the agency of individuals to form and realize their own goals. In the context of Wikipedia, editors who are motivated to fulfill particular roles like copyediting or vandal fighting (Geiger & Ribes, 2010; Welser et al., 2011), are socialized into sub-communities with like-minded collaborators (Bryant et al., 2005), and recognize or revert the contributions of other editors (Kriplean, Beschastnikh, & McDonald, 2008).

Studies of social roles are notable in two contexts: post-disaster organizing and online communities. In post-disaster organizing, unexpected shocks to or the sudden loss of social structure in the aftermath of a disaster and the re-emergence of structure vividly illustrates the origins and transformation of formal and informal roles in enacting that structure. In latter, mediation through technology provides members of online communities with explicit roles tied to material affordances (e.g., administrators have different

privileges than typical users) but this social structure emerges from behavior of participants and is generally negotiated within the confines of the system. However, studies of role negotiation in the sociology of disaster emphasis necessarily revolve around inquiries about formal status and the materiality of the work which is problematic in the context of online and peer-produced knowledge work which is generally lacks formal signifiers, tangible work, or serious consequences. Conversely, studies of role negotiation in online community emphasize communicative competence as well as custodianship of norms (Golder & Donath, 2004) which fails to capture the dynamism of tasks, domains, resources, and activities in organizations expanding, extending, and emerging after disaster and catastrophe (Bosworth & Kreps, 1986; Dynes, 1970; Kreps & Bosworth, 1993). Disaster sociology places a great emphasis on concerns for *process* in role enactment (e.g., "how does work get done?") while online community scholarship places a greater emphasis on concerns about *dependencies* in role enactment (e.g., "who works with whom?"). Studying responses to disasters and catastrophes in online knowledge collaboration thus provides an especially unique opportunity to integrate and synthesize these approaches.

Several previous studies have employed a social role framework to examine knowledge collaboration in Wikipedia (Gaved, Heath, & Eisenstadt, 2006; Gleave, Welser, Lento, & Smith, 2009; Welser et al., 2011; Yates, Wagner, & Majchrzak, 2010). Although Wikipedia has some formally credentialed roles such as administrator and bureaucrat, these are a tiny minority of the user population. The balance of users inhabit emergent roles organized around practices such as vandal fighting (Geiger & Ribes, 2010), copy editing, new page patrolling, content standardization, administration, article evaluation, tool development, and new user welcoming. Welser, Cosley, et al. (2011) identify four distinct social roles: technical editors correcting small style and formatting errors, vandal fighters reverting vandalism and sanctioning norm violators, substantive experts who specialize in improving articles within a particular domain, and social networkers who use the wiki as a platform for interpersonal relations rather than substantive contributions to content or administration. Yates, Wagner, et al. (2010) identified the

placeholders, completers, housekeepers, and shapers who contribute, integrate, and synthesize content on Wikipedia. Kane, Majchrzak, et al. (2009) identified “flitterers” who place ideas then leave, idea champions who ensure the kernel of idea is maintained and evolved, and defenders who use technology to respond to adverse changes in the content. These analyses of social roles in Wikipedia do not examine the roles used for high tempo knowledge collaboration, but they are instructive for identifying analogues and antecedents.

However, scholarship has only begun to examine how Wikipedians engage in high-tempo knowledge collaboration following unexpected and surprising events. Dynamic network analysis of articles about current events reveals distinct types of collaboration characterized by narrow or broad participation (Iba, Nemoto, Peters, & Gloor, 2010). The revision histories of Wikipedia articles about current news events are sites of collective memory, sensemaking, and commemoration (Ferron & Massa, 2011a, 2011b, 2011c; Keegan, 2011). Decisions about which news items to promote to the front page rely on egalitarian gatekeeping in which expert opinions are alternatively embraced and dismissed in different contexts (Keegan & Gergle, 2010). The only study which analyzed the structures, dynamics, and role which support high tempo knowledge collaboration examined only the Tohoku earthquake and tsunami, an unprecedented catastrophe which limits the generalizability of its findings (Keegan, Gergle, & Contractor, 2011).

In light of findings suggesting the growth of new editors, and articles has slowed (Suh, Convertino, Chi, & Pirolli, 2009) and indirect work increasingly occupies Wikipedians’ attention (Kittur et al., 2007), breaking news articles are a unique site to study how late-modern Wikipedians coordinate entirely novel knowledge rather than synthesizing stable knowledge about historical events. However, the articles upon which our theoretical understanding of knowledge collaboration in online communities been established is disconnected from our knowledge about the roles and practices which support knowledge collaboration around the articles that are most prominently produced and consumed. The behavior of users engaging in

information seeking, the sources from which knowledge is being synthesized, and the contribution patterns associated with articles about these incidents operate under very different dynamics than those of other Wikipedia articles. Faraj, Jarvenpaa, and Majchrzak (2011) highlighted the positive and negative consequences of time as time can both lead to the accumulation of effort while also disenfranchising members who are unable to participate as regularly or intensely. They suggest that online communities respond to these tensions by enacting temporary roles, channeling participation, shifting boundaries, and adapting technological affordances.

Integrated framework

The review above identified limitations of prior theoretical approaches to understanding high-tempo online knowledge collaboration. Understanding these collaborations represent not only the convergence of several distinct theoretical traditions, they also represent the convergence of editors and resources from prior and on-going collaborations. This framework proposes an ecological approach for understanding how these contexts influence breaking news articles collaborations. In particular, breaking news article collaborations not only provide an occasion for editors with diverse skills to temporarily come together for high profile and intense work, but the collaborations support the exchange and development of skills among collaborators that they then bring back to their other work, improving Wikipedia's overall capacity for knowledge collaboration. This duality implies communities lacking occasions for prominent high-tempo collaborations will fail to foster the development of diverse skills across its members and communities lacking diverse skills will face greater challenges coming together to collaborate on projects of importance.

High-tempo online knowledge collaborations are only possible because large online collaborations like Wikipedia contain users with diverse capacities to engage in varying types of tasks and coordination. Users may fulfill roles specializing in topical work such as editing articles about hurricanes, task work such as copy-editing, or administrative work such as protecting articles from vandalism. An article

requiring more content as well as protection from major vandalism has more tasks than an article that only requires expansion. These different types of tasks have different levels of complexity, modularity, and granularity, which demand more diverse responses from and will appeal to more diverse editor roles. Integrating and coordinating these types of work, especially in high tempo contexts, is a particular type of work that builds upon them and draws upon the themes outlined above about improvising responses, regenerating structures, and encoding information.

The ecological feedback loop consists of four components. First, high tempo online knowledge collaboration is only possible when there is a large pool of individuals with diverse capacities that can be brought to bear on tasks with diverse levels of complexity, modularity, and granularity. Some editors may specialize in editing articles about breaking news events, but much of the collaboration will be drawn from other editors who have other types of skills and knowledge that are demanded on these articles. Knowledge coordination may not depend on the regeneration and re-enactment of roles across the entire group, but rather regenerating specific roles for specific members of the organization with specialized expertise, credibility, and coordinating capacity to coordinate this work in contexts which demand particular types of coordination. Thus, the contributors to breaking news articles should have diverse topical and task backgrounds.

Second, there has to be an occasion for diverse users to self-assemble and focus their efforts outside of their typical contexts. Because breaking news event attracts substantial interest while being unfinished, Wikipedia editors are more likely to encounter these articles and find work that fits with their skillset. Wikipedia articles about breaking news events succeed in spite of their complex coordination demands precisely because they have such diverse task modularities and granularities that in turn attract contributors with diverse motivations, resources, and capacities: there is so much to be done so quickly and at so many levels that there is something to be done by anyone who wants to do it. Breaking news articles should have significant differences in the size of their membership and patterns of contribution.

The intense attention should lead to differences in the observed structures of the collaborations as breaking news articles occupy more central positions in the structures of the collaborations.

Third, these individuals employ and exchange practices at multiple levels to coordinate their work reflecting a confluence of themes from high-tempo organization, distributed knowledge collaboration, and online community. The convergence of these demands in the context of high-tempo online knowledge collaborations elicits responses and editor contributions that operate at multiple levels of modularity and granularity and draw from the crosscutting practices of improvisation in the face of uncertainty, regeneration of prior organizational structures, and encoding of knowledge into relationships and artifacts. Processes of improvisation should lead to similarities in the concentration of activity on articles, processes of regeneration should lead to editors occupying similarly central role in different collaborations, changes in task modularity and granularity should lead to changes in the structures of collaborations over time, and processes of encoding knowledge into artifacts should lead to artifacts influencing the structure of the network.

Finally, these individuals return to their prior collaboration contexts with new skills and capacities to collaboration. This contributes to the positive feedback loop as editors depart from the breaking article collaboration and they return to their prior collaboration contexts with higher levels of skill. This will manifest in the patterns of editors' participation: breaking news articles are tangential diversions from their prior collaborations and editors return to these prior milieux afterwards.

To model high tempo online knowledge collaborations in Wikipedia, this project employs a network theoretical approach to capture the structure of the relationships that bind editors and articles together and how these structures change over time. The nodes are editors and articles and the links represent revisions made from an editor to an article (Chapters 5 and 8), revisions made by an editor modifying another editor's version of an article (Chapter 6), or the editors' movement from one article to another article

(Chapter 7). These types of ties are not exhaustive of all possible relationships present in Wikipedia like communication ties between editors, co-membership in a project, or hyperlinks between articles, to name a handful. However, the topologies of these networks reveal positions of influence, inequalities in connectivity, and emergent clusters of activity that are both influenced by participants' behaviors and influence their outcomes (Borgatti & Halgin 2011).

Chapter 3: The Convergence of Encyclopedism and Journalism

Turning to an encyclopedia for the latest information on current events is a profound departure from centuries of philosophical rationales and editorial practices in the genre. To understand the emergence of Wikipedia and its ability to cover current events, it is valuable to re-trace the histories of encyclopedism and journalism. These types of knowledge work occupied distinct niches in their organizational environments but nevertheless had overlapping dependencies on each other and a variety of other social worlds. For the vast majority of Western intellectual history, it was "natural" that an encyclopedia would not cover recent historical events, new scientific discoveries, or contemporary biographies owing to the material limitations of printing, philosophical values of its editors, and consumers' preferences for obtaining this information through alternatives such as newspapers. However, demand from consumers for the latest information and the combination of new information technologies have complicated the distinctions between encyclopedism and journalism.

Wikipedia departs from traditional encyclopedic models along many dimensions, but its flexibility and capability to rapidly synthesize journalistic accounts of breaking and current news events represents a chimera of encyclopedism and journalism. This chapter is an "origin story" to trace the historical trajectories of encyclopedic and journalistic knowledge work and theoretically ground the emergence of news work on Wikipedia. These histories are part of the larger history of knowledge reflecting profound changes which have occurred over the past 500 years in the social, political, and technological arrangements for collecting, storing, and disseminating information (Burke, 2000, 2012; Gleick, 2011b). I examine encyclopedists' struggles to reconcile the need to establish authority by writing a stable knowledge artifact against the need for contemporary relevance during the profound scientific, social, and political change from the Renaissance through the twentieth century. Louis Heilprin, a literary critic and co-editor of the *American Encyclopedia,* summarized the problem in the encyclopedic domain while reviewing *Britannica's* fourteenth edition (Kogan, 1958):

> "No encyclopedia, however lavishly equipped with editors and paraphernalia, can ever hope to be quite up to the minute, for men will die, or resign, or be promoted or demoted, and things are not always today quite what they seemed to be yesterday…"

This chapter also outlines a parallel history of journalism following how technological change and professionalization contributed to the emergence of news values and other particular patterns of work. Encyclopedism and journalism shared a central tension over how to balance competing imperatives to be simultaneously authoritative, accessible, and relevant from the Renaissance through the re-emergence of the encyclopedic imaginary in modern computing cultures. Relevance in the encyclopedic context is having an up-to-date accounting of the world. Analogously, for all the timeliness afforded by newspapers, norms of the genre and professional practice limited the contextual detail included. Relevance in the journalistic context reflects the fact that no newspaper can provide the biographies, histories, and background on each of the actors, places, and concepts invoked in a story.

For the centuries in which encyclopedias and newspapers were bound by the limitations of paper these tenets were true, but the advent of information technologies fundamentally transformed what either an encyclopedia or a newspaper can be. The barriers to integrating new information and recent historical changes into a large, collaboratively-authored encyclopedia are no longer material, technical, or organizational. This confluence of genres, practices, and philosophies raises fascinating questions about the processes and consequences of classification in knowledge work as well as the adjudication and consecration of claims to legitimacy, authority, and reliability (Pachucki, Pendergrass, & Lamont, 2007). Encyclopedia articles can now be rapidly updated and revised within minutes of an unfolding event unlike paper encyclopedias and news articles can also be situated within a system of cross-referenced and hyperlinked knowledge networks providing essential contextual information unlike paper newspapers.

Encyclopedias are objects embodying their editors' epistemologies about valuable knowledge, parts of book collections communicating their owners' status and values, historical artifacts encoding the situated knowledge and cultural values of prior eras, and certainly educational tools for reference and research for students, citizens, and scholars. Similarly, newspapers are vehicles for promoting its publishers' moral and political values, outlets for communicating timely information to an audience, and collections for creating audiences that can be sold to marketers. The histories of encyclopedism and journalism suggests this confluence should not be surprising nor are Wikipedians' attempts to define this work new: the boundaries between newspapers and encyclopedias have shifted and expanded repeatedly in response to changing socio-cultural values and new technological innovations. Both types of knowledge work emerged into distinctive organizational forms in the late eighteenth century, assumed great social and cultural importance in the nineteenth century, and consolidated in the face of disruptive innovations in the twentieth century. Encyclopedias and newspapers are thus examples of objects used by and familiar to diverse communities, which leads me to introduce the concepts of "boundary work" and "boundary objects" as theoretical frames for understanding how these interrelated domains became distinct.

Boundary work and boundary objects

This chapter explores the emergence and change in the activities and structures of journalism and encyclopedism through the theoretical lens of boundary work. Rather than tracing the changes in practices and routines within a single organization (such as the use of treatises in the *Encyclopedia Britannica*) or the change of a single bounded organization over time (such as *The New York Times*), this chapter examines changes in the encyclopedic or journalistic work bounded by particular historical contingencies and social structures. Encyclopedias and newspapers both depended upon particular configurations of resources such as technologies, audiences, and identities to support their business models. Competition for these resources pushes them towards adopting similar forms as well as developing interdependencies and cooperative relations as they competed against organizations in other niches. However, as resources

such as labor, production technologies, and audiences in these organizations' environments changed, existing organizations engaged in boundary work to develop new types of competencies, structures, and legitimacy as well as responded to new and more agile organizations.

Encyclopedists and journalists engaged in particular kinds of work to adapt to their changing environments and define their identities and legitimacy in the face of disruption. The literature on boundaries and classification provides both a vocabulary and practices by which categories are invented, reinforced, redefined, and contested. These issues are central to scholarship on social structure and cultural production because categories and boundaries contribute to identity, membership, position, and legitimacy as groups and actors struggle to relate and define a social reality (Lamont & Molnar, 2002). The work of making both encyclopedias and newspapers required the erection of boundaries reflecting the topics of the day and articles of importance to be included.

Gieryn's (1983) concept of "boundary work" provides a useful vocabulary to understand the kinds of strategic action actors (like scientists) employ to differentiate their work from work from others (like amateurs, charlatans, pseudoscientists). Using science as an example, scientists employ three genres of boundary work: "expansion" to acquire intellectual authority over a domain, "expulsion" to deny these resources to non-conforming agents, and "exemption" to protect their intellectual province from outside social or political interference. In the context of encyclopedias, how did editors and publishers engage in analogous processes to expand the encyclopedic genre in light of new developments in the arts, sciences, and technology (expansion)? What actions did they take to differentiate themselves from competing approaches to encyclopedic knowledge (expulsion)? And how did they ensure this work remained independent from outside social and political forces (exemption)? Similar questions can also be framed for the history of journalism. This chapter explores how members of both the encyclopedia and journalist communities of practice dealt with the problem crafting institutions that could be simultaneously authoritative and relevant in the face of profound social, political, and technological change.

Expanding the boundary metaphor, Star and Griesemer (1989) introduce the concept of "boundary objects" to understand how actors residing across epistemic or institutional boundaries project their interpretations onto objects which cross these boundaries. They define boundary objects to be:

> "Objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. They may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable means of translation. The creation and management of boundary objects is key in developing and maintaining coherence across intersecting social worlds."

Because encyclopedias vary considerably in their philosophical predispositions, editorial motivations, and socio-cultural contexts, the work of enforcing and negotiating boundaries provides a framework to compare encyclopedias and the interpretations encoded within or projected onto them. Encyclopedias provide a rich history to examine the changes in symbolic boundaries and epistemological classifications distinguishing which types of knowledge warrant inclusion or exclusion as social contexts have changed (Darnton, 1984; Yeo, 2001). Likewise, the news work performed by the news media play a central role in modern business, politics, culture, and social life. However, these institutions have undergone profound transformations in their organization, technological adoption, and professional practices transforming the boundaries of participation and information dissemination (Klinenberg, 2005; Mitchelstein & Boczkowski, 2009). The confluence of both of these trends in Wikipedia's coverage of breaking news events reveals dynamics governing how some types of knowledge and information warrants inclusion. Generally, the predisposition to cover current events in an encyclopedia is in tension with a prevailing consensus that:

> "The response to the new must be tempered by a king of serenity, lest the encyclopedia turn insensibly into a hard-cover annex of the illustrated weeklies. The instructional role of [encyclopedias] should be to counter—not follow—the tendency already satisfied by periodical journalism" (Barzun, 1962).

Wikipedia's coverage of breaking and current news events offers a compelling entry-point to understand the social construction of boundaries for what types of knowledge warrant inclusion when a genre's historical conventions and cultural values collide with the affordances of new information technologies unconstrained by the former's material limitations. This collision is illustrative of how the expansion of information systems and their adoption by more kind of people for more different things complicates "pure" or "universal" ideas about representation and information (Bowker & Star, 2000). The boundary work of journalists and encyclopedists defining and defending the domain of their knowledge work provides a window into understanding more abstract theoretical concerns about how organizations transform themselves in the face of larger changes in their social, economic, and cultural environment as well as how these institutions create and defend their identity and legitimacy. The following sections highlighting the ways in which upstarts challenged prevailing norms and structures, transformed themselves in the face of change in their environments, and the work of legitimizing their positions through the creation and maintenance of boundaries of encyclopedic and journalistic identity.

New knowledge and the encyclopedia

Encyclopedias are exhaustive reference works meant to cover all knowledge and have precedents tracing back millennia across cultures (A. Brown, 2011). The Greek word "encyclopedia" does not refer to a written document or other object, but rather a metaphorical shape circumscribing a well-rounded education. Karl Popper categorized encyclopedias as a "third world" of objective, codified, and rational knowledge, separate from the first world of external phenomena and the second world of sensation and cognition (Popper, 1994). Plato's nephew Speusippus compiled an early encyclopedia which later

inspired Pliny the Elder's first century *Historia Naturalis*, a 37 chapter work covering natural history, architecture, medicine, and geography as an object for supporting cultural education. Pliny's work exerted considerable influence on encyclopedias in the European world between the first and seventeenth centuries which were consumed with classical ideals and knowledge such as the *trivium* and *quadrivium* as a complete system of learning rather than a book-object (Collison, 1964; Sullivan, 1990). Encyclopedias served as collections of knowledge that surpassed the capacity of any individual mind, a legacy stemming from medieval views of the world as a mirror of the divine mind. The collation and synthesis of knowledge from a larger whole was motivated by the desire to identify the most valuable knowledge about both the spiritual and natural universes so to prevent it from being lost as happened with both Adam and Eve's fall from grace as well as the post-Roman era (Yeo, 2001).

Encyclopedists and philosophers through the European Middle Ages and Renaissance organized and mapped the relationships and trajectories among faculties and subjects according to overriding philosophical principles. Descartes dreamed of a total encyclopedia containing systematic relations of all the sciences and complete knowledge, Bacon had plans to reform all of knowledge, and Leibniz imagined an encyclopedia that would systematically condense all knowledge into a universal symbolic language (Yeo, 2001). The rise of encyclopedias during this era can also be attributed to seventeenth and eighteenth century anxieties about how to store, retrieve, and organize the multitude of manuscripts, books, and other publications following the adoption of printing presses. The encyclopedia offered a mechanism by which the "explosion" of knowledge and information could be organized and digested making it the high point of Enlightenment intellectual traditions (Burke, 2000, 2012). However, early encyclopedists struggled to organize or structure on this type of knowledge. Encyclopedists like Béroalde de Verville sought to delimit their work to recording only that knowledge which was "necessary" but strained to establish criteria establishing these boundaries, to include knowledge that resisted the classifications they imposed, and to define schemas which were complicated by new Renaissance knowledge (Kenny, 1991). Classical

conceptions of the totality of knowledge as ordered, unified, hierarchical, and deductive clashed with knowledge which was complex, subjective, and expanding in the Renaissance era of exploration, cultural upheaval, and intellectual expansion (Darnton, 1984; Sullivan, 1990). These episodes suggest the structure of encyclopedias bear the traces of social and political contexts which influence the inclusion and framing of particular types of knowledge (Zimmer, 2009).

By the late eighteenth century, the size and scope of encyclopedias increasingly became symbols of national prestige. Authors and editors of encyclopedias like Ephraim Chambers's 1728 *Cyclopedia* and the *Encyclopédie* of 1751 by Denis Diderot and Jean d'Alembert shifted from a focus on developing a rational and systematic classification of the totality of knowledge to recognizing the impossibility of such an effort and instead promulgating a collection of knowledge which would be easily managed and accessed (Darnton, 1984; Yeo, 1991, 2000). The *Cyclopedia* included a preface enumerating the forty-seven arts and sciences (see Figure 1), each followed by subtopics indicating the order in which they should be read as well as a large network of explicit cross-references. The *Encyclopédie* attempted to compensate for the seeming disorder of alphabetical entries by employing a tree to map the relationships among "branches" of knowledge (Darnton, 1984). Chambers and Diderot legitimized their new approaches to encyclopedism with activist rhetoric: Chambers's work emphasized the accumulation of knowledge as a symbol of royal power and therefore national prestige while Diderot and d'Alembert framed their work as a challenge to the traditional order and affirmation of more cosmopolitan ideals of societies united by the power of science and reason against superstition (Hundert, 1986).

The impetus for these efforts to simplify knowledge were partially influenced by the development of dictionaries of the arts and sciences like Antoine Fretiere's *Dictionnaire universel* (1690), Thomas Corneille's *Dictionnaire des arts et des sciences* (1694), John Harris's *Lexicon Technicum* (1704), and William Owen's *New and Complete Dictionary of Arts and Science* which contained short articles arranged in alphabetical order (Loveland, 2006; Yeo, 2000). Encyclopedists' increasing pre-occupation

with accessibility and democratizing learning led new encyclopedias like the *Britannica* (1771 first edition) to eschew philosophical pre-occupations with reorganizing or mapping the intellectual world. Encyclopedias of this era replaced complex ontologies for tracing the connections throughout classical knowledge with the pragmatic alphabetized ordering combined with extensive cross-referencing. This alphabetization was also beneficial insofar as it provided the flexibility to simply insert new knowledge and recent discoveries without having to revise lengthy treatises or classification schemes. Alphabetization also had the effect of transforming the encyclopedia into a more user-friendly, quick-reference guide with a familiar and predictable arrangement of information (Yeo, 2001).

However, the growth of content in these new encyclopedias required a more systematic and less labor-intensive approach to organizing their content and publishers adopted the alphabetic orderings (Kafker & Loveland, 2009). Because these alphabetic encyclopedias evoked dictionaries, they were criticized for breaking up the unity of knowledge and traditional "courses of reading" by inviting readers to choose their own points of entry and subsequent trajectory (Sullivan, 1990). The English poet and critic Samuel Taylor Coleridge condemned alphabetization on philosophical grounds and attempted to emulate Francis Bacon's scheme for classifying all knowledge with his "classified encyclopedia" *Encyclopaedia Metrpolitana*. Although his encyclopedia failed, his "Preliminary Treatise on Method" reinvigorated the ancient motivation for the encyclopedia to be a tool for the revelation of harmony among morality, philosophy, arts, and sciences rather than simply an accumulation of knowledge (Sullivan, 1990). The publishers of *Britannica* responded to these critiques in part by publishing extended entries spanning several dozen pages called "treatises" on the arts and sciences to provide a comprehensive review of topics like "Anatomy", "Chemistry", "Medicine", "Law", and "Astronomy" so to save readers from "hunting for science through… a labyrinth." The first edition of Britannica contained forty-four treatises occupying around 58% of the total pages. However, many treatises about topics such as the mechanical arts were identical to content found in earlier encyclopedias and dictionaries like Owen's *New and*

Complete Dictionary and *New University History*. The number of treatises grew with every edition of Britannica (rising to 290 treatises composing 41% of the pages in the ninth edition), reflecting the proliferation and division of knowledge, but their length tended to decrease after 1800 reflecting growing uniformity and standardization being imposed by its editors (Loveland, 2006).

Encyclopedic treatises played an important role in the evolution of the encyclopedic genre through the nineteenth century. Encyclopedias of this era increasingly employed treatises to summarize and document rapid changes in new domains of science, technology, and other practical arts with secondary importance assigned to "classical" domains such as biography, history, geography, and literature (Loveland, 2012; Yeo, 1991, 2000). Treatises within encyclopedias and scientific dictionaries became vehicles for publishing and popularizing accessible and original content with subjects written by the discoverers and inventors themselves. Editors struggled to keep new editions current with the latest findings and theories and readers recoiled against the elitism inherent in esoteric topics like "Algebraic Forms" which appealed more to "men of science" than the audiences of "ordinary readers" or "higher [classes] of reader" imagined by the earlier encyclopedists (Hughes, 1951, 1952, 1953, 1955; Loveland, 2006; Yeo, 1991). Advances in math, physics, mechanical arts, and chemistry prompted editors of *Britannica* to issue recommendations to avoid "innovations in language" and only to describe these arts and sciences in the "prevailing and most intelligible phraseology." However, these scientific advances also marked a crucial shift in the division of labor and intellectual authority of encyclopedia writing from the generalist editor and encyclopedic preface to the specialist authors and well-demarcated topics as *Britannica* editors emphasized "the task of defining [sciences], with a greater degree of precision, properly devolves upon those to whose province it belongs…" (Yeo, 1991).

Furthermore, the material process of publishing encyclopedias meant volumes were serialized or published incrementally over the period of years (Loveland, 2012). Encyclopedias often ran into financial or publishing limits resulting in dramatically uneven coverage: the first edition of Britannica dedicated

697 pages to the letters A and B and the balance of alphabetically-sorted topics were squeezed into the remaining 2,000 pages. Many encyclopedias also bear witness to the exhaustion or frustration of their editors as there were more and longer articles for topics earlier in the alphabet while topics in the later part of the alphabet considerably waned (Yeo, 2001). This serial publishing model furthermore had the profound consequence that concepts earlier in the alphabet preceded the fundamental principles from which they were derived (e.g., “air pump” preceding “pneumatics”) which complicated disambiguation, cross-referencing, and pedagogy by referencing topics in volumes not yet published or which were subsequently never elaborated (Yeo, 1991). This resulted in internal contradictions as principles later in the alphabet were often revised or revolutionized by scientists in the intervening months or years between publication of encyclopedia volumes, effectively obsoleting preceding concepts from the succeeding principle within the very same edition (Kafker, 1981).

Whereas encyclopedias were initially envisioned as fundamentally conservative artifacts in which older material was simply carried over into new editions, the profound intellectual transformations in philosophy, science, industry, and culture throughout the eighteenth and nineteenth century outpaced the ability for publishers to keep prior editions up to date. Louis Moreri first attempted to publish new editions in rapid succession with his *Grand Dictionnaire Historique* of 1674 (Hammer & Zalta, 1997), but approaches like this perpetually undermined themselves by incentivizing buyers to wait for still more recent editions. A more conventional approach was to publish supplements. Supplements to Chamber’s *Cyclopedia* were published in 1753 and supplements to *Britannica* were released in 1824 and 1842 emphasizing the recent dramatic changes in sciences and resulting in wholesale rewriting of the previous treatises on chemistry, magnetism, and botany, among other fields. These supplements emphasized concerns about the boundaries between these new domains and substantial attention was paid to acknowledging the relationships among sciences while stressing the integrity of each subject as a discipline with its own intellectual domain (Yeo, 1991).

Editors, consumers, and critics alike began to lament that nothing "of any stable value could be added to the work" and by the ninth edition of *Britannica* (1889) the amount of material surviving from the previous version was less than in any previous edition (Hughes, 1951; Yeo, 1991). There were ultimately 11 supplementary volumes to this ninth edition of *Britannica* that were collectively called the "tenth edition". In response to rapid changes in tax and regulatory codes as well as accounting practices, "loose leaf services" were invented to allow subscribers to insert, remove, and replace pages in a binder to keep the work current (Wootton & Wolk, 2000). Sets of replacement pages called releases were published regularly and the replaced material could be inserted into a "transfer binder" to be stored for archiving (Petit, 2003). Encyclopedia publishers began to offer loose leaf encyclopedias like *Winston's Cumulative Loose-Leaf Encyclopedia* and *Nelson's Loose-Leaf Encyclopedia* between 1909 and 1920 which could be kept "perpetually new" by publishing batches of content every six months (see Figure 3), however these new technologies neither supplanted nor reformed practices used by major publishers like *Britannica*.

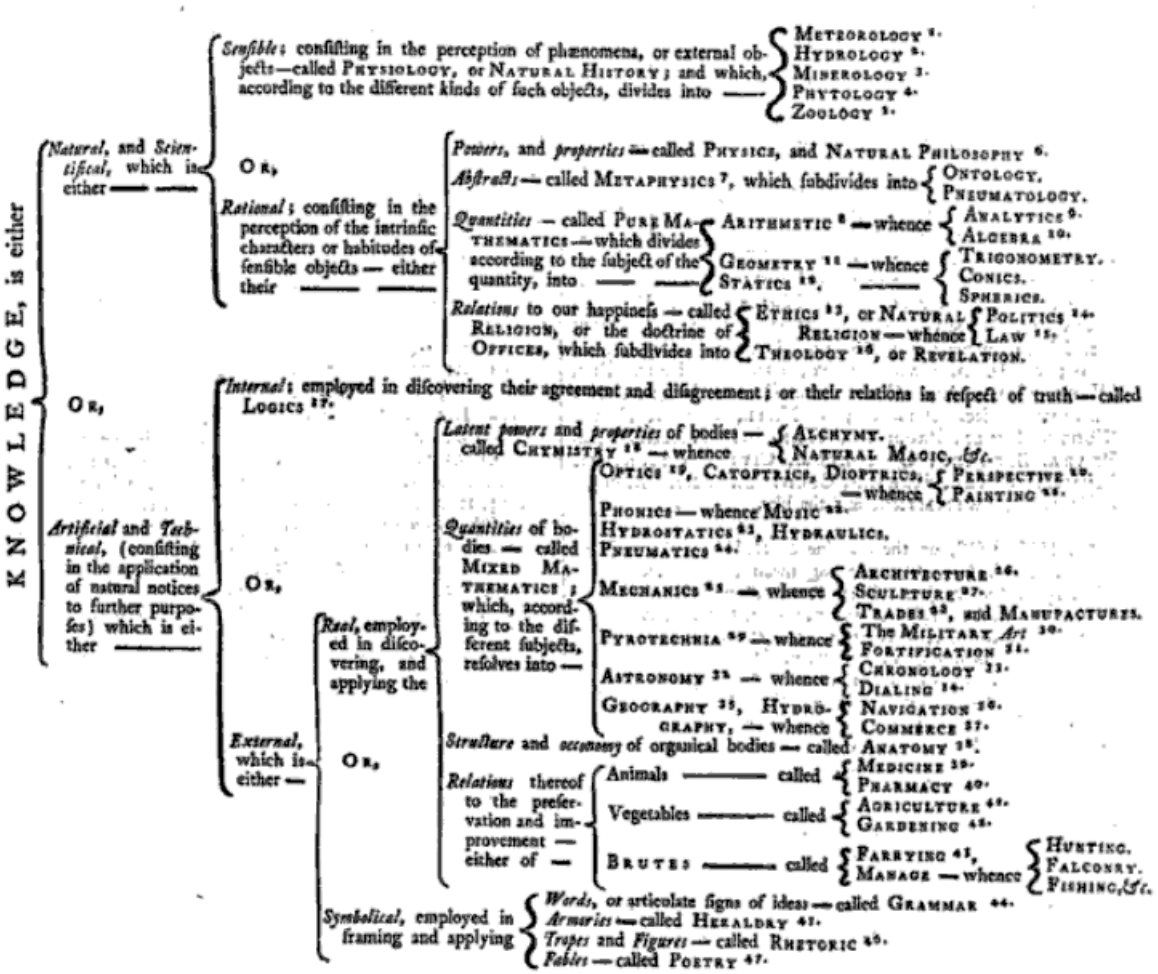


Figure 3-1: Chambers' preface to the Cyclopedia.



Figure 3-2: Advertisement in the December 1921 issue of National Geographic for Encyclopedia Americana's second edition.

Encyclopedias became an expression of the rising bourgeois class's desire to acquire, possess, and accumulate commodities including the entirety of humanity's intellectual estate (Dieckmann, 1961; Foucault, 2002). As tools of reference, commodities, and symbolic objects conveying social distinction, familial security, and national prestige, the growth of encyclopedias conveyed much about how editors made decisions about what is to be included or excluded. In their marketing, encyclopedias criticized one another for neglecting important content or rationalized their own exclusions (see Figure 3-3 and Figure 3-4), invoking a latent impulse to expand. The initial growth in the size of encyclopedias was largely a consequence of competition among titles as well as perceptions of the market for encyclopedias and the psychology of its readers. Increases in size were an obvious feature used to market new editions to subscribers by implying the incompleteness or superficiality of other encyclopedias. Moreover, the success of prior editions granted the editors and publishers the capital and experience to attempt larger and more lucrative editions. Despite technological advances such as machine-made paper, steam-powered printing, stereotyping presses, and changes in banking which should have combined to reduce the cost of capital, materials, and manufacturing, by the mid-nineteenth century the size and length of encyclopedias stabilized as standardization, the cost of compiling information, and financial conservatism of encyclopedia makers imposed constraints on this expansionary drive (Loveland, 2012).

Nevertheless, the number of encyclopedias continued to proliferate and more inexpensive and mass-market versions were developed from the late nineteenth century through the mid-twentieth century. Encyclopedias were hugely popular at the turn of the century in the United States with the sale of *Britannica* to American partners in 1901 and the release in 1902 of a major new edition of *Encyclopedia Americana* (1833 first edition) as well as the introduction of new general purpose encyclopedias like *Funk and Wagnalls* (1894), *Collier's* (1902), *World Book* (1917), and *Compton's* (1922) all marketing themselves as more up-to-date and relevant for "modern" or "industrial" times (see Figure 3-3) (Collison, 1964). The previous imperative for timeliness and relevance was made all the more acute by the major

conflicts and upheavals of the early twentieth century which rapidly dated the prevailing encyclopedic editorial cycle of decades. Like previous encyclopedia editors' responses to scientific and industrial revolutions of the nineteenth century, publishers responded to profound social change by issuing supplements to augment existing editions. In 1922 and 1926, *Britannica* published supplements to the eleventh edition summarizing the history of 1910–1926 that not only traced developments before, during, and after World War I but also included contributions from Harry Houdini, Albert Einstein, Marie Curie, Sigmund Freud, Henry Ford, and Leon Trotsky. In 1947, *Britannica* published a four volume encyclopedia titled "10 Eventful Years" (see Figure 3-4) covering the events preceding, including, and following World War II (Kogan, 1958).

The pace of change and market pressures challenged this prevailing model of issuing supplements to update editions. All 28 volumes of the eleventh edition of *Britannica* (1911) were released simultaneously as an integrated ensemble rather than as serial and potentially contradictory volumes under the prevailing encyclopedia-publishing model. After the publication of the fourteenth edition in 1929 and decline in sales through the Great Depression, *Britannica* President E.H. Powell revisited the assumptions of the prevailing encyclopedia publishing so to create a more consistent, accurate, and most of all, financially stable model. He enacted several significant reforms oriented towards keeping the encyclopedia timelier so to create more consistent demand, two of which merit discussion. The first was a system of "continuous revision" wherein limited runs of an edition were printed and updated every year to prevent flagging demand. A permanent editorial staff would monitor all the content of the encyclopedia to ensure its accuracy in such a way that each article was revisited approximately twice a decade and updates were included for the subsequent year's publication. Secondly, Powell popularized the practice *Americana* had pioneered in 1923 of publishing annual or yearbook edition reviewing current events by having *Britannica* also publish an annual "Book of the Year" volume reflecting developments in science, technology, culture, and politics over the preceding year as well as ample pictures from *Life* magazine

(Kogan, 1958). Both these changes significantly altered the practice of encyclopedic work and re-oriented it towards an ongoing mindfulness about evaluating and incorporating current events that prevailed through the 1970s. The advent and adoption of information and communication technologies would profoundly change the ontologies for classification as well as the means for publishing and updating it.

The Only One That's Always New

NELSON'S LOOSE-LEAF REVOLUTIONIZED ALL ENCYCLOPAEDIA MAKING

Because by means of **Nelson's** Patented Loose-Leaf Binding Device **Nelson's** Perpetual Loose-Leaf Encyclopædia is always **new** and therefore it can always be relied upon to supply the facts of to-day as well as the facts of yesterday. The first volumes of other encyclopædias are often five or more years old by the time the last volumes are completed. NELSON'S PERPETUAL LOOSE-LEAF ENCYCLOPÆDIA IS MADE NEW EVERY SIX MONTHS.



Nelson's Subscribers Are Insured Against Inaccuracies Because Articles Are Written by Experts and Reviewed and Approved by Official Authorities

READ THESE LETTERS

OFFICE OF THE PRESIDENT
CORNELL UNIVERSITY, Ithaca, New York
August 25, 1915.

Dear Sirs:—I beg to acknowledge yours of the 24th inst. enclosing the article on Cornell University which is included among the Revision Pages of Nelson's Perpetual Loose-Leaf Encyclopædia.

I am returning the article to you. Please note the changes which bring the article down to date. Very truly yours,
JACOB GOULD SCHURMAN, President.

CHAMBER OF COMMERCE of Paterson, New Jersey
Office of Secretary
September 20, 1915.

Dear Sirs:—Enclosed please find revised copy of your excellent article on Paterson which is to be used in Nelson's Perpetual Loose-Leaf Encyclopædia. Yours very truly,
JOHN J. FITZGERALD, Secretary.

FEDERAL TRADE COMMISSION, Washington
September 24, 1915.

Dear Sirs:—Your letter of the 7th instant, enclosing copy of material intended to be used as a part of your Loose-Leaf Encyclopædia, has been received. This copy has been revised in compliance with your request and it is herewith returned.
Very respectfully,
EDWARD N. HURLEY, Acting Chairman.

THE NEW YORK PUBLIC LIBRARY
ASTOR, LENOX, AND TILDEN FOUNDATIONS
Office of the Director, 476 Fifth Avenue, New York
September 22, 1915.

Dear Sirs:—I am returning to you herewith the copy you sent me of the article on the New York Public Library which is to be issued among the Revision Pages of Nelson's Perpetual Loose-Leaf Encyclopædia.
Hoping this article as received will meet with your approval,
Very truly yours, E. H. ANDERSON, Director.

Bound in Twelve Handsome Volumes

Treating over 1,000,000 topics with 7,000 illustrations, 500 new maps and plans, models, manikins, and colored plates. The Patented Loose-Leaf Binder makes a volume so practical that by simply turning a nut the pages are loosened, the obsolete pages easily removed, and the new pages substituted. The Publishers issue to subscribers not less than 250 revised pages each six months—in March and October—thereby making Nelson's Encyclopædia always new and abreast of the times.

Nelson's FREE Research Bureau

Nelson's permanent Research Bureau for Scientific Reports and Special Information is the only organization of its kind. It is prepared to answer fully and freely any question, to supply **free** any information that Nelson's subscribers may ask or may want to know. Any time a subscriber wishes a special report on any subject, large or small, old or new, he simply writes to Nelson's Research Bureau with the positive assurance that he will promptly receive the latest obtainable and most dependable information on that subject.

The World is Moving Fast. We Must Move With It

Sciences change overnight. Discoveries are of last week—inventions of yesterday. The man or woman who progresses is keeping abreast of the times.

Encyclopædias that served you well within the last decade are now practically obsolete. They tell you nothing of The Great War of Europe—destined to change the map and history of the world; Sinking of the Lusitania; The new U. S. Federal Trade Commission; Vaccine Therapy; Federal Reserve Board; Naval Advisory Board; The New Seaman's Law, and thousands of other subjects and current events that cannot be found in any other Reference Works.

EXCHANGE We are constantly receiving inquiries asking us to make an allowance for old encyclopaedias to apply as part payment on **Nelson's**. We have therefore prepared a price list, stating the amount allowed, which will be mailed upon request.

When All Other Encyclopaedias Fail
"LOOK IT UP IN NELSON'S"

THOMAS NELSON & SONS
Publishers Since 1798—Bibles, Prayer Books, and Hymnals
Dept. 270 381-385 Fourth Avenue New York City

Dept. 270
Thomas Nelson & Sons
381-385 Fourth Ave. New York City

Please send me portfolio containing sample pages—beautifully illustrated with Color Maps and Plates, also photographs and plans and full information how, by easy monthly payments, I can own Nelson's Perpetual Loose-Leaf Encyclopaedia and receive a FREE Membership to Nelson's Research Bureau for Special Information. This incurs no obligation on my part.

Name.....
Address.....
City..... **State**.....

Figure 3-3: Advertisement in the June 1916 issue of Southwestern Law Review for Nelson's Loose-Leaf Encyclopedia.

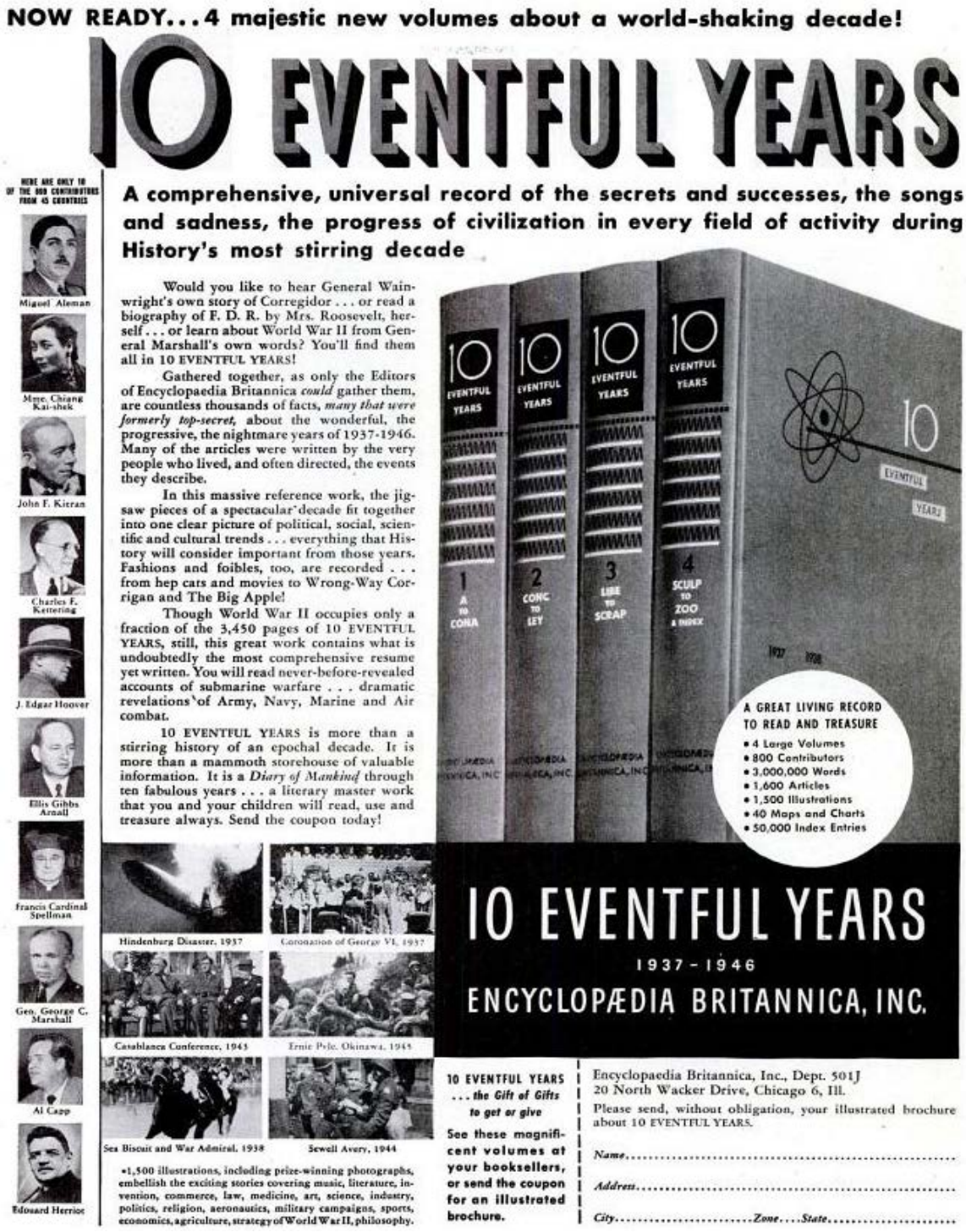


Figure 3-4: Advertisement in September 1947 LIFE Magazine for Encyclopedia Britannica's "10 Eventful Years".

Kogan, writing in 1958, describes the efforts necessary at the time to keep *Britannica* up to date:

> "With the explosion of the atom bomb at Hiroshima…some five hundred other articles, from 'Alchemy' to 'Uranium' [had] to be checked—and, in many cases, altered—by editorial workers before the onrushing deadline for a scheduled new printing. When Elizabeth II became the British queen in 1952, more than a few royal biographies and articles on English history were affected. Her accession added a Roman numeral to the first Elizabeth. Further, because of the British custom of referring to virtually all government offices, official celebrations, and prizes in the name of the sovereign, all through the work were scattered references to the high court of justice known as 'King's Bench Division,' 'His Majesty's Stationer,' and 'King's Scouts.' These had to be changed. …An example of how political changes can be vexatious is the checkered history of Vilnyus, the European city also known as Wilno and Vilna. Before 1938 it was in dispute between Poland and Lithuania; then it was seized on March 17 of that year by Poland, captured by the Soviet Union in September, 1939, ceded to Lithuania a month later, taken again by the Soviet Union on June 15, 1940, lost to Germany on June 22, returned in the summer of 1944 to Soviet control, and ultimately established as the capital of the Lithuanian Soviet Socialist Republic. Doggedly, the *Britannica* editors followed and recorded the shifting fortunes of this city and of others similarly affected. …As soon as the Presidential candidates of all parties are nominated, assignments are instantly made for full biographies of each nominees, and appropriate photographs are assembled. At the same time, shorter biographies are also prepared. When the results of the national election are known, the full biography of the winner and the pictures are sped to the printers, as is the shorter biography of the loser. Similarly, if and when Alaska and Hawaii are admitted to statehood, the *Britannica* will be ready; changes will be necessary in at least 514 articles and in the big full-color plate now showing the forty-eight-starred American flag. …When Franklin D. Roosevelt died suddenly, the editors were able to insert this event into

> the work along with information about his successor. When Pope Pius XII was critically ill, the printers were asked to hold pertinent sections of [the volume containing his entry] as long as possible." (pg. 284-286).

Nationalism of the late nineteenth and early twentieth centuries as well as the end of imperialism after World War II gave rise to the creation of "national" encyclopedias for the expressed purpose heralding ascendant national cultures and histories ignored or distorted by British, German, French, and American publishers who had dominated the genre. Examples of these new nationalist encyclopedias include the Italian *Nuova encyclopedia popolare italiana* (1842-1848), Greek *Megale hellenike enkyklopaideia* (1924-1934), Swedish *Svensk Uppslagsbok* (1947-1955), Swiss *Schweizer Lexikon* (1945-1948), Yugoslavian *Encicklopedija Jugoslavije* (1955-1962), and *Ensiklopedia Indonesia* (1954-1956). The Cold War's polarized ideologies permeated many dimensions of cultural life and encyclopedias were not immune. Calls to create a "world" or "universal" encyclopedia were couched in language of protecting the legacy of Western intellectual tradition against pervasive Soviet revisionism (Stecchini, 1962). Encyclopedism was re-imagined as being:

> "…responsive to the challenges posed by present and imminent changes in the civilization. The encyclopedist must sense the growing edge leading into the future and relate his material to it" (Stover, 1962).

These examples illuminate how encyclopedias transformed into artifacts contingent upon particular political and institutional interests and embedded within cultural values rather than objects encoding some ideal of universal knowledge. These national encyclopedias inherited the traditions and norms of previous encyclopedias such as alphabetization and distributed authorship, but they also reflected significant changes in the environment and the emergence of new niches which other encyclopedias were ill-prepared to fill. These new encyclopedias remained major collaborations bringing together disparate

scholars, leaders, and celebrities toward the purpose of articulating knowledge worth preserving and values worth disseminating and reflecting the values of the eras in which they were written (Foucault, 2002; Sullivan, 1990).

Emergence of news values

The history of encyclopedias above traced how editors attempted to reconcile the aspirations of recording and contextualizing rapidly changing world knowledge with the material limitations of a printed artifact. This section traces the parallel history of journalism as it developed from a distributed network of information exchange into a highly specialized profession with distinct institutional logics. Values such as objectivity, autonomy, and immediacy are explicit examples of boundary work meant to differentiate the domain of journalism from alternative approaches to information gathering. However, social, economic, and technological changes have reconfigured the boundaries of news work repeatedly throughout history. This history, as well as the ways in which Wikipedia complicates prevailing journalistic identities, reveals how classifications of journalistic versus non-journalistic practices are arbitrary distinctions constructed to advance particular moral, social, and political interests. Wikipedia's coverage of current events is emblematic of disruptive alternatives emphasizing new media and citizen participation, but the history of journalism also reveals interesting historical cases of alignment with encyclopedism itself.

News in medieval Europe consisted of letters exchanged among correspondents, intelligence agents, and other merchants exchanging information about military, diplomatic, and ecclesiastical affairs which following trade and financial networks (Rantanen, 2007). The practice of journalism at this time was one in which writers would walk from town to town sharing news, royal courts and literary circles would hire private writers for political intelligence and awareness of gossip, postmasters would collect news from their own district to be mailed to commercial centers, and publishers re-printed whatever news arrived along with travelers (Hoyer, 2003). News-sheets were first printed in Germany in 1609, further developed in Amsterdam throughout the seventeenth century, and had spread across the continent by the dawn of the

eighteenth century. Amsterdam was the center of this new genre and published newspapers in English and French with dispatches and intrigue from the Thirty Years War and the English Civil War as well as critiques of the Catholic Church and policies of the French government (Conboy & Steel, 2008). Following the lapsing of the Licensing Act in 1695, the number of newspapers in London had increased from nine to nineteen by 1709 and the center of the news industry shifted to London as it became the financial capital of Europe (Burke, 2000; Rantanen, 2007).

Two types of press emerged in the late eighteenth century and early nineteenth century. Commercialized news catered to lower-class tastes and espoused egalitarianism but was more open to innovation and concentrated in urban areas. Rural presses published more partisan content for elite audiences interested in complex arguments and debates. This "printer's paper" was culled from other newspapers, letters, and word of mouth reports and intended for consumption by the political and mercantile elite as a type of coffeehouse reviewing other news and sharing opinions with peers. The emergence of mass politics and national political parties in the early eighteenth century led to many of these parties co-opting the press system towards partisan ends embodying the ideal of a courtroom in which essayists would try to persuade the "jury" of readers on a narrow set of perspectives like abolitionism or abstinence. Editors had a prominent public role as both party activists expected to participate in the organization as well as selecting and composing the matter for the paper, managing the business, and running the presses (Nerone & Barnhurst, 2003). These distinct audiences led the newspapers to embrace divergent models of communication to alternatively share information among co-equal producer and consumers or alternatively as an outlet carrying the banner in the struggle against political and economic arrangements which demanded the need to "speak truth to power" (Hartley, 2008). However, between 1830 and 1880 newspapers underwent a profound transformation which led them to become depoliticized, reject ideological and partisan stories in favor of sensationalist and human interest stories, refocus their business towards commercial interests and advertising, and adopt new technologies like the telegraph, steam press,

pulp paper, and typewriter. These changes accelerated shifts in business models toward fewer and larger companies serving a “mass market” (van Tuyll, 2010) as well as a shift in style. Newspapers foregrounded simplicity and brevity, narrating particular versions of events crafted for specifically targeted audiences with distinct political orientations and class values (Conboy & Steel, 2008).

While newspapers differ from other discursive genres because of the “tacit promise to publish consistently” (Hoyer, 2003), the practice of periodical publication intersected with various other genres, including the serial models employed in both literature as well as encyclopedia. Periodicals included not only daily or weekly newspapers but also monthly and quarterly publications which would become magazines like *The Atlantic Monthly* as well as literary and scholarly journals like *The Transactions of the Royal Society of London* (Briggs & Burke, 2009). Publishing shops were focal points for many kinds of cultural and intellectual exchange as they brought scholars, artists, literati, foreign translators, and political refugees together (Eisenstein, 1980). Moreover, overlaps between newssheets, propaganda, handbills, and other printed materials were common. Publishers interacted with literary and scholarly societies publishing their proceedings, news editors reviewing the latest news, authors and scientists publishing their books, and encyclopedists compiling their tomes to create a tight-knit network of knowledge workers (Hoyer, 2003). While scientific articles published in periodical journals primarily emphasized book reviews, the death of scholars, and announcements of new discoveries, the boundaries between the types, authors, and publishers of scientific news articles and scientific treatises published within encyclopedias are clearly slippery. This historical precedent foreshadows encyclopedists’ on-going concern for keeping their content relevant as well as the emergence of news-work within new encyclopedia models like Wikipedia.

This network of news workers involved not only authors, editors, and publishers, but also compositors and press workers to perform the printing, the postal workers to distribute the information, and administrative work ranging from mundane tasks like accounting to historically-contingent work like

interacting with government censors or party apparatchiks. Differentiation of roles was initially limited by the costs of paper, limited speed of the Gutenberg-style press, and size of the audience which all constrained the size of the newspapers. Early printers were responsible for the production of the content of the paper as well as the material of the paper itself. They could not specialize into reporting or editorializing and their news work was often secondary to printing other works like pamphlets, books, or other periodicals. Among writers, compiling the news was a freelance activity and they drifted among other literary genres such as journals and books as well as "correspondents" who would frequent coffeehouses, bars, hospitals, police headquarters, and the morgue to collect the news (Hoyer, 2003). The advent of the telegraph accelerated the dissemination of information over geographic distances but also contributed to the centralization of newspapers in those cities from which telegraph and railway lines emanated (Rantanen, 2007). Steam presses, typewriters, pulp-based paper, improving postal service and transportation infrastructure, ballooning urban populations, improved lighting in homes, and growing literacy among the population also contributed to growth in the size and number of newspapers. The politicization and commercialization of newspapers (such as the rise of the penny press) required publishers to hire employees to go out and gather information. This displaced the freelance newsgathering model and triggered the development of a class of professional workers such as full-time reporters and dedicated correspondents, specialists to manage advertising and other business interests, essayists to author editorials, and technical and mechanical departments for maintaining presses and telegraphs. This occupational specialization was reinforced by unionization among typesetters, the adoption of codes of ethics among reporters, and emergence of advertising as distinct from editorial or other news work (Hoyer, 2003; Nerone & Barnhurst, 2003; Salcetti, 1995). While the newspaper emerged from a printing culture where it was a boundary object across several genres of knowledge work, this process of specialization lead it to become even more of an ideal type as various constituencies and communities of practice used it as a symbol of their own work and a conduit to communicate with increasingly distinct groups.

The emergence of distinct classes of news workers also led to a diversification of types of news sub-genres as publishers attempted to distinguish their papers from competitors. As newspapers evolved from partisan to commercial entities, the social, cultural, and political effects of industrial revolution began to galvanize the progressive movement (Baldasty, 1992), the newspaper reconfigured itself as a civic institution rather than a commercial interest. In the late the nineteenth century and early twentieth century, muckrakers like Lincoln Steffens, Ida Tarbell, and Upton Sinclair targeted corporate wrongdoing, government misbehavior and social injustice not as isolated entities but as three interconnected and systematic problems (Feldstein, 2006). By the early twentieth century, journalism reframed its mission as central to the democratic process and emphasized its practice is about "providing citizens with the information they need to be free and self-governing" (Kovach & Rosenstiel, 2001). Following World War I, the amelioration of the worst abuses of politicians and corporate trusts, the decline of the progressive movement, the economies of scale demanded by capital-intensive production, and the reactions against irresponsible "yellow journalism" lead to consolidation in the newspaper market. Furthermore, introduction and adoption of the radio fragmented the monopoly newspapers had over reliable accounts of the world and led to popular suspicion or outright hostility towards newspaper owners and reporters (Conboy & Steel, 2008). The post-World War II consensus emphasized conformity to an anti-communist political climate and deference to authority (Protess, 1992). Ownership shifted from family-owned chains of papers to publicly-held corporations focused on maximizing shareholder value by developing sophisticated audience segmentation and marketing methods focusing on high-income and highly-motivated readers rather than mass readership. However, these developments undermined the autonomy and ideological consensus of professional reporting values (Nerone & Barnhurst, 2003). Finally, the adoption of new information and communication technologies like cable television, the World Wide Web, and social media radically decreased the barriers to entry to disseminating information which challenges prevailing institutional arrangements over the production, practice, and profession of news work (Deuze, 2005; Mitchelstein & Boczkowski, 2009).

Reporters throughout much of the twentieth century embraced the independence their position afforded them, but they were initially oppressed by the industrialized labor conditions, ready and cheap supply of replacements, norms against voice or attribution, and the inability to unionize. Reporters responded by professionalizing. The establishment of journalism schools and trade organizations in the early twentieth century, erection of a barrier between reporting and advertising missions in the organization, and adoption of codes of ethics established a professional ethos emphasizing objectivity, credibility, expertise, and autonomy (Nerone & Barnhurst, 2003). This shift is a classic example of Gieryn's boundary work whereby journalism expanded itself to claim a privileged civic role, excluded competing approaches (e.g., yellow journalism) and ideologies (e.g., partisan press), and justified their privileged position from outside interference by emphasizing its adoption of professional roles and values. However, this professional identity emphasizing the ability to inform became a professional imperative for reporters as much as the ability to heal, to litigate, to design, or to lecture. However, this process was also complicated by the lack of formal methods, specialty knowledge, concrete barriers to entry, and financial dividends which clearly demarcated other professions like law, medicine, or science (Salcetti, 1995).

The professionalization of journalism contributed to the development of new but increasingly stable institutional norms, structures, and practices. In particular, the emergence of contemporary "journalistic logic" emphasizes the observance of particular news values, media formats, organizational contexts, and editorial positions (Waisbord, 2011). The development of "news values" merits particular attention since it is emblematic of the boundaries journalists erect and complicates their self-espoused ideals of autonomy and objectivity. Galtung and Ruge (1965) advanced the concept of news values as the criteria journalists apply to translate events into news. In addition to the twelve factors they proposed, Schlesinger (1987) and Bell (1991) proposed additional news values which increase the likelihood any event is selected as news:

1. Frequency – Events occurring at the same time or frequency as publication are more likely than irregular events or long-term trends.
2. Threshold – Events surpassing a threshold for intensity are more likely than less severe events.
3. Unambiguity – Event that can be interpreted without multiple meanings are more likely than complicated or ambiguous events.
4. Meaningfulness – Events that occur in culturally similar contexts are more likely than culturally unfamiliar contexts.
5. Consonance – Events which fit with pre-existing frames or expectations are more likely than events which are unprecedented or difficult to frame.
6. Unexpectedness – Events that are rare and out of the ordinary are more likely than events which are commonplace.
7. Continuity – Events that have received previous coverage are more likely than events which are unfamiliar or lack a precedent.
8. Composition – Events that provide balance to other items in the news are more likely than events which exaggerate existing biases.
9. Reference to elite nations – Events occurring in elite nations are more likely than events occurring in less consequential nations.
10. Reference to elite people – Events involving famous people are more likely than events involving less notable people.
11. Reference to persons – Events impacting human interests are more likely than events involving abstract social or natural forces.
12. Conflict – Events involving two clear opposing sides are more likely than events involving multiple or unclear factions.
13. Imitation – Events covered by rival organizations are more likely than events ignored by rival organizations.

14. Predictability – Events that can be anticipated are more likely than events which must be written from scratch.
15. Constraints – Events that can fit into budgets and production cycles are more likely than events requiring substantial time and resources.

While these values provide the deep structure and a cultural map that journalists use to make sense of the world and guide their actions (Harcup & O'Neill, 2001), they are not exhaustive. However, news values reflect a normatively "correct" way to handle new events and adherence to these values can be compulsive or ritualistic despite the limited relevance to core journalistic identities like objectivity and timeliness (Tuchman, 1972). They reflect a consensus within the profession of journalism itself, the institutional setting of other actors supporting this news work (i.e., advertisers, printers), and the broader cultural setting of other institutional actors interacting with the news workers (i.e., policymakers, businesspeople). These values are shaped by socially constructed dispositions about their own role and the preferences of their audience, material and cognitive limitations of specific kinds of organizations and organizing, and broader political economic context incentivizing particular types of content. As problematic as these heuristics may be, news values provide a framework of boundaries to compare the extent to which other new institutions engaging in news work like Wikipedia imitate or replicate these same values as they develop a consensus on how to distinguish events from news.

Other types of journalistic logic emphasize media formats which refer to the links between specific technological environments and reporting styles which prize particular visual, sensationalistic, storytelling, and statistical elements; the organizational environments and contexts in which reporters and other journalists work characterized by high pressure, immediate deadlines, and reactive reporting; and the particular editorial positions which emerge from the configuration of ownership interests and power of individual members of editorial boards and management to influence the frames, issues, and sources which are legitimate (Waisbord, 2011). Expanding on this framework, Deuze (2005) proposed a five-

pronged framework describing the "shared occupational ideology among newsworkers." This framework emphasizes the provision of a public service; commitment to impartiality, neutrality, objectivity, fairness, and credibility; autonomy, freedom, and independence in their work; immediacy, actuality, speed, and relevance in the material they produce; and a shared sense of ethics, validity, and legitimacy.

Promise of new mediums

As developments throughout the nineteenth and early twentieth century made encyclopedia editions obsolete with increasing frequency while demanding ever more space, thinkers began to reconsider the material limitations of traditional publications as well as the genre itself. In 1938, H.G. Wells published a collection of essays on knowledge and education titled *World Brain* in which he imagined the possibilities of microfilm creating a "permanent world encyclopedia" to facilitate international peace and co-operation (Reagle Jr, 2010; Yeo, 2001). Unencumbered by the need to print, bind, and ship thousands of pages, such an encyclopedia could be more easily distributed and updated. This system would revolutionize the production and distribution of information such that it could transcend the static form of the encyclopedia printed in books and instead be a "network" in a state of constant change – "a sort of mental clearinghouse for the mind, a depot where knowledge and ideas are received, sorted, summarized, digested, clarified, and compared" (Gleick, 2011a). More famously, Vannevar Bush anticipated the advent of hypertext with his 1945 *Atlantic Monthly* article "As We May Think" in which he introduced the idea of a "memex" as a means for developing "wholly new forms of encyclopedias…made with a mesh of associative trails running through them" as a mechanism for dealing with organizing and access a rapidly-growing body of information. The user of a memex could create and store arbitrary sequences or sets of microfilm frames (pages) semantically linking disparate yet related topics together while also inserting their own comments, photos, documents, or other content and sharing these associative trails and annotations with other users (Bush, 1945). Bush's argued the value of information was not how it was categorized or classified, but rather the connections it had to other data (Reagle Jr, 2010; Zimmer, 2009).

However, microfilm still had very real material limitations in the storage, retrieval, and transmission of information and neither of these systems was developed. After World War II, digital computers and computer networks offered new alternatives for storing, organizing, and disseminating knowledge. Project Gutenberg was launched in 1971 as a way of digitizing public domain books so they could be easily distributed to other computer users. While initially manually entered, the shift to optical character recognition (OCR) technologies still introduced errors and demanded a proofreading system capable of scaling to the thousands of works each containing hundreds of pages. Charles Frank launched the "Distributed Proofreaders" project for Gutenberg in 2000 allowing users to complete small batches of proofreading tasks that cumulated into projects where hundreds of users checked each page of a book at least twice. Transferring the entirety of *Britannica*'s public domain 11th edition into Gutenberg took on a mythic importance among the Gutenberg community but the complexities of how to preserve illustrations and peculiarities of printing as well as employing hypertext to link between content flew in the face of norms emphasizing that all Gutenberg content must be rendered in the simplest possible plain text formats (Reagle Jr, 2010).

While the printed word was seen to define the presentation and centralize the organization of knowledge (Burke, 2000; Eisenstein, 1980), the hypertext anticipated by Bush, articulated by Ted Nelson's Project Xanadu, implemented in the World Wide Web by Tim Berners-Lee, offered alternative and associative organizations of knowledge which allow information to escape the "straightjacket of hierarchical documentation systems" (Berners-Lee & Fischetti, 2000; Zimmer, 2009). However, the motivations for decentralizing knowledge traces its origins back to before Diderot and d'Alembert's *Encyclopédie* which used *renvois* to elude censorship by cross-referencing topics to juxtapose radical or subversive content (Zimmer, 2009). The World Wide Web as envisioned by Berners-Lee would be a symmetric medium simultaneously allowing users to browse and edit content. However, the developers of the Mosaic browser, which went on to become the core of popular software like Netscape and Internet Explorer,

focused their efforts solely on the downstream browsing effectively making the web an asymmetric, read-only medium for the vast majority of its users. As a result, publishing content to the web became limited to those users with the requisite skill for authoring the hypertext markup language (HTML) or users possessing specialized web publishing programs. In 1995, software developer Ward Cunningham developed a "wiki" as a way of editing web pages via an editor embedded within a web page form. The editing software for translating users' inputs into HTML resides on the server rather than the users' local computers and any user who can browse the page could potentially edit the page as well. Wikis were initially used for software developers to engage in "agile development" emphasizing flat organization and rapid ad-hoc collaboration over formal structures and documented requirements. Crucially, wikis undermined the decentralized structure which distinguished the WWW from other media by virtue of linking only to other content which resided on the same wiki rather than across the web (Reagle Jr, 2010).

While electronic encyclopedias like Encarta or *Britannica* allowed publishers to overcome the material limitations of printed books, the intellectual labor and licensing remained significant costs in the production of an encyclopedia. Moreover, these encyclopedias were still fundamentally static and required the publication of a new CD-ROM to update if their content was to be updated. The advent and adoption of hypertext and networking technologies like the World Wide Web offered an alternative route by which information, including encyclopedias, could be dynamically and instantly updated. One of the first online encyclopedias, the *Stanford Encyclopedia of Philosophy*, granted its authors File Transfer Protocol (FTP) accounts so that they could revise, expand, and update their entries whenever needed. The authors of this system argued that "any encyclopedia which [quickly adapts to reflect advances in research] will necessarily be more useful in the long run than those which don't" (Hammer & Zalta, 1997). This new type of dynamic encyclopedia also offered the possibility of expanding the potential audience, reducing the limitations on inclusiveness or size, eliminating the lag between writing and publication of entries, reducing expenses associated with publication, being able to adapt to new tools,

languages, and techniques, and providing a record of user data to inform decisions about expanding topics attracting substantial interest (Allen, Nodelman, & Zalta, 2002; Hammer & Zalta, 1997).

However, this entirely computer-mediated approach raised complex questions about version control, editorial oversight, tracking co-authorship, and structuring communication and administrative functions that the nascent domain of computer-supported cooperative work (CSCW) was only beginning to address. Among these details were how to establish a systematic method for evaluating both new entries and subsequent changes to entries, ensuring a uniform style across distributed work, educating authors and editors to have the skills to use technologies like UNIX, FTP, and HTML, automating routine tasks such as updating the table of contents and notifying editorial board members of updates, securing copyrights and licensing for both submitted content and related media like photographs, developing norms about the regularity with which articles should updated, securing the project from unauthorized access or malicious behavior, and issues of digital preservation when citations reference outdated content (Hammer & Zalta, 1997). Unlike electronic journals or pre-print exchanges, this new "dynamic reference work" (DRW) would require a highly customized work-flow system to allow authors and editors to collaboratively write and maintain a refereed resource while also tracking new ideas being introduced into the system (Allen et al., 2002).

The Stanford Encyclopedia of Philosophy was still fundamentally a project only open to participation from scholars. However, other volunteer-driven, online collaborative encyclopedia projects sought to elicit encyclopedic content from volunteers on the Internet, without compensating or attributing contributors, and would be available to members of the public via the Internet at no cost. These "proto-Wikipedias" were announced between 1993 and 2001 and used a wide variety of different technologies to attempt to solicit and distribute contributions (B. Hill, Forthcoming; Reagle Jr, 2010). These projects included "Interpedia" announced in 1993 which sought to develop a client-server technology for contributing and sharing this content, "The Distributed Encyclopedia Project" (TDEP) founded in 1997

which would have been an index to user-submitted topics, “h2g2” announced in 1999 which was affiliated with science fiction writer Douglas Adams and later acquired by the BBC, “The Info Network” announced in 2000, “Nupedia” announced in 2000 by Larry Sanger and Jimmy Wales and the immediate predecessor of Wikipedia, “Everything2” announced in 1998 which would host encyclopedic content alongside creative content, and “GNUpedia” announced in 2001 by free software developers inspired by an essay from Richard Stallman calling for a free encyclopedia. The reasons why these projects failed to gain any traction while Wikipedia succeeded are complex, but Hill argues Wikipedia succeeded because it was built around a familiar product, was focused on substantive content development over technology development, offered lower transaction costs for participation, and de-emphasized attribution (B. Hill, Forthcoming).

Information and communication technologies like digital computers and computing networks likewise transformed journalism. News organizations transitioning to online forms of journalism resisted the affordances offered by these new socio-technical arrangements and preferred linear text instead of hypertext as well as favoring traditional gatekeeping roles over participatory alternatives (Steensen, 2010). The introduction and stabilization of information production systems like Linux characterized by decentralization, peer-production, and anti-market access has played an increasingly major role in the production of the public sphere (Benkler, 2006). Traditional media players reacted by incorporating decentralized elements within their model. However, they competed for the same attention market and were unable to engage in gatekeeping or otherwise exercise control over access to information. Professionally-focused non-profits such as ProPublica, factcheck.org, the Sunlight Foundation, and WikiLeaks suggest the emergence of organizations leveraging technological affordances of computer-mediated media to fulfill traditional investigative journalist roles outside of established media organizations. The rise of online organizations leveraging technological platforms like blogs and diary-based content management systems (CMS) likewise serve as “party presses” for ideological coalition-

building and news aggregation. News organizations also relied on alternative forms of journalism such as "citizen journalism" in which users self-publish journalistic content without the institutional backing, professional accreditation, or formal affiliations which traditionally differentiated the work done by journalists (Bruns, 2008; Deuze, Bruns, & Neuberger, 2007). These examples echo back to nineteenth-century forms of newspapers which were exchanges rather than original sources of content and in which unpaid readers supplied much of the news, poetry, commentary, and minutes of meetings (van Tuyll, 2010).

Discussion

Encyclopedism and journalism shaped social structures and values as well as created an environment in which new populations such as online encyclopedias could emerge to compete or collaborate with pre-existing populations of paper encyclopedias and newspapers. Both populations of encyclopedias and newspapers exhibited features of interdependence as the viability of one affected the other and both depended upon common technologies like printing presses and were oriented to similar normative orders of sharing and disseminating information. Rising literacy, increasing political liberalization, and the rapid progress of scientific discoveries and technological inventions created new markets and audiences. Despite the varying philosophical predilections and political dispositions of their editors, encyclopedias and newspapers were also businesses that provided value to disparate communities' demands for a knowledge artifact that was authoritative, timely, and useable. Journalism and encyclopedism did not inevitably collide because of some intrinsic similarities, nor did Wikipedia come about solely because of the invention of new technologies. Rather, the actors and institutions invested in the production of increasingly overlapping types of knowledge redefined the boundaries of their work and established new identities in the face of technological and social discontinuities that demanded they transform and adapt. Both genres evolved as division of labor where work was initially concentrated in a few editors or

publishers and later became differentiated among different specialties such as reporter versus printer and section editor versus article author.

The case of how encyclopedians from Diderot through Wikipedia negotiate new knowledge reveals tensions in how boundaries and classifications are constructed, reinforced, crossed, or dissolved in knowledge production. The attempts by encyclopedists throughout history to balance epistemic authority with contemporary relevance and accuracy, is illustrative of the perils of classification work. News workers expanded their domain from gossip-mongering into a crucial "fourth estate" in democratic civic life, excluded related informational genres and discourses like radio, politics, or gossip by emphasizing their values of objectivity and autonomy, and justified its privileged position by adopting professionalized values and codes of conduct. The expansion of these genres to encompass larger and more diverse types of content led both genres to inhabit similar domains: encyclopedism strived to reflect current events and journalism sought to make authoritative and objective knowledge claims.

The histories of encyclopedism and journalism illuminate how the identity of these distinct genres of knowledge production and sharing have been challenged and transformed by profound social and technological changes leading both types of artifacts to inhabit larger and more diverse social worlds. The evolution of these genres and slippages and transformations in their identities and boundaries illuminate the permeability, salience, and durability of boundaries in knowledge work as well as the work necessary to activate, maintain, dispute, and dissolve boundaries as encyclopedists and journalists establish and reinforce social identities and roles (Lamont & Molnar, 2002). The boundary work encyclopedia editors and news workers engaged to expand the scope and possibilities of encyclopedic and journalistic knowledge work is emblematic of processes by which institutions are legitimized and develop shared meanings which become taken-for-granted. This history is instructive because the emergence and influence of Wikipedia's news work explored in the subsequent chapters illuminates the socially constructed nature of these distinctions and identities. The norms, values, identities, and legitimacy that

constitute institutions like encyclopedias or newspapers are the result of long historical processes of contestation and transformation. The fact that it has become *unnatural* to read an encyclopedia and find it out of date is emblematic of how organizations are transformed and new cultural values are socially constructed and institutionalized.

This chapter offers only a sketch for evaluating the substantive relations between journalism and encyclopedism within the larger knowledge community and the boundaries between these populations. Historical work like this is also subject to selection biases emphasizing the features of successful organizations over the features of organizations that did not persist and jeopardizing the validity of any inferences made. Further archival research examining primary and secondary historical sources should examine changes in the ideological alignment between journalists and encyclopedists, the cohesion of the identity within each population, the flows of individuals, artifacts, and symbols between populations, and the practices of how these artifacts were produced and consumed in relation to each other. For example, while encyclopedias and newspapers both made similar demands on the resources available in the environment such as the need to recruit authors or publish regularly, less is known about the direction or extent to which the consumption of newspapers or encyclopedias drove the consumption of the other.

The next chapter will explore how new socio-technical systems like Wikipedia have profoundly transformed the institutional arrangements governing access to and means of dissemination information which has in turn problematized existing classifications over what is or is not encyclopedic and journalism. As Star and her colleagues argue (Bowker & Star, 2000; Star & Griesemer, 1989), entities like Wikipedia's coverage of current events that explode these socially-constructed categorical distinctions reveal their artificiality. While the material forms of each limited their direct overlap—encyclopedias cannot be published daily nor can a newspaper be hundreds of pages—the advent and adoption new technologies required the re-definition of existing boundaries and the re-articulation of identities. Wikipedia differs from these organizations along several major dimensions such as its non-

profit business model, lack of compensation or attribution for contributors, autonomous membership, and absence of direct or formal control structures, but the patterns observed in the evolution of encyclopedism and journalism provide a framework to evaluate how Wikipedia's chimeric coverage of breaking news events within an encyclopedia evolved along analogous trajectories. Facing an environment saturated with established competitors like *Britannica* and *Encarta*, new encyclopedia projects sought out untapped niches and resources, exploited discontinuities in technology and norms, and developed new interdependent relationships with other populations such as journalism. This shift by Wikipedians to erect boundaries about the types of knowledge and information that warrant inclusion bears a remarkable resemblance to the boundary work encyclopedists and journalists employed through history. However, human predispositions to construct classifications, standards, and categories as means of making sense of the world suggests that the work we do in the breach inclusion to re-establish boundaries reveal the possibility of new forms of social action.

Chapter 4: An Encyclopedia with Current Events

Wikipedia is the online encyclopedia that "anyone can edit." But this encyclopedia also contains information that is not found in any other printed or electronic encyclopedias: the recentness of the Tōhoku earthquake and tsunami, protests in the Arab world, or death of Osama bin Laden prevents them from being included in printed editions of *World Book* or *Encyclopedia Britannica* despite their impact, severity, or historical import. However, these events have lengthy and complex entries on Wikipedia authored by hundreds of editors and viewed by hundreds of thousands of users in the span of hours.

As the previous chapter described, for the vast majority of Western intellectual history, it was "natural" that an encyclopedia would not cover recent historical events, new scientific discoveries, or contemporary biographies owing to the material limitations of printing, philosophical and cultural values of its editors, and consumers' preferences for obtaining this information through alternative genres and mediums. However the "naturalness" of this exclusion has been completely reversed in the last decade; it has become entirely natural to go to an online encyclopedia (in addition to going to newspapers and broadcast media) to find information about current events (Melanson, 2010; Montgomery, 2011; Siegel, 2007). While this shift has only become possible because of the development of new technologies, it also resolves a centuries-long tension: how should encyclopedia editors integrate new information and recent historical changes into a very large, collaboratively-authored document?

Wikipedia departs from traditional encyclopedic models along many dimensions, but this chapter is an "origin story" to trace the trajectory of Wikipedia's policies on the use of breaking news events back to a transformative case and illustrates how Wikipedia has come to inhabit the social world of news work. Despite socio-technical capabilities that could permit Wikipedians to rapidly coauthor articles about any breaking news event, blending the authority of encyclopedic knowledge with the immediacy of journalistic information about on-going events, Wikipedia has developed an extensive set of standards

and rules governing the scope and type of current events that may be included. While Wikipedia firmly locates its identity within the encyclopedic genre,[3] Wikipedia's inclusion of and concentrated focus on articles about current events required Wikipedia's users to re-demarcate the types of knowledge worthy of inclusion as well as practices to enforce these boundaries which have a remarkable correspondence with the institutionalization of journalistic news values. The expansion and articulation of these boundaries follows analogous processes of boundary work observed among encyclopedists and journalists in the previous chapter.

Just as news workers expanded their domain from gossip-mongering into a crucial "fourth estate" in democratic civic life, an analogous process of institutionalization is also found in the evolution of Wikipedia's inclusion of current events. Wikipedia's adoption of formal policies delineating the types of content which warrant inclusion which bear a remarkable similarity to the news values seen in traditional forms of journalism. Given this outward isomorphism between journalistic values and Wikipedia policies, a natural question is the extent to which Wikipedia's news work and workers have followed similar trajectories over time and have also become "specialized" in their own news work. This chapter reviews the early history of Wikipedia from the motivations of its founders up to the overwhelming outpouring of editorial effort in response to the terrorist attacks of September 11, 2001 (9/11). Because these debates and the evolution of precedents, rules, platforms, and tools occurred within the confines of Wikipedia itself, they have largely been recorded and preserved for analysis. This case provides a context to examine how this radical approach to encyclopedism still crossed distinct social worlds but also had to bridge the divide between the authority of encyclopedic ideals and the rapidity of journalistic practice. As the collective cultural trauma of 9/11 faded, the memorial content it contained became problematic and the community began to engage in complex boundary work to (1) isolate this "unencyclopedic" 9/11 content from the rest of the Wikipedia project by adopting regulations governing the type and extent of

[3] http://en.wikipedia.org/wiki/Wikipedia:Introduction

information permitted to be in the encyclopedia while simultaneously (2) promoting Wikipedia's coverage of current and breaking news events on its homepage despite (3) the existence of a dedicated WikiNews project. These rules about the notability of news events and precedents for featuring news content on the encyclopedia illuminate the permeability, salience, and durability of boundaries in knowledge work as the work necessary to activate, maintain, dispute, and dissolve the boundaries of encyclopedism and journalism (Lamont & Molnar, 2002).

Creation of Wikipedia

Jimmy "Jimbo" Wales and Larry Sanger, following the failure of their earlier "free encyclopedia" project called Nupedia, founded Wikipedia in January 2001. Wales was independently wealthy after a brief career as a futures and options trader in Chicago and participated in a number of Y2K-related internet message boards where he met Sanger. Wales described his idea of a free content, collaborative encyclopedia to Sanger in January 2000 and offered him a position as editor-in-chief and project manager of their new encyclopedia project. Sanger launched "Nupedia" in March 2000, but Sanger's onerous emphasis on soliciting content only from credentialed experts and extensive peer review severely bottlenecked content creation. Eighteen months later, only twenty-five articles had been published, many of them on eclectic and mundane topics such as Irish traditional music, SNOBOL 4 programming language, quasispecies model, and polymerase chain reactions.[4] Streamlining the content generation and vetting process consumed Wales and Sanger over the next year and in January 2001, Ben Kovitz proposed Sanger adopt a "wiki." This technology was developed by Ward Cunningham in 1994 and allows a webpage to be modified through a web interface rather than being a static file on a server.

This "Wikipedia" (a portmanteau of "wiki" and "encyclopedia") was launched on January 15, 2001 as the "ULTIMATE open and simple format for developing content" for the public to develop a content that would initially remain separate from Nupedia but could be fed into its editorial process. However, editors

[4] http://web.archive.org/web/20011205053814/http://nupedia.com/newest.phtml

of the Wikipedia ("Wikipedians") created nearly 20,000 articles by the end of 2001,[5] outpacing all expectations and capacity to process through the Nupedia editorial process. The failure of Nupedia to achieve a critical mass of content or contributors, the burgeoning success of the Wikipedia side-project, and implosion of the dot-com bubble in mid-2001 all contributed to Wales laying Sanger off in February 2002. Sanger briefly stayed on as a volunteer but left the Nupedia and Wikipedia projects entirely in March and by December 2002 Nupedia was shuttered and its limited content moved entirely to Wikipedia.

Wikipedia and the September 11 attacks

On September 11, 2001, 19 terrorists affiliated with al-Qaeda hijacked four airliners and deliberately crashed two of them into the World Trade Center twin towers in New York City, one of them into the Pentagon outside Washington, D.C., and a fourth intended for the U.S. Capitol into a field in rural Pennsylvania after passengers resisted. The attacks destroyed the tallest skyscrapers in Manhattan, severely damaged a wing of the largest office building in the world, killed nearly 3,000 people, and injured another 6,000. The audacity of the attacks and scale of the destruction resulted in profound and haunting images that were widely disseminated through traditional news and broadcast media as well as new forms of media such as web sites, email, and instant messaging applications. People flocked to these new forms of media to receive the latest information, check in with each other, and begin the process of trying to make sense of the calamity by joining discussions and coping (Rainie, Internet, & Project, 2002). The responses on the web suggested new forms of two-way information sharing, collective action, and civic engagement which challenged prevailing models of hierarchical organizing or centralized broadcasting (Foot, Warnick, & Schneider, 2005).

Wikipedia's existence coincided with the catastrophic terrorist attacks on September 11, 2001 and provides a profound example of how this socio-technical system was used to support high-tempo

[5] http://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia

knowledge collaboration and sharing as well as how it shaped subsequent consensus on the scope and role of this content within the system. Unfortunately, changes in Wikipedia's technological systems and software have censored changes made to any articles created before early 2002, making an authoritative history of editors' revisions to other articles about major events in 2001 (President George W. Bush's inauguration, merger of AOL-TimeWarner, death of Dale Earnhardt, Robert Hanssen spying scandal, Mir re-entry, Hainan Island incident, Dennis Tito space tourism) or the immediate aftermath of the attacks unavailable. However, cached versions of this content are available from the Internet Archive's "Wayback Machine" from as early as October 9, 2001[6] as seen in Figure 5 below.

The sheer scale of Wikipedians' responses in the four weeks following the September 11 attacks is profound. Wikipedia had approximately 13,000 articles in October 2001 and no fewer than 100 of these articles (~1%) were directly related to the attacks, aftermath, lists of news articles, and articles about casualties and survivors. As seen by the hyperlinks from the parent article in Figure 1, the topics of these articles are wide-ranging and include ostensibly encyclopedic content such information about the flights, buildings, military and economic responses but also decidedly non-encyclopedic content such as a list of casualties and survivors, personal experiences, and details about how to donate blood, money, or provide other assistance. The ease of editing content on Wikipedia quickly lead to it becoming a "clearinghouse" recognized by the content categorizers at Yahoo! and added to a list of "Online Memorials and Tributes."[7]

The lists of casualties, missing persons, and survivors are striking insofar as the authors invoked the policy "Wiki is not paper" to argue that each victim warranted a separate article.[8] One of the lead editors, "The Cunctator", called for editors to build memorial pages by adding content to personal experiences, building articles for casualties using a template, creating historical pages such as a timeline or critical

---

6 http://web.archive.org/web/20011009234151/http://www.wikipedia.com/wiki/September_11,_2001_Terrorist_Attack

7 http://web.archive.org/web/20011111045050/http://dailynews.yahoo.com/fc/us/memorials_and_grief/

8 http://web.archive.org/web/20011009025548/http://www.wikipedia.com/wiki/September_11,_2001_Terrorist_Attack/Casualties_Talk

background on Osama bin Laden or box-cutter knife, and administrative or technical tasks such as creating navigation and formatting templates.[9] By October 17, there was a list categorized by last names, location, and status (civilian vs. responder) containing several dozen articles about victims.[10] However, the task of writing biographies for more than 3,000 victims and survivors was beyond the capabilities of the few dozen contributing editors. Other editors began to raise concerns about the coherence of these articles' quality, lasting notability of these victims, or the importance of this project to authoring an encyclopedia.

By September 2002, there was a concerted discussion about what do with the "September 11 pages" as a result of several heated discussions about deleting non-notable victims' articles from Wikipedia. A consensus emerged to remove the content related to victim memorials, tributes, personal experiences, and general discussion to a "September 11 Memorial Wiki." The Memorial Wiki project was launched in March 2003 and hosted by the Wikimedia Foundation as a separate namespace akin to another language version of Wikipedia.[11] The Wiki described its mission as providing "additional resource[s] of personal opinions, individual experiences, memorials, and tributes" while the Wikipedia content was to be focused on a "neutral and complete history of the attack, including the background history, the aftermath, and more."[12] This decision had the effect of extending the feature of "interlanguage links" intended to provide encyclopedia pages about the same topic in another language, effectively creating a whole parallel English Wikipedia solely devoted to topics related to 9/11. The "9/11 Wiki" was immediately controversial as some editors felt the focus on memorializing the victims of the event distracted from the mission of writing about topics of encyclopedic importance, was an artifact of a non-neutral, chauvinist, and/or ahistorical perspective over-emphasizing the recent past, or would set a precedent that should

[9] http://meta.wikimedia.org/wiki/Sep11wiki
[10] http://web.archive.org/web/20011110011001/http://www.wikipedia.com/wiki/September_11,_2001_Terrorist_Attack/Casualties
[11] http://web.archive.org/web/20030413072718/http://sep11.wikipedia.org/wiki/In_Memoriam
[12] http://meta.wikimedia.org/wiki/Dealing_with_September_11_pages

extend to creating memorial spaces for victims of other attacks, genocides, and catastrophes. Discussions about which victims, topics, and articles were notable enough to remain on Wikipedia and which were to be moved to the new website were also fraught.

September 11, 2001 Terrorist Attack



HomePage | Recent Changes | Preferences | Receive an article a day!

In Memoriam, September 11, 2001

On the morning of September 11, 2001, what might well be the most devastating terrorist attack in the history of the world occurred concurrently in New York City, Washington, D.C. and near Pittsburgh. Four passengers jets were hijacked and then deliberately crashed into the World Trade Center and the Pentagon. Both towers of the World Trade Center subsequently collapsed and part of the Pentagon was destroyed in the ensuing fire. Casualties are expected to be in the thousands: 266 passengers; about 6600+ people, including hundreds of firefighters who had rushed in, at the World Trade Center; and 125 at the Pentagon.

Some passengers on the doomed flights were able to make phone calls reporting on events on board. They reported that there were more than one hijacker on each plane, and that they took control of the planes using box-cutter knives. It appears that the passengers on the fourth jet tried to overpower the hijackers and that the plane crashed in a sparsely populated area as a result, thereby missing its intended target which may have been the White House.

The attack had immediate and deep global political effects and economic effects; as well as an international outpouring of memorials and services. The massive undertaking of rescue and recovery, and providing assistance to the survivors and victims, is ongoing. There will be a great need for donations for a long time.

Though no group has claimed responsibility, the US government immediately launched a full-scale response, stating its intentions to go to war against those responsible. On October 7, a coalition led by the United States launched an attack in Afghanistan. See **2001 U.S. Attack on Afghanistan**.

Timeline: September - October - Back history

A minute-by-minute breakdown of events, and a history of events leading up to the attack. Where to [add] or read the latest breaking news.

Casualties - Missing Persons - Survivors - Personal experiences

Victims and survivors, and their personal stories. All together, these pages may serve as a memorial to those lost. Your help is welcome and needed.

Give Blood - Donations - Assistance - Rescue and Recovery Effort

How people across the world have helped, and can continue to do so. What is being done to assist victims.

Closings and Cancellations - Memorials and Services - US Governmental Response

How people have responded.

Responsibility - Hijackers - World political effects - World economic effects

The long-term history of the attack.

Airport security - Gasoline price gouging - Communications - Slogans and terms

Specific topics.

See also: World Trade Center -- Pentagon -- New York City -- Washington, D.C. -- AA Flight 11 -- UA Flight 75 -- AA Flight 77 -- UA Flight 93 -- U.S. Department of Defense -- FBI -- terrorism -- domestic terrorism -- terrorist incidents -- Osama bin Laden -- Taliban -- Islamism -- Afghanistan -- collective trauma -- September 11

External news sites and references - Misinformation and Rumors

/Talk - /Casualties Talk - Give Blood/Talk - /Hijackers Talk

/Casualties template - /Footer template

HomePage | Recent Changes | Preferences | Receive an article a day!
You can edit this page right now! It's a free, community project
Edit text of this page | View other revisions
Last edited October 8, 2001 11:25 pm (diff)
Search: Search

Figure 4-1: Screenshot of the Wikipedia article for the 9/11 attacks as of October 9, 2001.

Following the catastrophic 2004 Indian Ocean earthquake and tsunami, suggestions were again made to create an analogous memorial wiki for victims, topics, reflections, and responses to the catastrophe but dismissed by both the community and Wikimedia Foundation, the non-profit organization that operates the Wikipedia and other projects. The decisions cited both the reluctance to create a wiki following each disaster as well as refusal to adjudicate which disasters were sufficiently serious to warrant memorial wikis. Moreover, the failure of the 9/11 Wiki to independently thrive was obvious: there was little daily editing activity, content had stagnated (the logo remained unchanged years after Wikipedia adopted a new design), and it was victimized by newbies, spammers, vandals, and internet trolls while lacking a core of dedicated editors to police these changes. Following the emergence of policies regarding using Wikipedia as a news outlet or memorial (see next section) as well as a growing body of jurisprudence governing notability and neutral point of view (NPOV) from other articles, editors began to point the untenability of maintaining a memorial with personal and non- NPOV content under the aegis of the Wikimedia Foundation's mission.[13] A proposal to delete the September 11 Memorial Wiki was introduced in December 2005 with an overwhelming consensus to close the project.[14] Another proposal was introduced in early 2006 to move the content from under the Wikimedia Foundation to another site such as "MemoryWiki.org".[15] This content move may have happened, but never appeared in the cached versions of MemoryWiki available on Internet Archive. In May 2006, another extensive discussion occurred as a result of a proposal to close the Memorial Wiki with over 104 editors supporting closure and only 6 editors opposing.[16] In September 2006, the Memorial Wiki was made "read only" which effectively locked it from any further editing and then was taken completely offline from Wikimedia Foundation servers at a later date.[17] The Memorial Wiki content was hosted at "sep11memories.org" until at least June 2009 but as of May 2012, the domain name no longer resolves. A database dump of the content is still

[13] http://meta.wikimedia.org/wiki/Sep11wiki/Babel_thread
[14] http://meta.wikimedia.org/wiki/Sep11wiki/Request_for_deletion
[15] http://meta.wikimedia.org/wiki/9/11_wiki_move_proposal
[16] http://meta.wikimedia.org/wiki/Proposals_for_closing_projects/Archive/September_11_Wiki
[17] http://lists.wikimedia.org/pipermail/foundation-l/2006-September/023338.html

available[18] as well as cached versions through Internet Archive, but the Memorial Wiki has effectively ceased to exist as either a website or wiki.

[18] http://dumps.wikimedia.org/sep11wiki/20071116/

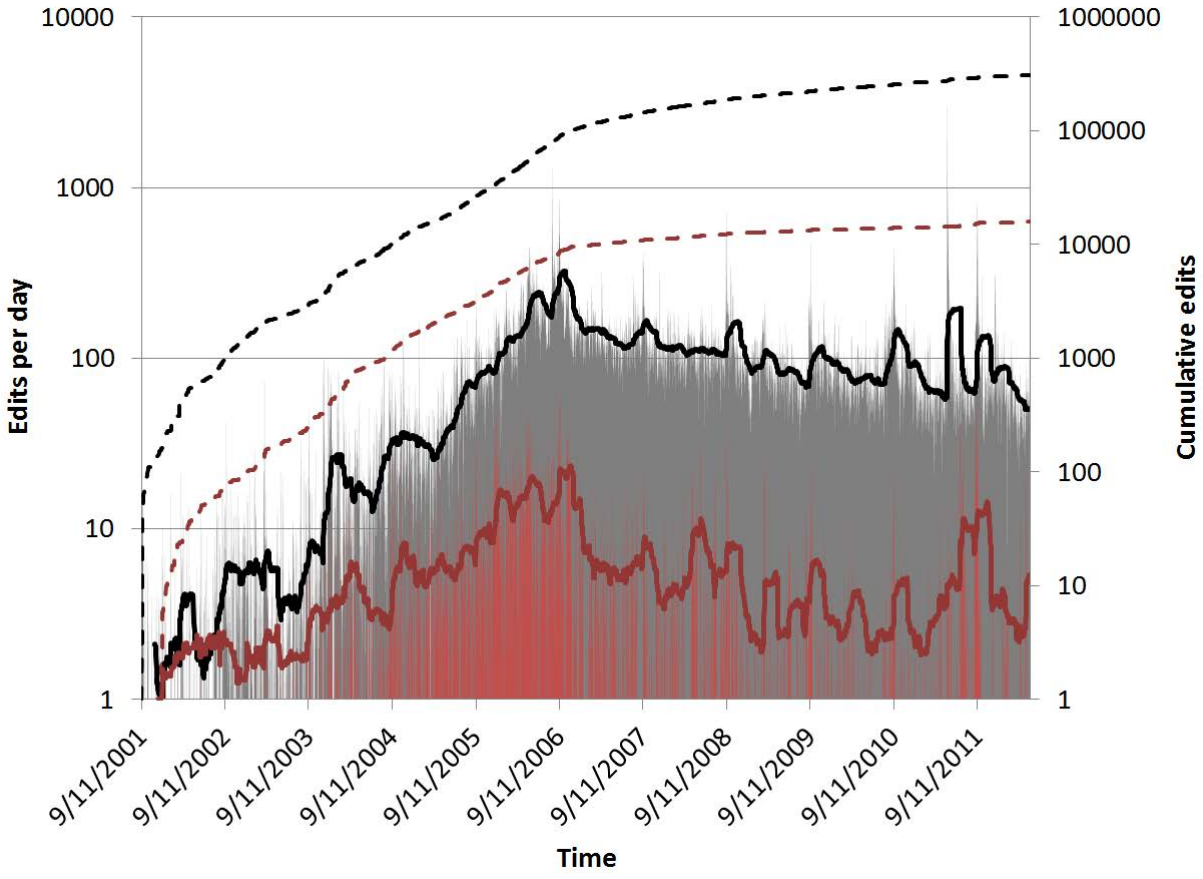


Figure 4-2: Edit activity on 606 articles in "September 11 attacks" category.

Despite the dissolution of the Memorial Wiki, a substantial amount of September 11-related content remained on Wikipedia and continued to be edited. Figure 4-1summarizes the contents of the "September 11 attacks" category. 88 articles are categorized in the category which itself contains eight subcategories about the aftermath, buildings destroyed, criticism of official accounts, people associated, memorials and flights. As of January 2012, 309,727 revisions from 65,032 unique editors have been made to the 606 articles in this category and its sub-categories. Articles about the timeline of events, information on the flights and hijackers, effects on airport security and economy, aftermath for closings and cancellations, government and military response, lists of memorials and services continue to exist well after their creation in 2001. Figure 4-2 shows articles in category and subcategories in grey, edit activity on the "September 11 attacks" article in red, 60-day moving averages in darker shades (respectively), and cumulative edit count in dashed lines (colored respectively). Editing activity peaked around the 5-year anniversary in 2006, which was also close to the peak of activity on Wikipedia itself (Suh et al., 2009), but bursts of activity are also observed near the anniversaries suggesting editing activity is a form of commemoration (Ferron & Massa, 2011a, 2011c). The 605 articles related to the "September 11 attacks" article continue to receive dozens of contributions per day and seven of these articles have been awarded "Featured Article" status reflecting their quality.[19]

The scale and controversy surrounding the 9/11 attacks informed the development of practices, policies, and precedents for subsequent articles about current events which I explore below. In light of Wikipedia's mission as an encyclopedia, the bursts of activity related to users' information seeking behavior following news events, and the possibilities of using wikis to rapidly generate new content, I argue Wikipedians developed new ways to highlight encyclopedic content in the news for users. However, the fallout of the September 11 memorial episode also motivated the development and adoption of policies which serves to explicitly demarcate the boundaries of Wikipedia's encyclopedic identity as neither a news source nor a

[19] 7 World Trade Center, American Airlines Flight 11, American Airlines Flight 77, Construction of the World Trade Center, Khalid al-Mihdar, United Airlines Flight 93, and Ziad Jarrah

memorial site. These initiatives revealed fascinating tensions the community was negotiating over the role and scope of current news events on the project.

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
What links here
Related changes
Upload file
Special pages
Permanent link

Print/export

Languages
العربية
Български
Česky
Dansk
Deutsch
Español
Euskara
فارسی
Français
한국어
Bahasa Indonesia
Italiano
עברית
Bahasa Melayu
Nederlands
日本語
Norsk (bokmål)
Polski
Português
Русский
Suomi
Svenska
ไทย
Türkçe
Tiếng Việt
中文

Category Talk
Read Edit View history TW Search

Our updated Terms of Use will become effective on May 25, 2012. Find out more.

Category:September 11 attacks

From Wikipedia, the free encyclopedia

Pages in this category should be moved to subcategories where applicable.
This category may require frequent maintenance to avoid becoming too large. It should directly contain very few, if any, articles and should mainly contain subcategories.

The main article for this category is ***September 11 attacks****.*

See also: Category:World Trade Center

September 11 attacks portal

Wikimedia Commons has media related to: *September 11 attacks*

Subcategories

This category has the following 8 subcategories, out of 8 total.

A
- ► Aftermath of the September 11 attacks (5 C, 74 P)
- ► American Airlines Flight 77 (11 P)

B
- ► Buildings destroyed in the September 11 attacks (9 P)

C
- ► Criticism of the official accounts of the September 11 attacks (2 C, 3 P)

P
- ► People associated with the September 11 attacks (3 C, 44 P)
- ► Proceedings surrounding the September 11 attacks (1 C, 6 P)

S
- ► September 11 attacks memorials (18 P)

U
- ► United Airlines Flight 93 (1 C, 23 P)

Pages in category "September 11 attacks"

The following 88 pages are in this category, out of 88 total. This list may not reflect recent changes (learn more).

- September 11 attacks
- Portal:September 11 attacks

1
- 102d Intelligence Wing

2
- 20th hijacker

9
- September 11 attacks advance-knowledge conspiracy theories
- 9/11 paintings

A
- A Manhã do Mundo
- Able Danger
- Al-Quds Mosque Hamburg
- Eyad Alrababah
- American Airlines Flight 11
- American Airlines Flight 77
- ATC Zero
- Ayub Ali Khan

B
- The Bathtub
- Bill Biggart
- Bin Laden Issue Station
- Osama bin Laden
- Bin Ladin Determined To Strike in US
- Jens Blauenfeldt
- Bojinka plot
- Jean-Charles Brisard

C
- Cantor Fitzgerald
- Communication during the September 11 attacks
- Cortlandt Street (IRT Broadway – Seventh Avenue Line)

D
- Delta Air Lines Flight 1989
- Detroit Free Press v. Ashcroft
- Deutsche Bank Building

E
- Emma E. Booker Elementary School

F
- The Falling Man

G
- Global Guardian
- Ground zero

H
- Abdallah Higazy
- Huffman Aviation
- Saleh Ibn Abdul Rahman Hussayen

I
- In God We Trust, United We Stand, September 11, 2001, Memorial Garden

J
- Jake (rescue dog)
- John J. Harvey

K
- Niaz Khan
- Korean Air Flight 85

L
- Let's roll

M
- Transportation 9-11 Medal
- Millenium Hilton Hotel
- Khalid Sheikh Mohammed
- Motives for the September 11 attacks
- Abdul Hakim Murad (militant)
- Abdelghani Mzoudi

O
- Offutt Air Force Base
- Outline of the September 11 attacks

P
- Park51 controversy
- The Pet Goat
- Phoenix Memo
- Sandra Priest
- Project Rebirth

R
- Radio communications during the September 11 attacks
- Responsibility for the September 11 attacks
- Transportation 9-11 Ribbon

S
- St. Nicholas Greek Orthodox Church
- Security Control of Air Traffic and Air Navigation Aids
- September 11 intelligence prior to the attacks

S cont.
- Shanksville, Pennsylvania
- Stairwell A
- Strategic Assessments Branch
- Survivors' Staircase

T
- List of tenants in One World Trade Center
- List of tenants in Two World Trade Center
- List of tenants in Four World Trade Center
- List of tenants in Five World Trade Center
- List of tenants in Six World Trade Center
- List of tenants in Seven World Trade Center
- Timeline for the day of the September 11 attacks
- Timeline of al-Qaeda attacks
- Timeline of the September 11 attacks
- Tourist guy
- Trakr

U
- U.S. military response during the September 11 attacks
- United Airlines Flight 93
- United Airlines Flight 175
- United Nations Security Council Resolution 1368
- United States government operations and exercises on September 11, 2001

V
- Vehicles of the hijackers in the September 11 attacks
- Verizon Building
- Visa Express

W
- Windows of Hope Family Relief Fund
- Windows on the World
- Winter Garden Atrium

M
- Template:September 11 attacks

T
- Template:9-11 hijackers

Figure 4-3: Screenshot of the Wikipedia category "September 11 attacks" as of May 2012.

Current events on the homepage

Like other peer-produced projects and online communities, the young encyclopedia urgently needed to recruit more contributors to generate content and sustain an active community. The burst of activity related to the production and consumption of 9/11-related content was impossible for administrators on the project to ignore and links to these oft-sought articles were placed on the Wikipedia homepage by early October with "Breaking News" and "Special features: current events." The choice to privilege this content on the front page reflects a usability imperative to channel readers to popular content, a socialization imperative to convert readers to contributors, and an institutional imperative to demonstrate the legitimacy and potential of the system to rapidly and dynamically author content.

Given the continued unfolding of news events, this front-page content evolved and expanded in scope from 9/11-related content to include a "Current Events and Breaking News" section with the note "Encyclopedia articles about topics behind the news. Develop articles about important events as they happen!" By November 5, 2001, this current events section which had been devoted to 9/11-related articles such as the aftermath of the attacks, the coalition campaign in Afghanistan, Osama bin Laden, and government agencies expanded and included a link to a dedicated "Current Events" article.[20] Other articles related to international, business, sports, and celebrity news were also included on this article, albeit in substantially less detail than the 9/11-related content.

By early January 2002, the Current Events article had evolved from a list of articles providing background about topics in the news to a temporally ordered list of topics appearing in the news. No data is available from the Internet Archive for the Current Events article between January and August 2002, but by that August, content related to terrorism, the 9/11 attacks, and war in Afghanistan had been removed entirely and replaced with a daily blotter of articles about events in the news. 9/11-related

[20] http://web.archive.org/web/20011105190446/http://www.wikipedia.com/wiki/Current_events

content disappeared from the “Current events and breaking news” section on the Wikipedia homepage by January replaced by a link to the “Current Events” article and an assortment of other salient articles. The Current Events section was re-labeled “In The News” (ITN) by November 2002 and was joined by other sections on “Recent deaths”, “Ongoing events”, and “Historical anniversaries”. These sections only included hyperlinks to articles while the “Current Events” article included sentence-length summaries of news events. Before the Main Page was protected so that only administrators could revise it, edit wars between users were common as they debated which articles were sufficiently important or still timely to merit inclusion.



Figure 4-4: The "In the News" section on the Wikipedia homepage before (top) and after 2004 redesign.

As seen in Figure 4-4, a major redesign of the Wikipedia homepage in February 2004 led to ITN and other homepage sections becoming specialized objects called "templates" rather than simple hyperlinks.[21] A template is a very specialized Wikipedia "article" which can effectively appear within other articles. Templates are generally used to standardize content by referencing a single, central object such as a navigation box of related articles without having to edit the content on each individual article. However, while the ITN section which had appeared on Wikipedia's homepage over two years had simply linked to the "Current Events" article and was followed by context-free hyperlinks, the creation of an ITN template differentiated practices in how current events were communicated on the homepage.

While the "Current Events" page continued to remain a daily blotter any user could edit with sentence-length and hyperlinked summaries of the news, the ITN template adopted the same sentence-length summaries but with some important differences in practice. Because of its prominence on Wikipedia's homepage, the template was a magnet for vandalism and tendentious editing that would result in intermittent page protections being imposed which limited editing to only users with administrative privileges. The growth of Wikipedia, on-going debates and edit wars about Anglophone biases, and increasingly severe misuse of the template resulted in it being permanently protected in February 2005.[22] Following this protection, non-administrative Wikipedia editors had to nominate current events candidates to be discussed until the group reached a consensus and successful candidates were later added by an administrator. The influence of editor tenure and experience has interesting gatekeeping effects as prolific editors' oppositions to nominations during these deliberations appear to be more influential on the final decision than their support of nominations (Keegan & Gergle, 2010). This nomination and deliberation process for ITN items has remained largely unchanged through 2012.

[21] http://web.archive.org/web/20040224004817/http://en.wikipedia.org/wiki/Main_Page
[22] http://en.wikipedia.org/wiki/Wikipedia_talk:In_the_news/Archive_2#Protected

The practices around how Wikipedia popularized its coverage of current events shifted dramatically as the project became more popular. What began as a feature to rapidly funnel users to the most-used content in the aftermath of the 9/11 attacks evolved to become a mechanism for highlighting a variety of timely content for information seeking readers, a tool explicitly intended to recruit new users to join the community, and recognition that wikis could enable encyclopedias to be reference works that are authoritative as well as dynamic and relevant. Despite the stability of the community consensus to prominently feature a regularly-updated current events section on the homepage, the ease of editing this content afforded by the wiki technology as well as the inherent subjectivity of determining the relative importance of events led to significant institutionalization of what had previously been a very open process. However, this institutionalization of the Current Events and ITN section reflects the stabilization and importance of the role current events content plays in the daily operation of Wikipedia and its community. Wikipedia's privileging of Current Events on its homepage illustrated the possibilities of creating an encyclopedia that was not only authoritative owing to its radical openness, but also using this inherent instability as a feature for matching users information seeking behavior with wikis' information production behavior by showcasing and rapidly updating suddenly salient knowledge.

Policies for news content

Wikipedia's policies are official rules developed by consensus reflecting the encyclopedia's mission (e.g., Wikipedia is an encyclopedia, not a dictionary), values (e.g., espousing a neutral point of view and employing reliable sources), and expectations of editor behavior (e.g., do not infringe copyright). Users who violate policies are warned, temporarily barred from editing, or permanently blocked depending upon the severity of the infraction. Like its encyclopedia articles, Wikipedia's policies are open to being revised by community members: changes can be introduced unilaterally or after extensive deliberation. Consensus around these policies is often constructed, expanded, legitimized, or challenged through a network of guidelines that are typically norms governing mundane topics such as style and formatting as

well as essays which can be humorous observations,[23] detailed expositions of a position,[24] or a synthesis of other policies[25] (Reagle Jr, 2010). Essays, guidelines, and policies on Wikipedia can be design standards specifying the properties and features of articles or technologies like templates to ensure their uniformity and compatibility, terminological standards over naming and formatting which ensures the stability and consistency of meaning across articles, performance standards over the features articles must exhibit to warrant recognition, and procedural standards delineating how processes such as protecting articles from editing should be performed (Timmermans & Epstein, 2010). Each of these standards are components of a larger administrative which allows editors to self-regulate their behavior and interrelate their actions with others in the absence of any overt coercive or central coordinator (Butler, Joyce, & Pike, 2008). This system of standards serves to depersonalize authority from agents with particular expertise or familiarity and abstracts this authority to a system of rules, norms, and conventions which are often implicit and learned by participation (Brunsson & Jacobsson, 2002; Bryant et al., 2005). However boring standards may be (Lampland & Star, 2009), their invisibility obscures the moral arguments and technical work that went into their development until attempts are made to change or ignore them (Timmermans & Epstein, 2010). Despite the amount of 9/11-related content and enthusiasm with which current news events were promoted on the Wikipedia homepage, the role of this content in the project raised complex questions about the identity and boundaries of the project itself. The outcomes of these debates influenced policies about Wikipedia's approaches to covering current news events and the types of content it would permit to be included.

Between 2002 and 2008, Wikipedia adopted several policies about news-related content, memorials, and the notability of events which emerged from a variety of precedents, but they are notable insofar as they

[23] http://en.wikipedia.org/wiki/Wikipedia:Fart
[24] http://en.wikipedia.org/wiki/Wikipedia:News_articles; http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_not_a_newspaper; http://en.wikipedia.org/wiki/Wikipedia:Big_events_make_key_participants_notable
[25] http://en.wikipedia.org/wiki/Wikipedia:WIDESPREAD

serve to exclude a substantial amount of potential content which is produced during an unfolding news event and moreover, would have excluded a substantial amount of 9/11-related content itself. An early policy on Wikipedia was a "What Wikipedia is Not" (also known as "WP:NOT", in the Wikipedia shorthand for policies) to enumerate organizational forms, identities, or missions in which the encyclopedia project would not engage. While policies reflect the consensus, like other parts of Wikipedia, they are open to being modified themselves. In September 2002, user "Toby Bartels" updated "WP:NOT" to include a passage about news reports:[26]

> *A news report. Wikipedia should not offer news reports on breaking stories. However, creating "encyclopedia" articles on topics currently in the news is an excellent idea. See "current events" for some examples. (However, the Wiki process lends itself to collaborative, up-to-the-minute construction of current events of historical significance, as long as these are written as encyclopedia articles.)*

This policy became known as "WP:NOT#NEWS" and since evolved to the following:[27]

> *As Wikipedia is not a paper source, editors are encouraged to include current and up-to-date information within its coverage, and the development of stand-alone articles on significant current events. However, not all verifiable events are suitable for inclusion in Wikipedia. Ensure that Wikipedia articles are not:*
>
> 1. ***Journalism.*** *Wikipedia should not offer first-hand news reports on breaking stories. Wikipedia does not constitute a primary source. However, our sister projects Wikisource and Wikinews do exactly that, and are intended to be primary sources. Wikipedia does have many encyclopedia articles on topics of historical significance that are currently in the news, and can be updated with recently verified information. Wikipedia is also not written in news style.*
> 2. ***News reports.*** *Wikipedia considers the enduring notability of persons and events. While news coverage can be useful source material for encyclopedic topics, most newsworthy events do not qualify for inclusion. For example, routine news reporting on things like announcements, sports, or celebrities is not a sufficient basis for inclusion in the encyclopedia. While including information on recent developments is sometimes appropriate, breaking news should not be emphasized or otherwise treated differently from other information. Timely news subjects not suitable for Wikipedia may be suitable for our sister project Wikinews. See also: "Wikipedia:Notability (events)"*
> 3. ***Who's who.*** *Even when an event is notable, individuals involved in it may not be. Unless news coverage of an individual goes beyond the context of a single event, our coverage of that individual should be limited to the article about that event, in proportion to their importance to the overall topic. (See Wikipedia:Biographies of living persons for more details.)*

[26] http://en.wikipedia.org/w/index.php?title=Wikipedia:What_Wikipedia_is_not&diff=next&oldid=295853
[27] http://en.wikipedia.org/wiki/Wikipedia:NOTNEWSPAPER#NEWSPAPER

> 4. ***A diary.*** *Even when an individual is notable, not all events he is involved in are. For example, news reporting about celebrities and sports figures can be very frequent and cover a lot of trivia, but using all these sources would lead to overdetailed articles that look like a diary. Not every match played, goal scored or hand shaken is notable enough to be included in the biography of a person.*

This policy, in turn, references another policy known as a notability guideline specific to events ("WP:EVENT"). This policy derives from the central notability guideline which tests whether a topic can have a dedicated article by asking "if a topic has received significant coverage in reliable sources that are independent of the subject, it is presumed to satisfy the inclusion criteria for a stand-alone article."[28] Discussions about notability and importance stem from early discussions on the Wikipedia mailing list, including debates about what to do with the biographies of "unimportant" people that lead to the spin-off of the 9/11 Memorial Wiki.[29] Following rancorous "inclusionist" versus "deletionist" debates about the scope of the content to be included in the project and developing consensus related to "WP:NOT" (Reagle Jr, 2010), efforts to write a policy specifically enumerating how to determine Wikipedia's encyclopedic notability began in September 2006.[30] Because different domains have different standards for warranting significant third-party coverage, specific notability guidelines were subsequently created for organizations, people, sports, numbers, music, film, books, and astronomical objects.[31] A notability guideline for news events was proposed in August 2008 following conflicting precedents on outcomes of article deletion debates over which news events are notable and promoted to full guideline status in December 2009 after a community consensus for its need.[32] The guideline has a five-pronged test to establish the notability of events:[33]

> 1. **Lasting effects**. An event that is a precedent or catalyst for something else of lasting significance is likely to be notable.
> 2. **Geographical scope**. Notable events usually have significant impact over a wide region, domain, or widespread societal group.

[28] http://en.wikipedia.org/wiki/Wikipedia:Notability
[29] http://lists.wikimedia.org/pipermail/wikipedia-l/2002-September/005458.html
[30] http://en.wikipedia.org/w/index.php?title=Wikipedia:Notability&oldid=74402065
[31] http://en.wikipedia.org/wiki/Category:Wikipedia_notability_guidelines
[32] http://en.wikipedia.org/wiki/Wikipedia_talk:Notability_(events)/Archive_2#Promotion_to_Guideline
[33] http://en.wikipedia.org/wiki/Wikipedia:Notability_(events)

3. **Depth of coverage**. An event must receive significant or in-depth coverage to be notable.
4. **Duration of coverage**. Notable events usually receive coverage beyond a relatively short news cycle.
5. **Diversity of sources**. Significant national or international coverage is usually expected for an event to be notable. Wide-ranging reporting tends to show significance, but sources that simply mirror or tend to follow other sources, or are under common control with other sources, are usually discounted.

The policy emphasizes that otherwise reliable sources' routine coverage of wedding announcements, obituaries, sports scores, crime blotters, or entertaining "man-bites-dog" events; sensationalist gossip or scandal mongering; and one-hit wonder people notable for only a single event generally do not meet the notability threshold and should be excluded. The policy also has guidelines specific to breaking news events:

If an event is still being widely covered in the press, editors may place the "currentevent" template on it to inform readers of the changing nature of the article.

It is wise to delay writing an article about a breaking news event until the significance of the event is clearer as early coverage may lack perspective and be subject to factual errors. Writing about breaking news may be recentism, and Wikipedia is not a crystal ball. It is recommended that editors start a section about the event within an existing article on a related topic if possible, which may later be split into its own article if the coverage suggests that the event is independently notable.

Many articles on events are created in anticipation of their notability. Anticipation is the creation of an article on a recent event with the expectation that it will meet inclusion guidelines, before the duration of coverage or any lasting effect is certain. For example, "June 22, 2009 Washington Metro train collision" was started just 60 minutes after the crash occurred. The rescue operation was still ongoing, an investigation was yet to begin, and the final death toll was unknown.

Anticipation of notability may be mistaken. Many events portrayed by the media as major on the day they occur quickly become only a footnote. For example, it was reported in January 2009 that a man was planning to travel to Washington to assassinate George W. Bush. It was reported several days later that he had no such plans and this event was shown to be nothing more than a routine arrest. However, articles about widely reported major unexpected or unprecedented events such as the 2004 Indian Ocean earthquake, the Assassination of Benazir Bhutto or the Death of Michael Jackson will almost certainly gain consensus to be kept even when created on the same day as the event occurred.

This test and related articulations of the policy have remarkable congruence with the news values proposed by Galtung and Ruge discussed in the previous chapter. "Lasting effects" is a type of threshold reflecting the intensity of an event, "geographical scope" is a combination of reference to elite nations and

meaningfulness, “depth of coverage” is a type of threshold and consonance, “duration of coverage” is an example of continuity, and “diversity of sources” is a type of imitation while the exclusion of routine events and anticipation of notability are examples of unexpectedness and consonance. The alignment of Wikipedia policies with extant journalistic schemas for establishing the notability of events as news is an example of a diffusion of standards. Neo-institutional theories of organizations suggest that organizations become more alike as they are exposed to similar rules, norms, and beliefs (DiMaggio & Powell, 1983). Organizations adopt values and standards from other domains and fields by transforming them into something that fits their local contexts. The diffusion of values, standards, and organizational forms can be driven both by organizations conforming to the rules and requirements of their environment as well the emulation of elites and other pioneers in the adoption of standards (Dahl & Hansen, 2006; Scott & Meyer, 1994). The demands and conflicts raised by the news work Wikipedians were engaging in made them receptive to the institutional standards news workers themselves employed which were seen to be more reliable and elite.

Wikipedia’s adoption of policies aligning with the standards and news values of journalists followed the development and adoption of other policies governing the types of content that warrant inclusion. The presence of memorial content related to the 9/11 attacks in particular created the precedent for other users to attempt to use the encyclopedia as a memorial. While the debates about how to deal with 9/11 Memorial content were on-going, in November 2004 user “MacGyberMagic” updated “WP:NOT” to reflect a new boundary:[34]

> *A memorial. It's always sad when people die, but Wikipedia is not the place to honor them. We're trying to build [an] encyclopedia. Of course, you're free to write articles about dead celebrities or other people with notable achievements. (9/11 victims should be entered on the 9/11 memorial wiki)*

[34] http://en.wikipedia.org/w/index.php?title=Wikipedia:What_Wikipedia_is_not&diff=7964872&oldid=7910671

This policy has since been absorbed into the "Wikipedia is not a blog, webspace provider, social network or memorial" thread within the "WP:NOT" policy and reads:[35]

> ***Memorials.*** *Subjects of encyclopedia articles must satisfy Wikipedia's notability requirements. Wikipedia is not the place to memorialize deceased friends, relatives, acquaintances, or others who do not meet such requirements. Note that this policy does not apply outside of the main article space. While using user space to create a memorial is generally not acceptable, limited exemption applies to the user space of established Wikipedians who have died. At a minimum it is expected that they were regular contributors, and that more than one tenured Wikipedian will have used the deceased user's page (or an appropriate sub-page) to add comments in the event, and after verification, of their death.*

The necessity of such a rule and consensus to adopt it reflected interesting ambiguities about the motivations of contributors, arrangement of technical capabilities and standards, and practices of analogous institutions. The web provides a medium for both the dissemination of information about major events as well a forum for discussion and social sharing of emotions in response to cultural traumas (Cohn, Mehl, & Pennebaker, 2004; Gortner & Pennebaker, 2003; Stone & Pennebaker, 2002). Wikipedia was a crucial component in this information sharing and sensemaking ecosystem and, moreover, explicitly encouraged users to participate in the construction of articles about news events.

The response of editors authoring memorial content and editors attempting to remove it are telling as they surface the implicit and negotiated order of the norms and standards of the community and an interesting break in the alignment of human interest news values with encyclopedic imperatives. Committed Wikipedians approached the issue of memorials to victims of news events through the lens of NPOV and notability policies: "if *Britannica* would not commemorate or eulogize the passing of this person, why should Wikipedia?" However, the authors of this memorial content had reason to believe it warranted inclusion given the importance assigned to it by otherwise reliable sources: "if CNN and ABC have content memorializing and commemorating these victims, why shouldn't Wikipedia?" This type of ambiguity necessitated the creation of the memorial policy effectively privileging the encyclopedic imperative over the commemorative imperative. This breach complicates any narrative emphasizing the

[35] http://en.wikipedia.org/wiki/Wikipedia:NOTMEMORIAL#MEMORIAL

convergence of encyclopedism with journalism as it highlights a bright line distinguishing the former as an ideal type of detached orientation from the latter as a coincident boundary between information dissemination and memorializing. Indeed, Wikipedians' reluctance to abandon their standards of neutrality and notability to engage in the memorialization found in major news outlets even comes full circle as a reflexive critique against journalism failing to adhere to the standards it sets for itself: "Wikipedians do a better job of neutrally sharing the real news than even journalists."

The emergence of these policies in response to the unintended precedent set by the 9/11 Memorial Wiki content reflects a particular kind of boundary work to expand the encyclopedic genre to include news-related content, exclude particular types of news content deemed to be unencyclopedic, and justification of this decision articulating these new boundaries through the values and standards of other elite institutions. Wikipedians, faced with similar decisions over which events warrant coverage in their work, emulated the standards and news values journalists had long employed. However these standards were contingent upon particular historical precedents and institutional configurations that granted journalists substantially more autonomy to simultaneously engage in commemoration and objective reporting. Facing equivocal institutional identities of encyclopedism and journalism without the authority apportioned to either, Wikipedians faced the irony of having their legitimacy undermined by the invisible moral and cultural arrangements of journalistic values they had appropriated because of their legitimacy.

Emergence and irrelevance of WikiNews

The success of breaking news coverage on Wikipedia as well as limitations imposed by the encyclopedic genre motivated members of the community to adapt Wikipedia's free-content, open collaboration model for news reporting. The first proposal came on January 5, 2003 on the Meta-Wiki and proposed creating a project having "news on a wide variety of subjects, unbiased and in detail."[36] By June 2004 a formal

[36] http://meta.wikimedia.org/wiki/Talk:Wikinews/Archive

discussion was started with the goal of starting the project and a demonstration wiki was launched in November 2004 to begin developing the technical and community infrastructure and the project was moved into a beta stage in December 2004 with multi-language editions eventually launching in more than 20 languages. WikiNews distinguished itself from collaborative or citizen journalism projects like OhmyNews by emphasizing the neutral point of view policy imported from Wikipedia, but unlike Wikipedia, it encouraged original reporting, interviews, and research. Major events such as the 2004 Indian Ocean earthquake and tsunami, 2005 Hurricane Katrina, and Iraq War provided the early impetuses for growth on the project, growing to 1,000 articles by March 2005.

The early discussions about establishing a WikiNews project wavered between concerns that the project would be a fork and distraction from Wikipedia's existing coverage of current events and a poor imitation of existing and highly institutionalized approaches to news coverage versus assurances the project would create an outlet and community for editors whose contributions are rebuffed because of Wikipedia's policies and would appeal to a larger set of the population who regularly consume news. Advocates believed editors already contributing to current events articles on Wikipedia would join the WikiNews project, the ease and style of contributing on Wikipedia would be easily adapted, Wikipedia's NPOV policy would broaden the pool of readers and contributors, and the lack of space limitations would permit broad topical coverage as well as deep investigative reporting with in-line referencing of source material. But important differences also existed. While some proponents envisioned a project serving as a secondary source curating a chronological index of articles, images, and other content elsewhere on the web, the majority backed the idea of a project where "WikiReporters" provided primary source material that could be synthesized and integrated with other accounts.

Like Wikipedia, WikiNews employs a host of rules and stylistic guidelines emphasizing the need for content to be sourced, to employ a neutral point of view, to not infringe on the copyright of other works, and to use a "news" style. Unlike Wikipedia, WikiNews also permits editors to engage in "Original

Reporting" by conducting and publishing interviews, eyewitness accounts, taking pictures, and abiding by a code of ethics. Most significantly, WikiNews articles are tagged as "In Development" until they have the appropriate levels of information, citations, and stylistic coherence. After editors are done developing an article, they can flag it for review and "hopefully a reviewer will review [the] article in short order."[37] Reviewers ensure that articles do not have copyright infringement, meet newsworthiness guidelines, have verifiable sources, employs a neutral point of view, and abide by the style guidelines before flagging the article for publication to the WikiNews homepage. This is a significant departure from Wikipedia's model where the vast majority of articles neither require formal vetting nor differentiate the status of articles as being "final" or "in development." This is significant because attempts to edit a published article on WikiNews are met with the warning in Figure 9 while no corresponding warnings exist for Wikipedia articles that are featured on its homepage. Emphasizing that only minor changes should be made, no new sources introduced, and major new developments require the creation of an entirely new article substantially diverges from the Wikipedia model which invites contributions of almost any type on any article at any time.

Because of this shift to primary accounts, concerns about establishing the reputation of and vetting the reliability of WikiReporters' contributions focused on restricting access to users who had demonstrated credentials, real identity, or trusted community status, protecting content from modification, and employing peer review processes to vet the content. Ironically, these discussions tracked many of the same concerns and design decisions in early discussions between Larry Sanger and Jimmy Wales as they developed Nupedia all of which were of course abandoned in favor of Wikipedia's radically open collaboration. WikiNews ultimately adopted an "anyone-can-edit" model analogous to Wikipedia, but rather than developing a new genre of news work particular to the participatory, high tempo, and hyperlinked online context, proponents set their sights too low and attempted to emulate the variety of

[37] http://en.wikinews.org/wiki/Wikinews:Writing_an_article, http://en.wikinews.org/wiki/Wikinews:Reviewing_articles

news genres such as reports, summaries, and analyses as well as editorial systems that bottlenecked the publication of news events. New developments could not be integrated into existing accounts but had to be spun off and developed from scratch. Of course, it is impossible to know *a priori* whether breaking news developments will require deep and sustained attention. This creates a chicken-and-egg problem in which existing articles forgo coverage of the new development out of deference to explicit rules that this content should only appear in follow-on coverage but at the same time the creation of new articles lags until more information comes to light and editors with the motivation to begin writing an entirely new article appear. As a result, WikiNews articles almost always contain stale information.

This emphasis on becoming a primary source for news was a significant departure from Wikipedia's existing norms centering on reliable secondary and tertiary sources, notability of people and events, and low barriers to participation. This departure was important to differentiate it from the work already being done on Wikipedia and proponents embarked on a familiar pattern of boundary work to expand the domain of open collaboration to journalism, exclude competing agents such as news espousing non-neutral points of view, and justifying this arrangement through the patterning institutional and technical infrastructures to establish reliability and authority. However, this boundary work introduced major path dependencies on the pool of readers and contributors and the types of contributions and content both of which erected barriers to participation from existing Wikipedians, failed to attract an audience or community of dedicated users, complicated attempts to collaboratively write stories in a coherent narrative voice, and often delayed the publication of events until after their newsworthiness had passed.

The site failed to thrive in comparison to the explosion of traffic, contributions, and new editors on Wikipedia at the same time (see Figure 4-6). Throughout this era, Wikipedia had orders of magnitude more editors and content than its sister project. WikiNews generated only 15 new articles a day at its peak in mid-2005 (in comparison to an average of approximately 1,500 new articles. Wikipedia at the same time) and by 2008 the average number of new articles per day was fewer than 10 and falling further

below 5 by 2011. In September 2011, a large portion of the English contributor community departed to join a fork called OpenGlobe that will employ a more lax publication policy without the detailed checks for copyright, neutrality, and reliability violations present on WikiNews.

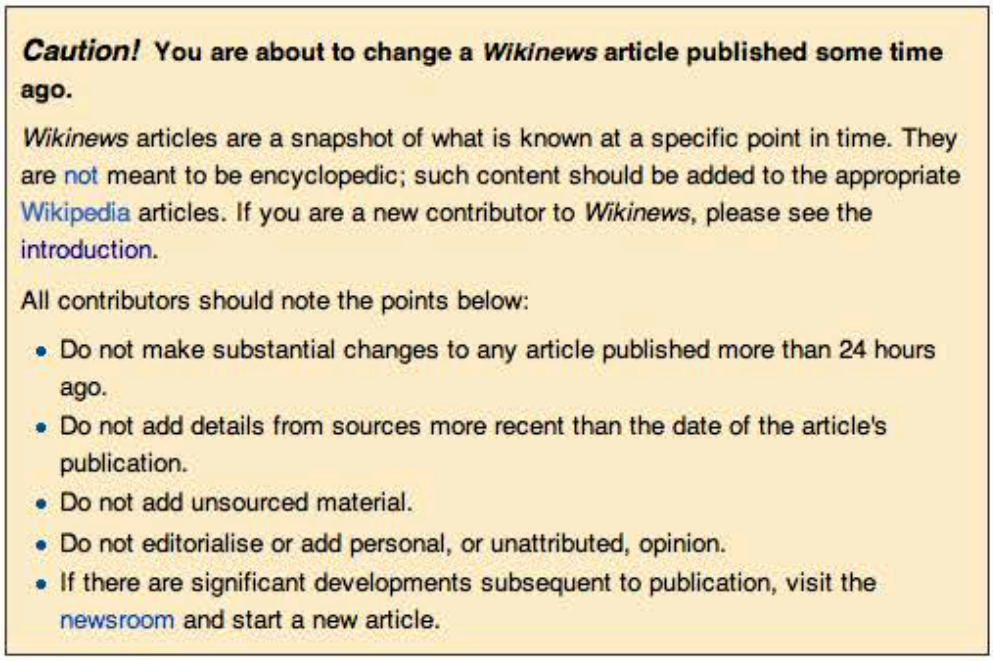


Figure 4-5: WikiNews editing warning for published articles.

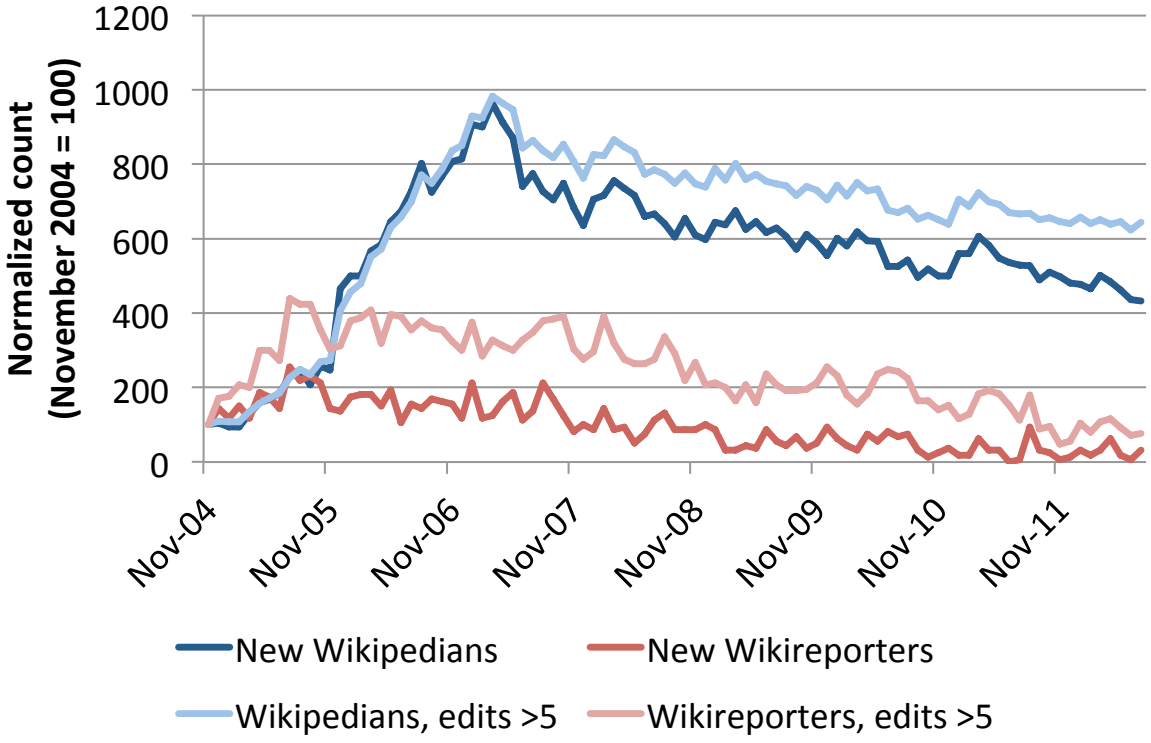


Figure 4-6: Comparison of growth between English Wikipedia and English Wikinews

The reasons for the failure of WikiNews relative to the success of Wikipedia are multifaceted. The resource dependency between the project and the environment is important. As Jonathan Dee noted, “So indistinct has the line between past and present become that Wikipedia has inadvertently all but strangled one of its sister projects…Wikinews. …On bigger stories there's just no point in competing with the ruthless purview of the encyclopedia" (Dee, 2007). Wikipedia received far more web traffic that resulted in a much larger pool of potential contributors as well as greater visibility for the content that in turn can motivate many users. Wikipedia also had a far larger editor base meaning there are more editors with the skill and experience that can be translated from other domains to writing news articles. While WikiNews linked extensively in the body of the text to Wikipedia articles to provide contextual details, links on Wikipedia articles to corresponding WikiNews articles are rare and require specialized formatting. Furthermore, many categories of encyclopedia articles about events such as earthquakes and hurricanes are not only highly formulaic but had highly-specialized sub-communities of dedicated editors whose experience with the style, sources, and vocabulary of writing articles about historical events readily translates to writing about breaking news articles about the same.

The normative style of writing news articles is also very distinct from writing encyclopedia articles. Whereas news stories emphasize parallel and iterative accounts of the latest developments on a single topic that often lack context, encyclopedia articles normatively require authoritative and integrated accounts. As a result, while WikiNews may invite potential editors to contribute to any of several dozen articles about the impacts of a major event such as the 2004 Indian Ocean earthquake,[38] Wikipedia offers a central and up-to-date page containing a complete and contextualized synopsis rather than a collection of chronological and static accounts. The articles promoted to the WikiNews homepage are also interesting because they do not resemble the stories being covered in other news outlets. For example, headlines appearing on the homepage in mid-August 2012 include “Reform Party of the United States

[38] http://en.wikinews.org/wiki/Portal:2004_Indian_Ocean_Tsunami

nominates fitness model Andre Barnett for president", "Students from Liceo Maria Luisa Bombal of Rancagua, Chile detained after taking control of school", and "Cities across Texas increase efforts to combat mosquitos."[39] Part of this reflects the adoption of unique news values attempting to appeal simultaneously to international and local audiences, but also is emblematic of the hazards of a strong peer review gatekeeping model also seen in Nupedia: content ends up reflecting the interests of editors who possess the motivation and skill to navigate the system, even if the topics are obviously provincial, rather than content meeting the demands of information seeking readers.

As large and complex as Wikipedia or WikiNews are, they are still embedded within other larger socio-technical structures of web use such as technical standards and search engines. This interaction of journalistic norms with technical features of these wiki-based projects had a profound impact. Former Google vice president Marissa Mayer noted that journalist norms emphasizing frequent and parallel publishing of content on the same topic results in web pages competing against each other in terms of authority and placement in search results. She asked, "how [might] the authoritativeness of news articles grow if an evolving story were published under a permanent, single URL as a living, changing, updating entity?"[40] Whereas WikiNews sought to emulate the traditional nature of journalistic storytelling emphasizing iterative, parallel, and static reports of an event which undermined its ability to develop a critical mass of effort or attention on a topic, Wikipedia's insistence on a single but repeatedly updated account boosted it to the top of search results. Moreover, the way algorithms like Google's PageRank assign importance to webpages meant that breaking articles on Wikipedia would inherit authority from other Wikipedia articles which are themselves highly ranked by Google. Differences between Wikipedia's and WikiNews's publication models thus have profound impacts on how users search, edit, and manage content. Wikipedia's homepage, current events page, in the news templates, and other highly-valued articles were updated to point to a single breaking news article, this article would inherit their

[39] http://en.wikinews.org/w/index.php?title=Main_Page&oldid=1519685
[40] http://www.niemanlab.org/2009/06/google-news-experimenting-with-links-to-wikipedia-on-its-homepage/

authority and quickly rise to the top of search rankings, creating a virtuous feedback loop which would drive more traffic, contributors, updates, and content to the article, effectively making it the clearinghouse on the web for information about an event. In contrast, WikiNews invited users to create multiple and parallel accounts of events, diluting attention and focus among readers seeking up-to-date information, contributors seeking collaborators with which to work, algorithms assigning importance and relevance for the larger web audience, and administrators trying to feature high-quality work.

Discussion

The 9/11 Memorial Wiki as well as the WikiNews projects are illustrative of the perils the vast majority of online communities face in acquiring a critical user base, developing a coherent identity, sustaining engagement as well as how the success of Wikipedia is truly rare and exceptional. The case of Wikipedia's embrace then rejection of 9/11 Memorial Wiki has important implications for archivists, historians, and media theorists as it vividly highlights how the ephemerality of social tastes can undercut the willingness to provide resources or infrastructure to maintain technologies. This suggests that digital and previously popular knowledge artifacts no more durable than the paper they replace. Wikipedia's adoption of rules governing the type and extent of news information to be incorporated reflects a balancing act between integrating motivated users' contributions and preserving a coherent organizational identity focused on encyclopedic work. The attempts to shoehorn norms and practices of journalism into an open peer-production system for news replicated profound limitations in the practice of traditional journalism while ignoring new ways that news was being consumed. As Bowker and Star (2000) argue:

> "…all information systems are necessarily suffused with ethical and political values, modulated by local administrative procedures. These systems are active creators of categories in the world as well as simulators of existing categories."

As socio-technical systems, wikis offer technological possibilities for knowledge collaboration and dissemination. But these technologies are also socially constructed and embedded within larger social, institutional, and cultural contexts that can emphasize notable and neutral encyclopedic content while delimiting alternatives like creating a space documenting cultural trauma and commemoration or emulation of prevailing practices of journalism. The frailties of categories imported into new contexts (incorporating news information into encyclopedias) can be opportunities for innovation while others (using wikis to reproduce prevailing approaches to news production) reveal the limitations of institutionalized practices and schemas.

The saga of the 9/11 Memorial Wiki, like the attempts by encyclopedists throughout history to balance epistemic authority with contemporary relevance and accuracy, is illustrative of the perils of classification work. The wiki provided a venue that allowed users with disparate motivations, dispositions, and membership in communities of practice to cooperate in documenting an event of profound historical importance and cultural trauma. However this cooperation proceeded without an agreement about the role or boundary of this new type of encyclopedia and the Wiki itself became a boundary object onto which various communities of practice projected distinct categories and meanings: current events-related content was unnatural to participants attempting to reproduce the “traditional” encyclopedic genre, 9/11 remembrance content was unnatural to participants attempting to reproduce objectivity and neutrality emblematic of either journalistic or encyclopedic accounts, and the unnatural lack of material limitations in the technology itself validated attempts from participants like “The Cunctator” to include detailed accounts and memorials of the event. The case of how encyclopedists from Diderot through Wikipedia negotiate new knowledge reveals tensions in how boundaries and classifications are constructed, reinforced, crossed, or dissolved in knowledge production as well as the dynamics by which socio-technical systems not only enable new forms of knowledge collaboration and dissemination, but the collapse and expansion of prevailing boundaries.

Conflicts over whether Wikipedia would be an encyclopedia covering only historical events or would reflect new knowledge about on-going events are certainly not new, as I discussed in Chapter 3. However, the development of new technologies that permitted the rapid authorship and dissemination of new content demanded a rapid re-appraisal of encyclopedic identity. Editors attempted to "naturalize" the actions necessary for maintaining and recreating the meaning of this new and highly flexible way of knowledge collaboration so that they would no longer seem strange or unexpected, but it required reconciling views from distinct domains (Bowker & Star, 2000). On one hand, the Wiki provided a forum to collaborate and rapidly disseminate up-to-date knowledge while providing rich contextual details and background. This would require reconciling encyclopedic authority with journalistic timeliness, distinct traditions which nevertheless share complimentary ideologies and parallel historical trajectories. On the other hand, the Wiki provided a commons to mediate, store, and share the outpouring of emotion, collective sensemaking, and other deeply-ingrained social behaviors following traumatic events (Cohn et al., 2004; Pennebaker & Harber, 1993; Solnit, 2009). This would require reconciling encyclopedism with commemoration, also distinct traditions but lacking the ideological alignment and parallel history of the former. It is then unsurprising that the domain with the more tenuous relationship with encyclopedism was the first discarded; Wikipedians naturalized the alignment of encyclopedism with journalistic objectivity more readily than they did with "journalistic" commemoration.

This alignment was enacted through various types of work such as Gieryn's "boundary work". The simultaneous imperatives for authority and relevance required Wikipedia editors to engage in boundary work to expand the possibilities of encyclopedic coverage to include and promote current events, exclude content and events which are insufficiently encyclopedic, and defend these choices by normalizing the type and scope of content to be included through formal policies and regulations. Early Wikipedians, seeing the traffic and contributions for 9/11-related content, saw the potential to expand the scope of encyclopedic action. However, this expansion introduced anomalies to expected categories such like

"non-notable victim", "unencyclopedic memorial", or "precedent for future catastrophes" which created problematic residual categories. Managing these anomalies arising from the expansion of the encyclopedia into current events coverage required Wikipedians to converge on some shared meanings about the types of content which warrant inclusion, exclude other alternative interpretations, and legitimize these distinctions by institutionalizing and adopting policies enumerating appropriate types of content. But the existence and influence of these types of work influencing the types of knowledge which are preserved and valued also suggests the need to maintain a critical eye towards socio-technical systems as this type of boundary work is clearly non-neutral and advances the views and interests of some to the detriment of others (Lanier, 2011; Van Dijck & Nieborg, 2009).

The case of WikiNews also provides an essential counter-factual lest technologically determinist thinking override a critical appraisal of why Wikipedia succeeded in adapting peer production to news production. Significantly, both projects employed fundamentally the same technical infrastructures of MediaWiki software and received similar operational and promotional support from the Wikimedia Foundation. However, the creators of WikiNews attempted to proscribe a distinct style of content and collaboration emulating professional Western journalism's periodic and parallel narratives rather than an on-going and coherent style found on Wikipedia. Moreover, the model for developing this content required a formal vetting process from other credentialed users and restricting major revisions that diverged significantly from Wikipedia's more *lassiez-faire* attitude for covering breaking news events by incorporating most users' changes to the main articles. These design decisions focusing on expanding the practice of news production however failed to consider the ways news consumption were being re-shaped by new forms of web use such as search engines and other forms of social media challenging the forms of journalism WikiNews sought to emulate.

This shift by Wikipedians to erect boundaries about the types of knowledge and information that warrant inclusion bears a remarkable resemblance to the boundary work journalists employed through history.

New socio-technical systems like Wikipedia have profoundly transformed the institutional arrangements governing access to and means of disseminating information that has in turn problematized existing classifications over what is or is not encyclopedic and journalism. The process by which current events coverage became naturalized within "encyclopedism" can both be interpreted as a part of a long historical process balancing competing tensions from journalism as well as an abrupt shift accelerated by disruptive new technologies. However, the fact that current events coverage has become naturalized within Wikipedia is difficult to dispute. Since 2003, the top 25 Wikipedia articles with the most contributors every month consists nearly exclusively of articles pertinent to current events (Wikipedia, 2011). For example, the articles which attracted the most contributors in February 2011 included "2011 Egyptian revolution", "Super Bowl XLV", and "2011 Christchurch earthquake" (Wikipedia, 2011). Similarly, articles receiving the most unique edits (Wood, 2011) and page views (Mituzas, 2011) in any given week or month likewise demonstrate a substantial bias toward articles about current events. Furthermore, Wikipedia is increasingly framed by traditional journalists as a neutral and reliable information source (Messner & South, 2011).

Wikipedia's coverage of current events is not only ordinary and routine, but it has inverted our expectations about the timeliness and relevance of reference works. Where it had previously been natural to assume contemporary reference works would not reflect late-breaking news events, it has instead become *unnatural* to find a Wikipedia article—or perhaps contemporary reference works of any kind—to be out-of-date because it highlights the circumstances of its production and the limitations of alternative forms of knowledge dissemination. This shift is made all the more poignant by *Britannica*'s announcement in April 2012 that it will be ceasing the publication of its print edition after 244 years to re-focus its efforts on its online encyclopedia.[41] The standard form or ideal type of an encyclopedia has become entirely digitized. The negotiation of boundaries over the types of content to be included, debates

[41] http://schoolsofthought.blogs.cnn.com/2012/03/14/encyclopedia-britannica-to-stop-printing-books/ ; http://mediadecoder.blogs.nytimes.com/2012/03/13/after-244-years-encyclopaedia-britannica-stops-the-presses/

over approaches to removing content violating these boundaries, and the very acceptance of current events-related content into the encyclopedia are examples of the messy and contingent work obscured and forgotten as this online and perpetually up-to-date encyclopedia has become naturalized. The remainder of the dissertation will unpack this naturalization and examine whether collaborations authoring the content about breaking news events are governed by similar structural arrangements as typical Wikipedia articles, the extent to which contributors to breaking news articles the exhibit characteristics of specialization analogous to news workers, the boundaries which are crossed as communities of users collaborate on these articles, and the extent to which these processes have changed over time.

Chapter 5: Coauthorship Network Analysis

The previous chapters traced the emergence of role specialization as a response of journalists and encyclopedists orienting themselves to coordinating increasingly complex and timely information for inclusion in their products. These changes were not only organizational, but as Chapter 4 outlined, they were normative as news values diffused from journalists to Wikipedia's decisions about the boundaries between news events warranting inclusion versions. However, the case of the 9/11 Memorial Wiki covered in Chapter 4 was unique because of its scale of its impact on knowledge claims as well as Wikipedia's lack of organizational and normative precedent to act on it. Were the patterns of Wikipedia's co-authors coordinating responses to other breaking news events found across all these breaking news articles while being distinct from "typical" Wikipedia articles? The coordination demands of high-tempo online collaboration around these breaking news events are rarely found on other articles and this should lead breaking articles' collaborations to exhibit structural features distinct from other Wikipedia collaborations. This chapter[42] examines the structure of many breaking news article collaborations on Wikipedia and how the structure and evolution of these breaking articles' coauthorship networks resemble or diverge from the structure of non-breaking news articles' coauthorship networks. The results suggests that breaking article collaborations have distinctive structural patterns when they are young and intensively edited, but these distinctions vanish in the long run as both types of articles converge to similar patterns of collaboration.

Background and approach

As is the case with many online communities, the majority of contributions to Wikipedia come from a fraction of the entire user base (Kittur et al., 2007; Panciera et al., 2009). Previous studies of coordination in Wikipedia suggest implicit and informal coordination mechanisms can support article development

[42] This chapter is forthcoming in *American Behavioral Scientist*, co-authored with Darren Gergle and Noshir Contractor.

when the article is young and intensively worked upon by a dense cohort of authors. As an article ages, a shift towards more explicit and formal coordination mechanisms such as discussion is needed for continued improvement in article quality (A Kittur & R Kraut, 2008; Kittur et al., 2009). However, scholarship has only begun to examine how Wikipedians engage in high-tempo knowledge collaboration following unexpected and surprising events. Dynamic network analysis of articles about current events reveals distinct types of collaboration characterized by narrow or broad participation (Iba et al., 2010). The revision histories of Wikipedia articles about current news events are sites of collective memory, sensemaking, and commemoration (Ferron & Massa, 2011a, 2011c; Keegan, 2011). Decisions about which news items to promote to Wikipedia's front page rely on egalitarian gatekeeping in which expert opinions are alternatively embraced and dismissed depending on the context (Keegan & Gergle, 2010). Previous studies of the structures or dynamics of Wikipedia breaking news coauthorship have only examined a single category of articles limiting the generalizability of their findings (Keegan et al., 2011; Keegan, Gergle, & Contractor, 2012). This study substantially expands the sample over prior studies by analyzing a corpus of 3,233 Wikipedia articles across a much wider range of genres for breaking and non-breaking events since 2001 to compare the structure and dynamics of both breaking and non-breaking articles.

In many online collaborations and communities, the majority of activity is centralized in a handful of users (Butler et al., 2002; Wilkinson, 2008). Teams operating under conditions of uncertainty and complexity are more effective when employing decentralized interaction patterns as this promotes more contributions and reduces the risk of overload for central integrators. However, teams operating under conditions of threat and time pressure are more effective when employing centralized interaction patterns as this promotes greater control and simpler information processing schemas. Prior work suggests the interactions between these constraints are complex and their influences on team performance are complicated (Argote, Turner, & Fichman, 1989; T. M. Brown & Miller, 2000). Unlike typical Wikipedia

articles, breaking news articles place demands on users to collaborate in contexts that are simultaneously uncertain as well as time-sensitive which demands editors adopt implicit coordination mechanisms that lead to either greater concentration or distribution of work across editors (A Kittur & RE Kraut, 2008; Kittur et al., 2009). Because of the differences in the coordination demands of breaking news articles versus typical Wikipedia articles, (1) the concentration/distribution of editing activity on Wikipedia's breaking articles will differ significantly from typical articles and (2) this concentration/distribution of work will be consistent across events occurring in different years. The metrics for measuring this concentration are discussed in further detail in the methods section below.

H1a: Breaking news articles will exhibit concentrations of editing activity that are significantly different from non-breaking articles' concentrations.

H1b: Breaking news articles' concentrations of editing activity will be consistent for events occurring in different years.

Organizational theory has grappled with coordination and self-organization in temporary teams by examining the extent to which participants regenerate, adapt, and improvise roles and routines used in previous projects and collaborations (Bechky, 2006; Bechky & Okhuysen, 2011; Birnholtz, Cohen, & Hoch, 2007; Pentland, Hærem, & Hillison, 2011). Because the re-enactment of structures among a cohort of regular participants may contribute to high performance of the collective (Schwab & Miner, 2008), coauthorship of breaking articles on Wikipedia may not be "one-off" but rather involve editors who have repeatedly worked together or even specialize in editing content about breaking articles. Wikipedia's ability to cover breaking news events may reflect the regeneration of organizational forms capable of coordinating high-tempo work through the self-organization of editors who have previously participated in breaking news collaborations. The presence of some of the same editors across collaborations suggests they have greater capacity for knowledge coordination such as group mind to manage routines and

delegate tasks, credibility to manage conflicts, expertise to adapt processes and information in high-tempo work. If breaking article collaborations adapt to task demands by relying on contributions from editors who have previously contributed to other breaking articles, they should exhibit a greater tendency for repeated participation among editors for incidents occurring in different years.

H2: Breaking news articles will have more editors who previously contributed to other breaking news articles than non-breaking news articles' editors contributing to other non-breaking articles.

The structure and regeneration of these collaborations may not be stable but change over time. Understanding these dynamics can provide insight into how large and connected networks balance access to diverse and novel information with the costs of reduced information flow and heightened coordination demands (Aral & Alstyne, 2011). The heterogeneity of connections within large and connected networks provides for some individuals with highly central positions, others with highly embedded and clustered neighborhoods, and still other boundary spanners who broker connections. Each of these positions plays a distinct role representing users possessing special resources, strong ties for action, and connections across diverse groups and all are pre-requisites for high levels of organizational performance (Burt, 2004). The success of organizations operating in high-tempo contexts necessarily depends on their ability to rapidly assemble the resources and create the network of communication and resource-sharing ties to respond to the event (Faraj & Xiao, 2006). The coordination demands which give rise to the concentration/distribution of activity on breaking news articles (H1) combined with the regeneration of these organizational membership across collaborations (H2) suggests the networks of breaking news articles and their editors should exhibit different patterns of connectivity and shared coauthorship over time compared to non-breaking or historical articles.

H3: The distribution of activity on breaking article collaborations will differ significantly from non-breaking

articles' distributions in their early histories.

Data and methods

Seven broad categories of Wikipedia articles likely to include breaking news events were identified including conflicts (e.g., wars, battles, political unrest), crimes (e.g., murders, kidnappings, and terrorism), fires (e.g., building fires, wildfires, and explosions), health disasters (e.g., disease outbreaks), industrial accidents (e.g., spills, mine collapses), natural disasters (e.g., hurricanes, earthquakes, and tornadoes), and transportation accidents (e.g., airplane crashes, train collisions, road accidents). These categories are also categorized by incident year such that the Tōhoku earthquake and tsunami will appear in the "2011 earthquakes" category, reflecting the fact that the incident itself occurred in 2011 even if the first revision to the article occurred at some subsequent time.

Using the English Wikipedia's Application Programming Interface (API),[43] the revision histories for every article in these categories and their subcategories were extracted in January 2012. Owing to their scale, salience, and impact, some events such as the wars in Iraq and Afghanistan, events in the 2011 Arab Spring, 2011 Japanese earthquake and tsunami, 2010 Haiti earthquake, and 2010 Deepwater Horizon explosion and spill had dedicated sub-categories containing dozens of related and daughter articles. As these subcategories included articles and lists about people, places, and events of relatively minor importance or limited similarity to the parent event itself (e.g., political leaders, recording artists releasing benefit CDs, non-profit organizations), these sub-articles were manually identified and removed to create a list of 3,233 articles focused on the events and incidents alone. This list contains 195,831 unique editor-article interactions from 114,153 unique users over a period of time from September 2001 to January 2012.

[43] http://en.wikipedia.org/w/api.php

Articles are categorized by the year in which the event occurred and revision history data includes editor name and ID (or IP address), article name and ID, and timestamp. Based on these data, I extracted article-level attributes for statistical analysis such as whether or not the article is breaking or non-breaking (described below). To construct co-authorship networks, the revision histories for articles in a given year are converted to weighted bipartite edgelists of editor ID and article ID. A single edge represents the number of times a single Wikipedia editor made contributions to a single article. I summarize basic descriptive networks in Table 3.

Breaking News Classification

To capture the extent to which the structures of high-tempo collaborations in the aftermath of breaking news events differ from traditional Wikipedia collaborations, I identify three classes of articles for our analysis: breaking articles, non-breaking articles, and historical articles. Breaking and non-breaking articles are about events that are contemporaneous with Wikipedia's existence since January 2001 while historical articles are about events between January 1990 and January 2001. Breaking and non-breaking articles are categorized to differentiate the temporal proximity between the article's creation and the date of the incident itself. For events such as a plane crash, terrorist attack, or earthquake, the time and date when the event began and ended are effectively instantaneous even if the event has on-going consequences. However, political unrest, hurricanes, and epidemics are examples of non-instantaneous events often with uncertain inceptions and cessations but with consequences that may not emerge and may continue to unfold over days, weeks, or months. For these events, the date associated with the cessation of the event was used for breaking and non-breaking classification. Examples of breaking news articles are given in Table 5-1.

Computing the difference between the date of the first edit and the date of the event itself, I observe the distribution of article creation lags plotted in Figure 5-1. Negative values of article creation lag are an artifact of the coding for non-instantaneous events (described above) in which the first edit to the article

occurred before the incident ended. Taking a 1-day article creation lag to be the cut-off that differentiates breaking articles from non-breaking articles 1,212 breaking articles and 2,074 non-breaking articles were observed and the distribution of breaking, same year, and non-breaking articles plotted in Figure 5-1.

Coauthorship network evolution

The structures of these collaborations are highly dynamic and change substantially over time. To capture the evolution of these collaborations at the level of the article, the edit histories of each article were transformed to make the first edit to each article $t = 0$. Following this alignment, the extent to which the coauthorship networks of users editing articles are similarly structured at the same relative time in each article's revision history was examined. This process is illustrated in Figure 5-2 which contains example timelines for two breaking and two non-breaking articles. Here, breaking articles have a greater concentration of activity in the early stages and non-breaking articles have a more even distribution of activity throughout time. Because the distribution of editor revisions in breaking articles are strongly left-skewed, the network is thresholded at 13 quasi-logarithmic time points. These thresholds correspond to one hour, twelve hours, one day, two days, four days, one week, two weeks, one month, one quarter, one year, two years, four years, and one decade after the first revision to an article and are used below to query the network structures at different points in time described below.

Two approaches were adopted to model changes in network structure over time. The first is a *cumulative network* which aggregates all edits made by all editors to all articles before a given time threshold. In Figure 5-2 for example, it would provide a representation of all previous contributions to a set of articles up to the third threshold line but exclude any revisions after the threshold. This approach effectively captures when the network began to resemble the network when all the data are included. The second approach is a *snapshot network* which only looks at edits that occurred *between* adjacent time thresholds. In Figure 5-2, it would include revisions made between the second and third threshold lines but exclude any other revisions that occur beforehand or afterwards. Thus, the snapshot network provides a window

into the structure of the collaboration shifts at different points throughout the history of an article category. This approach would capture differences in the structure of the network at different points in time while censoring data from earlier points in time that may wash out important changes in the network.

2001	Enron, American Airlines Flight 587
2002	UK firefighter dispute, 2002-2003
2003	Space Shuttle Columbia disaster, Northeast USA blackout of 2003
2004	Second Battle of Fallujah, 2004 Madrid train bombings
2005	7 July 2005 London bombings, Hurricane Katrina
2006	2006 Israel-Gaza conflict, Comair Flight 191
2007	2007 United Kingdom floods, Writers Guild of America strike
2008	2008 Tibetan unrest, 2008 South Ossetia war
2009	2009 flu pandemic, Air France Flight 447
2010	Deepwater Horizon explosion, 2010 Haiti earthquake
2011	2011 Mumbai bombings, Tōhoku earthquake and tsunami

Table 5-1: Examples of breaking news articles from each year.

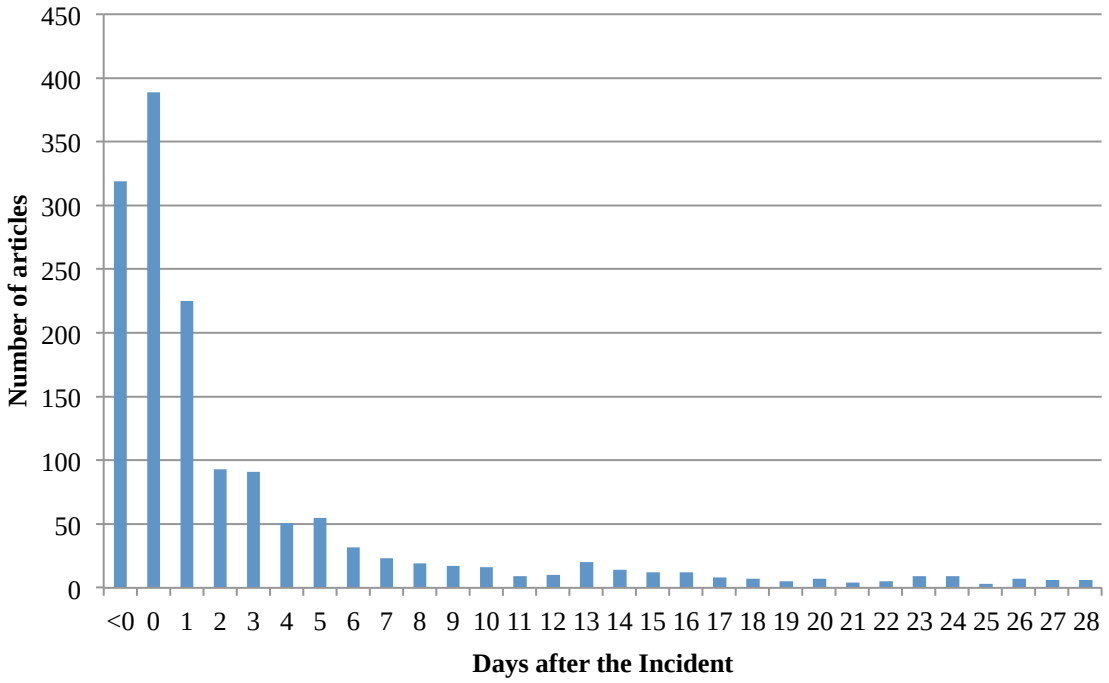


Figure 5-1: Distribution of article creation lags for articles with first revisions occurring the same year as the incident itself. Most articles' first revisions occur less than 1 day after the incident which I classify as "breaking news articles."

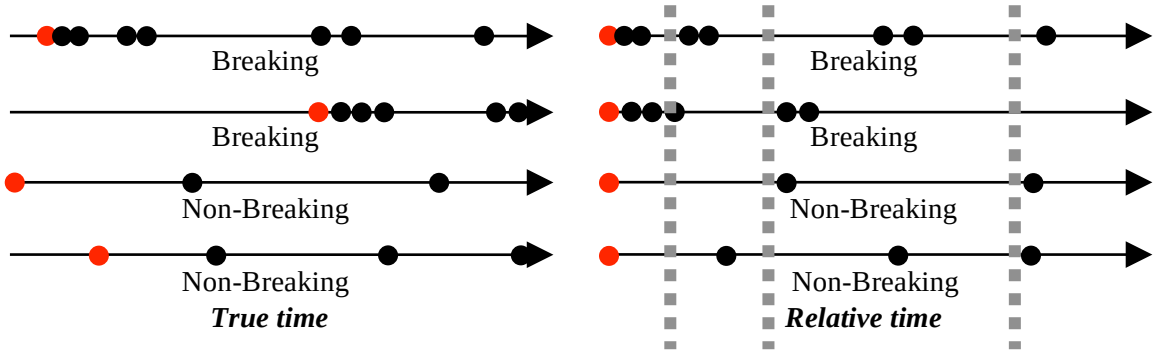


Figure 5-2: Schematic representation of edit history transformation and alignment for two breaking articles and two non-breaking articles.

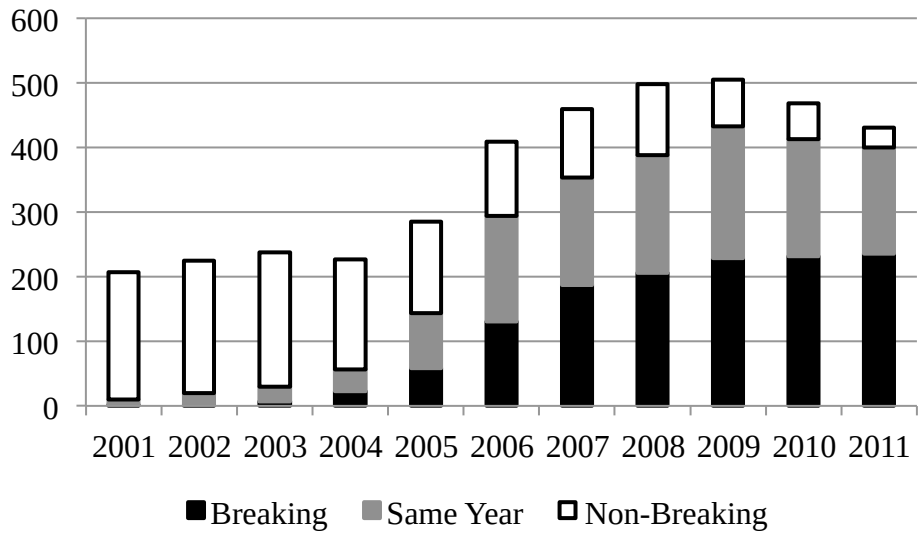


Figure 5-3: Frequency of article types per event year.

Results

64,272 unique editors had 82,254 distinct connections to 1,034 breaking articles and 61,571 unique editors had 113,577 distinct connections to 2,159 non-breaking articles. Figure 3 plots the frequency of different article breaking types by event year. First, an increase in the absolute number of articles is observed from 207 articles in 2001 to 456 articles in 2011 with a peak of 527 in 2009. Non-breaking articles about 2011 events are necessarily right-censored as the data was collected in late January 2012. While the number of disasters, accidents, and other catastrophic events certainly has not tripled around the globe in recent years, this article growth reflects both the increasing popularity of Wikipedia from 2001 through 2007 but also the growth of articles about current events despite slowing article creation and editor participation after 2007. This trend points to the growing role Wikipedia plays as a global memory place for documenting, encoding, and commemorating collective memories of traumatic events (Ferron & Massa, 2011c; Pentzold, 2009). Second, a distinct shift in the distribution in of breaking versus non-breaking articles is observed. Breaking articles made up less than 25% of all articles written about events in a given year from 2001 to 2005. In 2006, 48% of articles about events that year were written as breaking articles and in 2010 and 2011 that percentage had exceeded 70%. Removing breaking and same year articles, the trend remains: non-breaking articles decreased from 197 in 2001 (95%) to 41 in 2011 (8%). Using the 1-day lag to differentiate breaking from non-breaking articles, same year and non-breaking articles are aggregated as non-breaking articles in subsequent analyses. Structural similarity

Descriptive network statistics for the entire network are summarized in Table 5-2. The number of nodes for breaking and non-breaking articles reproduces the observations from Figure 5-3. There is a marked increase of contributions from unique editors on breaking articles while non-breaking articles exhibit substantially more stability in the number of editors over time. Both of these factors influence how the density of the collaboration networks change over time with breaking article collaborations becoming more sparse as they grow larger and non-breaking articles remaining stable across time. Breaking article

collaborations were initially much more dense than non-breaking articles suggesting less distribution of work, but collaborations around breaking articles in later years exhibit similar densities to non-breaking articles. Every single revision to breaking articles ends up in a single giant component for each year meaning these collaborations are all connected with each other through shared editorship despite being about vastly different topics. This contrasts to some extent with the non-breaking articles where a small amount of activity involving contributors editing articles no one else edits occurs in some years. Clustering coefficients for editors and articles capture the extent to which editors share articles in common or articles share editors in common, respectively (Latapy, Magnien, & Vecchio, 2008). The growth of the breaking article editorship over time has resulted in a decrease in editor "familiarity" as they are less likely to share breaking articles in common in later years. Conversely, non-breaking articles' editors have stable clustering patterns over time, a feature which is further unpacked in Figure 6. Non-breaking articles exhibit a greater tendency for articles to have editors in common than breaking articles. The lack of clustering suggests breaking articles exhibit less distributed editing activity than non-breaking articles.

In Figure 5-4 through Figure 5-9, the degree and weight distributions are plotted for the editors and articles for events occurring in each year. In both breaking and non-breaking articles, these distributions exhibit classic long-tail distributions. The fact that hundreds of editors edit only a single article while a handful of editors edit most or all of the articles is an example of *editor centralization*. Likewise, the observation that dozens of articles have only a few editors but a handful of articles have hundreds of editors is an example of *article centralization*. In the upper left of editor degree distributions (Figure 5-4, Figure 5-7), the top-ranked user edited nearly every article in a given year while in the lower right hundreds of users editing only a single article. Non-breaking articles in different years exhibit remarkable similarity in editor centralization. The distributions for breaking articles in earlier years from 2001 to 2006 is truncated which is largely an artifact of there being few breaking articles in these years thus a

smaller total population of editors. However, the editor centralization for breaking articles in later years has a similar intercept, shape, and slope as the non-breaking articles' distributions. This is evidence that breaking articles do not differ significantly from non-breaking articles' editor centralization.

Article degree distributions (Figure 5-5, Figure 5-8) exhibit similar patterns where the top-ranked article receives contributions from almost all of the editors in a year while the lowest-ranked articles only receive contributions from a few editors. Again, non-breaking articles exhibit remarkable similarity across the years while breaking articles' distributions are less centralized in early years than later years. However, the centralization of breaking articles in later years exhibits similar intercepts, slopes, and shapes as the centralization of non-breaking articles which is also evidence that article centralization does not differ significantly between breaking and non-breaking article types. Finally, the distribution of edge weights captures the *contribution centralization*, or how many times a single editor contributed to a single article (Figure 5-6, Figure 5-9). These also follow long-tailed behavior; a handful of editors make hundreds of contributions to a single article but the vast majority of editors make only a single contribution to an article when they edit. Non-breaking articles exhibit consistency in the distribution across years and again breaking articles in later years converge to the same distribution.

	Editors & Articles		Links		Density		LCC		Clustering	
	Breaking	Non-breaking	Breaking	Non-breaking	Breaking	Non-breaking	Breaking	Non-breaking	Breaking	Non-breaking
2001	2046 2	7415 154	2072	9863	0.506	0.000872	1.0 1.0	1.0 1.0	0.981 0.0127	0.724 0.0305
2002	84 1	8669 167	84	11372	1.0	0.0155	1.0 1.0	1.0 1.0	1.0 1.0	0.752 0.0443
2003	2336 6	11786 173	2482	15169	0.177	0.000674	1.0 1.0	1.0 1.0	0.9198 0.0166	0.773 0.0363
2004	3163 16	10803 175	3632	13718	0.0717	0.000675	1.0 1.0	0.999 0.989	0.838 0.0185	0.777 0.0314
2005	6628 45	10880 195	8295	14098	0.0278	0.000478	1.0 1.0	0.999 0.989	0.766 0.0201	0.765 0.0371
2006	12022 240	13137 155	16504	18042	0.0129	0.00360	1.0 1.0	0.999 0.995	0.731 0.0187	0.726 0.0292
2007	13137 155	13098 238	18042	16982	0.00886	0.00649	1.0 1.0	1.0 1.0	0.703 0.0211	0.771 0.0339
2008	12542 171	12055 240	17831	15647	0.008831	0.00754	1.0 1.0	1.0 1.0	0.685 0.0233	0.770 0.0308
2009	10270 191	10164 209	14241	12753	0.00726	0.00655	1.0 1.0	1.0 1.0	0.707 0.0284	0.787 0.0354
2010	8566 196	10467 195	12610	13054	0.00751	0.00636	1.0 1.0	0.999 0.995	0.660 0.0288	0.793 0.0294
2011	10532 189	4059 173	15261	5620	0.00766	0.00735	1.0 1.0	0.999 0.994	0.663 0.0257	0.701 0.0553

Table 5-2: Summary bipartite network statistics for breaking and non-breaking articles. For cells with two values, editor metrics are top and article metrics are bottom. "LCC" reflects the fraction of nodes in the largest connected component.

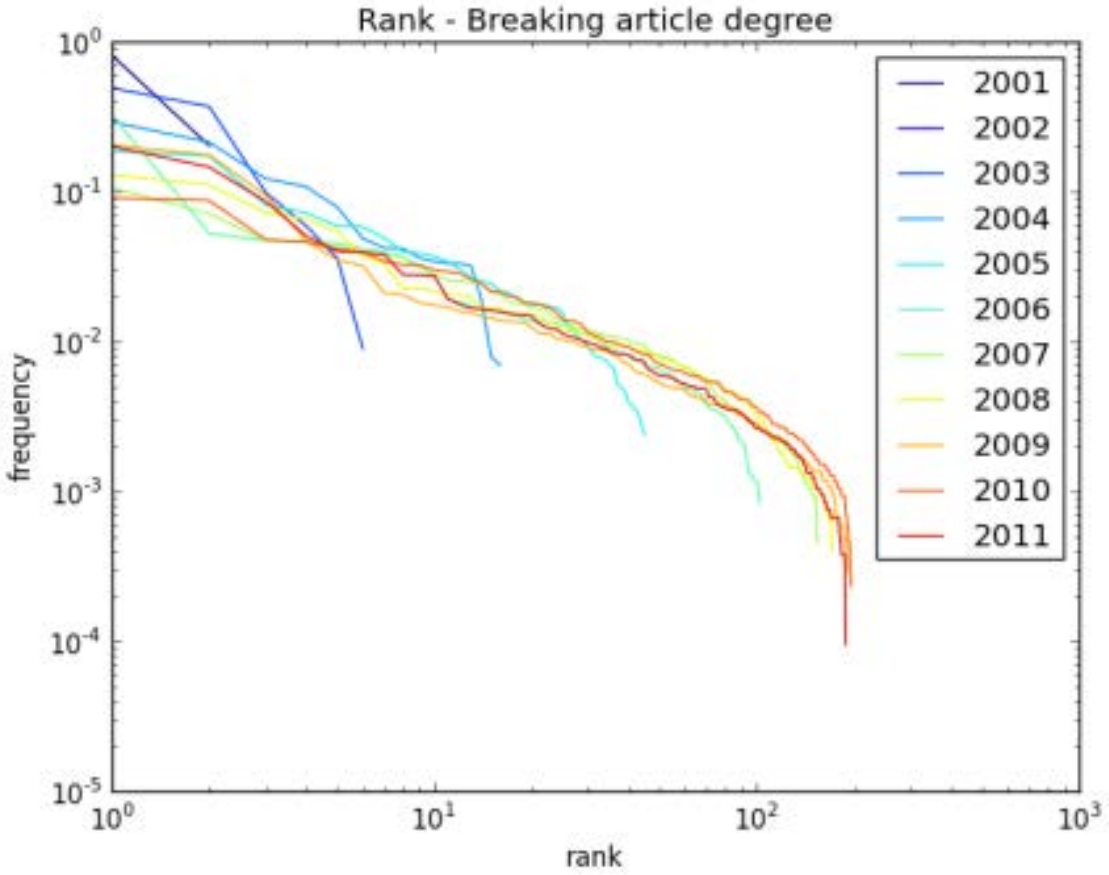


Figure 5-4: Editor degree distributions, by year for breaking articles.

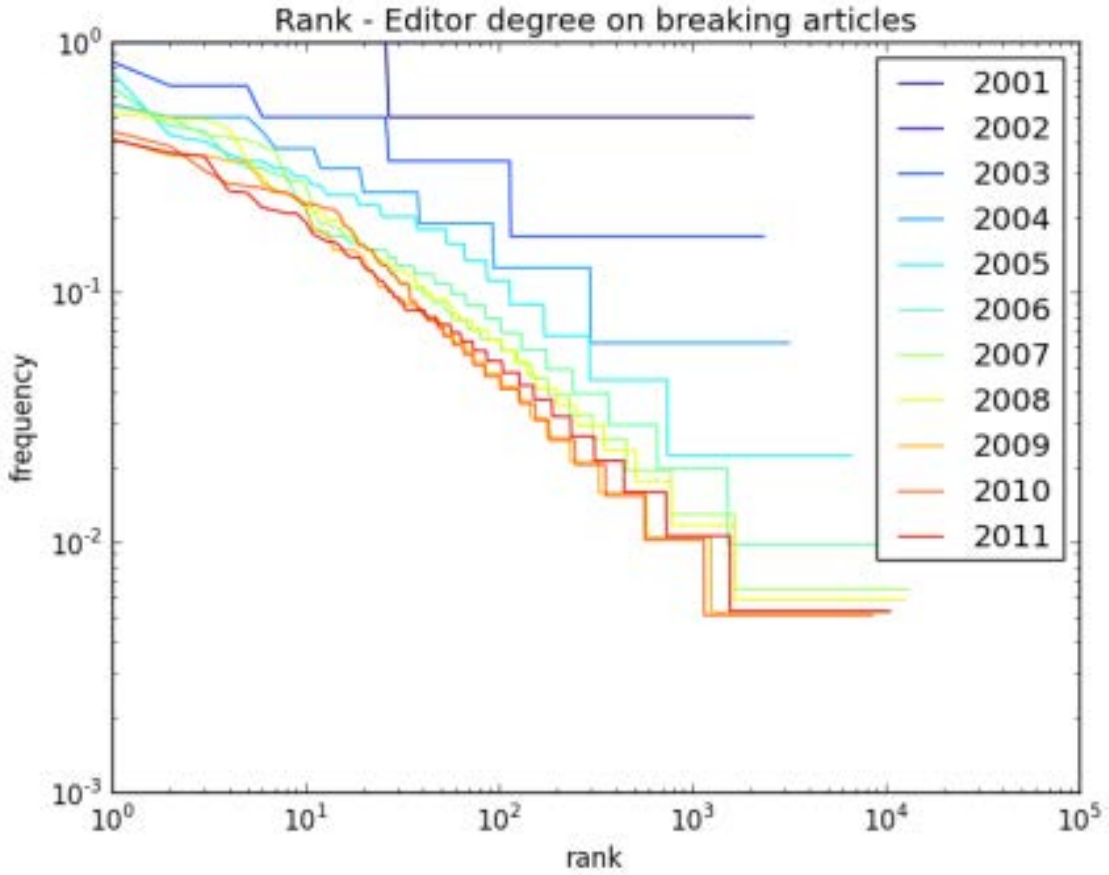


Figure 5-5: Article degree distributions, by year for breaking articles.

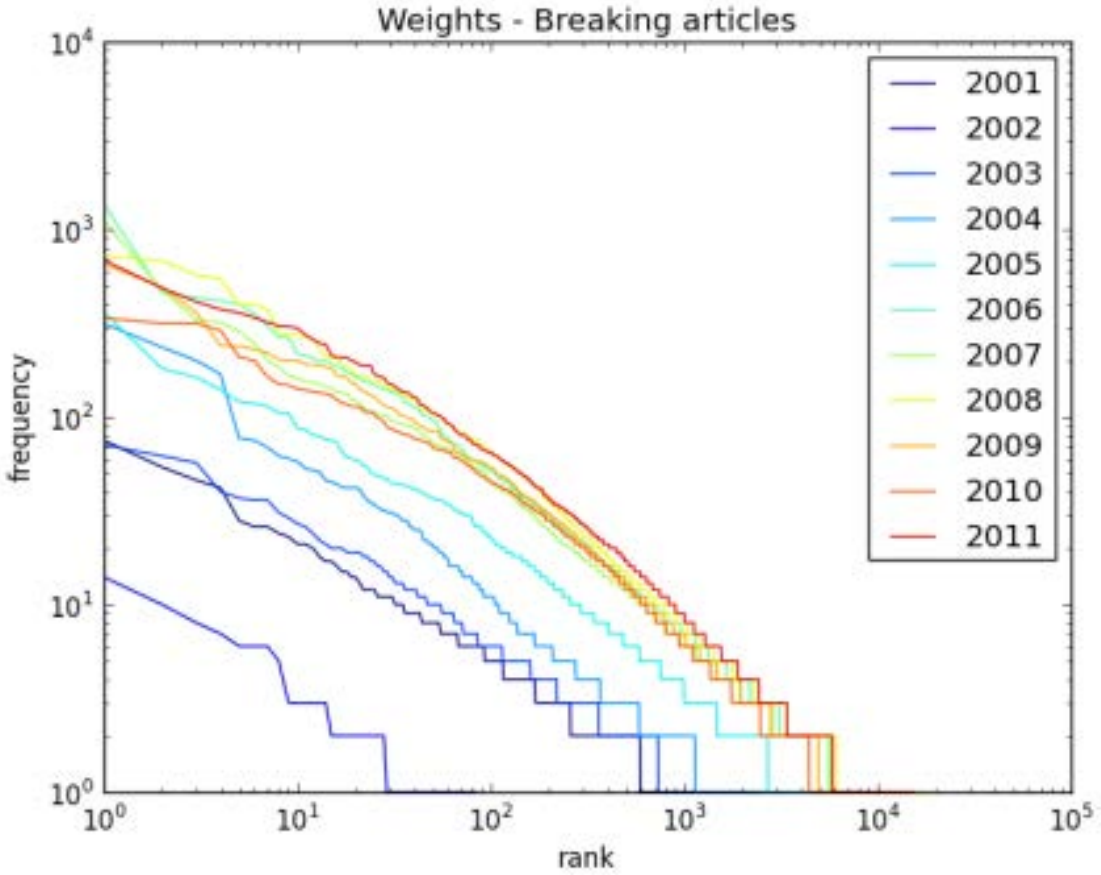


Figure 5-6: Edge weight distributions, by year for breaking articles.

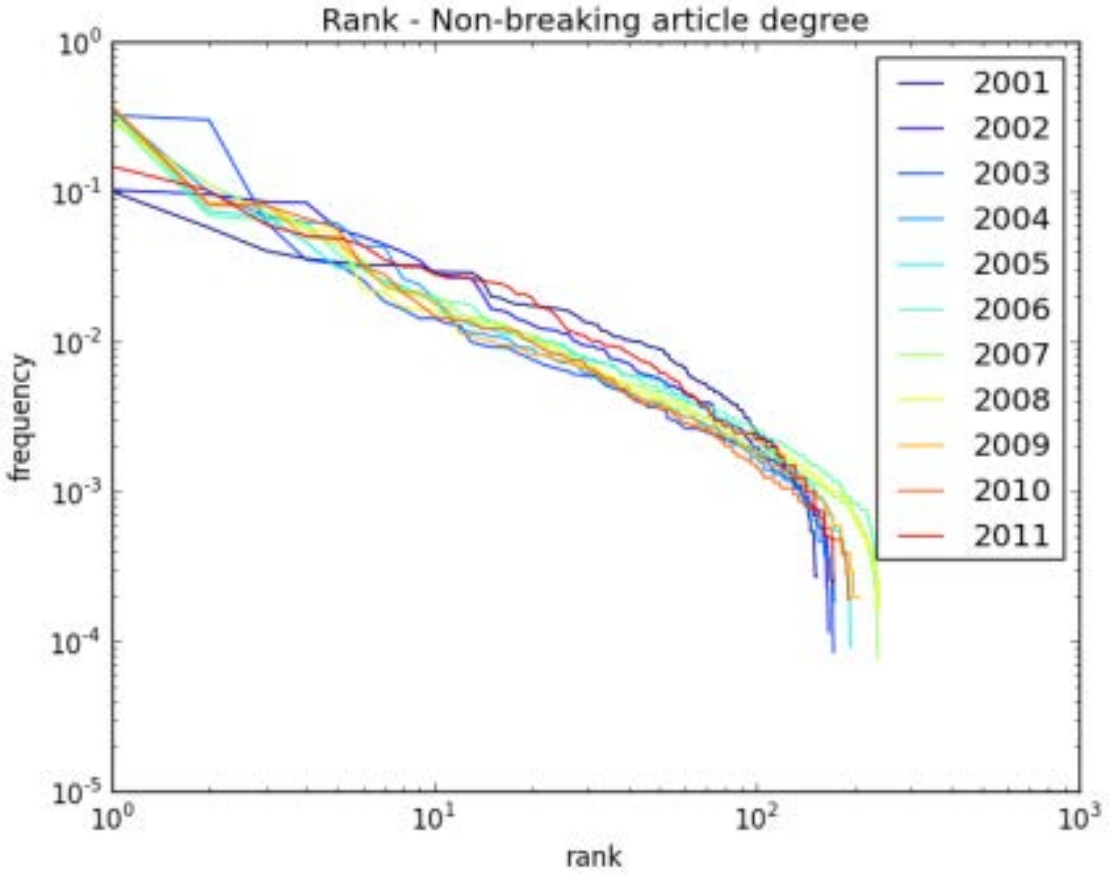


Figure 5-7: Editor degree distributions, by year for non-breaking articles.

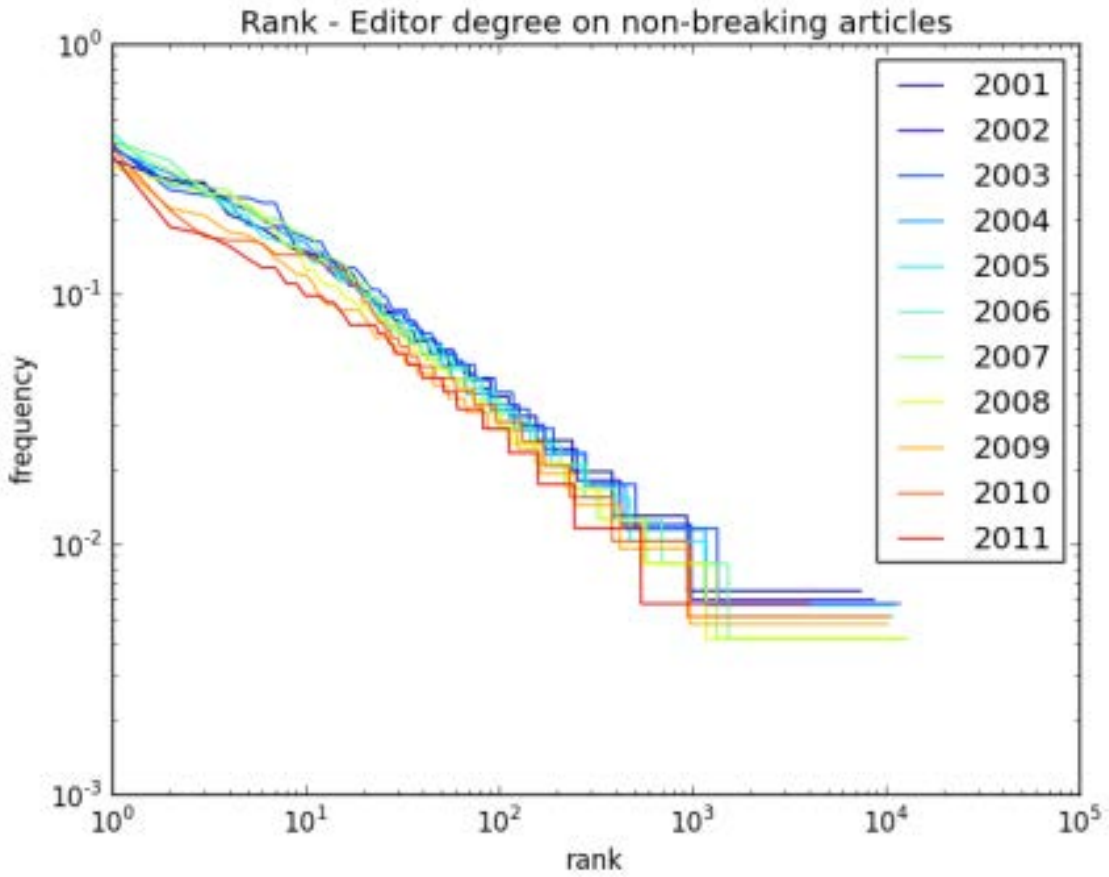


Figure 5-8: Article degree distributions, by year for non-breaking articles.

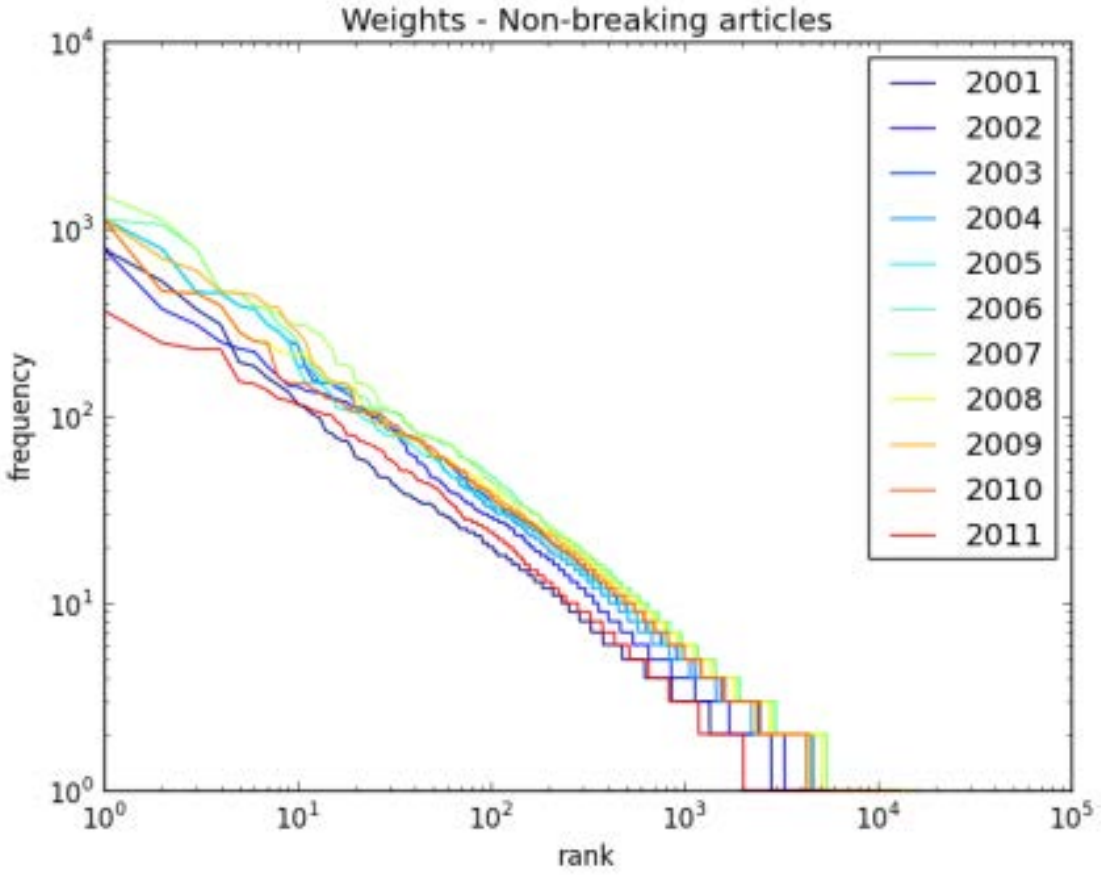


Figure 5-9: Edge weight distributions, by year for non-breaking articles.

Figure 5-10 and Figure 5-11 plot the degree correlations of breaking and non-breaking articles in each year. In bipartite networks, the article degree correlation reflects the extent to which articles having many contributors also have contributors who have revised many other articles. Both breaking and non-breaking articles exhibit positive correlation (also termed degree assortativity) as articles having many editors tend to be revised by prolific editors revising many other articles while articles with fewer editors tend to have editors revising fewer other articles (Newman, 2002). The degree correlations for non-breaking articles are relatively stable for all articles since 2001 while there has been an intensification and stabilization of this assortative pattern on breaking articles since 2001 (as seen by the increased level of assortativity over time).

Taken together, the results from Figure 5-4 through Figure 5-9 suggest mixed support for our hypothesis that predicted the concentration of editors on breaking articles would differ significantly from non-breaking articles. There is little evidence that collaborations involving breaking news articles coordinate this work by centralizing or decentralizing their work to a greater extent than non-breaking articles. However, as I show in the third analysis below, this observation is an artifact of breaking news articles becoming non-breaking articles in the long run; significant differences in edit activity centralization or distribution are present in the immediate aftermath of the event but these differences become diluted as the collaboration reverts to less high-tempo work later on. As I show in the third analysis below, there is also mixed support for Hypothesis 1b that predicted the breaking articles would exhibit similar concentrations for breaking articles in different years. Breaking articles have seen substantial growth in editorship that has dramatically changed the structure of their collaborations over the years while non-breaking articles are relatively stable. However, the distribution of well-connected articles, editors, and links within these networks are remarkably similar across time, and both types of articles exhibit strong assortative degree mixing patterns.

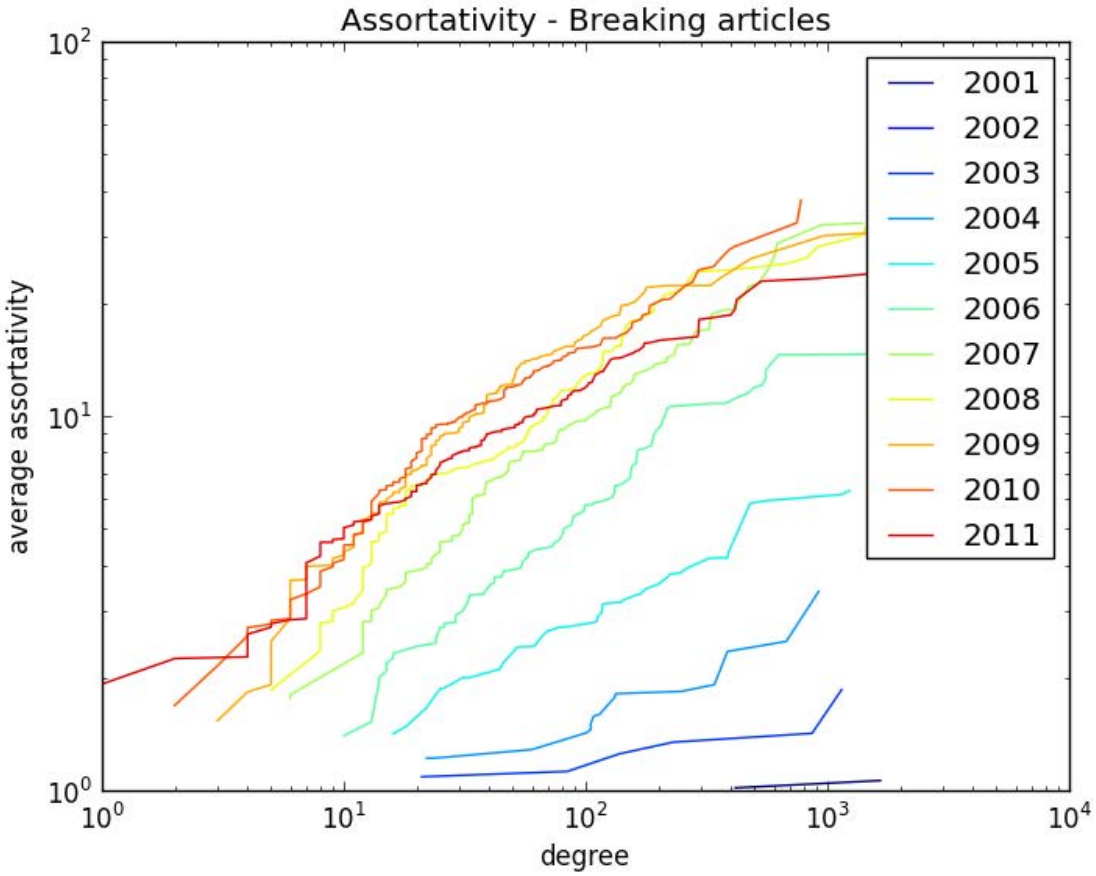


Figure 5-10: Article degree correlation, by year for breaking articles.

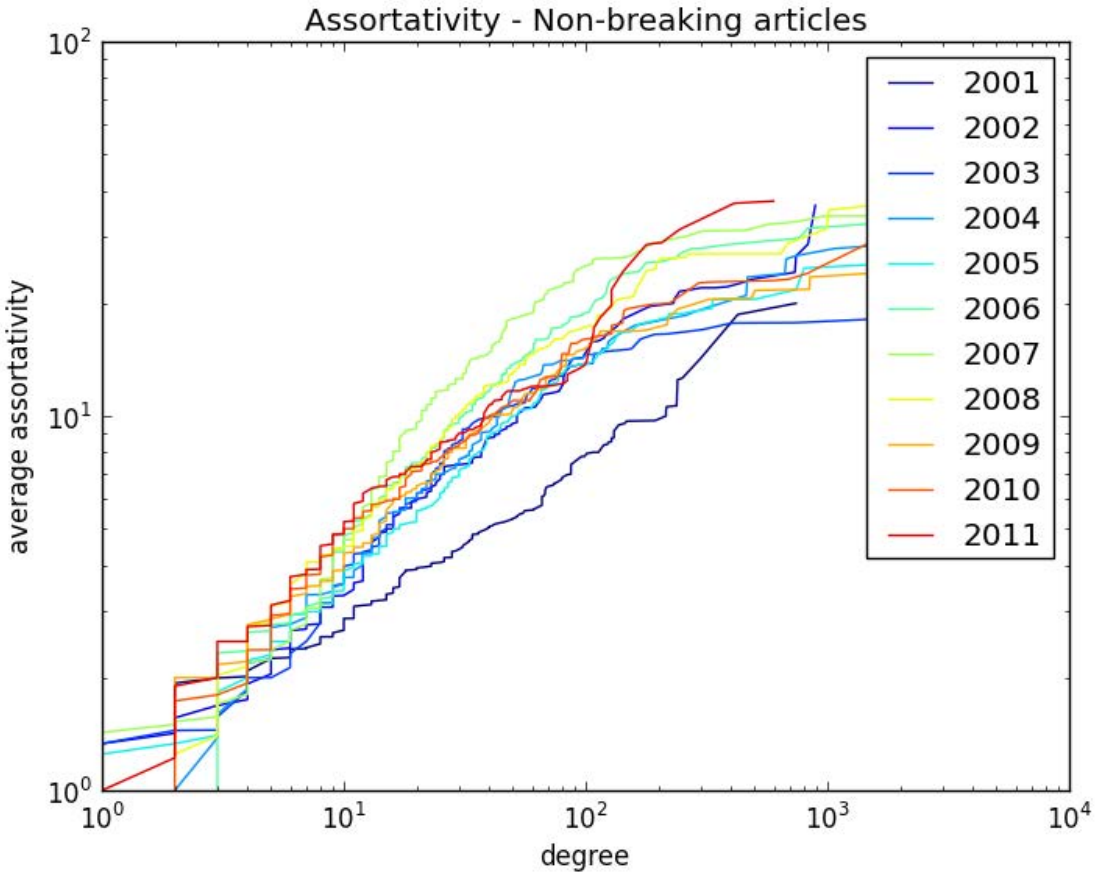


Figure 5-11: Article degree correlation, by year for non-breaking articles.

Organizational regeneration

These article collaborations may not be one-off assemblies of editors, but instances of prior collaborators coming together to regenerate the organizations for co-authoring breaking news articles. To test the extent to which breaking news collaborations are examples of this regeneration, this chapter examines whether editors active in articles one year were also active in articles for other years. Figure 5-12 and Figure 5-13 illustrate the overlap among editors between years for both breaking and non-breaking articles across all years. As expected, much of the editor overlap for breaking articles occurs among editors in adjacent years (cells adjacent to the diagonal): among the editors who contributed to breaking articles about incidents in 2008, 27.2% also contributed to breaking articles about incidents in 2007. Conversely, among the editors who contributed to breaking articles about incidents in 2007, 26% also contributed to breaking articles about incidents in 2008. The vertical axis for an incident's year can thus be interpreted as the *persistence* of that cohort across subsequent incident years while the horizontal axis for an incident's year can be interpreted as the *size* of the cohort relative to the other cohorts. 1.4% of the editors contributing to breaking articles about events in 2011 persisted from making changes to articles that had been breaking in 2001 while 7.3% of editors on articles that had been breaking in 2001 were still contributing in 2011.

Among breaking articles, there is a marked distinction in the overlap between editors who revised breaking articles about events before 2006 and editors who revised breaking articles about events since 2006. Historically, this distinction corresponds to the era of Wikipedia's broader adoption but it also reveals an interesting asymmetry in activity on these two types of articles. Editors of breaking articles about incidents before 2006 demonstrate substantial persistence and make up a non-trivial percentage of editors in subsequent years. This is largely attributable to the growth of the user population and concentration of activity on latter-day articles having hundreds or thousands of editors. However, Figure 5-13 suggests this asymmetry is notably absent among non-breaking articles across all years. The persistence of editing cohorts across incident years, as well as the stability of their presence on these

articles, suggests the existence of a core of editors revising non-breaking articles about incidents across different years as well as the relative immunity of non-breaking articles' collaboration structure from growing the number of Wikipedia editors over time. These results reinforce the previous observation that non-breaking article coauthorship is highly stable compared to breaking article coauthorship. These findings again provide mixed support for Hypothesis 2 that breaking article collaborations would exhibit stronger organizational regeneration than non-breaking article collaborations.

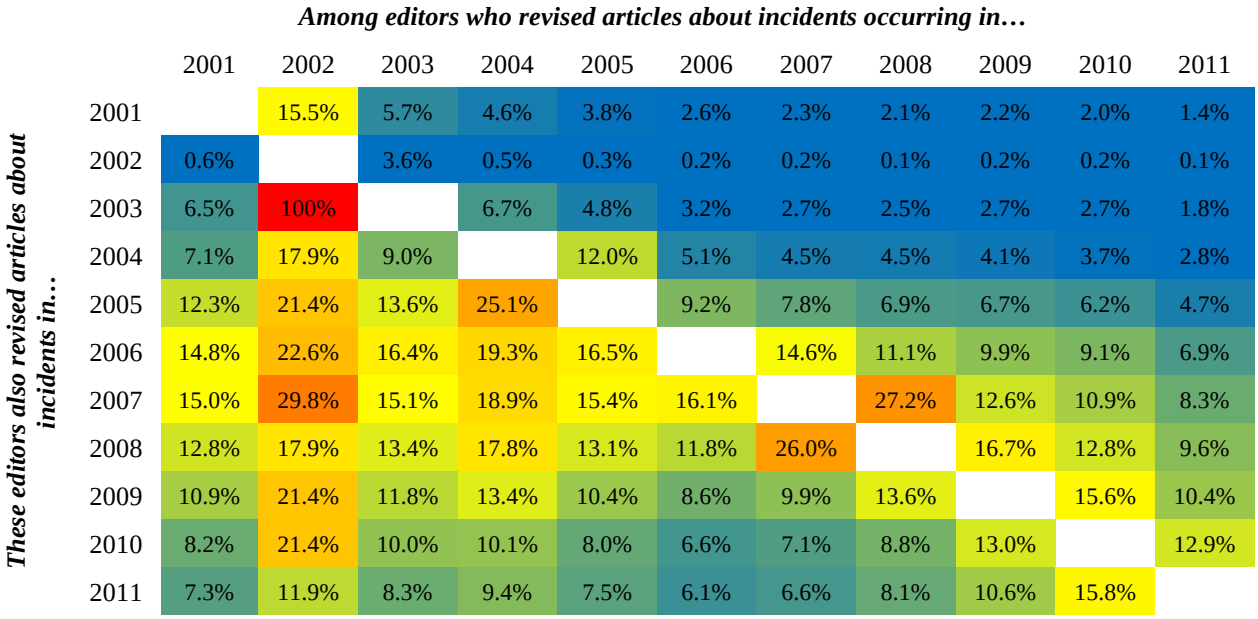


Figure 5-12: Editor overlap for breaking articles.

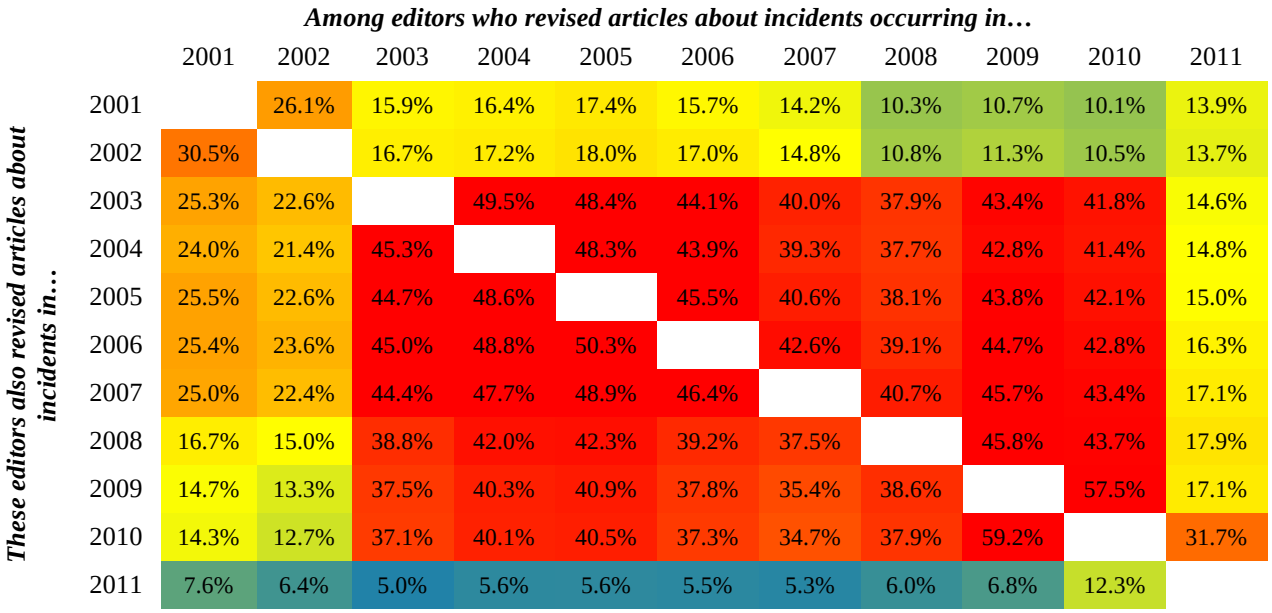


Figure 5-13: Editor overlap for non-breaking articles.

Coauthorship network evolution

The prior analyses compared features of the centralization and regeneration on breaking and non-breaking articles, looking at these on a cross-sectional basis. This cross-sectional perspective potentially occludes important differences in the structures of these coauthorship networks of both breaking and non-breaking articles as they changed over time. This section adds a third genre of articles called historical articles about events that occurred between 1990 and 2001 to the existing genres of breaking and non-breaking articles. As before, the coordination demands of breaking news articles should lead to differences in their observed network structures as compared to either non-breaking articles which happened at the same time but under different coordination mechanisms or these historical articles about events which happened before Wikipedia existed and thus precluded the possibility of them being authored as breaking news events. To test Hypothesis 3, the snapshot approach described above was used to examine the size of all three genres of articles at different stages of their development and track the evolution of their coauthorship networks.

The snapshot networks in Figure 5-14 plot the number of editors editing each of three classes of articles at each of the twelve time-thresholds. The number of editors active on breaking articles (in red) is relatively stable across time while non-breaking (blue) and historical articles (green) see substantial gains in the number of collaborators as articles reach about the age of one year. A similar pattern is reproduced for the degree of articles across time in Figure 5-15: breaking articles tend to have the same number of contributors across time but these articles have significantly more editors than either non-breaking or historical articles until approximately one year of age. Activity on breaking articles is initially much more distributed (more editors involved) than on other types of articles (few editors involved), but in the long run all types of articles converge on similar levels of editor activity which helps explain the similarity of the overall degree and weight distributions observed in the first analysis.

Changes in editor and article clustering patterns across time and article types in Figure 5-16 and Figure 5-17 are illustrative of patterns of regeneration as they point to substantial differences in the extent to which articles share editors in common and editors share articles in common. Clustering among editors in Figure 5-16 is initially significantly higher among breaking articles than other article types suggesting that the early editors of breaking articles have a tendency to collaborate together on many of the same breaking articles. In contrast, the early editors of non-breaking and historical articles are less clustered, but as time goes on these editors see dramatic gains in clustering and eventually exhibit similar levels of clustering as breaking articles after a year. Clustering among articles in Figure 5-17 is high in the first hour for all types of articles suggesting each type of article has a tendency to initially share many editors in common. However, the article clustering drops off significantly for breaking articles suggesting that editors making contributions between the ages of 12 hours and approximately one month are predominately new to editing articles of that type but is relatively stable for the remainder of the article's history. Substantively, this means early editors of breaking news articles are unlikely to contribute at similar stages of other breaking news articles. In contrast, historical and non-breaking articles see a more gradual attrition among articles sharing editors suggesting that editors making changes to one article are likely to make changes other similar articles. These different dynamics are what likely give rise to the differences in article overlaps for breaking and non-breaking.

Breaking article collaborations were also examined for the extent to which they are more cohesive at different stages of their development than either non-breaking or historical articles. Measuring the number of editors and articles in the largest connected component (LCC) reflects the extent to which editors and articles are indirectly linked to each other. For networks with few editors or articles in the LCC at a given point in time, this suggests collaborations are more atomistic as editors revise articles independently of other editors who have a history of contributing to other articles in this domain. Alternatively, networks with many articles and editors in the LCC at a given point in time reflect coherent collaborations in which

the editors work together on many of the same articles. The distributions of the editors in the LCC and articles in the LCC for each of the three types of articles across time in both the cumulative and snapshot networks are plotted in Figure 5-18 and Figure 5-19.

Unlike other article types, editing activity on breaking article collaborations (red) cohere into a large connected component within the first day of activity on these articles. The immediate emergence of this LCC for breaking articles requires the presence of editors who jointly revise many different breaking articles in the hours after these articles are created. Notably, this distribution includes breaking articles for all years between 2001 and 2011. This pattern contrasts with non-breaking and historical articles (blue and green, respectively) in which activity is initially isolated but coalesces into a giant component after a year or more. This same pattern is also borne out looking at the snapshot networks where breaking news collaborations reliably have more activity in the largest component at every time threshold until the articles are approximately a year old. These results suggest breaking articles' collaboration structures significantly differ from both non-breaking and historical articles' throughout the first year, after which all article types exhibit similar tendencies for their activity to cohere into a single giant component. The convergence of all article types after a year suggests the collaboration dynamics of breaking articles are largely driven by the proximity to the event itself but all types of articles cohere into large, connected collaborations after a year. These findings provide strong evidence for Hypothesis 3 that the dynamics of breaking articles are distinct from non-breaking and historical Wikipedia articles at different stages of their life cycle.

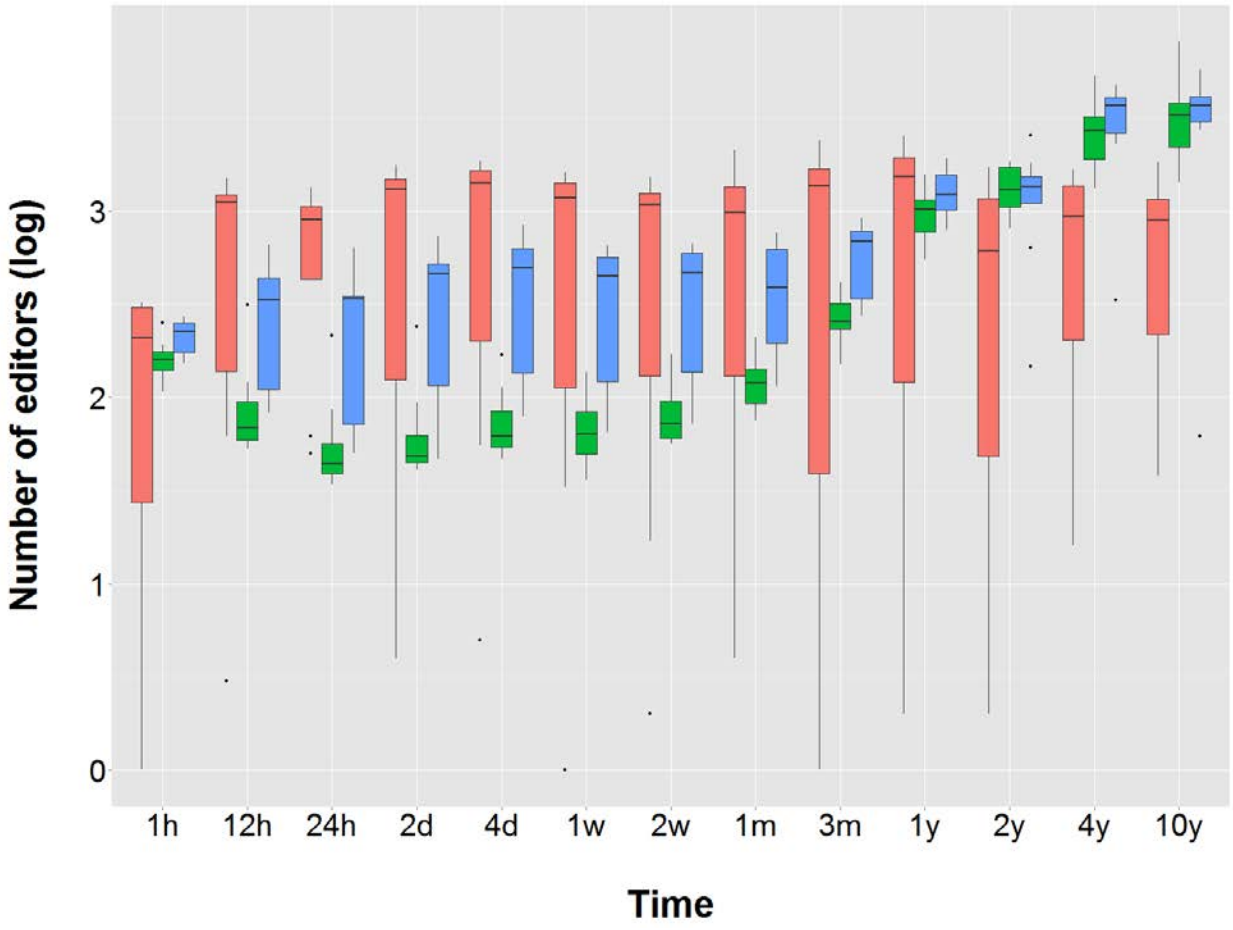


Figure 5-14: Number of editors in the snapshot networks for each time threshold for breaking articles (red), non-breaking articles (blue), and historical articles (green).

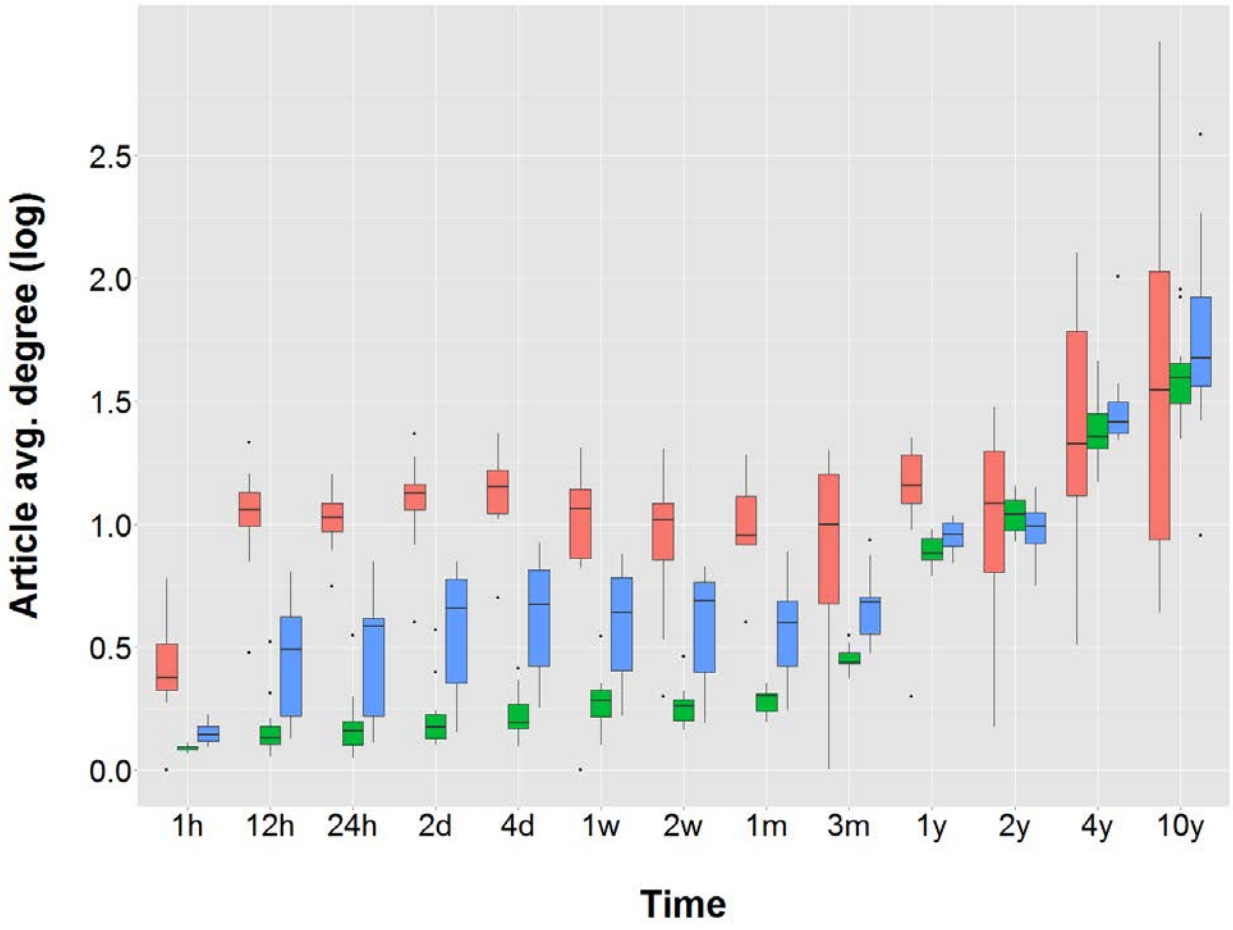


Figure 5-15: Average article degree in the snapshot networks for each time threshold.

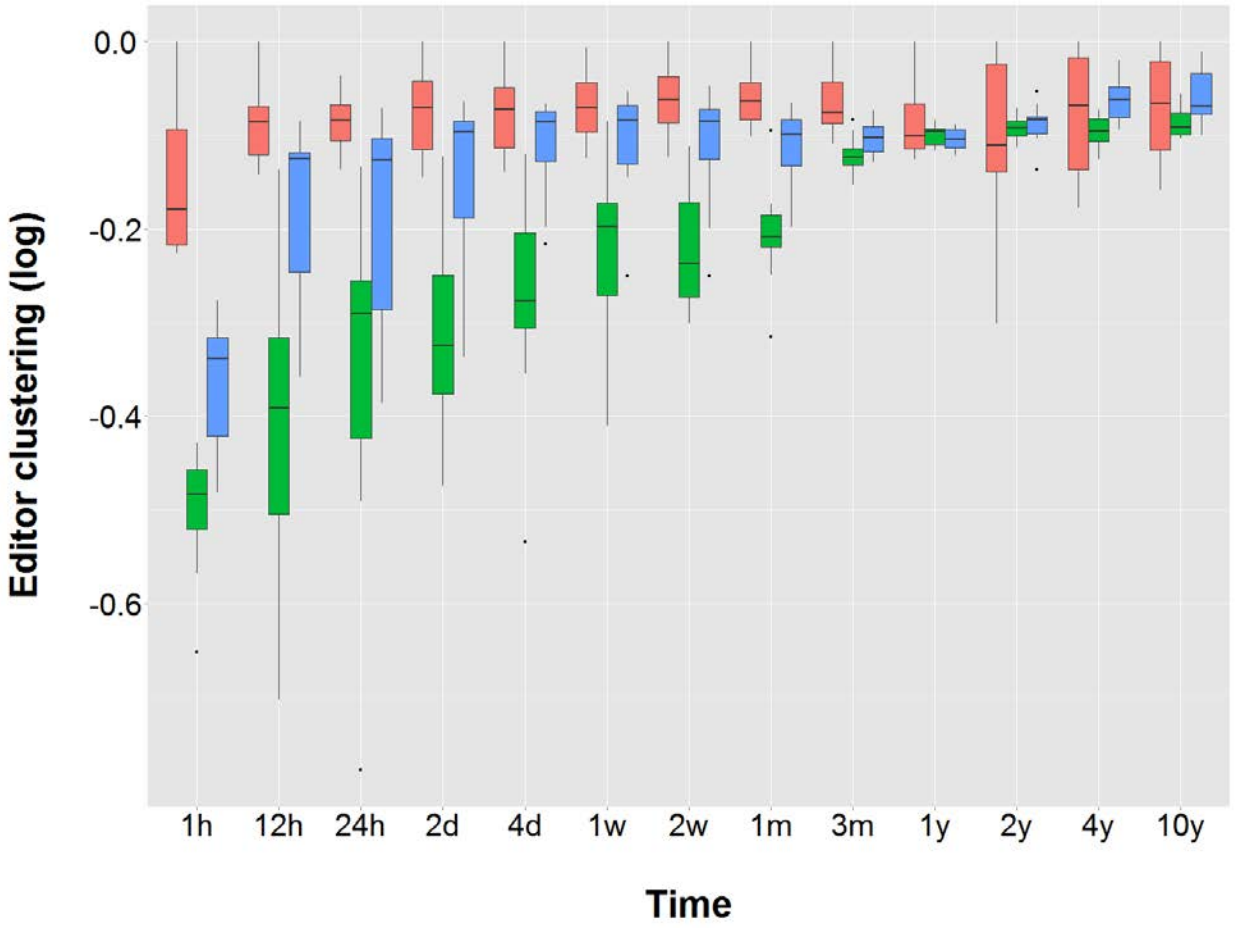


Figure 5-16: Editor clustering in the snapshot networks for each time threshold for breaking articles (red), non-breaking articles (blue), and historical articles (green).

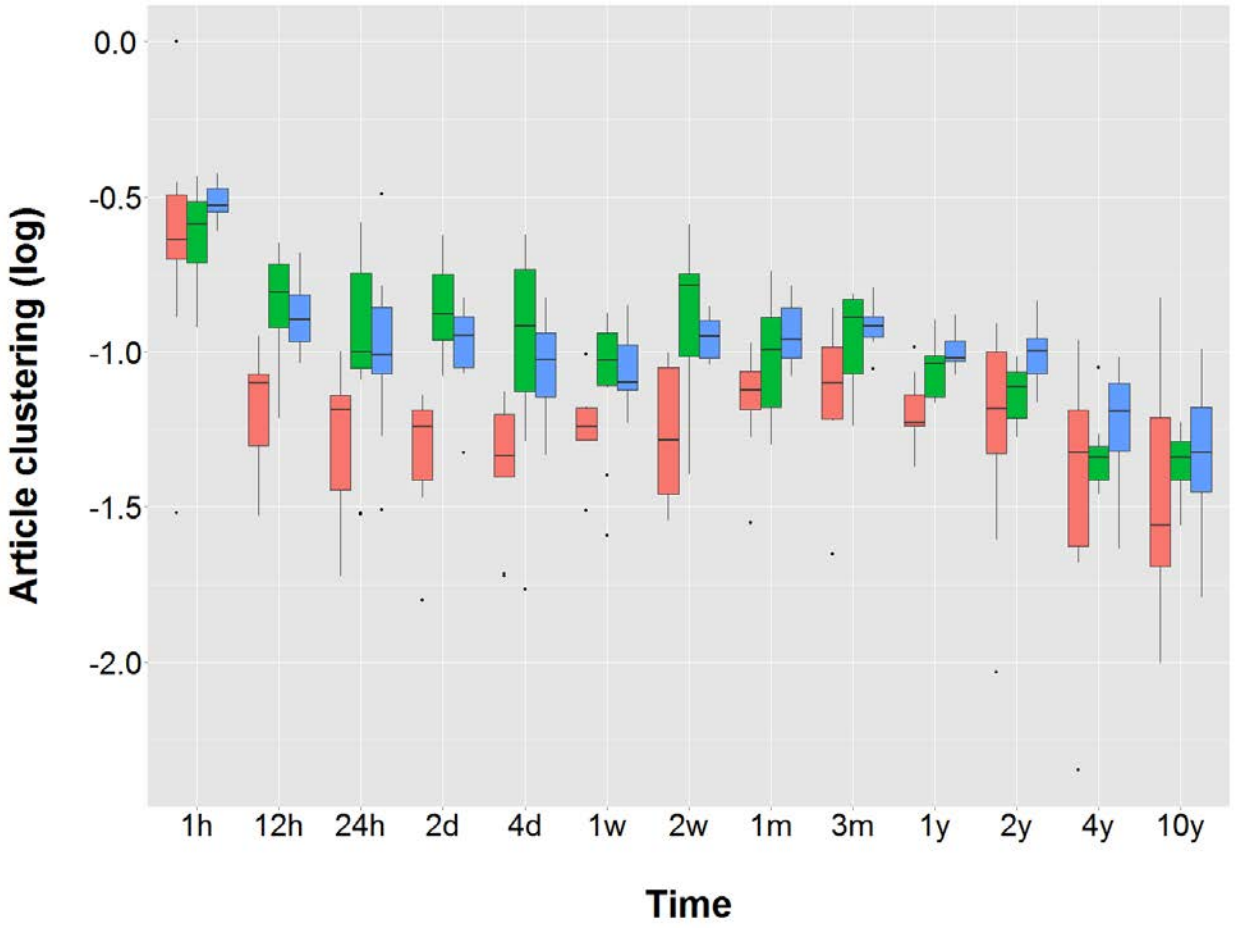


Figure 5-17: Article clustering in the snapshot networks for each time threshold.

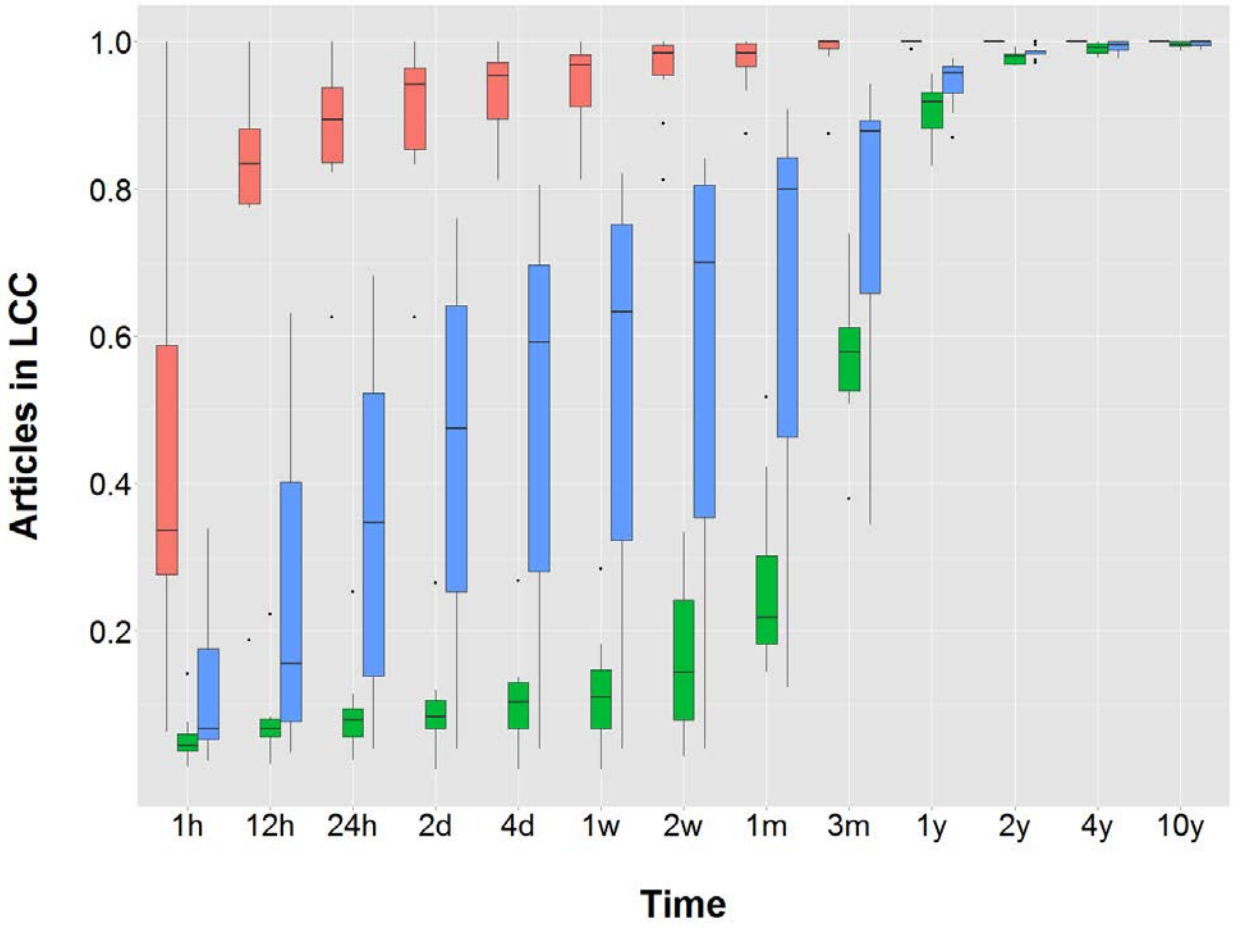


Figure 5-18: Articles in the largest connected component (LCC) for the cumulative network for breaking articles (red), non-breaking articles (blue), and historical articles (green).

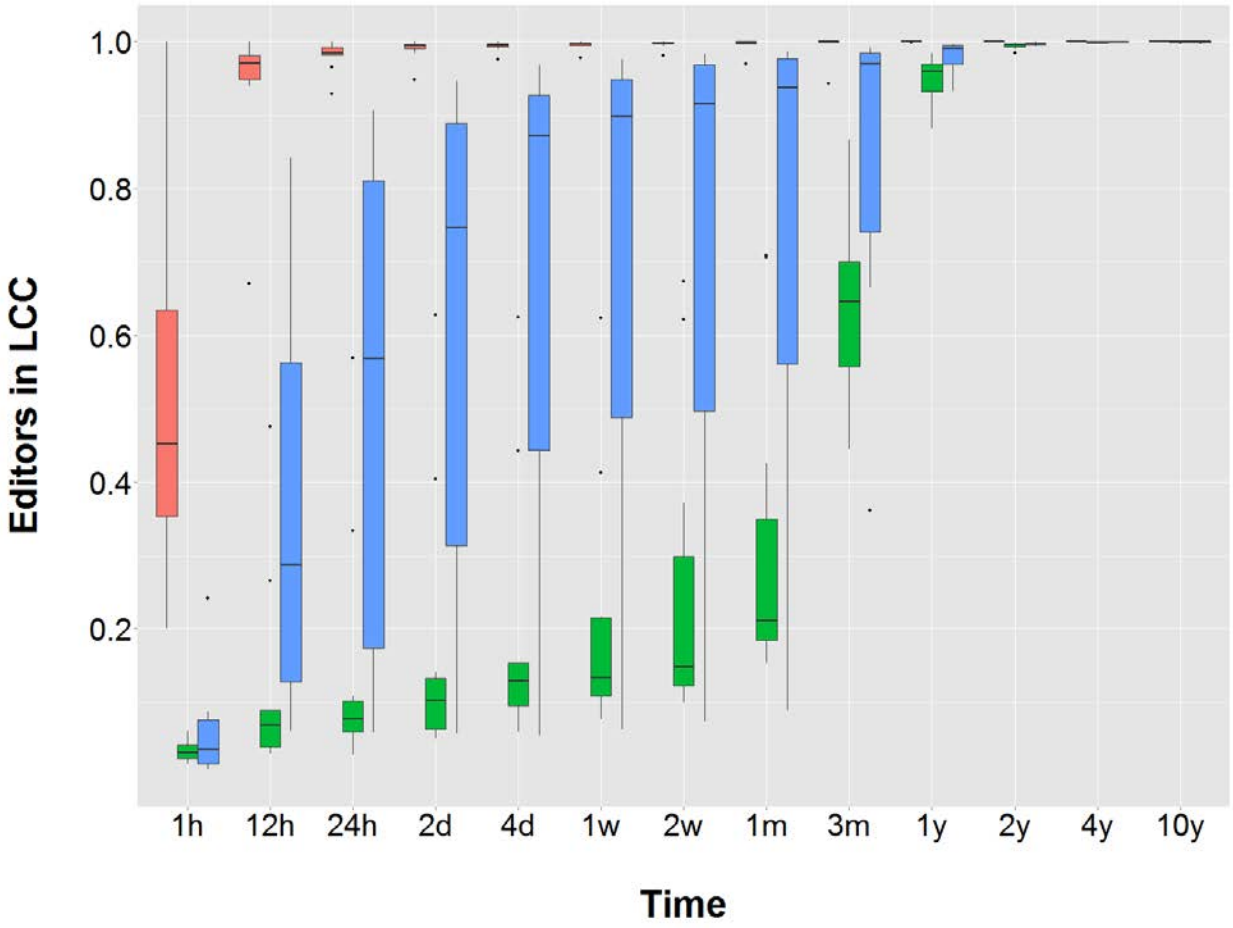


Figure 5-19: Editors in the LCC for the cumulative network for each article type and threshold time.

Discussion

This analysis of a large corpus of Wikipedia articles offers new insights into how the self-organization of collaborations around Wikipedia articles about breaking news events differs from traditional patterns of collaboration. Comparing the centralization of editors and articles within breaking and non-breaking article collaborations, both types of collaborations exhibit remarkable similarities in their density, clustering, and distribution of editor and article connectivity in the long run. There was evidence of repeat editorship across events in different years but non-breaking articles had a much greater tendency to rely upon similar sets of editors while repeat co-authors made up a substantially smaller portion of the population on breaking articles. While these findings contradicted the hypotheses to some extent, examining the time evolution of breaking and non-breaking articles' collaborations revealed significant differences in their structures at different points in time but long-term convergence which explains the similarities in the cross-sectional data and relative prevalence of editor overlap.

Wikipedia's breaking articles provide exhaustive large-scale and longitudinal data logs on user behavior and interaction with material artifacts allowing us to analyze the structure and dynamics of high-tempo, online collaboration. Breaking news article collaborations operate under conditions of simultaneous uncertainty and time pressure that creates tensions over whether to centralize or distribute activity among other group members. Whereas previous experimental studies of task performance have examined small groups completing arbitrary tasks, knowledge collaborations involving breaking news articles occur "in the wild" with groups involving dozens or hundreds of individuals on substantially more complex and emergent work. Nevertheless, this was evidence suggesting the decentralization of activity is a prevailing tendency of online peer production groups during the most acute phases of these high-tempo collaborations followed by a regression to the mean of interaction patterns typically found on non-breaking and historical articles.

This study provides a theoretically motivated empirical basis for understanding the structure and dynamics of rapid online self-organization in socio-technical systems like Twitter and Wikipedia. These findings challenge assumptions in prevailing organizational theories about high-tempo collaboration which predominately examine physically co-located teams in which roles can be assigned and tasks coordinated by encoding specialization into material artifacts such as differentiated uniforms and routinized through shared professional norms. However, this study provides evidence that group members in high-tempo online collaborations differentially pattern their interactions during the most acute phases of article development as compared to non-breaking and historical articles. The immediate emergence of connected components in the early stages of breaking article collaborations and the stability of breaking article editors' interactions with each other over time points to the presence of role specialization in editing breaking articles.

The size of the data in this corpus precluded any systematic examination of the situated practices editors employed to negotiate and coordinate this work. Although this study did not qualitatively examine the practices adopted by editors of breaking news articles, subsequent chapters will examine if differentiated social roles, found in prior work on Wikipedia (Gleave et al., 2009; Welser et al., 2011), are also employed to coordinate this work. The processes by which roles are negotiated and enacted, dependencies and expectations are negotiated and distributed, and how practices are translated and adapted across different breaking article collaborations are rich domains for subsequent inquiry. Chapter 6 introduces a method for quantitatively tracing the trajectory of an article through the editors who contributed to it so to identify highly central social roles or novel behavior. Chapter 7 extends this trajectory method to trace the trajectory of editors through the articles they edit and to qualitatively identify the types of work these editors perform as well as how the work of one editor intersects with the work of other editors.

Chapter 6: Trajectories

The previous chapter identified differences in the temporal evolutions of breaking and non-breaking article collaborations. While aggregate structures appeared to converge to similar patterns in the long run for both types of articles, major differences in the dynamics of article collaborations suggest it is important to model the temporal dependencies within collaborations. This chapter proposes a conceptual framework modeling the temporal relationships between social agents and technological artifacts. Understanding these interactions has implications for theorizing about technologically mediated social behavior, designing social computing systems, and making sense of complex behavior in large data sets.

This framework draws on prior work on information provenance and stigmergy but is grounded in the classical mechanical idea of *trajectories* which allows us to understand the progression of users and artifacts moving from one state to another over time. This chapter adapts a method for discovering the structural signatures of these sociotechnical trajectories based on sequences of behavioral interactions within large and complex digital behavioral trace logs like revision histories on Wikipedia or commit logs to free/libre open source software (F/LOSS) development projects. This method is evaluated using a case involving the editors, articles, and rules governing Wikipedia's coverage of breaking news events to identify social roles, communities of practice, and institutional routines which I then map back into this framework of sociotechnical trajectories as velocities, fields, and orbits.

These findings demonstrate that article revision patterns immediately following unexpected, catastrophic incidents differ from the revision patterns of similar articles about historical events. Although these patterns of interaction on breaking news articles have intensified as Wikipedia has become more popular, the revision networks of breaking articles also "regress to the mean" as the event becomes more distant. The chapter concludes by discussing the implications these findings have for analyzing and theorizing about the success of online communities when participation is intense, popular, and brief as well as

recognizing that collective intelligence systems create very different types of information artifacts under various contexts for production.

Background

The conceptualization of user behavior as a "trajectory" is not new. Prior scholarship has examined the sequence of stages through which people, resources, or teams move towards a goal (Bardram, 2000). This conceptualization of trajectories emphasizes the management of both participants' and tasks' successive interdependencies, often across multiple concurrent projects (Fussell, Kiesler, Setlock, & Scupelli, 2004). These approaches draws upon coordination theories locating the management of interdependencies as the need for reciprocal quality and timeliness of work among team members (Schmidt & Simonee, 1996), as the need for managing activities having prerequisites, simultaneous demands, and shared resources (Malone & Crowston, 1994), or as the need to manage related activities through artifacts and shared mental models (Bardram, 2000). However, in contexts like F/LOSS development or Wikipedia co-authorship where projects are typically in a constant state of development, coordination is substantially more complex and potentially more reliant on implicit or informal mechanisms (Espinosa, Lerch, Kraut, Salas, & Fiore, 2004). Tasks in these systems are on-going and emergent rather than moving sequentially toward a final stage and participants have substantially more volition to scope and select their contributions rather than being assigned to defined roles and responsibilities.

Other scholarship has approached analysis of sequential user behavior within information spaces by modeling the intentions and goals of users as following a path or "information scent". These models attempt to simulate users' information needs patterns as a flow over existing hyperlink paths linking documents. Users possess information needs and cues within the document guide their decisions about the value and cost of following links. The similarity between the present cues and their needs thus predicts the links that users will employ (Chi, Pirolli, Chen, & Pitkow, 2001; Pirolli & Card, 1999). However, this type of approach assumes the cues within documents are stable over time and users are simply consumers

navigating an information space, incapable of producing or altering these cues. Artifacts in peer production sociotechnical systems like Wikipedia complicate these assumptions because they are highly mutable to participants' uses and needs as they revise the content of cues and connecting links.

Approaches like information physics and interaction histories have articulated how interaction histories encoded within artifacts can be used to surface their patterns of use (W. C. Hill & Hollan, 1994; W. C. Hill, Hollan, Wroblewski, & McCandless, 1992). The history of social interactions with an artifact can reveal their value for other users and can be used to navigate complex collaboration networks containing information about authors, documents, and contexts (Dunne, Riche, Lee, Metoyer, & Robertson, 2012; Indratmo & Vassileva, 2009). Patterns of user behavior in persistent, threaded conversations can also be structured to reveal regular behavioral patterns indicative of distinct and stable social roles within online communities (Fisher, Smith, & Welser, 2006; Gleave et al., 2009; Smith & Fiore, 2001). However, these approaches locate the interaction histories of interest in either the user or the artifact, but not both. This suggests the need for a generalizable approach to provenance that can be applied to changes in behavior of both sociotechnical artifacts and users.

Biologically-grounded theories of indirect and distributed coordination employ "stigmergy" to explain self-organization and collective intelligence in distributed collaborations. Stigmergic processes involve the local modification of a shared environment that in turn provides the stimulus and resources for other agents to continue work (Heylighen, 2007; Susi & Ziemke, 2001). Shared environments like Wikipedia or code repositories contain artifacts like articles or files which are not only the articulations of prior agents' work, but also mediate interactions with other agents to support on-going cooperation. Stygmergic approaches suggest the work done in sociotechnical systems is not only conditioned on the preceding work but also serves as a model for on-going work which allows systems to become self-reinforcing as high-quality material attracts more diverse participants which increases the pool of expertise who in turn further improve the artifact (Heylighen, 2007; Howison, Østerlund, Crowston, & Bolici, 2012). Users in

sociotechnical systems like GitHub monitor other users and make rich inferences about their social roles and the action to be taken without explicit prompting despite relatively sparse behavioral cues (Dabbish, Stuart, Tsay, & Herbsleb, 2012). Articulating the progression of users and artifacts through each other provides a way to illuminate regular behavioral patterns of this mindfulness and mediated coordination.

Approach

"Trajectories" provide a conceptual toolkit and vocabulary for understanding the progression and dependencies of an entity moving through time and space. In classical mechanics, a trajectory is the curve a body in motion follows over time as it moves through space. The conic sections a planet traces as it orbits a center of mass is governed by variables such as its velocity and distance from the center while the parabolic path a launched projectile follows is determined by its inclination and drag. I extend prior conceptualizations of the information provenance to argue the interactions between users and artifacts, the changes in their states over time, and influence from outside forces can analogously be understood as *sociotechnical trajectories*. Trajectories will both capture the temporal ordering of work done in sociotechnical systems and reveal complex network structures of dependencies and clustering among users and artifacts that are occluded by traditional collaboration networks which simply look at whether a user contributed to a document.

A sociotechnical trajectory is the movement of a user or artifact through different states in the system over time. The trajectory reflects the source and sequence of actions through a system over time and reveals the influence of both intrinsic and extrinsic forces acting on it. Sociotechnical trajectories of users and artifacts are articulated through each other: a *user trajectory* reflects the sequence of artifacts a user has acted on over time and the *artifact trajectory* reflects the sequence of users who have acted on the artifact over time. Within an artifact trajectory, the same user interacting with the same artifact at different points in time will be linked to other users who have interacted with the artifact in the intervening time. Artifact trajectories thus trace the temporal path of artifacts "moving through" the users with whom they

have interacted. Within a user trajectory, a user's interaction with the same artifact at different points in time will be linked to the other artifacts the user has interacted with in the intervening time. User trajectories thus trace the temporal path of users "moving through" the artifacts with which they have interacted. The trajectory of any single user or artifact reflects its provenance from a source through a sequence of ensuing actions. The articulation of a user trajectory through the temporal sequence of articles or an article trajectory through the temporal sequence of users captures important temporal contexts and dependencies that are omitted through other approaches such as collaboration or user-to-artifact collaboration networks (Brandes, Kenis, Lerner, & van Raaij, 2009a), artifact-to-artifact citation networks (Bellomi & Bonato, 2005), or user-to-user discussion or revert networks (Suh, Chi, Pendleton, & Kittur, 2007). This approach adopted in this chapter uses the entirety of the article revision history as a trajectory of sequential revisions from contributors and implies relationships from users to users within the artifact of the article as they revise one another's contributions. By aggregating the patterns of these relationships to a network of edges and nodes, it becomes possible to use network analysis methods to understand and compare structural features such as centrality, clustering, and similarity across many articles and users.

Using Wikipedia as an example, a "revision" is the state of a Wikipedia article at a given point in time as authored by an editor. A Wikipedia article's revision history is an artifact trajectory consisting of a series of instances where editors have made changes which update a prior state of the article which itself was authored by an editor. A Wikipedia editor's contribution history is a user trajectory consisting of the series of articles the editor has updated after editing another article. I describe a method for structurally extracting artifact and user trajectories from digital trace data logs below. By examining multiple and overlapping user or artifact trajectories, it is possible to identify structural signatures and extend the trajectory analogy to reveal temporal patterns (*velocity*), pervasive forces (*fields*), recurring patterns of actions (*orbits*), and actions preceding abrupt changes (*collisions*) within sociotechnical systems. The

behavioral signatures which emerge from this approach offers one way to structure and interpret the flux of digital trace data in sociotechnical systems as well as providing contexts and visual patterns to compare behaviors.

Artifact trajectories

An “artifact trajectory” is derived from the temporal sequence of actions performed by users towards a single artifact. Because artifact trajectories are articulated through the time-ordered progression of user interactions with the artifact, I represent this trajectory as a directed graph. The nodes in the artifact trajectory are users and a directed edge between nodes represents an interaction (such as a commit) that transitions the artifact (such as a Wikipedia article) from one user’s state to the next. This approach captures a structural view of a single article’s revision history allowing us to view and analyze complex relationships between editors and article versions.

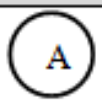
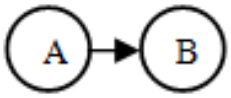
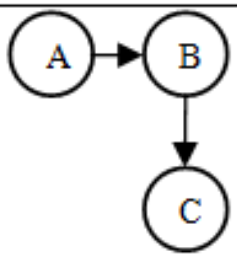
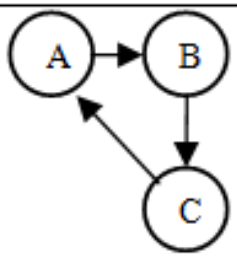
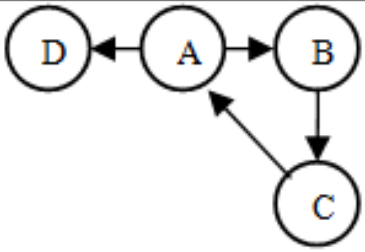
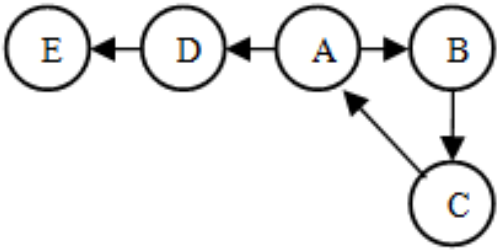
Revision number	Revision editor	Revision history chain graph
1	A	
2	B	
3	C	
4	A	
5	D	
6	E	

Figure 6-1: Trajectory construction example.

The Wikipedia example presented in Figure 6-1illustrates the construction of an artifact trajectory for a single article. Editor A creates an article by making the first edit at time 1. Then, at time 2, editor B makes some change to the article, replacing editor A's version. Thus, the A→B dyad can be interpreted as "control of the document passed from editor A to editor B." At time 3, editor C makes a change to the article which could reflect new content being added or removed or the modification of content previously introduced by editors A or B. In this view, the early stages of a "chain" forms: new editors modify prior editors' revisions, but earlier editors never revise subsequent edits.

At time 4, editor A returns to the article and commits a new revision after editor C. This additional commit by an editor who has previously contributed appears in the network representation as a "loop." At time 5, editor D saves a revision after editor A's *second* revision which indicates editor A has saved more than one article revision in this article's history, making her increasingly central in the network. At time 6, editor E modifies editor D's revision. Consider if editor A had not made her second contribution: editor D would have instead modified editor C's revision, and a "chain" of single editor revisions would have grown. A "loop" occurs when editors repeatedly revise an article after each other causing a chain to fold back on itself because of the repeated contribution of a user (e.g., editor A's second revision). A "chain" appears when subsequent editors only commit one revision and never contribute to that article again. A "self-loop" when an editor modifies her own revision in succession (A→A) are ignored. This article trajectory approach treats all revisions equally—major, minor, vandalism, reversion—because editors' activity in response to something like vandalism potentially captures unique structural positions and social roles such as anti-vandalism as well as editors' general engagement with and mindfulness about the article.

User trajectories

Analogous to artifact trajectories, "user trajectories" are the temporal sequence of actions performed by a single user to one or more artifacts. User trajectories are articulated through the progression of user

interactions with artifacts over time. Again, this time-ordered progression is represented as a directed graph. In the context of Wikipedia, the nodes in the user trajectory are articles, and directed edges between articles reflect an editors' change from editing one article to editing another article at a subsequent point in time.

Returning to the Wikipedia example in Figure 6-1, the editor makes her first contribution to article A at time 1. At time 2, she edits article B. The A→B dyad can be interpreted as "the user revised article B after article A." The user edits a third article C at time 3 but then returns to the first article A at time 4 to make a second revision to it, creating a loop. The editor's subsequent revisions of articles D and E begin to create a new chain. Like the user trajectory approach previously outlined, this method treats all revisions equally because articles may be loci for practices like discussion or arbitration as well as revealing the general breadth of editors' interest in modifying articles. User trajectories are explored in more detail in Chapter 7.

Research questions and hypotheses

This chapter evaluates this proposed methodology by revisiting the question raised in the previous chapter: do breaking news articles exhibit different patterns of coauthorship and revision than non-breaking articles? Immediately following a breaking news event, collaboration should be characterized by intense interaction in which individuals iteratively read and modify each other's contributions to construct a joint account of the event. The resulting structures of these articles' trajectories should manifest as highly centralized and dense interactions as early contributors and newer editors revise each other's accounts and process information.

Second, given significant differences observed in collaborations over time in the previous chapter, are similar changes in the structures of article trajectories also observed. In other words, how do the structural features of an article trajectory change over the article's lifecycle? As the article develops and more

reliable and authoritative information is revealed about the incident, a breaking article's content and style should stabilize and exhibit trajectories more similar to non-breaking articles over time.

Finally, the previous chapter also examined the features of large aggregations of article collaborations while this proposed article trajectory methods has only examined structural features at the article level. If article trajectories for multiple articles were aggregated, would it still reveal the editors who were so central in the coauthorship networks were themselves central in the individual article collaborations themselves or that they were only peripherally involved in many articles? What features of editors' interactions with other editors predict their re-engagement on other articles? In particular, do editors interact on subsequent articles?

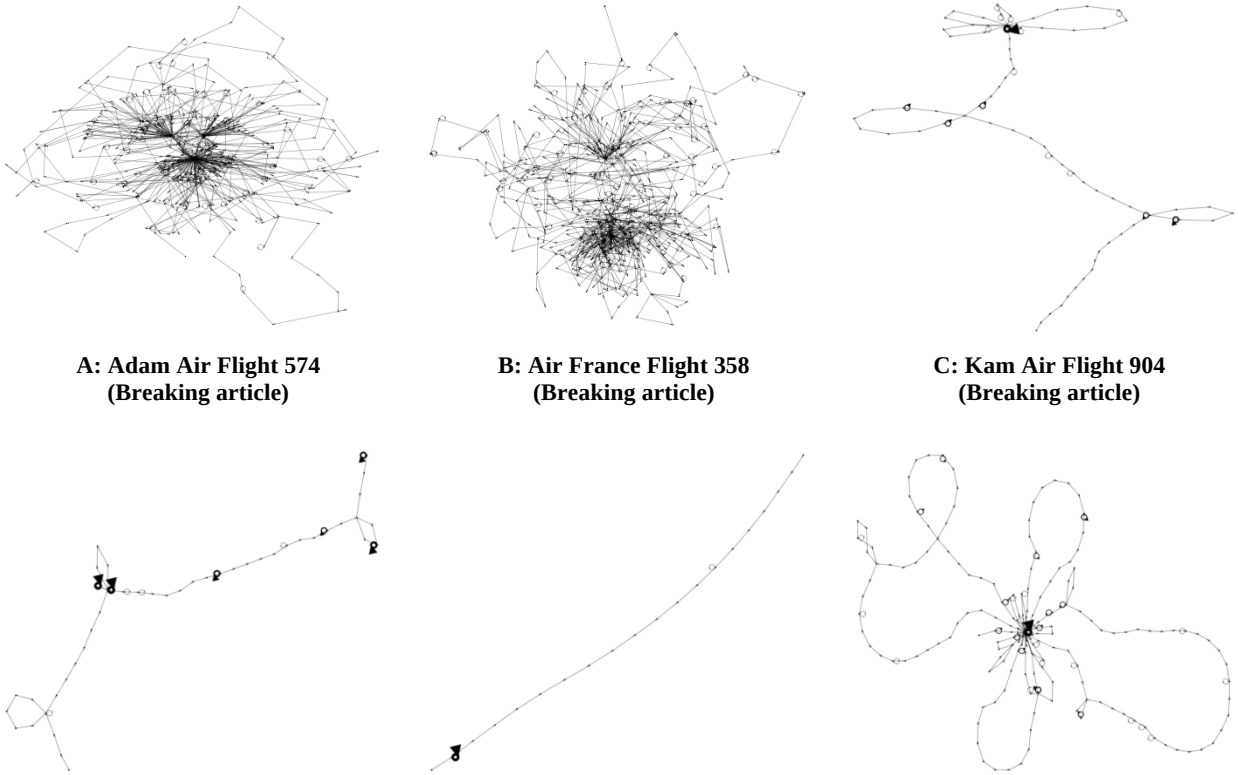
The re-emergence of interactions among editors and symmetries of arrival and departure patterns of editors on breaking articles would likewise suggest the coherence and similarity of action over time are instances of organizational regeneration (Birnholtz et al., 2007; Pentland et al., 2011). Repeated and central editorship across collaborations would suggest editors manage dependencies and self-organization in high-tempo collaboration by relying on social roles to constrain and enable action (Gleave et al., 2009; Welser et al., 2011).

Study 1: Structure and dynamics

The first study is a sample of 229 articles selected from Wikipedia's "List of accidents and incidents involving commercial aircraft". These articles were written about incidents that took place in a 20-year time window from January 4, 1990 to December 25, 2009. I excluded incidents attributable to hijacking or terrorism from the sample (like the flights involved in the 9/11 attacks), because these are outliers with respect to salience and visibility of the incident. Seventy-six articles in the second corpus were started within 96 hours of the incident which I categorized as "breaking" articles. I categorize the remaining 153 articles in the second corpus as "non-breaking". The non-breaking articles include 44 incidents that took

place after Wikipedia was founded in January, 2001, and 109 incidents that occurred before Wikipedia was founded and thus could not have been authored as a breaking article.

Figure 6-2 depicts six representative article revision networks; three for breaking articles and three for non-breaking articles. The thickness of the lines is proportional to the frequency of directed interaction between editors. In many cases, editors modified their own revisions in succession creating self-loops. By visual inspection, some breaking articles appear to have “tighter” patterns of revision activity, with fewer “chains” and more “loops”. However, other breaking articles like Kam Air Flight 904 exhibit the “loose”, chain-like structures I expect to find in non-breaking articles (see Figure 2c), and some non-breaking articles such as China Airlines Flight 611 exhibit the “tight” loop-like structures I expect to find in breaking articles (see Figure 2f). I therefore used linear regression models to analyze in a principled fashion the relationship between creation lag, fatalities, OECD location, number of editors and the four descriptive network statistics: diameter, closeness, betweenness, and clustering, across the entire corpus of articles.



A: Adam Air Flight 574 (Breaking article)

B: Air France Flight 358 (Breaking article)

C: Kam Air Flight 904 (Breaking article)

D: Lufthansa Flight 2904 (Non-breaking article)

E: Air Philippines Flight 541 (Non-breaking article)

F: China Airlines Flight 611 (Non-breaking article)

Figure 6-2: Article trajectory visualizations for three breaking and three non-breaking articles.

Metrics

Four network statistics were calculated to capture the structural characteristics of the article revision networks to identify "tighter" or "looser" patterns of editor activity. "Tight" patterns of editor activity exhibit many loops and few chains, as editors who have previously saved a revision return to the article later to contribute additional revisions. Conversely, "loose" patterns of activity exhibit few loops and many chains, as editors contribute single revisions and do not return.

The following network measures are dependent variables while incident and article attributes such as fatalities, location, incident date, and first edit timestamp are independent variables or controls in the regression models. The distributions for closeness, betweenness, clustering, creation lag, fatalities, and number of editors were all positively skewed and were log-transformed to generate quasi-normal distributions for the analysis.

- The *diameter* of a network is the length of the longest shortest path (geodesic) between any two nodes in the revision network. Articles with a "loose" revision history structure manifest greater final distance between editors' revisions than articles with a "tight" revision history structure. The longest geodesic on a breaking article such as Air France Flight 358 is 15 (Figure 6-2B) while the longest geodesic on a non-breaking article such as China Airlines Flight 611 (Figure 6-2F) is 33. Motivated editors who repeatedly contribute on breaking news articles are likely to make contributions after more diverse editors which should in turn bring these otherwise peripheral editors closer together.
- The *average closeness* of a network captures the tendency for nodes on the network to be able to reach each other with short paths. Closeness is formally calculated as the inverse of the sum of the lengths of the geodesics to every other node on the network (S Wasserman & K Faust, 1994). Breaking articles where many editors are repeatedly contributing and interacting with many other editors in a short period of time should lead to trajectories in which editors' modifications of other editors' contributions are highly proximate to each other. In a "tight" article revision network, editors'

saved revisions have a higher average closeness than in a "loose" revision network. Therefore, breaking articles should have higher closeness centrality than non-breaking articles.

- The *average betweenness* of the network represents the tendency for a network to have many "brokers" or connecting nodes that would otherwise be unconnected (S Wasserman & K Faust, 1994). Editors who make only a single contribution or do not repeatedly edit an article create structural holes with every other editor in the network, inflating their betweenness centralities. Alternatively, editors who revise many other editors' contributions close these structural holes and create dense network patterns with little opportunity to "broker". Revision networks of articles like Air Philippines Flight 541 (Figure 6-2D) that exhibit a "loose" pattern of activity, should have high average betweenness centralities because removing any editor would make it impossible for others to connect since there are no loops (perhaps indicating a lack of central coordinating editors). Article revision networks with a "tight" pattern of activity like Air France Flight 358 (Figure 6-2B) have low average betweenness since there are several editors connecting the entire network together providing multiple paths for interaction. Breaking articles should have lower betweenness centralities than non-breaking articles.
- The *average clustering* of the network captures the tendency of the immediate neighbors (alters) of an editor to edit each other's revisions as well. As before, editors making only few contributions or contributions in a short period of time are likely interacting with only a single other editor rather than interacting with many editors. The lack of these connections to other editors will create sparse networks with low average clustering. Trajectories in which editors make contributions modifying many other editors contributions should create denser networks with higher average clustering. Article revision networks with a "tight" pattern of activity and more neighboring editors making revisions with each other have higher clustering coefficients than those with a "loose" pattern.

The independent variables used to model these dependent variables reflect the size of the collaboration, the breaking nature of the collaboration, and potential contextual confounds about the severity or

importance of the news event. Specific attributes like the number of fatalities, location, and incident date were manually recorded for each of these articles.

- *Editors* is the number of unique contributors to the article—effectively the number of nodes in the network. I acknowledge that several of the dependent variables (network statistics) are affected by the number of vertices on the graph; this variable is included as a control rather than explanatory factor.
- *Creation lag* is the difference between the timestamp of the first contribution and the date of the incident. This captures how long after the incident occurred the article appeared in Wikipedia. Breaking articles have small creation lags and non-breaking articles have large lags.
- *Fatalities* is the number of individuals killed as a direct result of the incident. The severity of an incident may affect the amount of attention paid to that incident, leading to an increase in editor contributions for incidents with more fatalities.
- *OECD* is a dummy variable that codes for whether the crash occurred in one of the 32 member nations of the Organization for Economic Cooperation and Development (OECD) as of 2010. Because Wikipedia editors display a self-focus and local bias of article topics (Hecht & Gergle, 2010), events that occur outside of developed nations may receive different levels of attention. With the fatalities variable above, this variable is a control on the news values found in both the Wikipedia policies governing events.

Results

The article revision networks exhibit strong relationships between creation lag and how “tight” or “loose” the pattern of revision activity is in each network.Model 1 in Table 6-1 reveals that revision networks for articles with greater elapsed time between the creation of the article and the incident date (creation lag) have significantly longer diameters, even after controlling for the number of editors in the network. Conversely, articles that are created immediately after an incident exhibit much “tighter” revision patterns, resulting in denser networks. In Model 2, longer creation lags should result in revision networks

with less closeness centrality. However, the effect was not significant and therefore I did not find support for this claim. Model 3 examines average betweenness centrality, and articles with higher creation lags have fewer ties based on revisions by editors who had previously contributed to that article. Higher betweenness scores suggest the presence of more and longer chains in the revision network. Finally, Model 4 suggests articles about incidents that occurred farther in the past exhibit lower average clustering coefficients. In other words, the groups of editors whose revisions are modified by a single editor are less likely to mutually modify others' revisions. In contrast, the breaking news editors exhibit a greater tendency to revise in groups or clusters when engaging Wikipedia's editorial process.The number of fatalities had little effect on the revision network statistics and was only significant for average closeness. However, airline crashes that occurred in OECD nations had smaller diameters and smaller betweenness, suggesting that editorial engagement may take different forms across levels of socio-economic development.

Like the dynamics observed in Chapter 5, the revision network structures of breaking news articles should change over the lifecycle of the article. In particular, (1) what phases in breaking articles' histories contribute to the characteristic "tight" patterns, and (2) do the network structures found throughout these articles' revision histories become "looser" over time and look more like their non-breaking counterparts.

Examining the set of 76 articles authored within 96 hours of the incident, and construct three article revision networks for each article, one for each "breaking phase". *Breaking phase* is a categorical variable that codes each revision of a breaking article for whether it occurred in the first phase of the article's existence (days 0-7), the second phase (days 7-187), or the third phase (after day 187). This variable captures changes in the editor interaction patterns over the lifecycle of an article. Averages of 55.8% of all revisions are made in Phase 1, 17.3% in Phase 2, and 25.9% in Phase 3. As before, to control for exogenous factors such as the severity of the event, systematic biases in editor attention based on geography or socio-economic status, and the number of contributors to the article, control variables

similar to those used in the previous analyses are used. However, a hierarchical approach is necessary to account for repeated observations for each article at different points in time (breaking phase).

Network statistics (diameter, closeness, betweenness, clustering, editors) change as the network changes over time, while variables at the article level (wiki-age, fatalities, OECD) remain constant. Thus, the primary effect being measured is differences in articles' diameter, closeness, betweenness, and clustering at each of the three points in time, controlling for the age of Wikipedia at article creation, fatalities, location, and number of editors. The regression coefficients for this multilevel model are reported in Table 2 and the changes across time for the corresponding dependent variables are shown in the "Phase" parameters. A Hausman test comparing a random intercept model to a random coefficient model (allowing the coefficients for phase to vary over time) caused neither significant changes in valence or significance nor showed a significant improvement in fit; the results for the simpler random intercept model are reported instead.

Examining the breaking phase coefficients, significant differences are observed within articles between the first phase of the collaboration and the second, and between the first and third phases. The diameters of breaking articles' revision networks are significantly larger in both the second and third phases than in the first phase, holding the other predictors in the model constant. Similarly, the average closeness centralities of the networks decrease for the later phases as compared with the initial collaboration. The article revision networks exhibit significantly higher average betweenness centralities in Phase 2 and Phase 3 as well. Finally the average clustering coefficients of breaking articles are significantly smaller in latter phases of the collaboration. These results suggest breaking articles exhibit unique revision network patterns in the aftermath of an incident, but over time the activity of editors contributing to the revision network increasingly resembles the revision network patterns of non-breaking articles.

	Model 1: Diameter	Model 2: Closeness‡	Model 3: Betweenness‡	Model 4: Clustering‡
Creation lag‡	0.164* (2.47)	-0.045 (-0.75)	0.094*** (3.49)	-0.370*** (-4.99)
Fatalities‡	0.017 (0.24)	0.164** (2.79)	-0.036 (-1.37)	0.027 (0.38)
OECD	-0.142* (-2.14)	0.097 (1.62)	-0.067* (-2.52)	0.051 (0.69)
Editors‡	0.585*** (8.58)	-0.690*** (-11.17)	0.990*** (36.06)	-0.084 (-1.10)
R^2	0.284	0.412	0.884	0.112

N = 229, ‡ log-normalized * p<0.05, ** p<0.01, *** p<0.001

Table 6-1: Standardized beta regression coefficients (t statistics) for all articles

	M5: Diameter	M6: Closeness‡	M7: Betweenness‡	M8: Clustering‡
Wiki-age	-2.9e-3*** (-4.1)	1.26e-4* (2.16)	-2.18e-4** (-2.76)	2.26e-4 (0.44)
Fatalities‡	0.0170 (0.29)	-1.02e-3 (-0.20)	2.72e-3 (0.40)	7.66e-4 (0.02)
OECD	-1.46 (-1.53)	0.103 (1.27)	-0.086 (-0.78)	0.291 (0.41)
Editors‡	3.33*** (10.59)	-1.38*** (-57.78)	1.51*** (45.48)	2.17*** (8.43)
Phase 1 (days 0-7)	.	.	.	.
Phase 2 (days 7-187)	2.37** (2.98)	-2.68*** (-4.95)	0.358*** (4.65)	-1.78* (-2.41)
Phase 3 (days 187+)	6.23*** (8.15)	-0.45*** (-8.69)	0.603*** (8.18)	-4.36*** (-6.08)

N = 229, ‡ log-normalized * p<0.05, ** p<0.01, *** p<0.001

Table 6-2: Regression coefficients (t statistics) for breaking articles.

Study 2: Repeat Collaborations

The previous study analyzed articles as atomistic collaborations, but the coauthorship network analysis in Chapter 5 made it clear that editors contribute to many different articles. This study attempts to integrate both these approaches by looking at the patterns of editor interactions across multiple article collaborations. This second study analyzes 1.03 million revisions made by 158,384 unique users to 3,233 English Wikipedia breaking articles (n = 1538) and non-breaking articles (n = 1695) and expands the genre from airline disasters to a much larger variety of articles about disasters, catastrophes, and conflicts from 1990 through 2011. These articles include conflicts (e.g., wars, battles, political unrest), crimes (e.g., murders, kidnappings, and terrorism), fires (e.g., building fires, wildfires, explosions), health disasters (e.g., disease outbreaks), industrial accidents (e.g., spills, mine collapses), natural disasters (e.g., hurricanes, earthquakes, tornadoes), and transportation accidents (e.g., airplane crashes, train collisions, road accidents). These categories are categorized by incident year such that the Deepwater Horizon oil spill will appear in the "2010 industrial accidents" category, reflecting the fact that the incident itself occurred in 2010 even if the first or subsequent revisions to the article occurred at subsequent times.

Two analyses of article trajectories were conducted using this data. The first constructs a *repeat interaction multigraph* to assess the structure of editor interactions across all the articles. This multigaph reveals the extent to which editors not only collaborate on many of the same articles, but repeat their interactions with each other across multiple (predominately breaking news) articles. These patterns of repeated interactions across multiple articles are themselves potential features of organizational regeneration as editors reconstitute patterns of coordinating with each other across distinct articles. Do these repeated interactions occur on predominantely breaking or non-breaking articles and who are the editors who occupy the most central positions in these networks? Visualization and descriptive statistics of this complex multigraph demonstrate the coherence of editors' interactions across arrticles of particular types as well as the embeddedness of these interactions within emergent patterns of community action.

Using structural features derived from this multigraph analysis as well as other behavioral and contextual attribtues, this analysis then develops a statistical model to predict what leads to editors contributing to multiple articles. The combination of these multigraph structural features with other behavioral and contextual attributes unpacks whether editors' experiences within article trajectories lead them to edit more articles. Structural features such as degree, fraction top neighbors, and average neighbor degree capture the extent to which the connectivity of the editor and the editors whose work she modifies across articles influences the tendency to edit multiple articles. For example, editors with connections to many unique editors may make contributions that allow them to participate at a shallow level across many articles while editors. Alternatively, editors with a few connections to many other unique editors may prefer to work in conjunction or response to a particular editor, which therefor constrains the number and types of articles she edits. Editors modifying the contributions of other editors who are themselves well-connected (revising many other editors' contributions) may be exposed to unique practices or information that can then motivate editors to implement those practices on more articles than editors who were not exposed to well-connected editors. Contextual and behavioral features of editors such as the how early they join collaborations, how much time they elapse between edits, or how much content they add to articles may also reveal tendencies for involved editors to contribute to many article collaborations.

To test the extent to which contextual features of revision histories and the structure of multiple article trajectories influences how many articles editors revise, a linear regression model for each set of breaking and non-breaking articles is specified. Each observation is a record of an editor's revision history behavior and structure in the repeat interaction multigraph. The revision history of each article as well as its article trajectory encodes several important attributes about the structure and dynamics of the collaboration. With the exception of *bytes added*, each of these metrics is log-transformed to correct for strong skews in distributions. For each editor in each article's trajectory, I record the following metrics:

- *Article Count* is the count of the unique articles an editor has revised. Editors who have made at least one change to only one article will have an article count of 1, while editors who have made at least one change to 10 articles will have a count of 10.
- *Degree* is the number of connections an editor has to other editors in the article trajectory. Because the article trajectory is directed, degree is the sum of an article's in-degree and out-degree and the value is normalized to the size of the graph. A normalized degree of 1 indicates the node is connected to every other node in the network.
- *Timestamp* records age of an article by measuring the time elapsed since the article was created. The first revision to an article necessarily has a timestamp of 0. For editors who have made several revisions, I use the minimum value of timestamp to capture how early editors join collaborations.
- *Edit lag* is the amount of time elapsed since the prior revision to the article was made. An edit lag of 1 reflects a revision made one second after the previous article was committed. Editors fully engaged in the collaboration will have smaller edit lags as they make changes immediately following others' revisions. For editors who made several revisions, I use the median value of edit lag to capture a central tendency of editors' engagement for skewed distributions.
- *Bytes added* is the change in the length of an article compared to the previous revision. Revisions removing content add negative bytes. Editors who add large amounts of content to articles will have large positive bytes added. I use the median value of bytes added to capture the central tendency in a skewed distribution of positive and negative values.
- *Top editor* is a dummy variable coding whether or not the editor is among the top 10% of editors as measured by degree.

- *Fraction top neighbors* is the fraction of a node's neighbors who are among the top 10% of editors by degree. A fraction of 1 means all of a node's neighbors are in the top 10% of editors while a fraction of 0 means none of a node's neighbors are in the top 10% of editors.
- *Average neighbor degree* is the average of an editor's neighbors' degrees. High values of average neighbor degree mean the editor it connected to other well-connected editors.

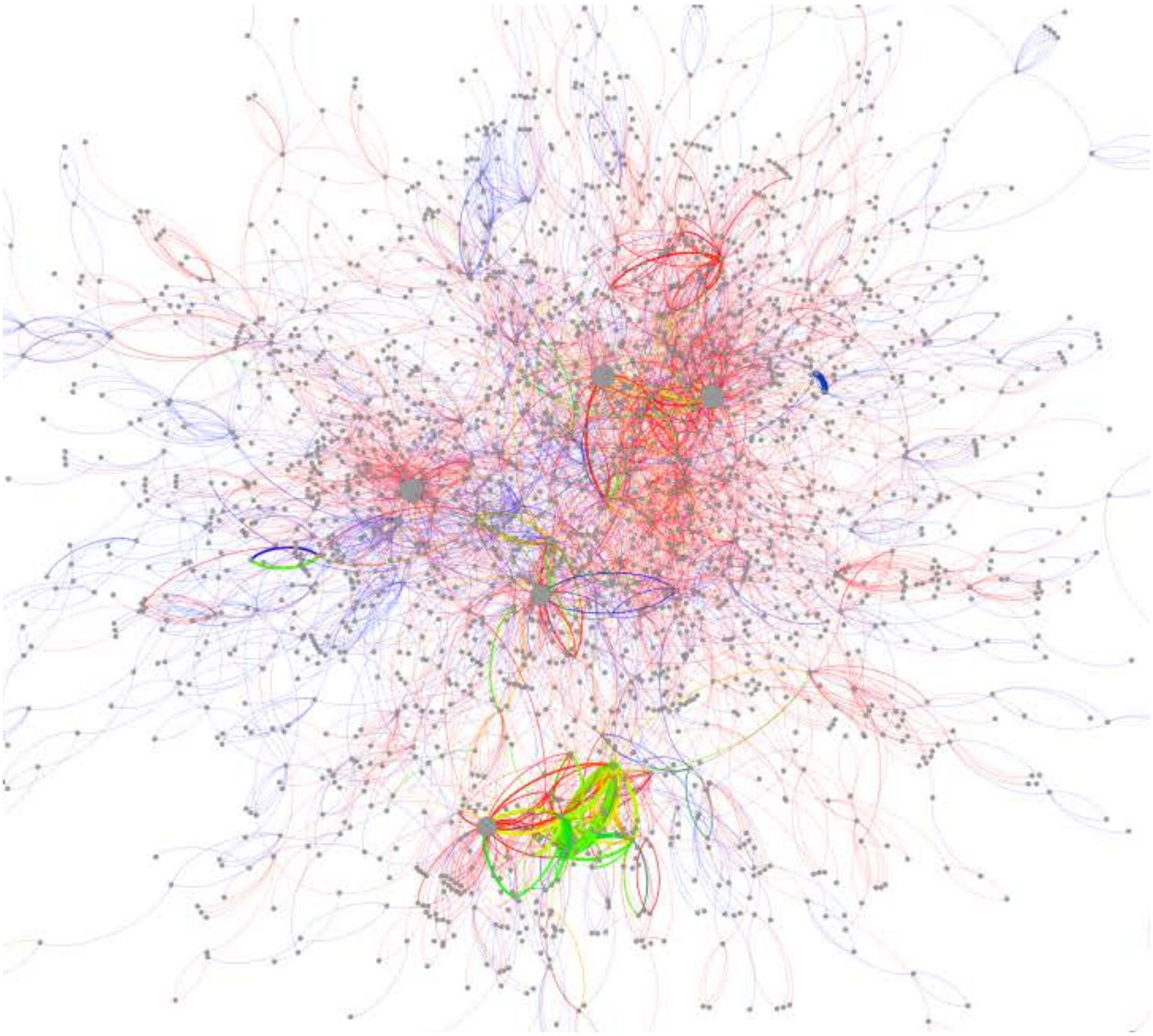


Figure 6-3: Largest component in repeat interaction multigraph. Editors (in grey, sized by degree) are linked with weighted edges if their interactions with another editor appear on two or more distinct articles. Bluer edges have greater fractions of non-breaking article interactions, redder edges have greater fractions of breaking article interactions.

Results

Article trajectories for each article were constructed and aggregated into a directed multigraph where pairs of editors can have multiple edges, each edge corresponding to interactions between the same editors on different articles' trajectories. Figure 6-3 visualizes the largest component of this repeated interaction multigraph. The graph contains 3,314 editors and 6,451 edges where the dyad interacted the same way on two or more articles. Several features of this article trajectory are notable. First, the largest component includes 92.3% of the nodes and 98.4% of edges in the total repeat interaction graph. Because editors' interactions across collaborations are highly connected and embedded rather than isolated and atomistic, interactions in article trajectories are not spurious but rather, many editors jointly monitor and edit articles with each other. Second, the vast majority of interactions between the same editors across different articles involve breaking articles (redder edges) rather than non-breaking articles (bluer edges). The dominance of repeat interactions on breaking articles in the repeat interaction multigraph implies breaking articles are not one-off collaborations but involve repeated interactions among the same editors across different articles. Third, there is a distinct cluster of thick, dense, and green interactions at approximately 6 o'clock in Figure 6-3 removed from (but still connected to) the larger collaboration. These editors' interactions reflect their specialization in a particular genre of articles; tropical storms and hurricanes, in this instance. The thick and yellow to green colors reflect these editors' repeated interactions as they attend to each other's revisions as they collaborate on a specific category of articles containing a mix of breaking and non-breaking articles. The article trajectory approach suggests a novel way to mine large-scale digital behavioral traces to surface interactions among collectives of individuals engaged in mutually shared practices.

Turning to the statistical model of whether structural features in these repeat interactions influence editors' tendencies to contribute to multiple article collaborations above and beyond behavioral and contextual patterns. A random intercept negative binomial model captures the count-level nature of the dependent variable as well as the multi-level nature of the intependent variables in which editors are

embedded within articles. The eight independent variables are estimated as fixed effects. AIC and log-likelihood statistics confirm the multi-level model provides a better fit than a simple linear regression model and correlation among the fixed effects is negligible. The estimates for each model are summarized in Table 6-3.

	M9: Breaking Article Count‡	M10: Non-Breaking Article Count‡
Degree‡	1.656 (26.13)***	0.6195 (11.763)***
Minimum Article Age‡	-0.02159 (-20.26)***	-0.006957 (-6.492)***
Median Edit Lag‡	0.03350 (36.12)***	0.02537 (25.272)***
Median Bytes Added	1.471e-06 (7.99)***	3.983e-07 (1.555)***
Top Editor‡	0.3826 (44.55)***	0.2105 (19.114)***
Fractoion Top Neighbors‡	-0.1216 (-18.38)***	-0.06261 (-7.214)***
Average Neighbor Degree‡	0.005765 (2.07)*	-0.03378 (-6.717)***
Intercept‡	0.9063 (48.81)***	0.6705 (29.064)***

‡ log-normalized; * p<0.05, ** p<0.01, *** p<0.001

Table 6-3: Regression coefficients (t statistics) for editors in breaking article (n = 173,292) and non-breaking article (n = 81,682) article trajectories.

The later an editor joins collaborations (*Minimum Article Age*), the less likely he or she will contribute to separate articles. Editors who join the collaboration after much of the early work shaping and framing the article has been done are potentially less engaged. This engagement deficit, in turn, manifests as a lack of participation revising other articles. This engagement interpretation is complicated by the estimates for median edit lag: positive coefficients suggest the presence of deliberative editors who wait longer to make a revision after previous revisions are in fact more likely to edit many articles. Editors who add more content to articles (*Median Bytes Added)* are also more likely to participate in many article collaborations. Taken together, these results suggest editors who join article collaborations early, are not intensively engaged with the editing of the article, and make larger changes in their revisions are more likely to participate in many article collaborations. These editors are potentially "wordsmiths" who are not involved in the substantive coordination and management of content, but move fluidly between many articles without becoming engaged and leave the integration of their contributions to others.

Turning to the structural context of editors' interactions, in both breaking and non-breaking articles, the more connections an editor has to other editors within an article trajectory (*Degree*) is correlated with the number of articles the editor revises. The most central editors (*Top Editor*) who interact with many other editors necessarily make many revisions over the course of an article but this behavior is also correlated with a tendency to edit many articles. Editors surrounded by many other top editors (*Fraction Top Neighbors*) are less likely to edit many articles and the effect is approximately two-times stronger for breaking articles than for non-breaking articles. This could be an artifact of vandalism, inexperienced, or contentious editing in which the interaction between editors is one of highly active editors reverting every contribution made by an editor that leads them to being disciplined or disheartened which prevents him or her from participating in other collaborations.

Finally, breaking and non-breaking articles diverge in how the connectivity of an editor's neighbors (*Average Neighbor Degree*) influences the number of articles that editor revises. Editors of breaking

articles who interact with other well-connected editors are more likely to revise many articles, while interacting with well-connected editors diminishes editors' tendencies to work on many non-breaking articles. Determining the direction of this effect is difficult to disentangle with the model as it may be the case that interacting with prolific editors on breaking articles provides encouragement to pursue other collaborations or alternatively, editors who contribute to many articles fulfill particular brokerage roles like administration, mediation, or copy-editing which leads them to interact with very-active editors in a collaboration.

Discussion

Unlike prior work looking at co-authorship patterns *across* Wikipedia articles (Bellomi & Bonato, 2005; Brandes, Kenis, Lerner, & van Raaij, 2009b; Keegan et al., 2012), article trajectories are a method for capturing the collaboration structure *within* Wikipedia articles. The trajectories of Wikipedia articles about breaking news events reveal they have unique structures compared to non-breaking articles reflecting their coordination demands. Study 1 found breaking articles' trajectories are initially characterized by highly clustered and centralized interactions with other editors but these features diminish over time and regress to structures found on non-breaking articles. Looking across different articles' trajectories in Study 2 found a well-connected latent collaboration structure of editors interacting with each other across articles. Finally, this approach identified significant behavioral and structural features correlated with the tendency for editors to participate in many article collaborations.

These results illustrate how breaking articles are a unique type of collaboration context that differs from traditional Wikipedia articles that center on activities, events or topics that can rely on pre-existing knowledge. While these findings complicate the coordination models outlined by Kittur, Kraut, and colleagues (A Kittur & RE Kraut, 2008; Kittur et al., 2009), they also reaffirm the importance of understanding the interdependence between task demands and coordination methods. This sociotechnical trajectory approach extends this work by allowing a more critical examination of the particular interaction patterns editors adopt in response to the coordination demands of high tempo collaborations and provided new insights into how editors structure these time-dependent knowledge collaborations. The characteristic "tight" revision patterns observed on breaking news articles may be attributable to different coordination processes that are invoked following an incident but as the salience of the event fades and more information comes to light, the pattern of revisions begins to revert to the looser revision networks more typical of non-breaking articles. These structures may be an artifact of editors' joint attention as they cope with the influx of new information and collaborators. As more accurate and consistent accounts of the

event begin to surface in other reliable outlets, this impetus subsides and the editors can begin to focus more on the structuring that occurs with more typical article types.

As Study 2 demonstrates, breaking and non-breaking article collaborations are not isolated but embedded in a larger community or collectivity of practice (Bryant et al., 2005; Lindkvist, 2005) where editors' interactions with each other occur repeatedly across breaking articles, in particular. This article trajectory approach also identified editors mutually engaged in a shared practice of editing articles about hurricanes as distinct from the rest of the repeat interaction multigraph. Taken together, this suggests the possibility of leveraging large-scale behavioral data analysis methods like sociotechnical trajectories to highlight individuals and groups for follow-on qualitative studies of their identities, motivations, and practices in Chapter 7.

Because this approach using article commits to the revision histories did not capture changes in the content of the articles, this analysis did not examine interactions between editor contributions and content (Antin, Cheshire, & Nov, 2012; Brandes et al., 2009a; Halfaker, Kittur, & Riedl, 2011; Rzeszotarski & Kittur, 2012). Contributions related to changes in punctuation or reverting vandalism are certainly less demanding tasks than authoring and integrating novel content but are nevertheless lumped together as equal revision processes in this approach. This analysis also did not examine the success of these collaborations (e.g., article length, rating, page views) that would allow us to separate out intense collaborations resulting from tendentious editing or controversial topics versus substantive coordination. This approach likewise only examined the English Wikipedia. Future work could establish whether these temporal and structural correlates are also found in other genres, languages, and cultures.

Chapter 7: Editor trajectories and social roles

The analyses thus far have focused on articles as the unit of analysis. Chapter 5 examined the structure of article collaborations and found breaking article collaborations rapidly emerged into connected components unlike non-breaking or historical articles. Chapter 6 found that the patterns of work and editor interactions within these articles are also different on breaking articles as well as some interesting evidence of editors having substantive patterns of interaction across articles. This chapter turns the attention away from the articles themselves to examine the editors participating in breaking news collaborations. To scope this analysis, the 2011 Tōhoku earthquake and tsunami is used as a case study to understand the trajectories of the influential editors preceding their involvement in this collaboration as well as unpacking the roles these editors played within the collaboration itself. The Tōhoku disasters are a unique case to employ because they are simultaneously an extremely large-scale and unexpected disaster, but also an event that comprised of several significant and overlapping events involving the nuclear disasters at Fukushima power plants over the following days. This case is illustrative of the fluidity with which collaborators on breaking news articles move between topics. Using the exhaustive longitudinal data in the revision histories of Wikipedia articles about breaking news events as well as the contribution histories of editors themselves will help us to understand the types and changes in the roles inhabited by collaborators on a breaking news event.

This chapter extends the trajectory method in the previous chapter to examine "user trajectories" as the progression of user interactions with artifacts over time. Within a user trajectory, a user's interaction with the same artifact at different points in time will be linked to the other artifacts the user has interacted with in the intervening time. User trajectories thus trace the temporal path of users "moving through" the artifacts with which they have interacted. The trajectory of any single user or artifact reflects its provenance from a source through a sequence of ensuing actions. The articulation of a user trajectory through the temporal sequence of articles or an article trajectory through the temporal sequence of users

captures important temporal contexts and dependencies that are omitted through approaches which only look at ahistorical collaboration relationships like those in Chapter 5.

This chapter begins by tracing the evolution of the main article's[44] development as well as the similarities and dissimilarities in the trajectories of other articles related to the disaster. The changes in the attention and editing behavior over the course of the collaboration will provide the context for identifying editors involved in distinct parts of the collaboration. Using a variety of editor-focused metrics such as the number of contributions made, articles contributed to, and concentration of edits, several influential editors are identified. Using the trajectory approach articulated in the previous chapter, the trajectories of these important editors are constructed and examined to identify features which pre-dispose editors to work on breaking news articles and understand how patterns of users' contributions change over time. This editor trajectory analysis of their previous work is thematically examined for engagement in particular types of tasks to help us understand whether users are enacting similar or dissimilar social roles across articles. Throughout these analyses, I qualitatively examine how users' inhabit distinct roles based on the patterns of their editing behavior over time. This mixed methods approach integrates descriptive statistics, network analysis like socio-technical trajectories, and qualitative examination of the routines, practices, and dispositions of editors captured from the digital traces of their contributions.

Data and methods

The revision histories for articles related to the 2011 Tōhoku earthquake and tsunami were extracted from the Wikipedia API using the "python-wikitools" library and custom scripts to access and parse output (see Appendix B for code). The data contained the article's name, article ID, revision comment, user ID, username, and timestamp. The seed list of 74 articles were drawn from the "Tōhoku earthquake and tsunami" category on the English Wikipedia on April 3, 2011. Articles were assigned binary "old/new" categories based upon their timestamps. Articles which existed prior to 05:30 UTC on March 11, 2001

[44] http://en.wikipedia.org/wiki/2011_T%C5%8Dhoku_earthquake_and_tsunami

were classified as “old” articles existing before the earthquake and articles created afterwards were classified as “new” articles reflecting their creation after the earthquake. Following the earthquake and tusnami, these old articles needed to be updated while the new articles needed to be written. This division of labor between the old and the new introduced interesting coordination problems about which editors would make the necessary updates to the old pages, who would focus on integrating the information on the new pages, and who would ensure the consistency of information across both.

Out of the 74 articles that were categorized by Wikipedians as affected by or related to the earthquake, tsunami, or nuclear incidents (including the articles about the events themselves), 50 articles existed before the earthquake. Most of this subset are articles about the Japanese towns, cities, prefectures but also includes articles about the nuclear power plants, seismic features of Japan, and other points of interest such as railways, stadiums, and airports. Between July 26, 2001 and March 10, 2011, 2,997 unique editors had made 8,074 revisions to these articles. As impressive as the collaborative effort over the decade may seem, in the four weeks (March 11 – April 2) following the earthquake 1,140 unique editors made 3,792 revisions to these pre-existing articles. The 24 articles that were created after the March 11 earthquake and tsunami include the articles about the event itself, nuclear accidents, humanitarian response, international reactions, radiation effects, evacuations, event timelines, and a list of damaged cities and towns. 2,439 unique editors made 11,709 revisions to these articles alone. As will be discussed below, there is some substantial overlap between the editors contributing content to both the new and old articles. In total, 3,217 unique editors had made 15,379 revisions to the articles affected by or related to the earthquake and tsunami from March 11 to April 2.

The socio-technical trajectory method introduced in Chapter 6 was used to understand the progression and context of an editor’s contributions over time by evaluating the distribution and regularity with which they immerse themselves in breaking news article collaborations. Analogous to artifact trajectories, “user trajectories” are the temporal sequence of actions performed by a single user to one or more artifacts.

User trajectories are articulated through the progression of user interactions with artifacts over time. Again, this time-ordered progression is represented as a directed graph. In the context of Wikipedia, the nodes in the user trajectory are articles, and directed edges between articles reflect an editors' change from editing one article to editing another article at a subsequent point in time.

Background

On March 11, 2011 at 5:46 UTC, a 9.0-magnitude undersea mega-thrust earthquake occurred approximately 72 kilometers east of the northeastern Tōhoku region of Japan. The earthquake vertically displaced 5-8 meters of seabed that in turn created a tsunami in excess of 10 meters in height. The tsunami washed over an extensive network of seawalls 10 to 30 minutes after the earthquake and inundated more than 470 square kilometers of Japanese coastline. Two separate power plants housing ten boiling-water nuclear reactors in the Fukushima prefecture suffered severe damage as a result of the tsunami. Over the course of the next week, the loss of cooling systems at the plants caused the reactor cores and spent fuel rods to overheat and resulted in partial core meltdowns and the release of radioactive materials at three of the reactors.

Japanese Prime Minister Naoto Kan called this cascade of disasters "the most difficult crisis" for Japan since World War II. As of April 3, 2011, over 12,000 people were confirmed dead, more than 2,800 injured, and more than 15,000 missing. Preliminary estimates of the damage exceed US$309 billion making the Tōhoku earthquake and tsunami the most expensive natural disaster ever recorded. In the face of this staggering human and environmental catastrophe, Wikipedia's coverage of this complex and on-going sequence of events is notable for both the timeliness of integrating coverage from disparate sources (and languages) as well as the sustained and intense collaborative effort to document the event as it unfolded. The Japanese Wikipedia created an article about the event by 5:57 UTC, just minutes after the 6-minute earthquake stopped. An English Wikipedia article was created at 6:18 UTC which included two

citations to the US Geological Survey and the Integrated Tsunami Watch Service warning of a potential tsunami.

At 6:29 UTC, the article was nominated as a candidate for the "In the News" template to appear on the English Wikipedia's homepage. By the time the article was promoted at 7:58 UTC, the nomination itself had been vetted by at least 12 editors. In those 90 minutes, the English Wikipedia article alone attracted 220 revisions from 82 editors. As a point of comparison, although wire services had reported the earthquake within minutes, *The New York Times* did not file a full story until 7:35 UTC. Articles about the catastrophe were also available on Wikipedia in Arabic, German, Spanish, Korean, Norwegian, Russian, Finnish, Swedish, Turkish, Ukrainian, and Chinese by 8:00 UTC. The remainder of this analysis will focus on the English Wikipedia article alone.

Bipartite coauthorship network

While traditional network analysis emphasizes the relationships amongst a single type of actor (i.e., people-to-people), two unique types of actors exist within the Wikipedia corpus: editors and articles. Because it is nonsensical to assert that an article may contribute to another article or an editor may edit another editor, we model the collaboration as a bipartite graph in which the nodes can be partitioned into exactly two mutually exclusive sets of actors (S. Borgatti & M. Everett, 1997). Thus, a link exists between an editor node, *E*, and an article node, *A*, if *E* made a contribution to *A*. However, neither *E*-*E* nor *A*-*A* links may exist. This bipartite structure is also alternatively referred to as a two-mode or affiliation network.

All 15,379 revisions made after 05:30 UTC on March 11, 2001 to the 72 Tōhoku articles were integrated into a weighted bipartite coauthorship network. As was the case with the coauthorship networks in Chapter 5, the edges between the two distinct types of nodes (editors and articles) represent the number of times an editor contributed to that article, also known as the edge weight. Although methods exist for

“projecting” or decomposing two-mode networks into traditional one-mode networks (S. Borgatti & M. Everett, 1997), these conversions result in the loss of information about editor-article edge weights and are not appropriate for our analysis.

The weighted bipartite network for the 72 articles, 6,003 editors, and 8,395 unique links between editors and articles is visualized in Figure 7-1. The sizes of the nodes are proportional to the number of connections that article/editor has to other editors/articles (degree). The widths and opacity of any link is proportional to the number of times an editor contributed to that article after March 11 (edge weight). The eight most-edited articles are labeled as well as the three in-line casualty templates clustered together at approximately 11 o’clock. The visualization highlights several important features of the collaboration. First, most of the unique editors contributing to any article tend to edit that article and nothing else in the corpus. This is seen with the “halos” of blue editor nodes surrounding the articles like the “2011 Tōhoku earthquake and tsunami.” This heterogeneous distribution of connections is characteristic of many complex networks.

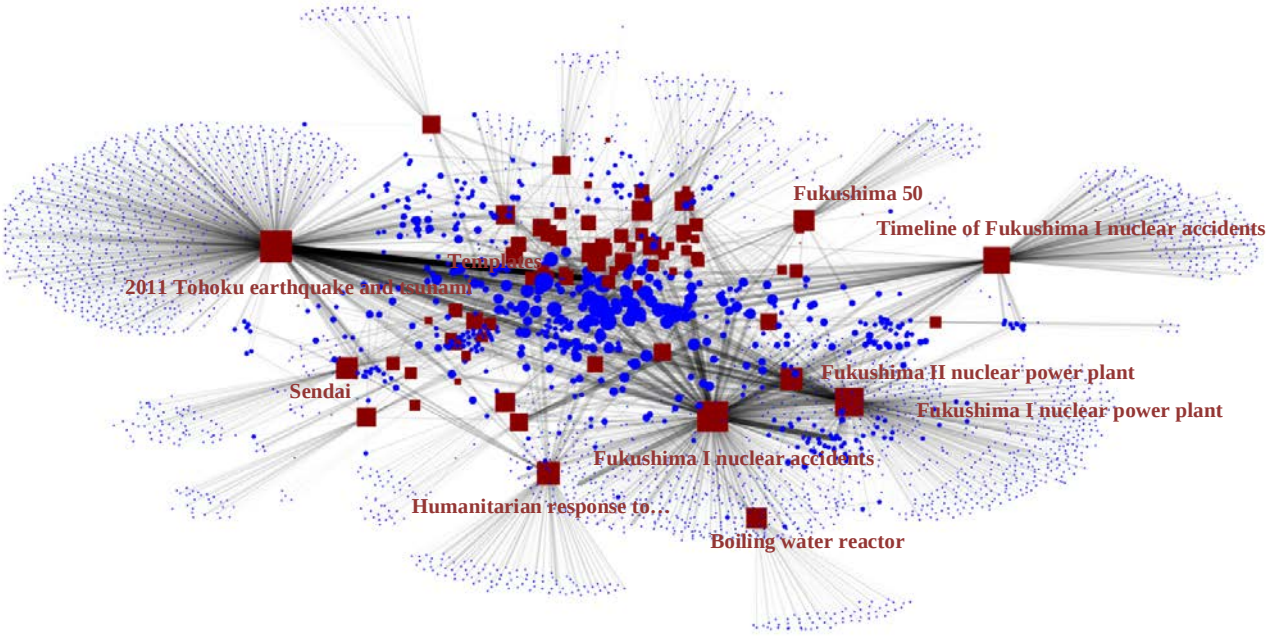


Figure 7-1: Two-mode network visualization of the 72 articles (red squares) and 6,003 editors (blue circles). Black links indicate an editor contributed to that article after March 11, 2011.

Second, the articles themselves tend to cohere thematically because similar subgroups of editors contributed to many of these articles. For example, the articles “Fukushima I nuclear accidents”, “Fukushima I nuclear power plant”, “Fukushima II nuclear power plant”, and “Boiling water reactor” share many of the same editors and as a result are clustered together at between 4 and 5 o’clock on the visualization. Similarly, the cluster of red nodes between 11 and 12 o’clock are the dozens of articles about the towns, cities, and prefectures damaged by the earthquake and tsunami. These patterns suggest that rather than editing articles randomly, some contributors exhibit a tendency to specialize in a particular subtopic such as nuclear-related issues or Japanese geography. Examples of these specialist editors can be seen above and below the two Fukushima nuclear power plant articles where editors contributed to these topics but not other corpus topics.

Third, the cluster of blue nodes at the center of the visualization represents the subgroup of editors who made many contributions to many articles. This strong central core of generalist editors consists of the individuals who were not only closely monitoring and editing content across many articles but also intensively editing articles by making many contributions as well. There was a strong and significant correlation (r=0.588, p<0.001) between the total number of revisions made by editors and the number of articles they edited (see Figure 7-2). Examining the editors who contributed to the most articles, the editors who made the most contributions, and the editors with the history of distributing their editing activity evenly reveals interesting patterns. In the tables below, the degree reflects the number of unique articles to which the author contributed, the Gini coefficient reflects the concentration of an editors efforts on a single or multiple articles, and the revisions reflect the total number of contributions made. A Gini coefficient of 0 implies that an editor’s contributions are exactly evenly distributed across all the articles she edits while a Gini coefficient of 1 implies that an editor’s contributions are wholly concentrated in a single article. Table 7-1 below summarizes the 33 editors who made contributions to 12 or more articles.

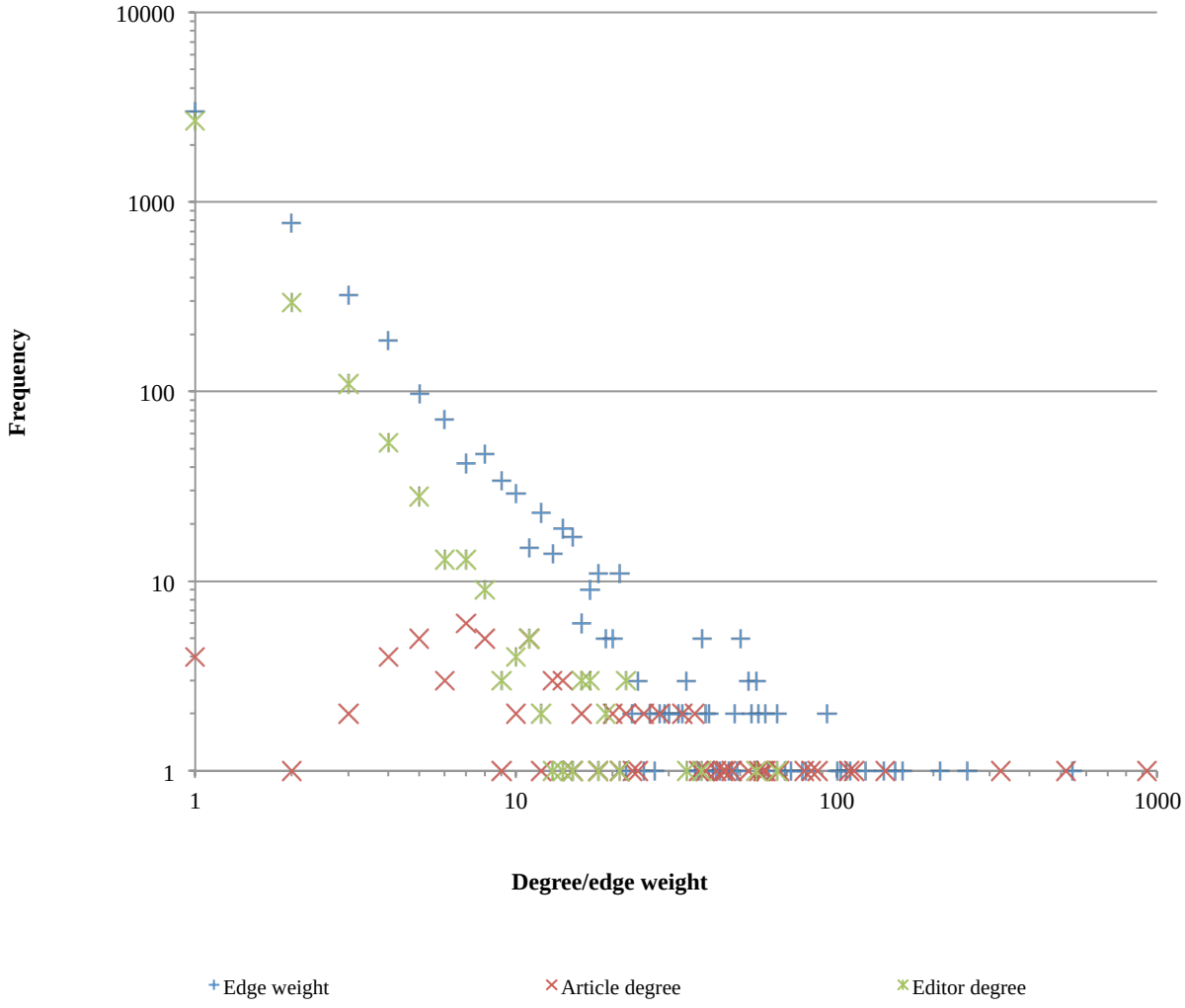


Figure 7-2: Degree and edge weight distributions for the bipartite network in Figure 5 on log-log scales.

Name	Degree	Gini coefficient	Total revisions
SmackBot	35	0.66	172
Luckas-bot	34	0.47	117
ACSE	34	0.76	356
TexasAndroid	32	0.18	40
RjwilmsiBot	24	0.37	54
AnomieBOT	23	0.62	103
Yobot	23	0.56	63
Cydebot	23	0.16	28
The Anomebot2	22	0.24	39
EmausBot	22	0.4	50
Andycjp	20	0.57	142
Lightbot	20	0.13	24
Cla68	19	0.74	264
WikitanvirBot	18	0.57	73
Johnfos	18	0.68	403
Neier	17	0.49	134
66.220.113.98	17	0.69	170
ChuispastonBot	16	0.39	33
Jllm06	15	0.06	16
Rich Farmbrough	15	0.34	25
Ohconfucius	15	0.55	55
184.144.160.156	14	0.6	96
ZéroBot	14	0	14
Theanphibian	14	0.53	115
ClueBot	14	0.69	115
Hmains	13	0.57	38
Xqbot	13	0.31	23
Rjwilmsi	13	0.35	27
TXiKiBoT	12	0.37	24
Colfer2	12	0.69	285
FrescoBot	12	0.3	20
Ripchip Bot	12	0.56	38
Materialscientist	12	0.74	112

Table 7-1: Editors with most connections.

Automated bots make up 17 of the top 33 editors, reflecting the growing influence these algorithmic actors have on Wikipedia's collaborative and editorial processes. Bots make about 16% of all edits to the encyclopedia and they occupy 17 of the top 20 most prolific editor positions across the encyclopedia (Geiger, 2009). These bots fulfill a variety of roles such as detecting and reverting vandalism, performing repetitive tasks, and increasingly complex functions such as reporting human users to administrators for violating rules. These bots operate autonomously and independently focused on distinct and mundane tasks. SmackBot was designed to perform over 50 distinct tasks related to automatically reformatting or cleaning up content in articles,[45] Luckas-bot created interwiki links to connect topics across languages, RjwilmsiBot changes hyphens in particular contexts to en-dashes, and AnomieBOT doing a variety of cleanup tasks. Their prevalence across many articles related to the Tohoku catastrophes reflects their role as enforcers of normative and stylistic guidelines that are ignored or broken in both everyday patterns of editing as well as the rush to provide content about breaking news topics (Geiger, 2011).

These bots differ substantially in the number of revisions made as well as the concentration of their effort: SmackBot's 172 revisions were concentrated in a handful of the most heavily-edited articles while ZeroBot repeated similar revisions on 14 distinct articles. SmackBot was an early editor of on the "earthquake and tsunami" article, making its first contributions at 7:05 UTC, about 45 minutes after the article was created. The work it performed involved revolved almost exclusively around adding dates to "citation needed" templates that other editors had added when claims in the article needed sources for verification. The act of adding the date, while mundane and not specific to breaking news articles, was nevertheless important because it was often overlooked and signaled to other editors when the request for additional information had been made. Citation needed templates are an impetus for other (human) editors to go out and seek this information, but the act of adding time-stamped information to these templates

[45] http://en.wikipedia.org/wiki/User:SmackBot#Tasks_and_authorisations

were also important indicators in the context of a breaking news article that the content is unreliable or out of date. SmackBot's efforts also track the shifting focus of many events as editors shifted their attention (discussed in more detail in subsequent sections) and errors to other articles about the "nuclear disasters", affected cities and towns, and related articles. If bots exist to fix the gaps and errors made by non-expert or hasty editing, the concentration of bots' activity on articles is potentially a reliable proxy for the expertise of the editors on that article. 66 of SmackBot's contributions to the corpus were concentrated on the "nuclear disaster" article suggesting that editing there was particularly problematic. As before, the work the bot performed here was focused on updating templates to include date tags. However, the tags it went to work on were both more prevalent and more varied. These included templates warning about unreferenced sections, content lacking temporal information, clarification, or unreliable sources.

Given the prominence of bots in this breaking article, are there significant differences between bots activities across breaking and non-breaking articles? Using the airline crash corpus from Chapter 6 (which will also be used again in Chapter 8) and classifying articles as breaking, non-breaking, and historical as was done in Chapter 5, the distributions of activity for bots and non-bot users are plotted in Figure 7-3 through Figure 7-6. Bots are identified based upon their membership in categories such as "All Wikipedia bots".[46] As Figure 7-3 shows, bots make similar changes in both historical and non-breaking articles, but their activity on non-breaking articles is substantially more skewed towards very large changes (hundreds of bytes per edit). While this suggests bots make substantially larger edits on breaking news articles than other types of articles, it is largely an artifact of anti-vandalism bots reverting large edits defacing or blanking articles, but also reflects the work done bots like SmackBot repeatedly adding changes to templates, which generally are not visible to readers. Nevertheless, this suggests bots engage in distinct kinds of work on breaking articles responding to real users' changes that lead to substantial differences in

[46] http://en.wikipedia.org/wiki/Category:All_Wikipedia_bots

the changes bot users, in turn, make. This disparity is more obvious when looking at the changes made by non-bot users in Figure 7-4 where users make similar changes across all types of articles.

Because older articles are more likely to have attracted incidental edits simply existing longer than younger articles, normalizing for article age and patterns of user and bot behavior across article types also reveals interesting differences. Figure 7-5 plots the number of unique users normalized by article age, revealing that breaking articles have a tendency to attract more non-bot users than non-breaking or historical articles. In absolute terms, there are substantially fewer bots editing across all types of articles than non-bot users, as seen in Figure 7-6. Again, historical and non-breaking articles exhibit similar patterns in bot behavior while breaking articles have more variance demonstrating greater prevalence of bot activity per unit time than other article types. Because bots generally operate autonomously in response to changes made by others, the over-representation of bot activity on breaking news articles suggests bots are responding to disproportionately poor editing that occurring there. Certainly the popularity of the topic can attract vandals, but the intensity of contributions in the aftermath of the event might result in users making less heedful changes that break existing parts of the article or introduce errors. The activity of bots may thus serve to reduce the overhead on other editors to police these changes and may free them to focus on more substantive contributions.

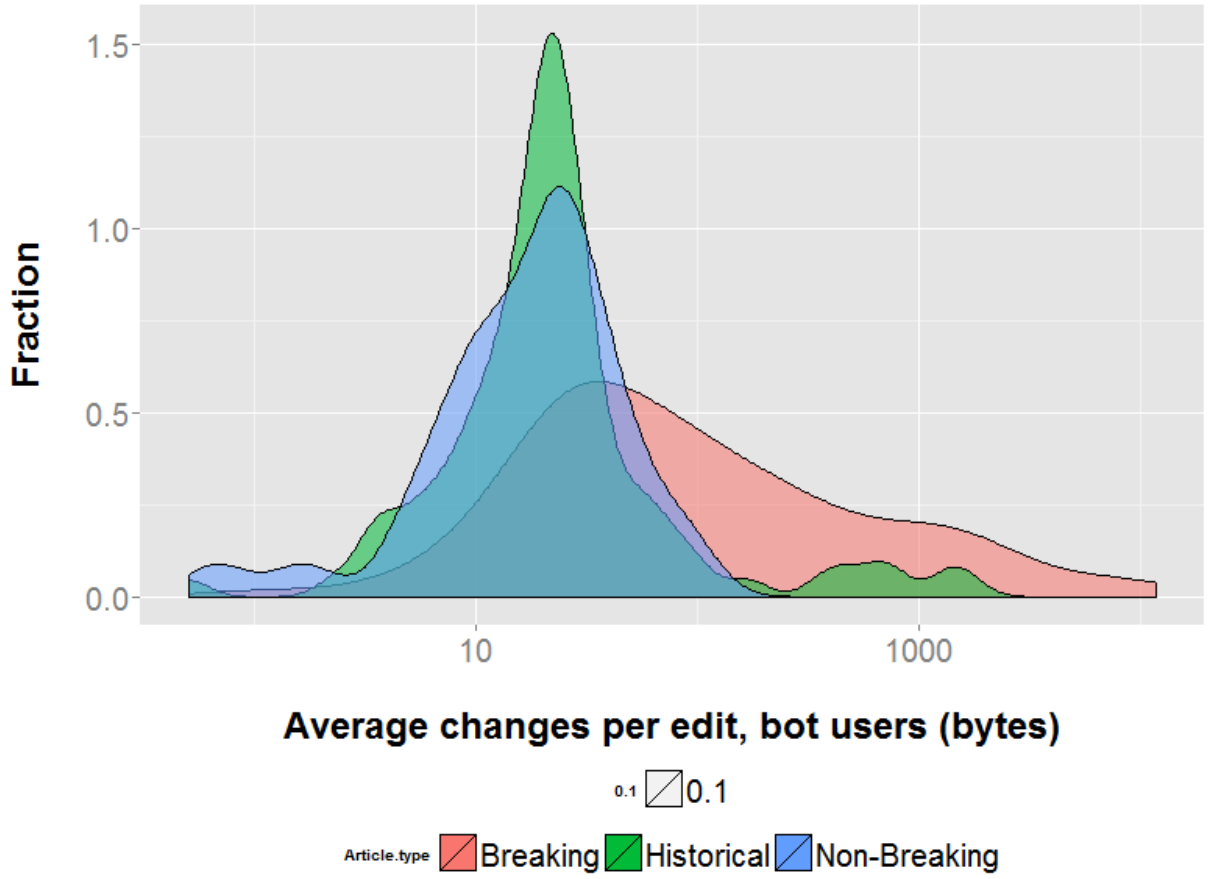


Figure 7-3: Distributions for average bytes changed per edit from bot users.

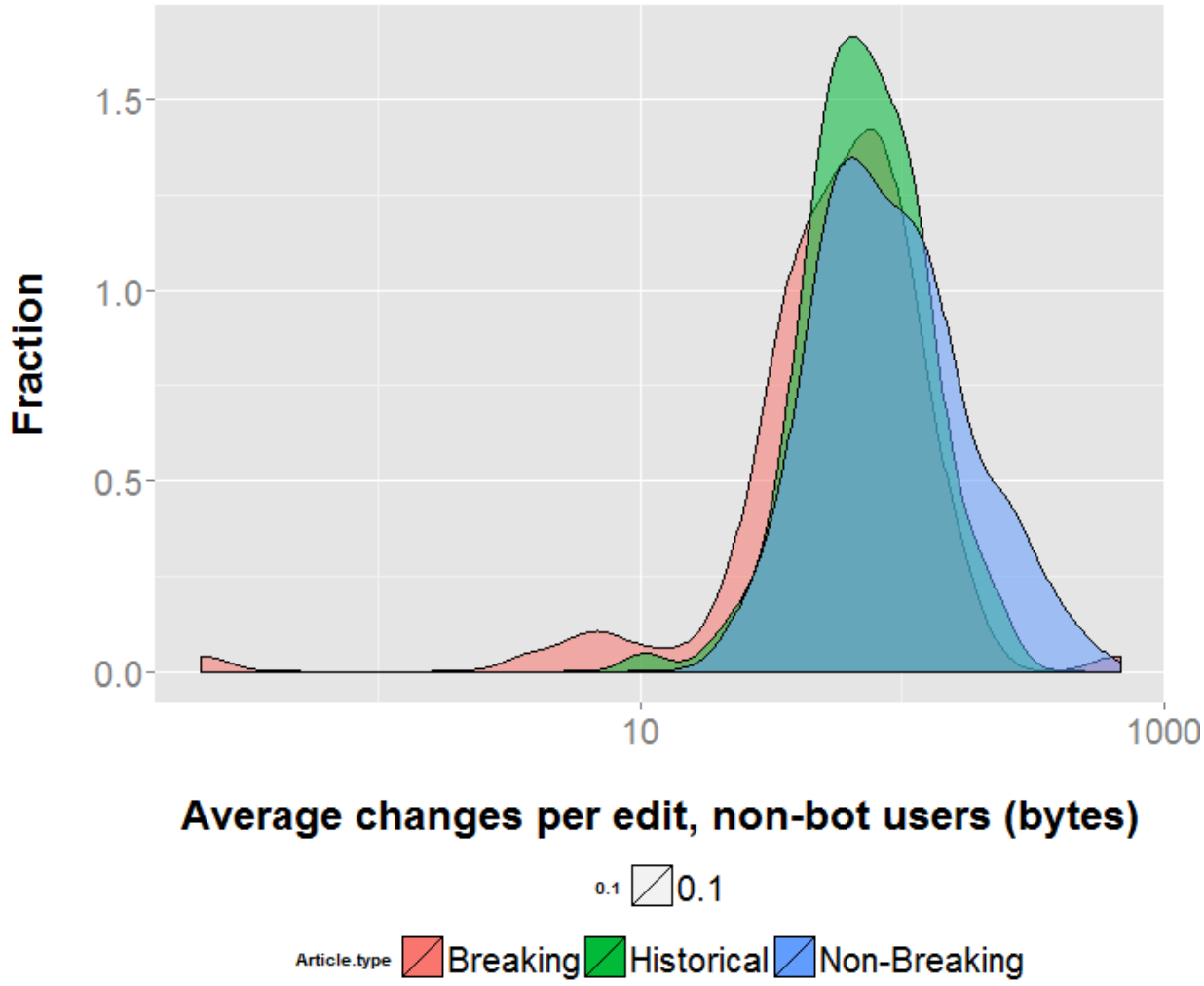


Figure 7-4: Distributions for average bytes changed per edit by non-bot users.

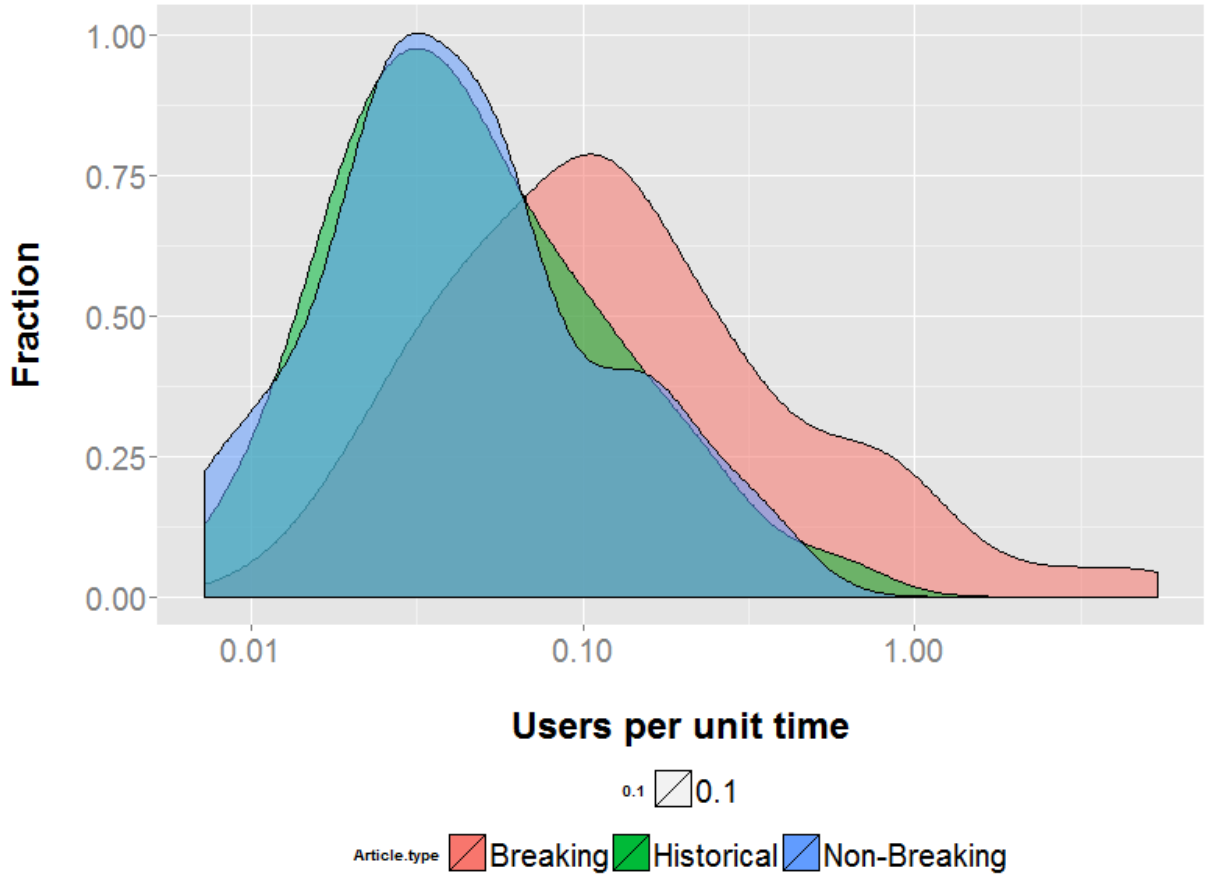


Figure 7-5: Distributions for unique non-bot users by article type, normalized by article age.

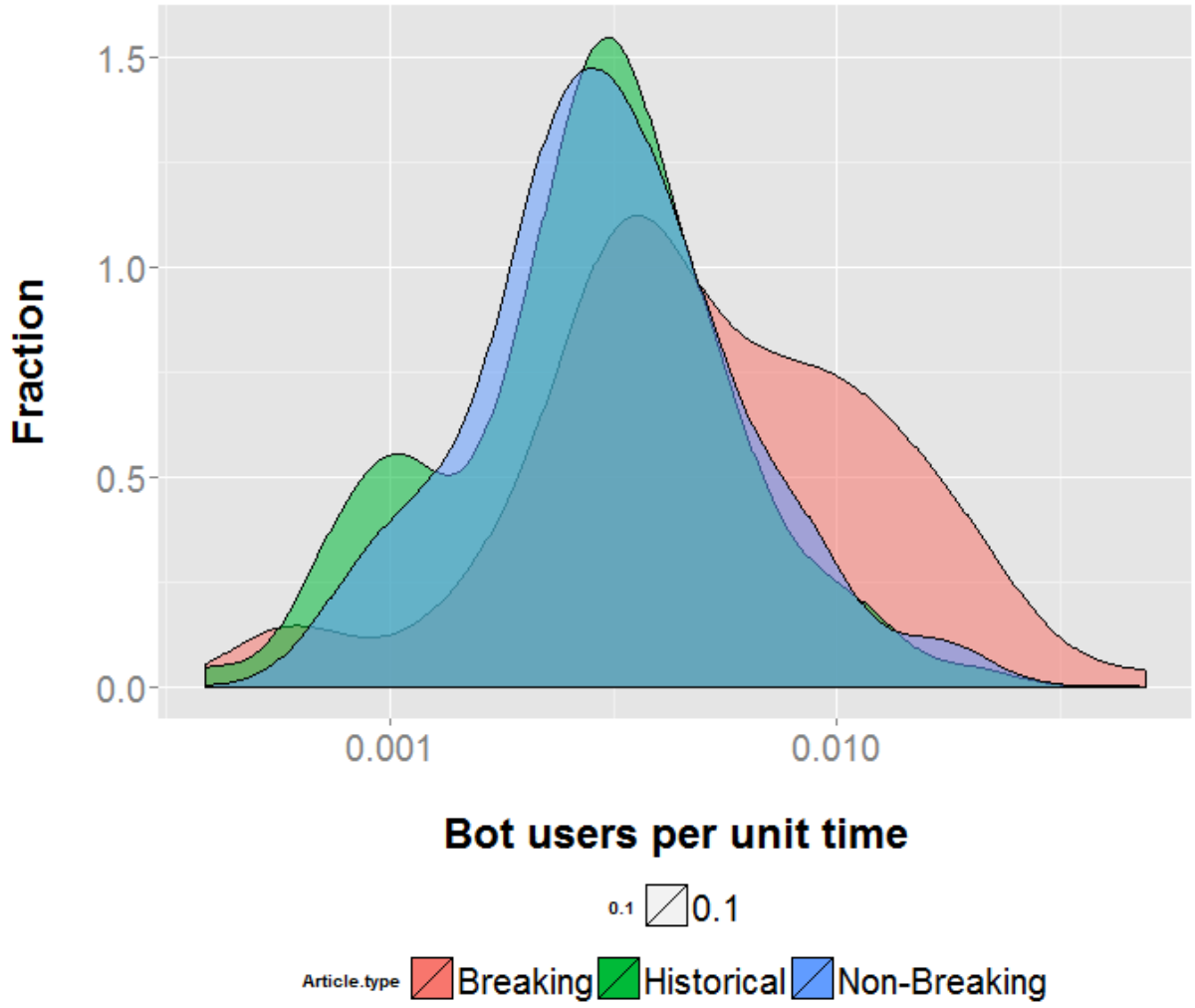


Figure 7-6: Distributions for unique bot users by article type, normalized by article age.

Name	Degree	Gini coefficient	Total revisions
L.tak	11	0.66	606
Flodded	11	0.89	560
Johnfos	18	0.68	403
ACSE	34	0.76	356
1947enkidu	9	0.57	341
Sandpiper	3	0.58	321
Colfer2	12	0.69	285
Cla68	19	0.74	264
66.65.191.165	8	0.63	230
Kolbasz	5	0.5	213
Shinkolobwe	3	0.4	195
Pontificalibus	8	0.64	192
Geofferybard	10	0.48	183
SmackBot	35	0.66	172
66.220.113.98	17	0.69	170
Peace01234	1	0	160
Andycjp	20	0.57	142
Tenmei	11	0.6	139
Neier	17	0.49	134
Rtc	2	0.3	133
WWGB	4	0.59	129
Goodvac	5	0.76	125
Obankston	11	0.58	119
Luckas-bot	34	0.47	117
Qrfqr	7	0.66	116
ClueBot	14	0.69	115
Theanphibian	14	0.53	115
Pol098	4	0.69	113
Materialscientist	12	0.74	112
Orcano	1	0	106
AnomieBOT	23	0.62	103
Beagel	7	0.71	103
Boardhead	2	0.46	103

Table 7-2: Editors with most revisions.

Table 7-2 summarizes the statistics of the 33 editors making more than 100 revisions to the corpus. These users' contributions are almost exclusively focused on a single article or handful of articles. User L.tak was extensively involved several articles related to the "Fukishima Daiichi nuclear disaster" and "Fukushima Daiichi Nuclear Power Plant" articles, talk pages, and related pages beginning on March 11. He (or she) continued editing these articles on a daily basis until April 1 ultimately making more than 211 revisions out of the 6,165 revisions on the article. User Flodded edited the main article about the earthquake and tsunami exclusively approximately 14 hours into the collaboration and continued to edit daily until March 23 making 542 of the article's 6202 revisions. L.tak's contributions were also wide-ranging and varied, the most active editor on both the articles for the "power plant" article and talk page as well as the second most active editor on the "nuclear disaster" article and seventh most active on its talk page. Like Flodded, L.tak's involvement was extensive but temporary and appears to have stopped contributing to either article after early April.

Remarkably, neither of these editors ever crossed paths: they worked on their "own" articles independently of each other despite the similarity and timeliness of their topics. Alternatively, a user like ACSE edited many articles, but concentrated attention on a pair of articles, editing the "nuclear disaster" article 160 times, the "earthquake and tsunami article" 83 times, and the other articles no more than 13 times. Thus, highly active users appear to occupy distinct social roles as either specialists focusing solely on a single article (like Flodded) or highly related topics (like L.tak) or as something like generalists moving between several or articles like ACSE. This specialization of prolific editors contributing to only a single article or subtopic is startling as it suggests substantive coordination or collaboration in coverage proceeds through other channels and mechanisms than coauthorship of articles. Of course this corpus only includes the revisions related to the articles themselves and does not include activity on article discussion pages, user talk pages, or other forums where important coordination occurs. These features and these users' interactions with them will be explored in editor trajectory sections below.

Name	Degree	Gini coefficient	Total revisions
Flodded	11	0.89	560
Goodvac	5	0.76	125
ACSE	34	0.76	356
Materialscientist	12	0.74	112
Cla68	19	0.74	264
Sexandlove	8	0.73	97
Beagel	7	0.71	103
Mikenorton	7	0.71	70
MWadwell	4	0.7	75
NortyNort	6	0.69	72
Pol098	4	0.69	113
ClueBot	14	0.69	115
Colfer2	12	0.69	285
66.220.113.98	17	0.69	170
Johnfos	18	0.68	403
Cybercobra	8	0.67	81
Spitzl	6	0.67	71
L.tak	11	0.66	606
SmackBot	35	0.66	172
Qrfqr	7	0.66	116
SunCountryGuy01	5	0.65	40
Wipsenade	4	0.65	57
Henk Poley	4	0.65	45
Pontificalibus	8	0.64	192
66.65.191.165	8	0.63	230
Moxy	4	0.63	64
P.T. Aufrette	5	0.63	49
AnomieBOT	23	0.62	103
Ryulong	7	0.62	39
MartinezMD	5	0.62	59
Fancy-cats-are-happy-cats	5	0.62	96
Jal11497	3	0.62	47
Ghostofnemo	3	0.61	74
Gold Hat	4	0.61	62
184.144.160.156	14	0.6	96
Tbhotch	8	0.6	72

Table 7-3: Editors with greatest Gini coefficients.

Table 7-3 summarizes the statistics of editors in the bipartite network having Gini coefficients above 0.60. The top users in this list necessarily concentrate their revisions in editing only a handful of article and for the top 5 editors here, it was the main “earthquake and tsunami” article. However, other editors such as NortyNort focused on other topics such as the “Fujinuma Dam”, Johnnos on “Japanese reaction to the Fukushima Daiichi nuclear disaster”, SunCountryGuy01 on “Timeline of the Fukushima Daiichi nuclear disaster”, and Ryuman on “You & Me (Superfly song)” a charity fundraising song. The unregistered editor identified by the IP address 66.220.113.98 is also interesting because unregistered editors are traditionally perceived as unengaged, novice, or suspicious.[47] The editor (or editors) associated with this IP address began editing in late March several weeks after the bursts of activity around the events itself and continued editing through June making edits related to images on the main “earthquake and tsunami” article as well as articles about the affected towns across 17 distinct articles.

[47] The IP address geolocates to an internet service provider in Bend, Oregon. http://whatismyipaddress.com/ip/66.220.113.98

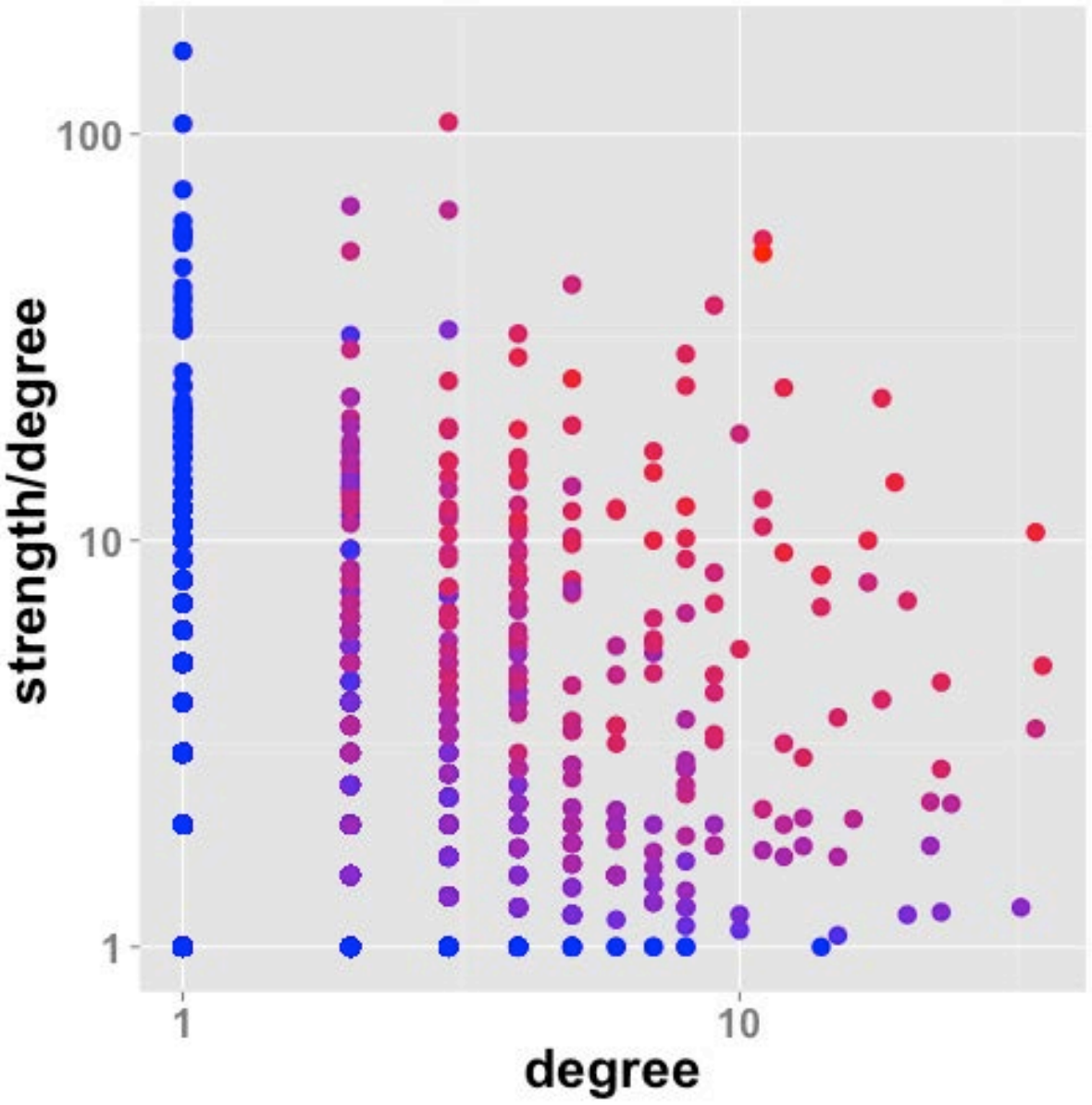


Figure 7-7: Normalized total revisions (strength/degree) plotted against degree, points colored by Gini coefficient.

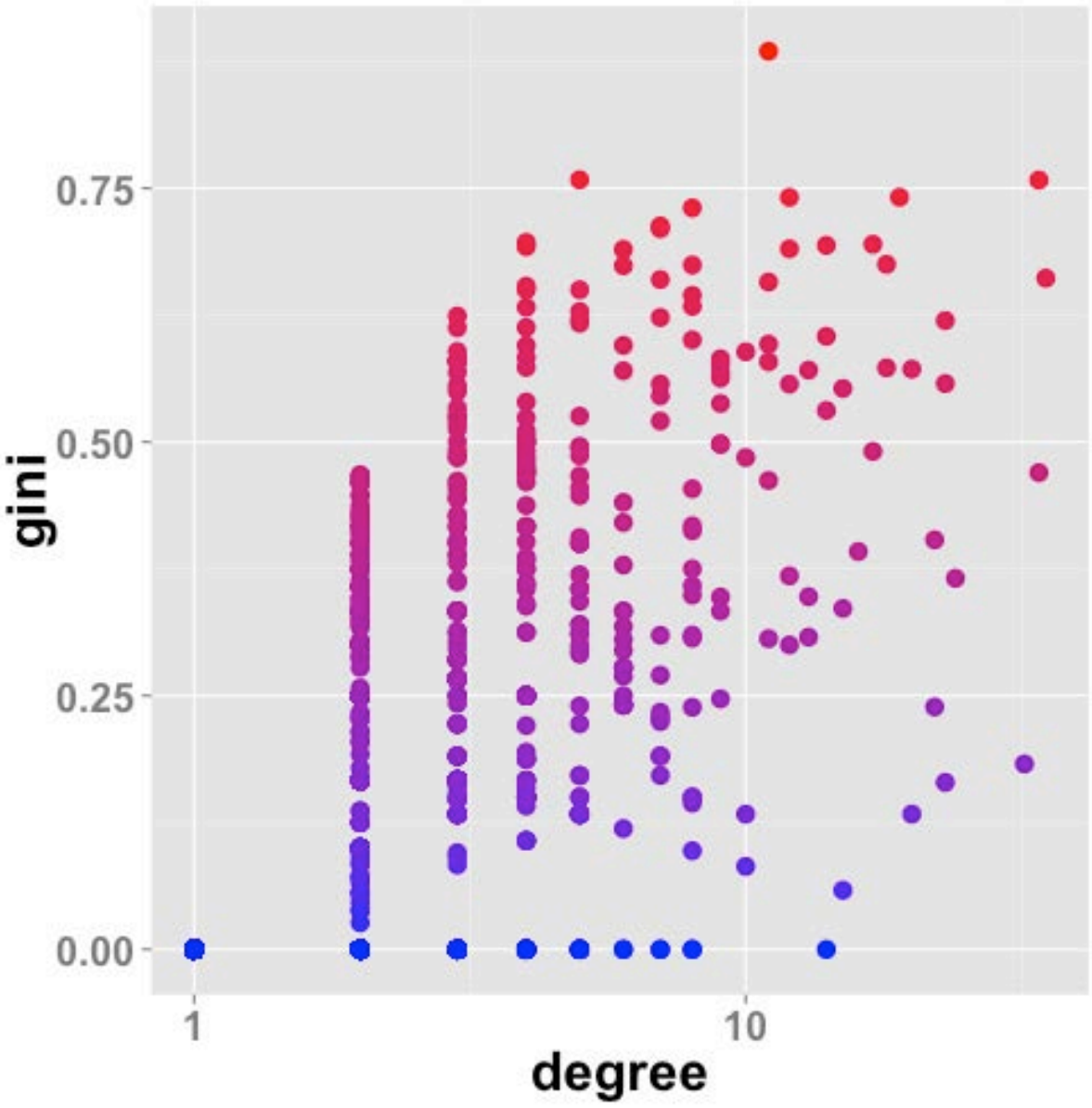


Figure 7-8: Gini coefficients plotted against degree, points colored by total revisions.

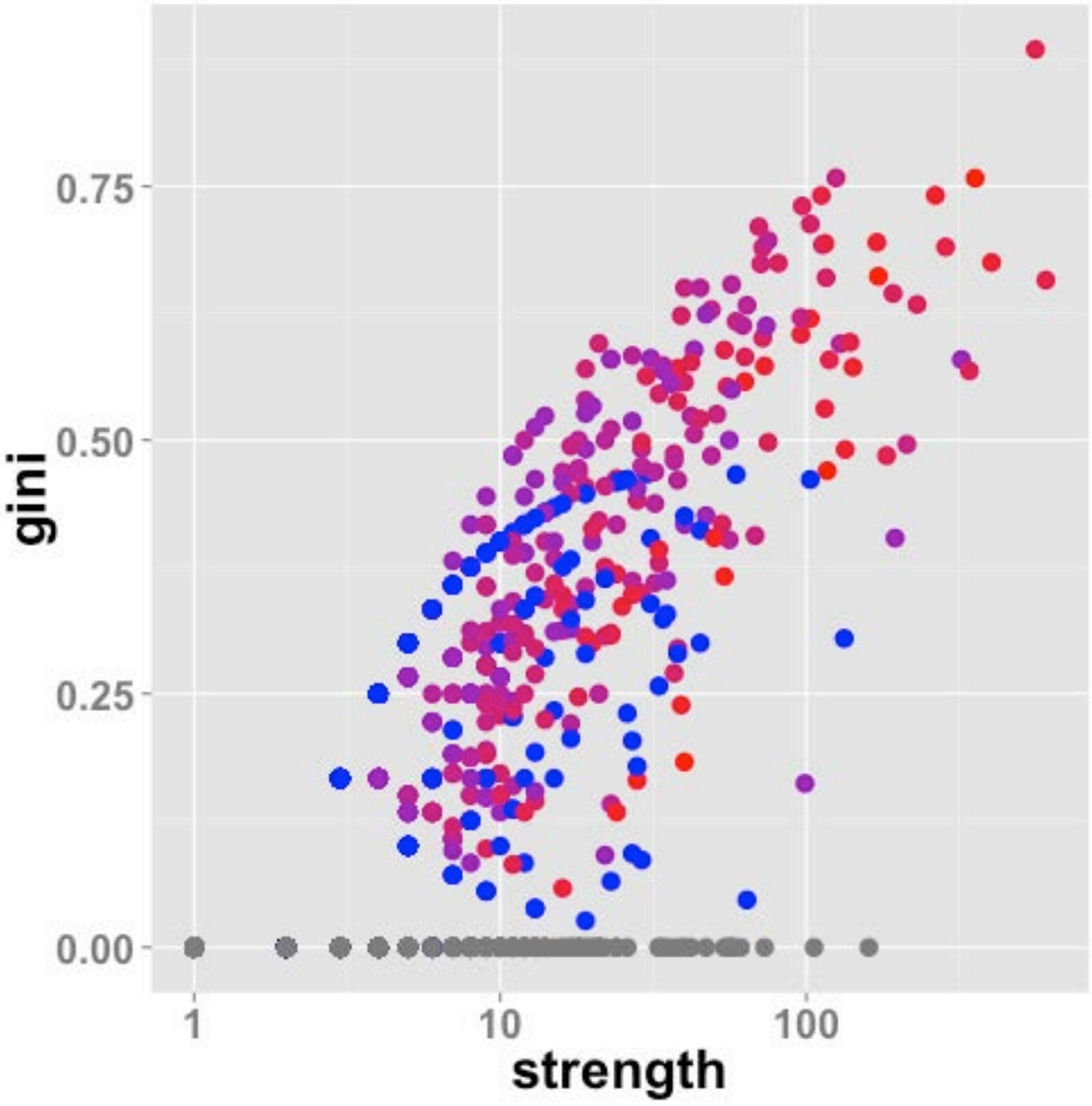


Figure 7-9: Gini coefficients plotted against total revisions (strength), points colored by degree.

As one would expect, in Figure 7-7 demonstrates the more articles an editor contributes to, he or she makes fewer average contributions per article (strength/degree). A similar trend is repeated in Figures Figure 7-8 and Figure 7-9 measuring the Gini coefficient of editors' contributions to each article. In Figure 7-8, the more unique articles any user contributes to tend to increase the concentration of their contributions; an editor may make contributions to other articles but the majority of their contributions are focused on a single article. In Figure 7-9, the more edits a user makes across all articles also tends to increase the inequality of contributions. Again, the more articles an editor contributes to (degree) or the more edits made in total (strength), the more likely an editor is to concentrate their efforts in a few articles.

Out of 6,003 editors in the corpus, 2,265 editors made more than a single contribution, the 919 editors contributed to more than a single article, and 619 unequally split their contributions across multiple articles. However, the vast majority of these multi-article editors nevertheless have extremely distributed editing behavior: 839 of the 919 multi-article editors had Gini coefficients below 0.5 suggesting their work was highly distributed and their contributions to other articles largely one-off and incidental. 300 editors contributed to more than one article, but only made one contribution on each of these articles, giving them a Gini coefficient of 0 as well. In fact, only 19 editors had a degree of more than 3 and a Gini coefficient above 0.66. These editors made contributions to many articles, but largely focused their contributions on one. The statistics for these editors are summarized in Table 6-3.

The question of whether prolific editors simply make relatively minor edits to many articles rather than major substantive changes in content or are largely focused on a single article with incidental edits to others warrants further attention. Qualitative analyses of these editors' contributions are explored later in this chapter. The next section examines the article trajectories of several of the major articles to identify common structural features and editors.

Article trajectory networks

Using the method described in the previous chapter, this section explores the article trajectories for selected articles like “2011 Tōhoku earthquake and tsunami”, “Fukushima Daiichi nuclear disaster”, “Sendai”, and “Fukushima Daiichi Nuclear Power Plant” which were the focus of substantial editing activity in the bipartite network above. Features of these articles’ trajectories are compared below, significant actors are identified from both the article trajectories and bipartite networks, and these actors’ positions compared. The identification of actors occupying structurally significant positions motivates a qualitative analysis of the work they performed in the following sections.

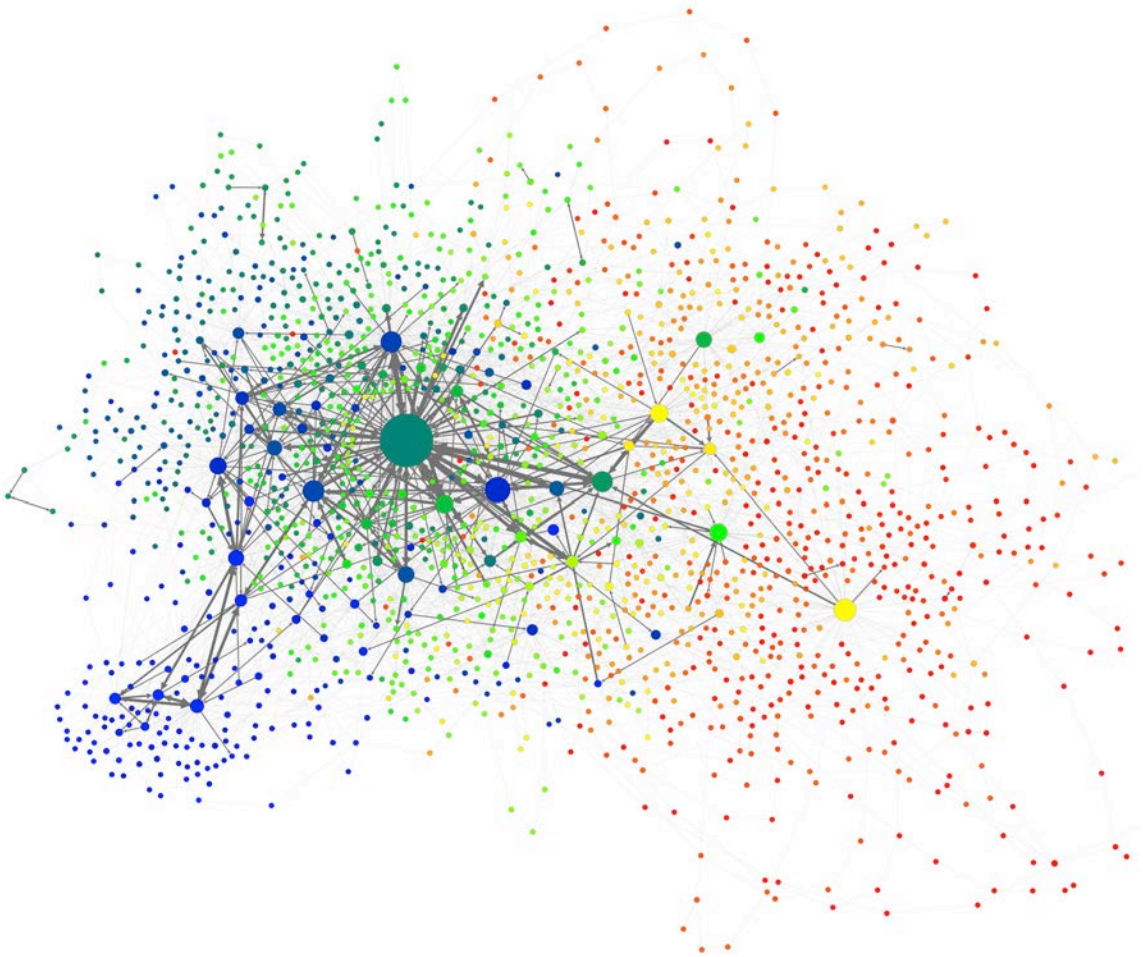


Figure 7-10: Article trajectory for "2011 Tōhoku earthquake and tsunami." Nodes are editors, sized by in-degree (number of unique editors modifying their revisions) and colored by minimum ordinal revision history position (bluer nodes joined collaboration earlier, redder nodes later). Isolate nodes are consequence of thresholding tie strengths above 2 contributions to reveal core network.

Several features of this article trajectory are notable. First, there is a strong core-periphery structure in this breaking news article's trajectory as observed among the trajectories of breaking articles in the prior chapter. There is a dense, cohesive core of nodes with many links as well as links to each other and a relatively sparse, unconnected periphery of nodes laying on the uninvolved "loops". Second, the strongest ties in this trajectory (repeated interactions between editors) occur predominantly in this core. Figure 6-5 colors the nodes by their minimum ordinal revision history position; effectively the stage of the article's history when they made their first contribution with bluer nodes first contributing earlier in the article's history and redder nodes first contributing later in the article's history. Third, following these node attributes, there is little evidence of mixing: bluer nodes are predominately connected to other blue nodes, green to green, and red to red. This is evidence that editors' involvement in the article collaboration is strongly temporally bound. Early editors necessarily edit with other early editors but leave the collaboration before interacting with later editors. Fourth, these early responder editors are also on the relative periphery of the article trajectory with the dense core of highly active editors first joining the collaboration at a later stage. The green editors comprising the core of the graph have minimal mixing with the later collaborators in red making up the majority of nodes. These later editors exhibit substantially less clustering and degree centrality (with notable exceptions explored later). This replicates the findings from the previous chapter that breaking articles patterns of collaboration regress to the patterns of collaboration and activity found in traditional articles' trajectories in the weeks after the event.

The article trajectories for "Fukushima Daiichi nuclear disaster", "Sendai", and "Fukushima Daiichi Nuclear Power Plant" are plotted in Figures 6-6, 6-7, and 6-8 respectively. The "Sendai" article is an example of an article which predated the catastrophe, the "Fukushima Daiichi Nuclear Power Plant" article is an example of an article which could have predated the catastrophe but was created in response to it, and the "Fukushima Daiichi nuclear disaster" is another example of an article about a breaking news event. Like breaking articles in the previous chapter and the earthquake article above, these articles exhibit core-periphery structures. Again, the strongest ties representing repeated interactions between

editors occur in the core. There is also little evidence of node mixing; early editors work with other early editors but do not persist in the later stages of the article collaboration. On one hand, this is a consequence of the trajectory method whereby the links necessarily imply a temporal dependency. However, because the nodes' colors reflect the earliest edit they made, the absence of mixing for prolific nodes suggests their participation in the collaboration does not persist or extend to modifying the contributions of later editors. Consider the "earthquake and tsunami" article where early editors were relatively peripheral compared to the early editors of the "nuclear disaster" and "power plant" articles who are very much embedded in the core of the article. This is necessarily not the case for the "Sendai" article as its first editors made their contributions seven and a half years before the event while the editors "ACSE" and "DAJF" were involved in the updating the article in response to the earthquake and tsunami. But in each of these cases, the prolific editors were connected to other editors who joined around the same time. This lack of persistence among editors complicates the idea that a core set of editors possess the institutional memory about the article through time and instead suggests that collaborators are highly transient.

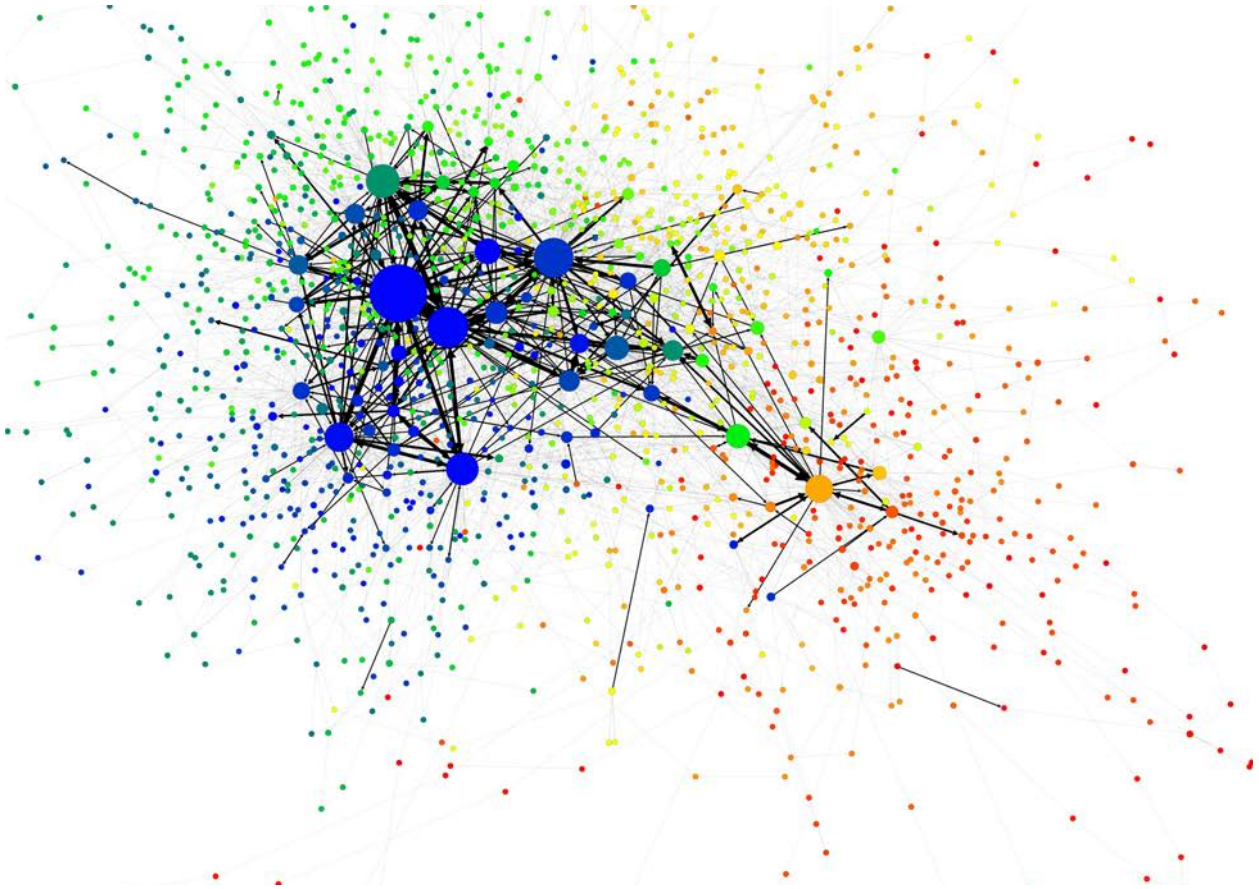


Figure 7-11: Article trajectory for “Fukushima Daiichi nuclear disaster”.

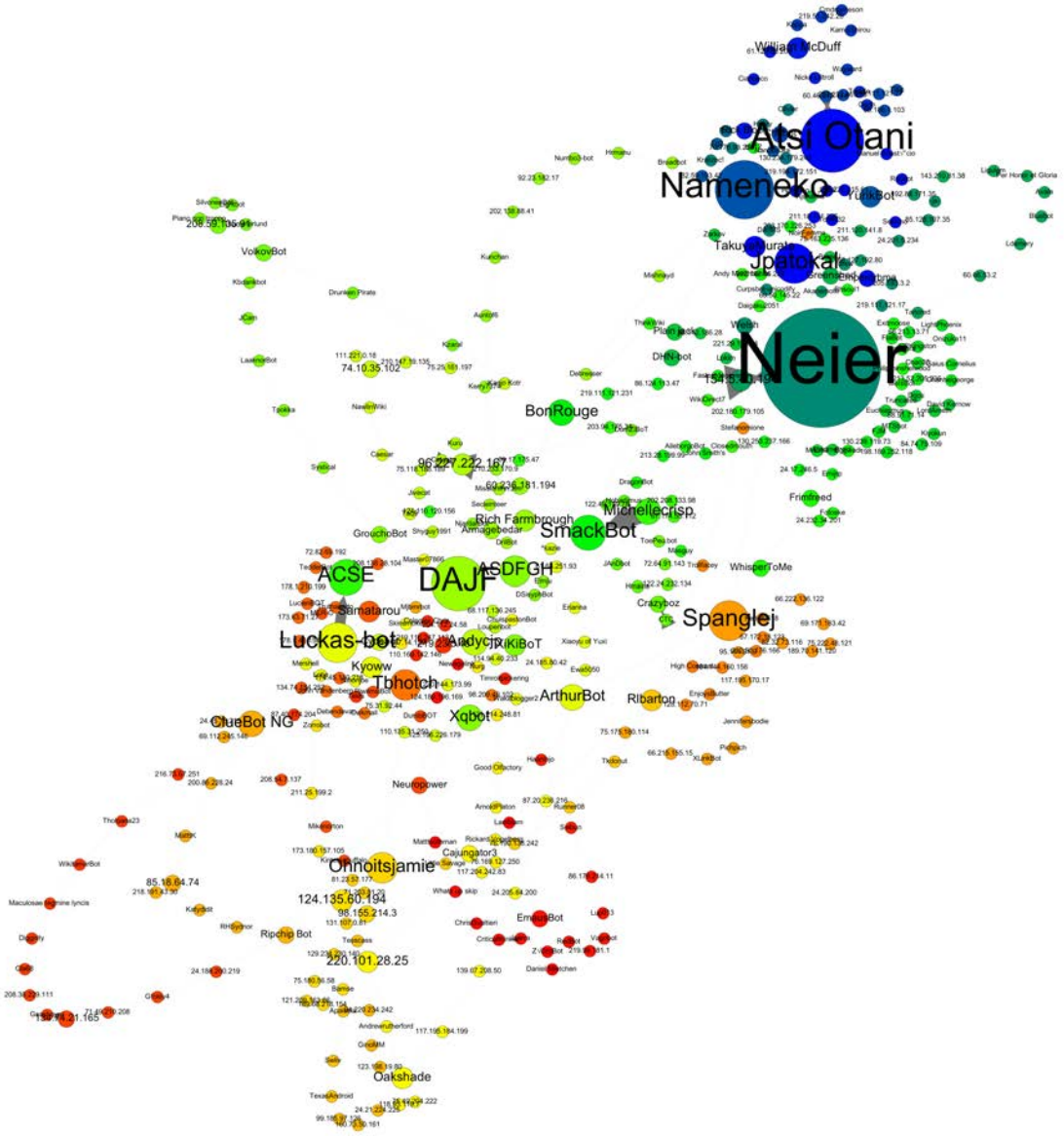


Figure 7-12: Article trajectory for “Sendai”.

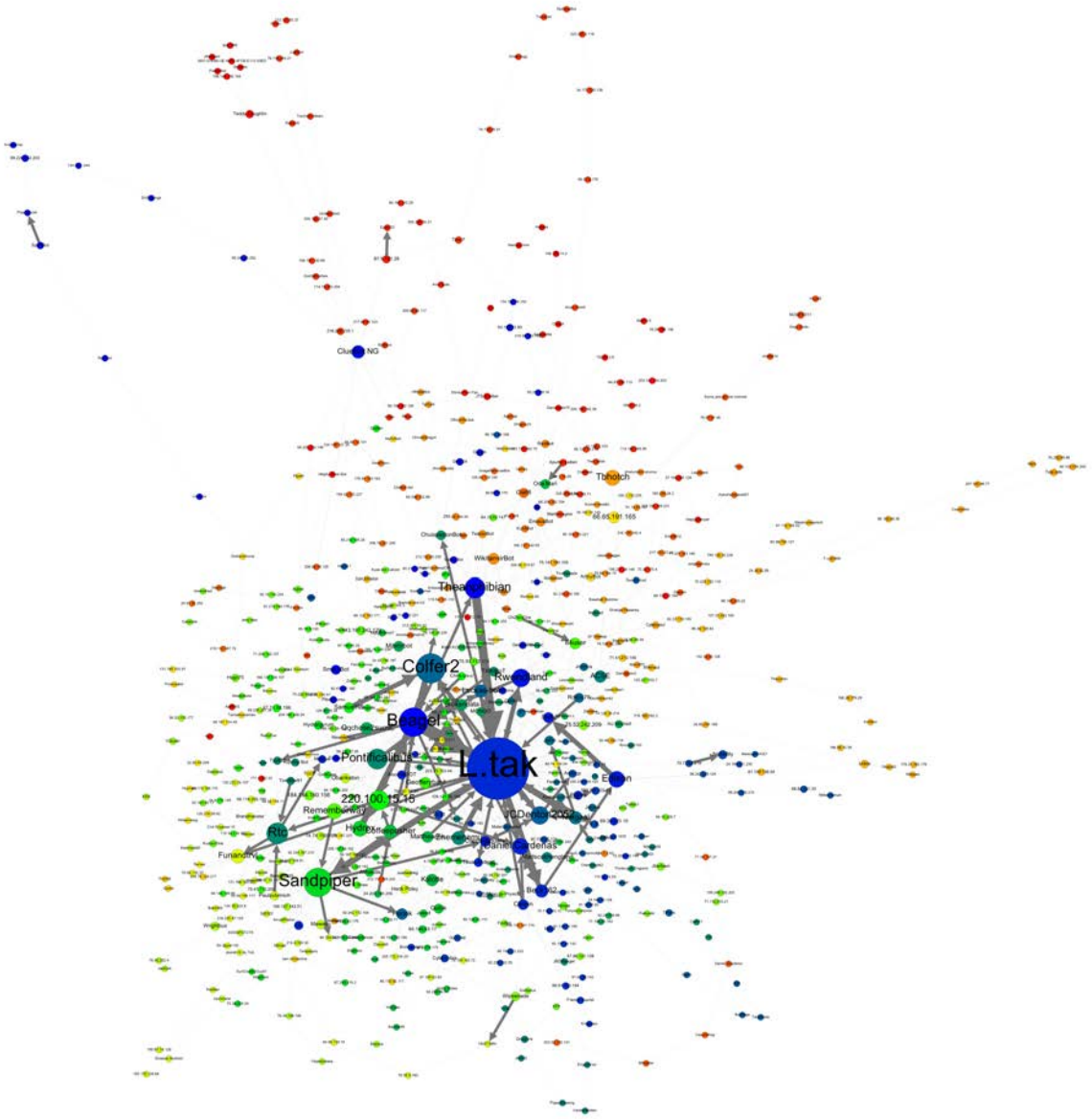


Figure 7-13: Article trajectory for "Fukushima Daiichi Nuclear Power Plant".

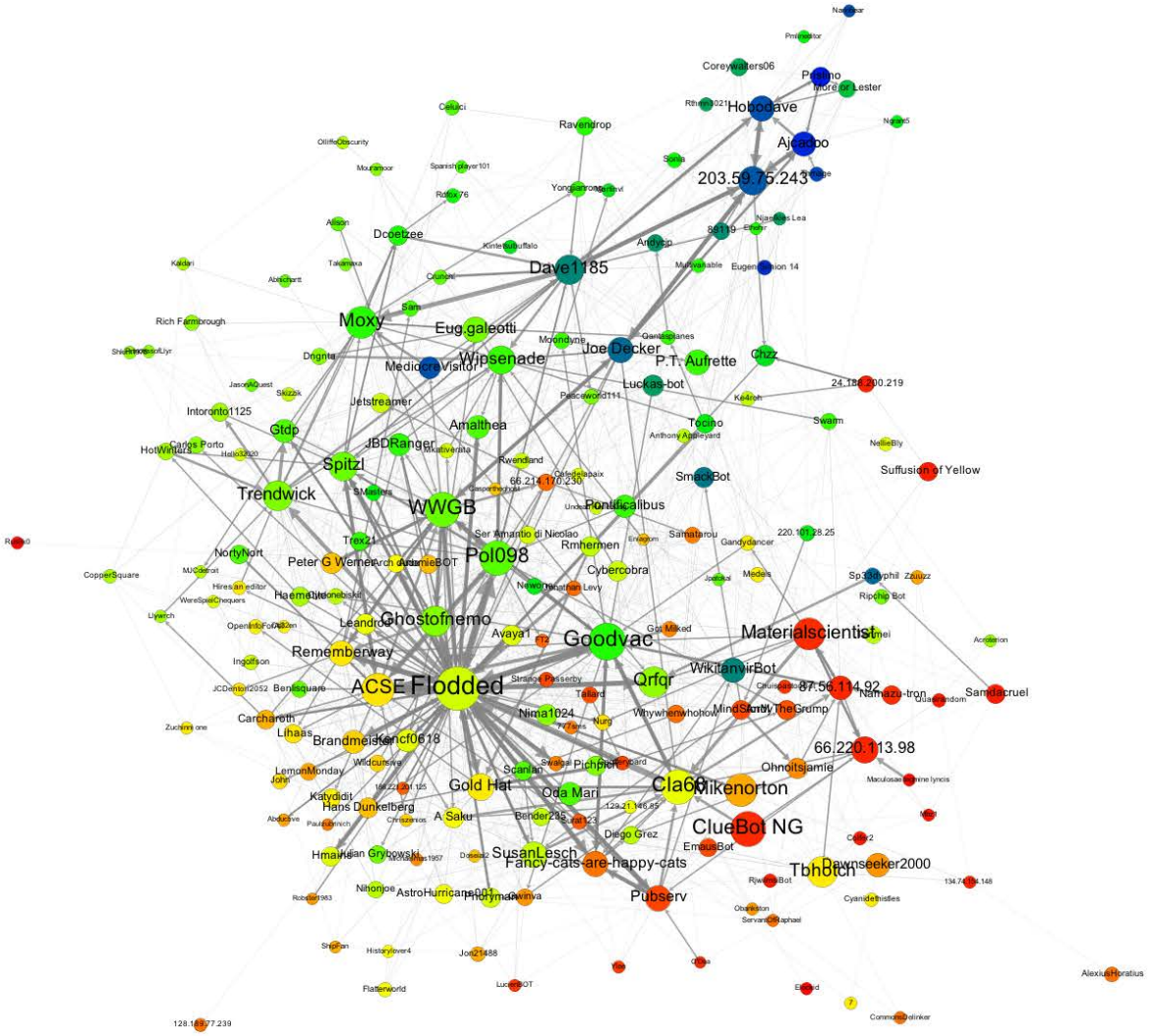


Figure 7-14: Subgraph of the article trajectory for "2011 Tōhoku earthquake and tsunami" for nodes with in-degree greater than 2 with nodes labeled by editor name.

Figure 7-15 replicates the article trajectory multigraph introduced in the previous chapter. As before, an edge represents an editor modifying another editor's revision, but in this case, we only examine the subgraph of editors who have interacted this way on two or more articles. The nodes are colored by minimum position in the revision histories across all articles in the corpus to again reflect the relative stage at which the editor began participating; bluer nodes joined article collaborations earlier and redder nodes joined article collaborations later. The edges are colored to reflect the number of distinct articles across which the editors shared interactions; lighter grey edges are on two articles and black edges are on five articles. Finally, edges are also weighted reflecting the total number of interactions these editors had across all articles, the floor being the number of articles jointly edited.

Among the top editors by articles edited, revisions made, and concentration of effort discussed above, there are some notable interactions and absences in the core of this multigraph. This core of the multigraph contained 372 edges and 216 nodes, giving it an intermediate density (7.97e-3) compared to the other article trajectories examined here. There are a number of bots present which is unsurprising since these algorithmic actors cast a wide net and make hundreds of thousands of contributions across all articles. However, it is interesting to see that these bots' revisions include modifications of prolific editors' revisions as well as other bots' revisions. For example, AnomieBot and Luckas-bot modified each other on three articles ("power plant", "earthquake and tsunami", and "Japan trench" and TheAnomebot2 and Yobot modified each other on two article about geographic features. Several users and bots also shared very strong ties across multiple articles. Johnfas and Anomiebot as well as L.Tak and Smackbot had strong and reciprocated interactions with each other on five articles reflecting their deep involvement in the same articles. However, these editors' contributions to articles preceding or following bots do not involve changes to the same content.

Turning to editor-editor interactions, Johnfas and 1947enkidu had the strongest relationship in this core network as both editors directly modified each other's revisions no fewer than 45 times across 5 articles, largely the nuclear disaster and power plant articles. L.Tak and Sandpiper, as the most active editors on

the nuclear disaster and power plant articles, also had a strong and reciprocated interaction on these articles. However, there is only a weak tie between L.Tak and Johnfas suggesting these editors contributed at very different stages of the article development or had little reason to interact.

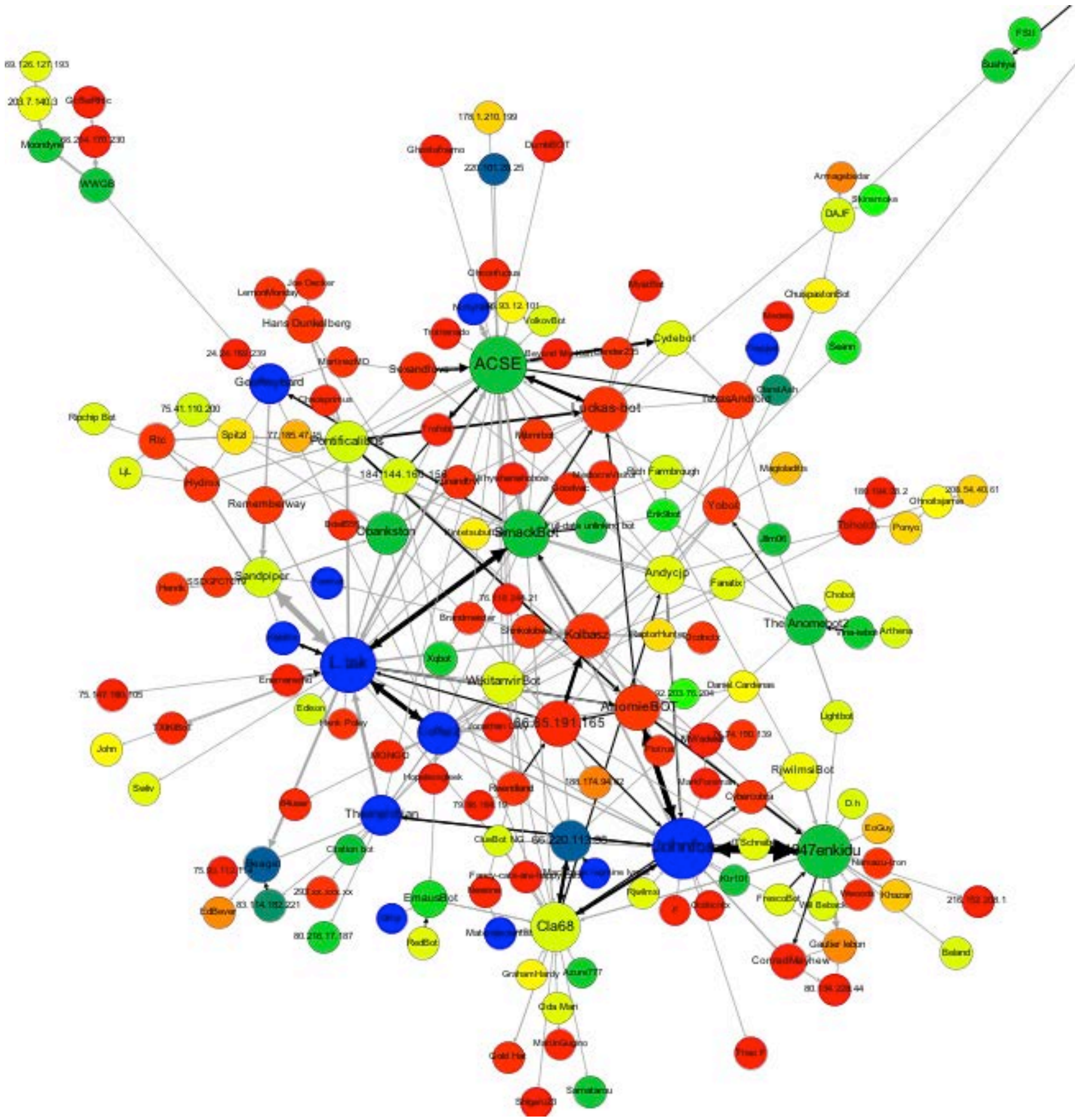


Figure 7-15: Multigraph for the corpus of revisions to all 74 articles. Nodes are colored by minimum position in any article from blue (joining collaborations early) to red (joining collaborations later). Edges are colored by the number of distinct articles editors.

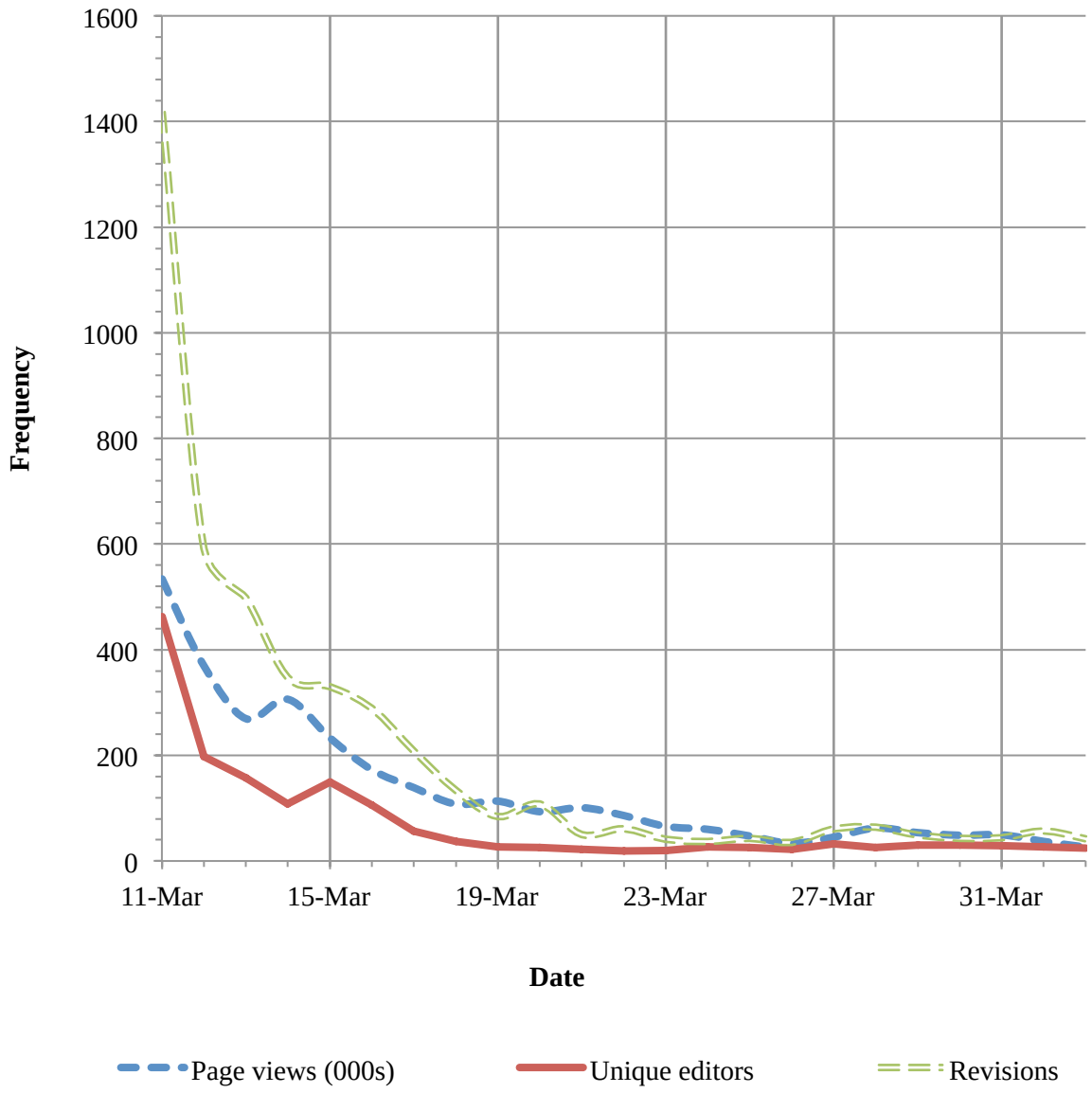


Figure 7-16: Page views, number of unique editors, and number of revisions per day for the "2011 Tōhoku earthquake and tsunami" English article's revision history.

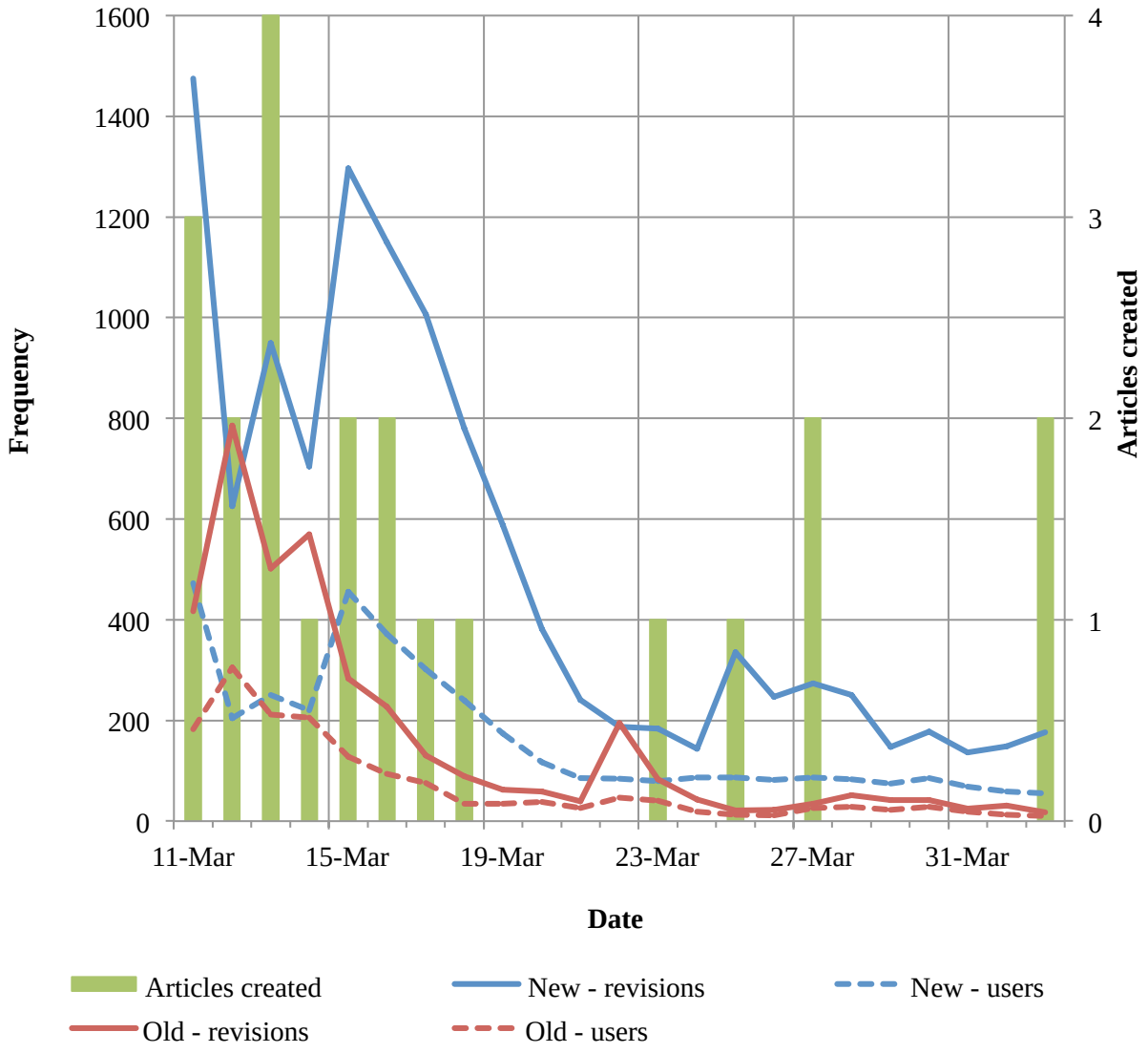


Figure 7-17: Revisions (solid lines) and unique editors (dashed lines) per day for the 62 "old" articles predating the earthquake and the 22 "new" articles created afterwards.

Shifting attention

Figure 7-16 plots the changes in page views, unique editors, and revisions by day for the primary article about the earthquake and tsunami. The article received over 533,700 pageviews on March 11 alone. It is clear that the bulk of the attention and activity for this primary article is concentrated in the immediate aftermath of the earthquake. The number of editors and number of revisions track each other relatively constantly suggesting the number of edits per editor remained stable over time. However, it may also be the case that concentration of editing has changed over time such that edits were relatively evenly distributed among editors initially but have become highly concentrated since that time. Activity by editor involvement and revisions decays rapidly and stabilizes at a few dozen users and revisions per day around March 23. However, this is only the activity on a single, albeit very central, article.

Looking at all 74 articles affected by the earthquake, Figure 7-17 plots the changes in revisions and unique editors by day for "old" articles that predated the earthquake as well as "new" articles created after the earthquake. As before, there is a clear concentration in activity across these articles immediately following the event. New and old article categories both attracted similar levels of attention and activity in the four days after the earthquake. However after March 15, activity on the older articles diminished rapidly while revisions and editor attention increases and remains high as attention is focused on the possibility that the nuclear reactors were melting down and spent fuel rods were burning. This drop-off reflects a shift in attention on the part of editors from updating extant articles to reflect the damage caused by the earthquake and tsunami and a move towards contributing to the newly created articles. The second burst of activity on new articles is largely focused on the articles related to the nuclear incidents such as "2011 Japanese nuclear incidents" and "Fukushima I nuclear accidents" over the next three days before declining as the plants cooling systems are brought back online by March 20. The shift in editor attention and activity from old articles to new articles remained stable after the first week. The March 22 spike of activity on old articles is an artifact of a bot re-categorizing several dozen pages in parallel with coordinated vandalism and edit warring from a group of unregistered editors.

As content is added to a Wikipedia article, it may become too long to be readable or coherent and it may be appropriate to "spin out" some content into a daughter article. Figure 7-16 plots new article creation by day as a bar chart along the bottom. Article creation and spinout is concentrated in the week following the event, but six articles are also created or spun-out more than twelve days after the event itself. There is a strong correlation between the number of articles created and same-day revisions (r=0.684, p<0.001) and unique editors (r=0.631, p<0.001) for new articles as well as same-day revisions (r=0.611, p<0.01) and unique editors (r=0.660, p<0.001) for old articles.

However, there are several risks to spinning articles out in the context of a breaking news event. The first risk is ensuring that content such as number of fatalities and timelines are quickly and accurately summarized or duplicated across the articles. A second risk is the possibility that spun-out pages can become "content forks" where controversial, overly specific, or trivial material is accumulated but never summarized or integrated elsewhere. A third risk involves editors' attention being spread too thinly across articles to meaningfully monitor or participate in them all. This loss of "critical mass" of editors' contributing to and discussing an event may lead to stagnation of content on the daughter pages. We examine this third point below. As more articles are associated with the catastrophe and others are spun-out or created to cover new aspects, the number of articles that can be modified in a given day increases. However, as Figure 7-17 demonstrates, the number of editors paying attention to the whole body of work is also diminishing. While the number of unique editors available to contribute to each of the articles modified on March 11 was high, there were substantially fewer "eyeballs" to monitor many more articles by March 31. Similarly, the number of revisions being made per article modified on that day was also high. As the number of articles increased and the number of editors monitoring or contributing to articles decreased, the attention of fewer editors was concentrated on more articles.

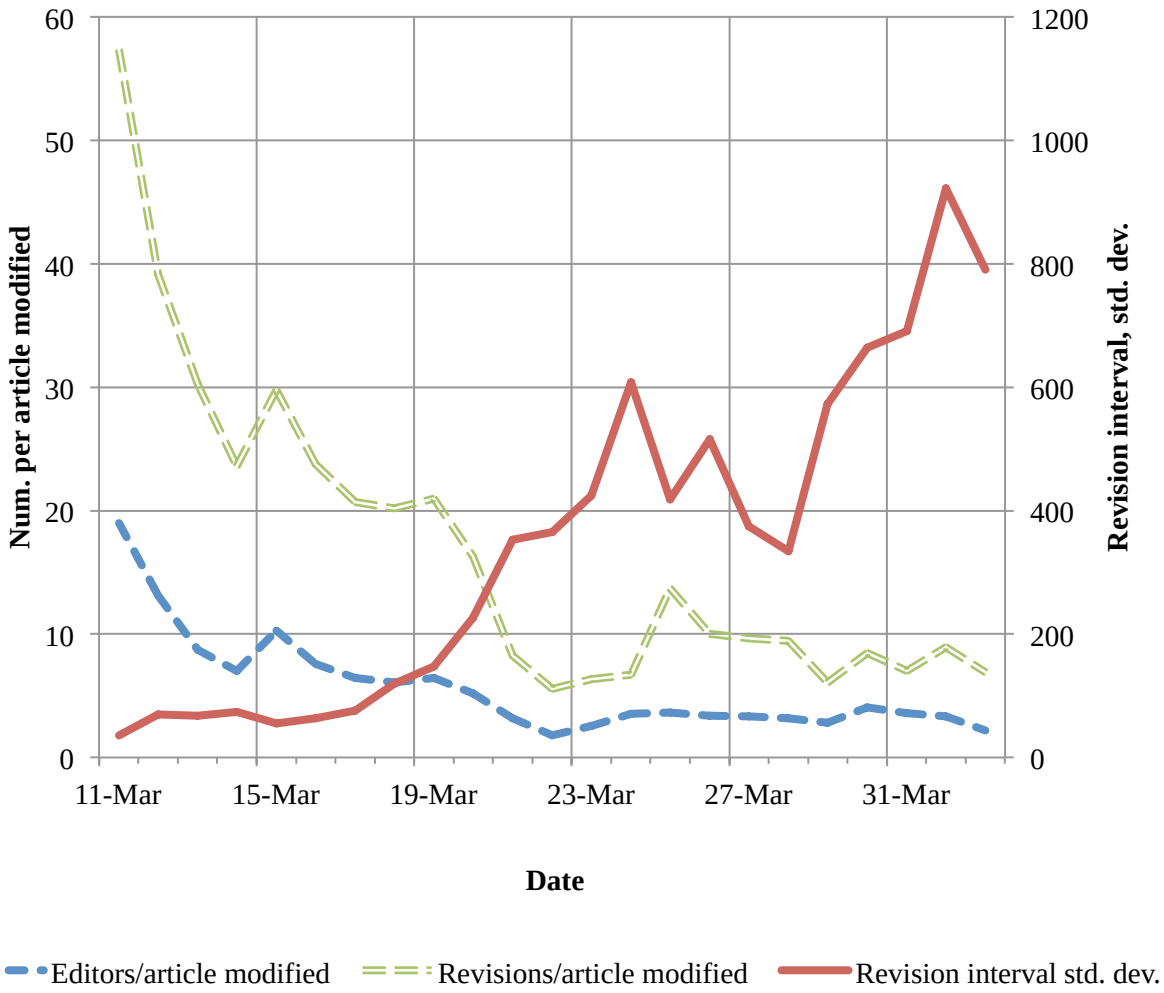


Figure 7-18: The variance in time between edits to any Tōhoku article increases dramatically in parallel with the dilution of editor attention and activity among actively-edited articles.



Figure 7-19: A screenshot of an in-line casualty template. Only the text "12,087 deaths,[1][2]" would be transcluded (appear) in the body of the article when this template is called.

In Figure 6-13, the standard deviation of the interval of time that lapsed between edits to any of the 74 articles in the category is plotted against two measures of attention for a given day: total number of contributing editors per article modified and total number of revisions per article modified. This fall-off in attention is also associated with an increase in the variance of time between modifications—the more articles available to edit is associated with an increase in time between edits made to any article. Strong negative correlations are observed between the total number of contributing editors per article modified that day (r=-0.668, p<0.001) and the total number of revisions made per article modified that day (r=-0.729, p<0.001).

Thus while the diminishing salience of the event certainly drives changes in the intensity of the collaboration as time goes on, these results suggest the dilution of editor attention among active articles is also implicated in changes in collaboration intensity. This “phase transition” is not necessarily a bad thing and may in fact reflect a shift from reactive tasks to systemic or maintenance tasks as described by (Wattenberg, Viégas, & Hollenbach, 2007). As the immediate urgency of developing an account and the intensity of dozens of revisions being made per hour subsides, experienced editors may be able to establish the highly interdependent routines and ordinary patterns of action and functions associated with high quality collaborations (Kittur et al., 2009).

One example of highly interdependent work is the adoption and updating of templates. Templates are Wikipedia pages that are called from within other pages. Templates are typically used for repetitive material such as boilerplate messages, navigation among thematically similar articles, or standardizing the presentation of information among articles. In addition, templates are often used in a directive or managerial role to highlight improvements that need to made to an article or to acknowledge contributions and achievements (Gaio, Besten, Rossi, & Dalle, 2009; Kriplean et al., 2008; Loubser & Den Besten, 2008). However, templates were adopted and implemented in unique ways by the Tōhoku earthquake editors to support their high-tempo collaboration across articles.

While the spin-out of articles imposes significant demands on editors to monitor and update content so that it remains consistent both within and across articles, important information like casualty numbers are constantly in flux as new information is released and various outwardly reliable primary and secondary sources present conflicting information. In the absence of a central location to discuss these disputes or a clear hierarchy to decide which figures should prevail, editors of previous breaking news events were often consumed with the task of debating changes to every article on which these figures are cited whilst failing to keep both the figures themselves and their corresponding references up-to-date. Although "bots" exist to automate repetitive tasks such as updating figures across many articles, broadcasting out changes to the dozens of potential pages every time a revision needed to be made to casualty counts was an untenable solution.

Wikipedia administrator "Dcoetzee" created a unique workaround by creating separate template pages for the number of missing and number of dead casualties. As shown in Figure 6-14, this template page only contains the number reflecting the current number of fatalities and two citations to the Japanese National Police Agency, which is the most reliable source for casualty figures. When this template is called from another Wikipedia page, the number of deaths and the two references are the only information that is inserted into that page. Editors thus only need to update the template with the latest figures and the changes are immediately reflected in every part of the page as well as any other article calling the template without the need for revisions on each page. User Flodded articulated the need for these templates:

> "Please USE them, and please stop removing them in favor of putting plaintext casualty counts in the article. We cannot keep numbers consistent between sections nearly as easily otherwise. I

> think I've changed raw numbers back to templates 2 or 3 times now in different locations on the page."[48]

In-line casualty templates for dead, injured, and missing were all created in addition to a template for navigating articles directly related to the catastrophe. But as unregistered editor 220.101.28.25 noted "[will] many non 'newb' editors will likely go [to these templates]? Perhaps a link in the hidden note to direct editors to the template page?"[49] While the numbers on the article appeared as plaintext, these casualty numbers were nowhere to be found in the body of the article. While the solution was elegant, it was not simple. Editors wishing to change or update these numbers would have to know that the template was calling these numbers, how to search for the template, that templates themselves could be edited, and how to edit these templates. Surprisingly, these templates were sites of substantial amounts of activity: the "dead" template was revised 208 times by 57 unique users, the "injured" template was revised 88 times by 24 users, and the "missing" template was revised 139 times by 37 users all engaged in the macabre task of revising casualty figures. Following (Kriplean et al., 2008), this approach of abstracting the work of casualty number updating to a template both creates a unique and well-defined class of work around which editors can specialize as well as instantiating this work for very specific roles. By calling the template to transclude its content onto other articles, an editor is able to simplify the complex dependencies and interrelations among articles so that only one "object" needs to be updated instead of dozens of content instances. The practice also insulates crucial information from the vicissitudes of passing editors and the high-tempo parent article while privileging deliberations and consensus formation for those editors skilled and motivated enough to seek it out. In so doing, this practice of in-line casualty templates also abstracts and complicates the presentation of basic syntax thereby raising the bar for participation from novice users for better and worse.

---

[48] http://en.wikipedia.org/w/index.php?title=Talk:2011_T%C5%8Dhoku_earthquake_and_tsunami&diff=418568425&oldid=418565206

[49] http://en.wikipedia.org/w/index.php?title=Talk:2011_T%C5%8Dhoku_earthquake_and_tsunami&diff=418779751&oldid=418779534

Editor trajectories

The analysis thus far has provided the contextual background to understand the patterns of coauthorship, examine the dynamics of editor contributions over time, and identify influential editors. This section exmaines the trajectories of these influential editors. While this structural method invites the application of existing network analytic methods to understand centrality, clustering, and similarity when tracking multiple user or artifact trajectories together, as I did in Chapter 6, the focus here will instead be on qualitatively examining features in these editors trajectories' that predispose them or uniquely qualify them to participate in the earthquake and tsunami breaking news article collaboration.

To review, in an editor trajectory, each node is an article or other page on Wikipedia the editor has contributed to. The nodes are colored by their namespace, or the category of page on Wikipedia. There are at least 14 distinct namespaces on Wikipedia,[50] but activity is primarily concentrated in a handful of these. "Main" namespace is where the articles themselves reside, "Talk" namespace is the discussion pages associated with these articles, "User" namespace is where editors post information about themselves, "User talk" is where editors communicate with other editors, "Wikipedia" namespace is for administrative and policy-related content, "Wikipedia talk" is for discussions about these policies and procedures. The remainder about files, MediaWiki, templates, help, categories, and portals are highly specialized and make up a tiny fraction of total contributions to the entire project. The extent to which editors' contributions are concentrated in any one of these namespaces reflects some social role or specialization on the part of the editor as a contributor, copywriter, consensus-builder, vandal-fighter, policy-enforcer, or other roles. Because these patterns of contributing to specific namespaces are important for understanding users' roles, I color-code the nodes in the network by namespace. The namespaces, their descriptions, and color-coding are summarized in Table 7-4 below.

[50] http://en.wikipedia.org/wiki/Wikipedia:Namespace

Namespace	Description	Color
Main	Article content	Blue
Talk	Article discussions	Light Blue
User	User information	Red
User talk	User-to-user discussions	Light Red
Wikipedia	Policies	Green
Wikipedia talk	Discussion of policies	Light Green
File	Images, videos, audio	Orange
File talk	Discussion of images, etc.	Light Orange
MediaWiki	Server software and interface	Purple
MediaWiki talk	Discussion of features	Light Purple
Template	Template content	Brown
Template talk	Discussion of template content	Light Brown
Help	Support documentation	Black
Help talk	Discussion of documentation	Black
Category	Groups of articles	Black
Category talk	Discussion of categorizations	Black
Portal	“Homepages” for topical areas	Black
Portal talk	Discussion of topical areas	Black
Special	Pages containing non-wiki text, generated on demand by software	Black

Table 7-4: Namespace summary

The edges in this graph also encode information related to the delay or lag between an editor's consecutive edits. Because an editor can potentially shift from editing article A to article B many times, this edge can contain multiple lag values that can vary dramatically in their values. To simplify this array of lags, only the median value reflecting a central tendency for the editor to wait before editing the next article is used. Some lags may be very short, on the order of seconds or minutes, reflecting a highly engaged editor moving quickly to update several articles in rapid succession while other lags may be very long, on the order of months or years, reflecting an editor who went on hiatus. These time lags are reflected in the trajectory by adjusting the darkness or opacity of the edges such that darker lines indicate shorter (median) lags reflecting immediate engagement while fainter or whiter lines indicate longer (median) lags reflecting incidental relationships. These distinctions are especially important in the context of a breaking news collaboration as the rapid engagement of editors across a variety of articles may reflect important coordination work responding to problematic users, standardizing information across articles, or executing a decision made in discussion with others.

User Flodded was the prolific contributor to the "earthquake and tsunami" article, making the most contributions (560) in the corpus and is the first editor trajectory (Figure 7-20). Flodded's first edit was made in August 2009 to the article "Shellfish" and involved updating and adding citations. He[51] edited an article about a failed dot-com company "AboveNet" (see phase 1 in Figure 7-20) and then went on a lengthy hiatus until January 2011. Flodded's renewed editing activity was related to another breaking news event, Jared Lee Loughner's assassination attempt against Gabrielle Giffords in Tuscon, Arizona. Flodded edited the articles "2011 Tuscon shooting", "Jared Lee Loughner", "Gabrielle Giffords", and "United States Congressmen killed or wounded in office" in rapid succession over an 11 hour period on January 11 (see phase 2 in Figure 7-20). Flodded was initially involved in copyediting the articles to remove unverifiable speculation and unencyclopedic content. As is often the case with breaking news articles, this article was "semi-protected" by administrators to limit the changes made by novice or

[51] For the sake of brevity and simplicity, I will use "he" as a gender-neutral pronoun to refer to editors unless they otherwise disclose their gender as female.

unregistered users. Unregistered editors or editors who have been active for fewer than 4 days and 10 edits are blocked from editing, but may make requests for edits on the talk page. Flodded was involved in responding to several of these edit requests and then became involved in an intense discussion about whether Loughner identified as an atheist on both the discussion page and "Biographies of living persons" administrative noticeboard. He continued to perform copyediting duties on the Loughner article, fixing capitalizations, ensuring the consistency of names and styles, and correcting grammatical mistakes as well as remaining involved in the article's discussion page. Despite the marathon 11 hour editing session, Flodded abruptly stopped editing the article and did not make another contribution until February 21, performing daily anti-vandalism work on unrelated articles about "Extremes on earth", "Bell Mobility", "Lowest temperature record on earth", and other topics on a daily basis. However, as phases 3 and 4 in his trajectory suggest, he was not deeply involved in the on-going maintenance of these articles but simply made a single contribution and moved on to other topics. In early March 2011, he edited the article "Cheiracanthium", a genus of spiders, to update information implicating them in a recall of Mazda vehicles.[52]

As discussed above, Flodded was a relatively early editor of the "earthquake and tsunami" article, the large article at the center of the trajectory at bottom, but as the article trajectory above demonstrated, he was not among the first editors. His initial edits focused on removing over-specific information relating to areas where minor tsunami alerts had been issued justifying these edits on the talk page:

> "We could list out thousands of places with tsunami warnings or that received a few extra cm of water. Obviously this is not feasible, nor is it encyclopedic. I suggest a good balance would be to only list places that have reported more than minor damage, have reported casualties, have reported large-scale evacuations in mainstream media, or are otherwise notable."[53]

---

[52] http://articles.latimes.com/2011/mar/03/business/la-fi-mazda-spider-20110304

[53] http://en.wikipedia.org/w/index.php?title=Talk:2011_T%C5%8Dhoku_earthquake_and_tsunami&diff=prev&oldid=418355602

Flodded was also an extremely active editor on the discussion pages, making 257 revisions between March 11 and March 22 on topics like the looming nuclear disasters, finding sources to verify the extent to which the island of Honshu had been displaced, and increasingly on the topic of establishing reliable numbers about the casualty tolls. Flodded went on a remarkable 24-hour editing marathon; between 19:35 UTC on March 11 and 19:47 UTC on March 12 Flodded made several edits per hour presumably precluding the ability to sleep during this time frame. After a seven hour break, he embarked on another 24-hour editing marathon stretching from 3:34 UTC on March 13 to 2:47 UTC on March 14 in which he made several changes per hour.

Returning to the editor trajectory, several structural features merit discussion. First the graph is comparatively small, having only 149 nodes and 285 edges, but very dense (1.27e-2). The halo of light red points around the central "earthquake and tsunami" article (phase 5 in Figure 6-7) represents the talk pages of other users Flodded communicated with about the article, warning them to stop reverting his changes or providing boilerplate welcome messages to new users cautioning them about the norms of editing on Wikipedia. This halo structure of pendants with reciprocated ties to the core article reveals that Flodded would be working on the earthquake and tsunami article, go to these users' talk pages to warn them, and then return immediately to editing the central article again. Several articles are also present in this halo such as articles with alternative titles for the event ("2011 Sendai earthquake", "Japanese earthquake and tsunami", "2011 Tohoku earthquake") that each redirect to the main article. The strong tie between the main article and the light blue dot reflects that a substantial amount of his total activity involved shuttling between the main namespace article and the article's talk page in rapid succession, 107 transitions in total with a median edit lag of 4 minutes and 7 seconds. Flodded was also involved in a variety of administrative processes related to requesting page protection as well as filing reports related to user misbehavior which are the peripheral green nodes near the central node.

Flodded's intense editing sessions became shorter and more infrequent and he began to shift attention to editing the casualty templates on March 16, (phase 6 in Figure 7-20). As previously discussed, this is

highly specialized and technical work involving knowledge of how to identify and locate templates, format them appropriately so they appear correctly in the rendered pages, and update the information contained within them on a regular basis. As the Japanese authorities released information about causualty numbers at the beginning and end of each day, Flodded would take these reports and update the numbers in the corresponding templates. Despite these contributions to the casualty templates, Flodded remained involved in many other aspects of the article, a “jack of all trades” involved in many discussion threads, communicating with users on their talk pages, performing copyediting, updating information on related articles such as “Lists of earthquakes by magnitude”, and participating in administrative discussions. His final edits on the topic were March 23, and apart from 3 revisions to the Libyan civil war on April 3, Flodded has not made a single contribution since then.

Flodded fulfills an interesting role of as an editor demonstrating a latent interest in editing articles about current events throughout his history but also unusually dedicated by contributing for 48 hours in a 55-hour period of time and making a substantial number of edits in the successive weeks. Although his edits were highly concentrated, he nevertheless played a crucial coordinating role discussing a variety of topics with editors on the talk and their user pages. Despite the apparent lack of an editing history which would qualify him for this type of work, Flodded fluently engaged in a variety of tasks, demonstrating knowledge of Wikipedia policies justifying his editing decisions when challenged by other editors, participating in arcane bureaucratic proceedings about protecting pages and notifying administrators of trouble, and actively developing and modifying highly specialized templates about casualty numbers.

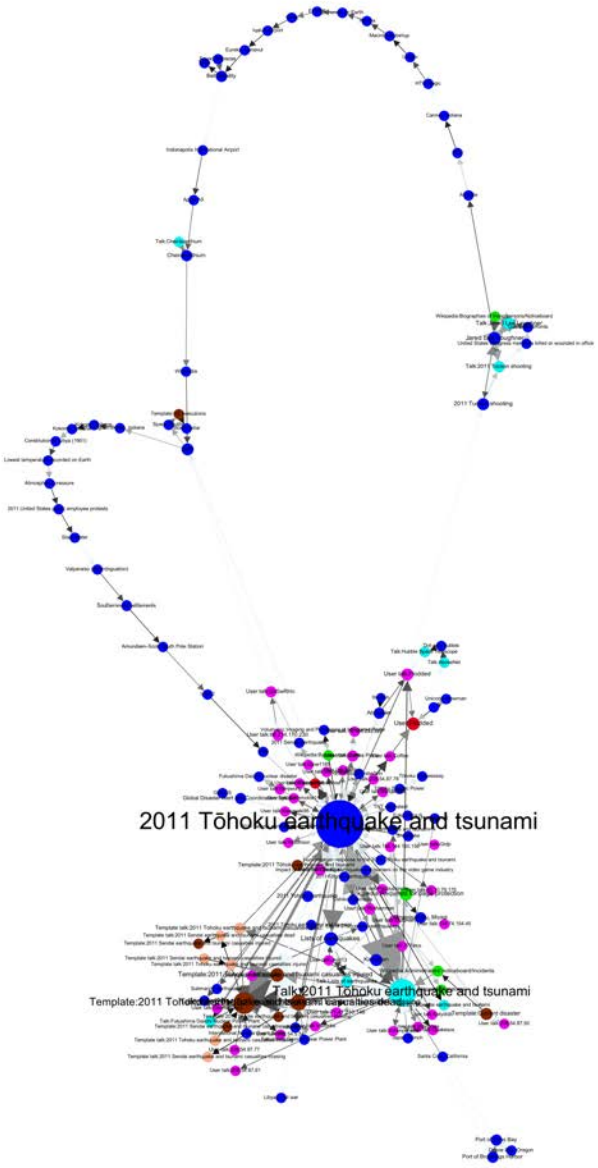


Figure 7-20: User Flodded's edit trajectory. Numbers in red boxes refer to distinct sub-graphs or phases of Flodded's editing career.

User L.Tak was the second most prolific editor in the corpus, the most active editor of the “Fukushima Daiichi Nuclear Power Plant”, and the second most active editor of the “nuclear disaster” article with 211 edits (after User Sandpiper’s 281 edits). L.Tak’s editor trajectory is plotted in Figure 7-21. This trajectory reveals several significant differences from Flodded’s structure that in turn have implications for understanding the role ecology of users responding to breaking news articles. First, it’s clear that L.Tak has a substantially deeper and more varied editing history than Flodded, making 9,907 revisions since making his first contributions in late October 2007 and then beginning the contribute regularly in May 2009 on the article “European Parliament election, 2009”. 63.9% of L.Tak’s contributions are in the “Main” article namespace, 14.5% in the “User talk” namespace, and 11.0% in the “Talk” namespace for article discussions. With 3,206 unique pages edited and 6,105 unique edges, L.Tak has a substantially larger but also less dense (5.93e-4) trajectory than Flodded.

While Flodded had a predilection for contributing to articles about events in the news, L.Tak’s extensive editing history is more complex. The most central article is his own talk page suggests much of his activity involves responding to other editors’ queries and concerns. The history of this talk page suggests a problematic debut and struggle with the learning curve of Wikipedia norms and rules initially but more recently becoming a backchannel with other editors soliciting his opinion and asking for elaboration on actions performed elsewhere. Other central articles in his trajectory concern international trade, visa, and labor agreements as well as environmental organizations. L.Tak’s intense involvement in and extensive contributions to the “nuclear disaster” article motivating this analysis are, incidentally, very peripheral in his trajectory residing in the dense outlying subgraph at approximately 1 o’clock. The articles preceding his involvement in the nuclear disaster article are a variety of copyediting tasks and linking to other concepts on a variety of outwardly mundane topics like provincial and colonial governance in the Netherlands and the articles following his involvement are about the foreign relations of European countries and nuclear treaties.

The work L.tak performed was initially focused on the “nuclear plant” article copyediting to ensure the consistency of times and timezones, removing alarmist predictions, and plagiarized material. While L.tak did not have the marathon 24 hour editing sessions of Flodded, he nevertheless made regular contributions over 6, 8, and even 14 hour periods of time between March 11 and March 15, with contributions slowing thereafter. L.Tak also fulfilled an essential coordinator role, with his contributions shuttling between the article page, discussion page, and user talk pages. The contributions L.Tak made during this time largely involved copyediting and removing duplicate information as well as adding information about the timeline of events and reliable sources.

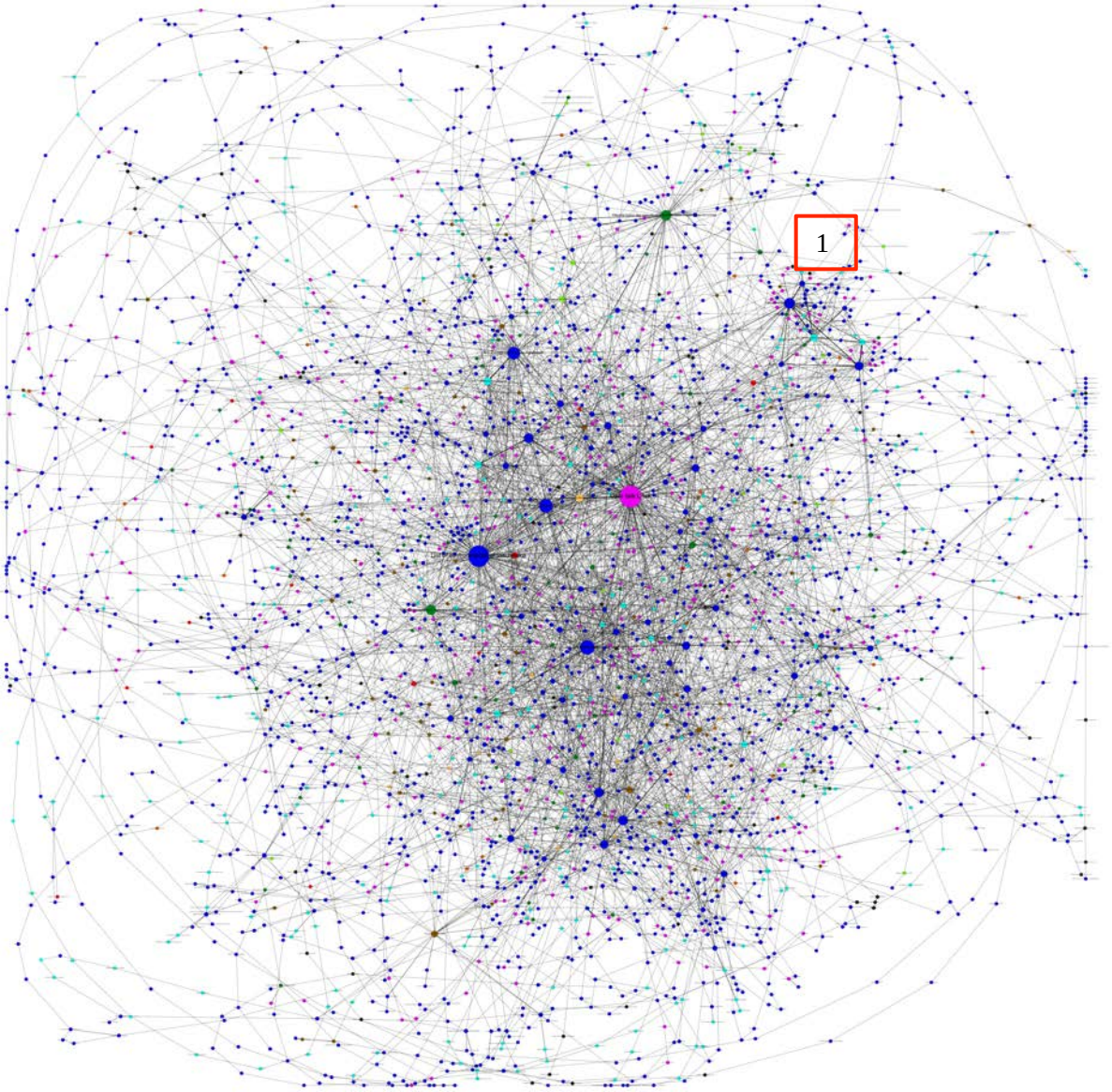


Figure 7-21: User L.Tak's edit trajectory. Numbers in red boxes refer to distinct sub-graphs or phases of LTek's editing career.

User Sandpiper was the sixth most active editor in the corpus and the most active editor of the "nuclear disaster" article. Sandpiper made 9,240 revisions since starting in June 2005 editing articles about Sussex and Harry Potter (phase 1 in Figure 7-22). Like L.Tak, his editing trajectory is also substantially more complex than Flodded but Sandpiper's trajectory also has distinct subgraphs corresponding to distinct phases of his editing history. Dalliances with unrelated topics are also apparent with a burst of editing relating to articles about English radio transmitting station towers (phase 2 in Figure 7-22), "Cutty Sark" (phase 4 in Figure 7-22), and a large amount of activity on the 1916 "Battle of Jutland" (phase 5 in Figure 7-22). Like L.Tak, Sandpiper's participation in the "nuclear disaster" is not embedded within a larger subgraph of breaking news events, but a tangent from his typical edits (phase 3 in Figure 7-22).

This lack of embeddedness in a larger context of current events editing occurs in many other editor trajectories as well. User ACSE's 41,778 revision editing history focused predominately on a strange pair of topics, Japanese pop and Japanese serial killers. But ACSE was also the editor who contributed to 34 articles in the Tohoku corpus, updating information on many of the pre-existing articles about towns, villages, and other points of interest that had been affected by the tsunami as well as editing the "earthquake and tsunami" as well as "nuclear disaster" articles extensively. User Johnfos's 38,134 revisions focus extensively on renewable energy and anti-nuclear movements, but he was the third most active editor by revisions with extensive contributions to articles related to the nuclear disaster. These editors are not regular editors of breaking news articles, but their editing trajectories and histories reveal they have specific and important types of expertise about Japanese culture and geography as well as the social and cultural history about nuclear technologies that are essential for these collaborations.

However, there are also users like Goodvac who participate to a great extent, not in user-facing articles in the "Main" namespace, but in administrative and support pages like WikiProject Resource Exchange where users can request and provide access to rare resources, anti-vandalism projects, and users with offensive names. Nevertheless Gooodvac also has a history of editing articles about current events such as the "Death of Osama bin Laden" and "Costa Concordia disaster" in addition to the "earthquake and

tsunami" article. On the "earthquake and tsunami" article, he initially focused on reformatting and populating the fields of references so that more contextual source information was available and then began to actively expand the article adding information from various sources about the economic and financial impacts. After Flodded departed in late March, Goodvac picked up the slack by updating information on the casualty templates and continued to provide these updates until mid-April. On other breaking articles like the "Death of Osama bin Laden" and "Costa Concordia disaster", Goodvac continued to do basic copy editing and reference formatting tasks, but did not exhibit patterns of central coordinators like Flodded or L.Tak shuttling between article discussion pages and user talk pages.

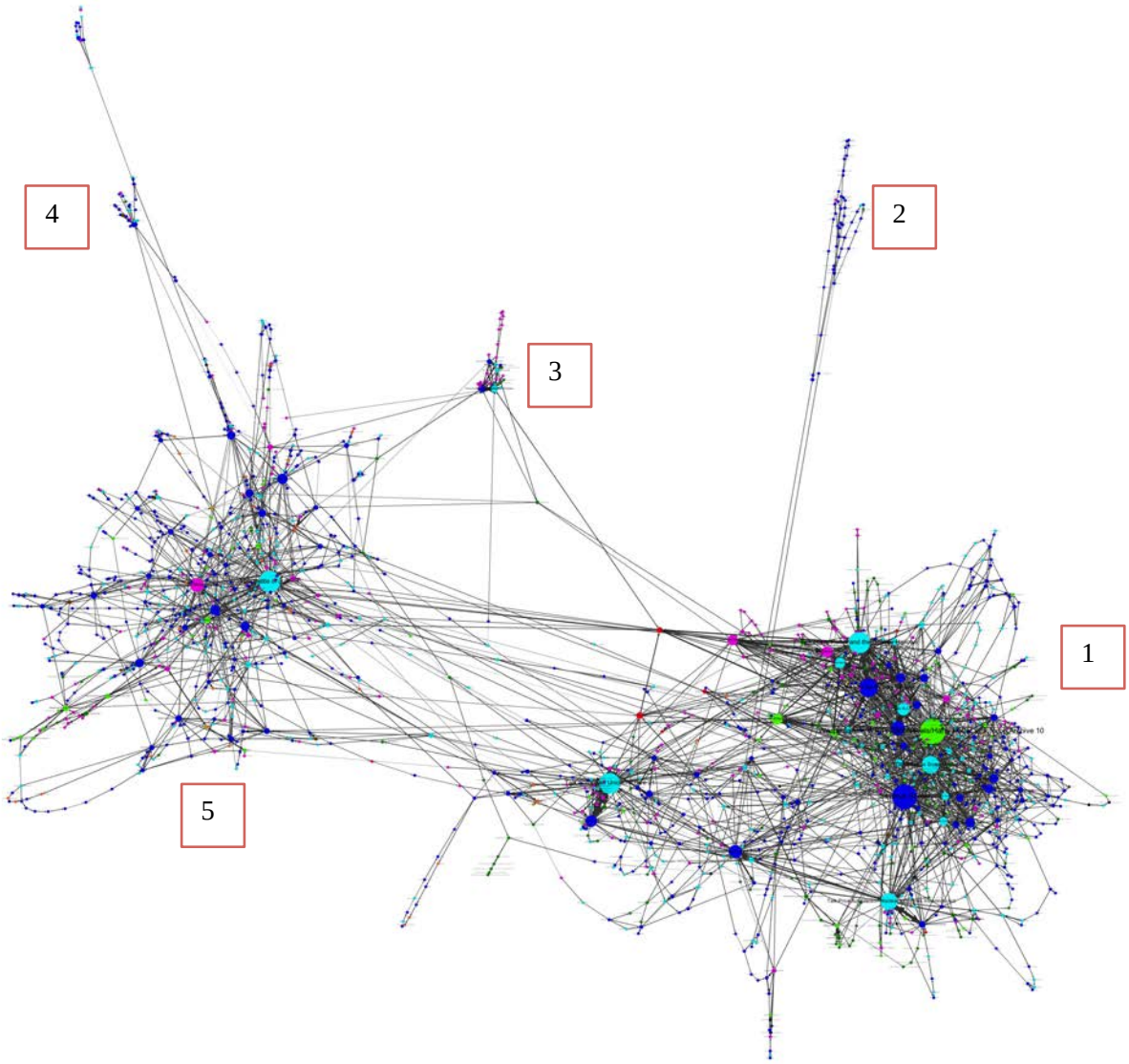


Figure 7-22: User Sandpiper's edit trajectory. Numbers in red boxes refer to distinct sub-graphs or phases of Sandpiper's editing career.

Discussion

This chapter examined several features of structures and dynamics of the editors' participation in a breaking news collaboration using the 2011 Japanese earthquake and tsunami. Like other breaking news events, its surprise and severity elicited an immediate and large-scale response in both the traditional news media as well as from the Wikipedia editors. But unlike collaborations around other breaking news articles that are largely retrospective and focused on a single event, this case was interesting because the Fukushima nuclear disasters following the earthquake and tsunami became major breaking news events in their own right and in parallel with the on-going response to the earthquake and tsunami. This case thus provides insights into how self-organized collaborations operate in the limit when there are multiple parallel demands on participants' attention as well as how to identify the ways in which this highly complex and contingent parallel work was coordinated. The findings from this analysis suggest that the work on these articles is highly centralized in specialists undertaking heroic and one-off efforts to coordinate this work rather than being distributed among teams dedicated to covering breaking news.

Bipartite network analysis suggests highly differentiated roles in the network owing to substantial differences in the connectivity of editors. Like many complex networks, this collaboration exhibits highly unequal distributions of connections; a tiny fraction of editors makes the lion's share of contributions and edits to multiple articles. Examining the trajectories of several articles reveals that these highly central and prolific editors have a tendency to join and leave the collaborations relatively early, but not after investing dozens of hours of effort. This "temporal homophily" suggests that breaking news article collaborations are not sites for sustained editor engagement on the articles about the event. Moreover, this rapid churn in editors suggests that tacit knowledge about events, sources, and consensus are possibly lost despite the records in revision comments and discussion pages.

A characteristic feature of these breaking news article collaborations are shifting attention across articles as collective effort was initially focused on central article (the earthquake and tsunami article) but then

diffused to other articles and re-centralized again on another related breaking event (the nuclear disasters). Despite the opportunity for a single editor to make substantial contributions to each of the articles about parallel breaking news events, the most prolific editors on many articles like the nuclear disasters had negligible activity on others like the earthquake and tsunami. Examining the editing trajectories of several top contributors suggests that prolific editors' investments in breaking news articles are at once novel but also reflect a latent interest or expertise in the topic. Editors of the articles about the nuclear disasters are drawn not from a cohort of editors dedicated to editing breaking news events, but rather editors like L.Tak with a background in international trade, ACSE's familiarity with Japanese pop culture, or Johnfos's background in the anti-nuclear movement. These editors' backgrounds conferred the collaborative competence, editing skills, and norm familiarity to extend and expand their repertoire of practices and routines necessary to manage a complex collaboration even if they had limited or no prior experience working on breaking news articles. This suggests the capacity to engage in the intense coordination demanded on these articles can be acquired and learned in-situ rather than developed from peripheral participation on prior breaking news articles or reliance on other editors with whom they have previously collaborated.

However, it is important to note that the findings in this chapter are based upon only a single case study and may not generalize across other breaking news article collaboration. While I used mixed methods to identify editors occupying central roles and ground their practices based on their contribution patterns and content, these editors' contributions are the exception rather than the rule when examining the vast majority of other editors performed across the histories of these articles. Alternative approaches such as random or weighted sampling of editors rather than rank order examination, surveys and interviews with editors to establish dispositions and motivations, or developing a grounded theoretical account of the evolution of these articles may have better surfaced representative and generalizable themes. The data I used for this analysis omits any account of the content that these editors viewed or interactions they had with others outside of the narrow confines of Wikipedia. Reconciling how to develop thick accounts

using qualitative and interpretive methods on very large sets of distributed social behavior remains a pernicious challenge, but approaches such as iterative coding using human computation methods such as Mechanical Turk offer exciting possibilities to distribute tasks and aggregate qualitative and interpretive responses from large pools of human raters (Little, Chilton, Goldman, & Miller, 2010; Quinn & Bederson, 2011).

Previous chapters examined the large-scale structures and dynamics across many Wikipedia articles but these analyses obscured the work and identities of the editors participating in these collaborations. This chapter found that Wikipedia's collaborations on articles about current and breaking news events brings together a unique cast of characters with disparate backgrounds who fulfill distinct roles in these collaborations. This analysis suggests breaking news article collaborations rely to a great extent on functional roles of motivated editors self-selecting into these articles rather than structural roles such as news editors wholly dedicated to editing breaking news articles. While editors exhibited considerable variability in the structure of their editing trajectories reflecting their diverse backgrounds, trajectories within breaking articles follow regular structural patterns reflecting the presence of a highly centralized coordinators and substantial churn in contributor cohorts. Across breaking articles, these central coordinators appear to be unique as well as otherwise inexperienced breaking news collaborators. This complicates attempts to frame these collaborations as communities of practice because they lack the deference to tenure and peripheral participation and instead appear to embody the improvisation and adaptation found in other high tempo and emergent response groups. The emergence and expansion of collaborative infrastructure on these breaking articles thus appears to be more like the contexts and responses of disaster response rather than the regeneration of collaborative infrastructures as found in other temporary teams like emergency medical care.

These findings have theoretical implications for understanding the origins and transformation of social roles and structures. As other authors have noted, roles in Wikipedia are highly informal but these breaking news articles appear especially flexible given the variance in participants' backgrounds.

Breaking news articles about major news events will inevitably attract a large number of editors making only passing contributions. The responsibility for synthesizing, copy-editing, and integrating these contributions falls to everyone in an open peer-production system, yet editors with some contextual background but wholly lacking the experience of working on other high tempo articles nevertheless appear to thrive and invest themselves heavily. As Baker & Faulkner (1993) as well as Bechky (2006) found in their studies of role adoption, roles are not a consequent of position in a structure but resources that are claimed, negotiated, and enacted. Editors do not operate in a vacuum but continually encounter collaborations in the midst of their unfolding development complete with dependencies on synthesizing content across articles, copyediting new content, and explicitly coordinating efforts with other editors working in parallel. These overlapping dependencies constitute a dynamic environment of opportunities and resources which results in an ecology of roles which editors adopt and negotiate in response to others' actions (as seen in the article trajectory) as well as their own background (as seen in their editor trajectory).

Given the fluidity with which editors inhabit and shed roles in breaking news article collaborations, this suggests a number of opportunities and implications for technical design of participatory systems in other high tempo contexts like disaster response and national security. Reputation and identity signals appear to play minor roles as unregistered editors and editors with little experience contributing to breaking news articles nevertheless played highly central roles in these collaborations. The openness of these platforms to revision coupled with tools to quickly and effectively revert undesireable changes also played a major role in supporting the velocity of the collaboration. However, technical limitations with Wikipedia's revision system imposed significant barriers on the collaboration. Editors had to work in fits and starts to alternatively introduce a new large body of text and then engage in a substantial amount of small-bore copy edits to maintain it as changes which took longer than a few seconds or minutes would become outdated as other editors made changes in the interim. These findings also raise questions about how to improve the design of large-scale socio-technical systems operate in the context of high-tempo online

collaboration. If one were to make a recommendation system to match users with relevant backgrounds and levels of experience to articles about events in the news, would it be more successful recommending editors based on the features and demands of the article or recommending articles based on the backgrounds and characteristics of editors? The next chapter explores this duality and attempts to unpack whether features of breaking news articles attract editors or the features of editors cause them to specialize in editing articles of particular types.

Chapter 8: Statistical Network Analysis

Previous chapters explored the patterns governing the self-organization of breaking news article collaborations in a largely descriptive way. Chapter 7 began to tie together threads over how editors' positions within an article trajectory are also functions of their background and history of revisions in their editor trajectories. This chapter develops a statistical model of the processes underpinning the self-organization of these collaborations to understand whether features of the articles or editors exert more influence over the evolution of these collaborations.

The motivations of editors, design features of the community, affordances of peer production, and other latent social processes interact in complex ways at multiple levels to make Wikipedia work. Previous research makes clear that editors of Wikipedia articles fulfill distinct and diverse collaboration roles and different types of articles employ different forms of coordination (Kittur et al., 2009; Welser et al., 2011). However, extant scholarship has not examined the interaction between these features: how do editors with particular skills self-organize around articles requiring different forms of collaboration? To answer this question this chapter examines a "boundary condition" for self-organization on Wikipedia. I compare the coauthorship of Wikipedia articles about current and breaking news events such as commercial airline disasters to topically similar articles about historical airline disasters. Articles about breaking news are coauthored under "high tempo" conditions that demand unique forms of coordination to manage interdependencies. This chapter analyzes the coauthorship networks of high and low tempo articles as well as the attributes of editors who contribute to them.

This chapter reviews existing scholarship on the duality of Wikipedia as both user action embodied in artifacts and processes that support collaboration. Integrating this review, a multi-theoretical, multi-level model describing how features of editors, articles, and interactions between both influence collaboration structure is proposed (N. S. Contractor, S. Wasserman, & K. Faust, 2006; P. R. Monge & N. S. Contractor, 2003). These findings suggest that while the features of articles and attributes of editors both

influence structure, editors' experience more strongly governs the types of editors they collaborate with and the types of articles they work on.

Background

As is the case with many online communities, the majority of contributions to Wikipedia come from a fraction of the entire user base (Kittur et al., 2007; Panciera et al., 2009). Despite this disparity in effort, online communities like Wikipedia are able to escape traps such as the tragedy of the commons and social loafing owing to members' uses and gratifications (Lampe et al., 2010), diverse motivations (Rafaeli & Ariel, 2008), affordances of peer production (Benkler, 2006), and design features of the community itself (Ren et al., 2007). These approaches generally emphasize the agency of individuals to form and realize their own goals. In the context of Wikipedia, editors who are motivated to fulfill particular roles like copyediting or vandal fighting (Geiger & Ribes, 2010; Welser et al., 2011), are socialized into sub-communities with like-minded collaborators (Bryant et al., 2005), and recognize or revert the contributions of other editors (Kriplean et al., 2008). Peer-production communities are oriented around producing and maintaining an artifact such as an operating system or encyclopedia article. Although these artifacts are the agglomeration of individual users' actions, artifacts like Wikipedia articles are embedded within socio-technical systems which imbue them with innate agency. This material agency enables them to operate outside of the control of any single person and emerges from the network of user and system interactions (Contractor, Monge, & Leonardi, 2011). In the context of a Wikipedia article, content added or reverted by other editors, markup language interpreted into style formatting by the MediaWiki software, and page protections enforced by administrators are examples of articles "acting" on their "own" outside of any one user's direct control.

Coauthorship patterns on Wikipedia thus foreground the duality of persons and material artifacts (Breiger, 1974; Contractor et al., 2011; G. C. Kane & Alavi, 2008): collaborations occur around articles exhibiting particular features but articles also emerge from the contributions of editors with distinct traits. The material agency of articles such as topic, quality, age, or number of contributors influences the types of

editors who are capable of making further contributions to an article. For example, a featured article about Barack Obama probably will not preserve contributions from newly registered editors identifying with the Tea Party movement. However, the properties of editors also influence the types of articles they choose to edit and maintain. The human agency of these actors manifests in attributes such as varying expertise, editing experience, and roles. For example, a college freshman who uploads photos about soccer players will be unlikely to take up correcting formulae on general relativity. This suggests the features and attributes of *both* articles and editors influence the self-organization of collaboration on Wikipedia.

While scholars have articulated rationales for how and why collaboration and social action emerge from both internal human agency (what I refer to as editor-focused attributes) as well as external social foci (what I refer to as article-focused features), the interaction between these two approaches has not been studied. Several themes from each of these literatures motivate a statistical analysis to "decouple" this duality and model the individual influences as well as interactions between editor attributes, article features, structural pattern of editors' contributions to articles, and structural patterns of an article's contributions from editors. I use a statistical approach to simultaneously examine how the features of articles, attributes of editors, and these network structures *all* influence the organizations of Wikipedia collaborations. While processes which occur at a single level of analysis (such as 1 and 2 below) certainly play a role, the *interactions* between each of (1) an editor's attributes and his or her history of editing other articles and (2) an article's features and its history of revisions from other editors will provide a more complete account of self-organization on Wikipedia.

1. Editor based attributes. For example, do experienced editors contribute to more articles than non-experienced editors?
2. Article based features. For example, do breaking news articles involve more editors than traditional articles?

3. Editor-focused interactions with article features. For example, is an experienced editor more likely to contribute to breaking news articles if they previously contributed to *other breaking news articles* than a non-experienced editor?
4. Article-focused interactions with editor attributes. For example, are breaking news articles more likely to attract contributions from experienced editors if *other experienced editors* have also contributed than non-breaking news articles?

The following sections examine the prior literature that has typically focused either on article-focused features or editor-focused attributes and consolidate this work using a multi-theoretical, multi-level modeling approach that allows me to describe how both features of articles, editors and the interactions between them influence the structure of collaborations on breaking news events about commercial airline disasters.

Article-focused: Task Coordination and Social Foci

The demands of coordinating coauthorship on Wikipedia articles vary substantially with the age of the article and the number of contributors to it (A Kittur & RE Kraut, 2008; Kittur et al., 2009). An additional dimension is the contemporary salience of an article. Like other forms of social media (Palen & Vieweg, 2008), current and breaking news events uniquely motivate editors to contribute to and collaborate in Wikipedia (Keegan et al., 2011). However, co-authoring an article about breaking news events like commercial airline disasters involves complex, time-sensitive, and highly interdependent tasks. This section reviews how the features of breaking news articles not only influence patterns of coauthorship, but how these article-focused features *interact* with the attributes of the editors who contribute to these articles.

Although prior research suggests the compounding coordination costs of many editors engaged in interdependent work will inhibit the development of high quality Wikipedia articles (A Kittur & RE Kraut, 2008; Kittur et al., 2009), articles about breaking news complicate this assumption. On one hand, Wikipedia articles about breaking news events are often perceived to be exemplars of timeliness, breadth,

and reliability in the immediate aftermath of an event like the Virginia Tech massacre (Cohen, 2007). On the other hand, the concentration of new editor activity is also densest while the article is less than 24 hours old and being intensively developed (Keegan et al., 2011). The popularity and quality of these articles in spite of these constraints poses a paradox in which "breaking articles" remain high quality in spite of the number of editors attempting to make simultaneous contributions with incomplete information and no centralized coordination. Examining the features of only articles or editors may present an incomplete picture and the novel coordination processes which enable these articles to be rapidly authored but also high quality likely emerge from the interactions between features of the article and attributes of its editors.

As high-tempo collaborations, Wikipedia articles about breaking news events are characterized by non-routine and extremely urgent work, abrupt consequences, and intense attention. As previous chapters demonstrated, editors have diverse backgrounds, varied resources to contribute, and often leave as suddenly as the collaboration itself came together. Breaking articles differ significantly from non-breaking articles given the intensity of attention, concentration of effort, and other complex coordination schemes. As Chapter 5 demonstrated, article features such as being breaking or non-breaking are correlated with major differences in the structure of co-authorship networks. One explanation for this is the topical salience and demands of coordinating interrelated tasks on breaking news articles makes them *foci* that actively bring people together and shape their collective action much more than articles about historical events which do not demand high-tempo collaborations (Feld, 1981). Therefore, articles about breaking news events may attract more editors than Wikipedia articles about non-breaking news events.

H1	Breaking news articles will attract more editors than non-breaking articles.

There are latent tendencies for breaking news articles to receive many contributions (discussed above) or experienced editors to simply be prolific or engaged in many articles (discussed in the next section). However, considering the features of the article absent the attributes of the editors who contribute to it is necessarily incomplete. For example, the "sink or swim" coordination demands of a breaking news article

or attempts to limit "dysfunction from diversity" (Chen, Ren, & Riedl, 2010) may predispose editors to only want to collaborate with other editors who exhibit similar characteristics or qualifications as themselves on a breaking news article. If other experienced editors are contributing this may be social proof about the collaboration and may attract other experienced editors. This would manifest as an article-focused homophily in which the attributes of the article result in similar kinds of editors collaborating (McPherson, Smith-Lovin, & Cook, 2001).

H2	Experienced editors will co-author with other experienced editors on breaking news articles.

Editor-focused: Social Roles and Identity

An alternative rationale why collaborations around breaking news articles exhibit different processes of self-organization revolves around the attributes of the editors who contribute to these articles rather than the features of the articles themselves. Editor-focused attributes like experience can potentially explain why some editors contribute to more articles than others, but also how these editor attributes interact with article features and lead users to fulfill distinct roles contributing to particular types of articles.

The previous chapter reviewed the theories about roles in social media and Wikipedia which can consist of substantive experts, technical editors, counter-vandalism, and community building (Gleave et al., 2009). Other typologies have identified the placeholders, completers, housekeepers, and shapers who contribute, integrate, and shape content on Wikipedia (Yates et al., 2010). Because dedicated Wikipedia editors have distinct but stable behavioral patterns (Panciera et al., 2009), editors can be classified into distinct roles based on the distribution of their activity (Welser et al., 2011). The patterning of interactions among editors inhabiting particular social roles across different breaking news article collaborations can potentially explain how breaking news articles are co-authored in spite of steep coordination costs. Editors may have a particular interest in contributing to topical areas such as airline disasters. Experienced editors in this domain may have deep expertise about the appropriate vocabulary and style for describing an accident or knowledge about relevant citations (Bryant et al., 2005). Thus, these editors'

may be fulfilling “caretaker” roles in which they edit many articles while less experienced editors specialize on fewer articles.

H3	Editor attributes like experience will co-vary with the number of articles edited. Experienced editors contribute to more articles than other editors.

Again, considering the attributes of editors separately from the features of the articles they edit presents an incomplete account by divorcing the role editors play from the resources to which they contribute. Experienced editors may self-identify as “breaking news editors” who preferentially edit these articles out of novelty or immediate gratification of contributing to an in-demand information artifact. Experienced editors preferring to edit only breaking news articles would manifest as a pattern of editor-focused homophily in which the attributes of editors cause them to contribute to similar kinds of articles.

H4	Editor attributes like experience will co-vary with the types of articles they edit. Experienced editors will be more likely to contribute to similar types of articles than dissimilar types of articles.

Alternative Explanations for Collaboration Structure

In addition to the hypothesized explanations, a variety of alternative processes could account for differences in collaboration structures between articles covering breaking, contemporary and historical events. I control for the influence of these factors by including them as parameters in the model in addition to the hypothesized variables of interest (breaking news article and editor experience).

Article features such as the severity of a catastrophe, proximity to developed countries, evaluated quality, and article length are also likely to influence collaboration patterns. For example, in the corpus of breaking news events centering on airline disasters the number of fatalities and survivors of an accident will be strongly correlated with the amount of attention an article receives from Wikipedia editors. Wikipedians also exhibit a "self-focus bias" in which geographic proximity influences why articles appear in one language but not another (Hecht & Gergle, 2010). Accidents occurring within or near developed countries should receive more attention from editors than accidents occurring in developing countries. Finally, a number of studies have identified that the number of editors, length of an article, and article quality are all correlated (Blumenstock, 2008; Hu, Lim, Sun, Lauw, & Vuong, 2007). Both the article quality and article length are controls on the number of editors who contributed to the article.

Editor attributes such as tenure in the community and whether or not the editor is registered are also likely to influence the likelihood of editors making contributions to articles of various types. Editors who started editing earlier in Wikipedia's history likely have greater familiarity with best practices and may be more involved in editing many articles (Bryant et al., 2005). Registered editors should be both highly motivated and more committed to the community, both of which lead them to make more contributions than non-registered editors (Ren et al., 2007).

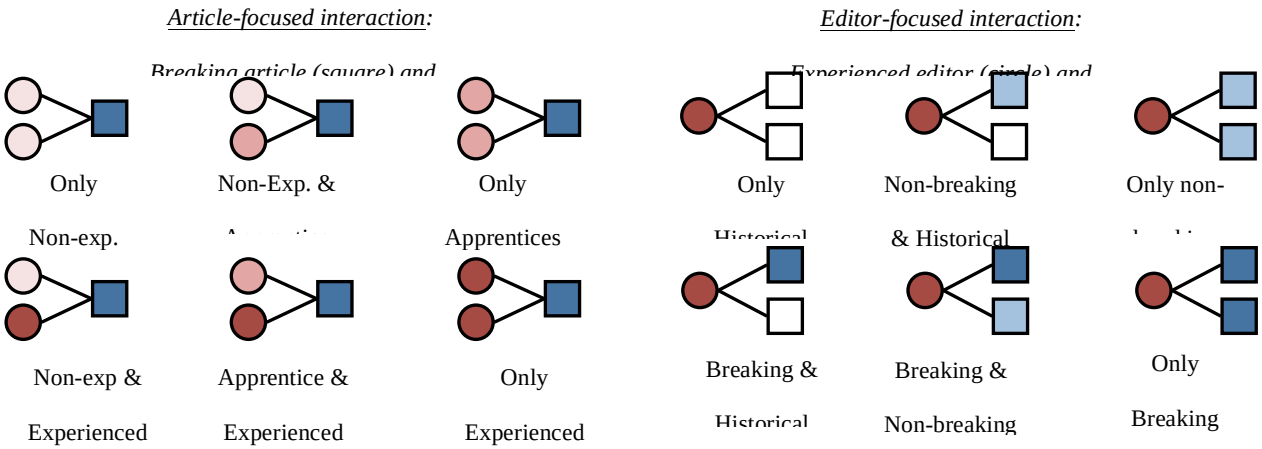


Figure 8-1: Visualization of p*/ERGM attribute interaction parameters that capture varying attributes of both the editors and the articles in Wikipedia. Dark-red circles are expert editors, medium-red circles are apprentice editors, and light-red circles are non-e

Approach

Establishing which of the collaboration processes is most influential requires a model accounting for the simultaneous contribution of each of these processes. However, owing to the methodological limitations of common types of network analysis, studies often only examine one level of analysis that accounts for little of the overall variance in the network. Analyzing the effects of network parameters interacting at different levels of analysis allow an integrated test of complimentary and competing theories of how network structure emerges (N. S. Contractor et al., 2006; P. R. Monge & N. S. Contractor, 2003). A statistical approach to network analysis uses generative approaches to simulate similar network structures, evaluate the differences, and thus confirm whether particular structural processes influenced the configuration of the network more than by random chance.

While descriptive approaches and regression analyses serve valuable purposes for answering particular research questions, these approaches are limited to analyzing the properties in a "snapshot" of a network. Techniques for measuring local and global properties of a social network have existed for decades. Metrics such as centrality, clustering, density, diameter, cliques, and equivalence fulfill valuable roles for describing the properties of nodes, subgroups, and the whole network. Nevertheless, these approaches are only descriptive in nature. These kinds of analyses cannot explain the endogenous processes of how the network structures itself nor the simultaneous influence of exogenous actor-level attributes on the network structure. Just as computing means, distributions, and pair-wise significance tests do not themselves provide meaningful insight into the relationship between gender and income (for example), simply calculating the aforementioned descriptive network metrics does not permit inferences about the underlying social and behavioral processes that govern how social networks are created or maintained.

Furthermore, because network data is fundamentally relational, these data violate the fundamental assumptions of traditional statistical approaches such as linear or logistic regression. A regression model of whether or not a tie exists between a pair of actors would assume the presence of a tie is independent from the occurrence of other ties in the network and probability for each tie being created is identically

distributed. In reality, this is an unreasonable assumption as the presence of a friendship tie between two actors in a social network may in fact be strongly influenced by a variety of exogenous factors such as whether the pair of actors have a friend in common, share the same race or gender, and the number of other friends they have.

Developing a statistical model of social networks thus requires a framework that accounts for the intrinsic interdependencies in network data. Over the past 25 years, a class of statistical models called P* or exponential random graph models (p*/ERGM) have been developed which explicitly incorporate dependence assumptions for network analysis (Anderson, Wasserman, & Crouch, 1999; Frank, 1981; Frank & Strauss, 1986; Pattison & Wasserman, 1999; Garry Robins, Pattison, Kalish, & Lusher, 2007; G. Robins, Pattison, & Wasserman, 1999; S. Wasserman & K. Faust, 1994; Wasserman & Pattison, 1996; Wasserman & Robins, 2005). Following Robins, Pattison, Kalish, & Lusher (2007), there are at least five distinct motivations for developing statistical models of network data.

1. Social behavior is a highly complex system that implies that even small changes can dramatically change how the system evolves and stabilizes. Rather than trying to make perfect deterministic predictions, the models are stochastic and this “noise” allows one to understand the distribution of possible outcomes for a model and the associated uncertainty of its estimates.
2. Statistical models of networks require the specification of statistical parameters and the estimates for these parameters indicate whether or not the structure occurs more or less often than would be observed by chance. Thus, analogous to multivariate regression, a statistical modeling approach allows for hypothesis testing about multiple social processes that may lead to the over- or under-representation of structural configurations of theoretical interest.
3. The presence of ties, clusters, or other structures in the network may emerge as a result of distinct processes occurring in parallel. As such, simply counting frequencies of these structures potentially confounds a variety of processes which could have led to their emergence. Again analogous to multivariate regression, a statistical model of networks should include controls

which allow the analyst to evaluate the relative contribution of each main effect to the network's structure as well as the possibility that these independent parameter variables also interact with each other to generate particular network configurations.

4. An accurate and parsimonious model of the processes which govern the structure of a network can be a highly efficient representation of the network. This is substantively important because accurate statistical models permit comparative network analysis by assessing the similarities or differences between the models themselves rather than either single-level descriptive statistics or large, complex complete representations.
5. Because global network structure emerges from local processes, statistical models of networks should rely upon parameters that are specified at a local level.

Statistical models like p*/ERGMs permit tests of hypotheses about the network without losing information about interdependencies among relations by estimating the probabilities that local network substructures are observed more frequently than by chance. These models extend the logic of multivariate logistic regression to relational data: the presence or absence of a binary tie in a network is an outcome variable predicted by a vector of independent variables called parameters. The range of possible networks and their probability of occurrence under the model are represented by a probability distribution on this set of all possible graphs. The model is constructed by specifying parameters corresponding to the structural characteristics(s) that are more likely to occur in the distribution. There are numerous structural tendencies to characterize these networks and the particular parameters employed must be theoretically relevant such as reciprocity, transitivity, cyclicality, and star tendencies. These structural tendencies, in turn, can be interpreted within a multi-theoretical framework that accounts for various endogenous and exogenous processes occurring at multiple levels of analysis that potentially govern the evolution of the graph (N. Contractor, S. Wasserman, & K. Faust, 2006; P. Monge & N. Contractor, 2003).

P*/ERG models make several assumptions. First, the number of nodes is fixed: models are compared against networks of the same size and do not account for the introduction or removal of nodes. Second,

the models only predict the presence or absence of ties, rather than the value (i.e., strength or frequency) or valence (i.e., positive or negative affect) of a tie. Third, each tie in the network is regarded as a random variable. Fourth, the model assumes a self-organizing and stochastic process of generating ties based on the surrounding context (what ties already exist). Fifth, a model is composed of parameters that correspond to the structural features of interest and apply homogenously across the network. Taking a friendship network as an example, one might expect that processes of reciprocity would occur over and above the chance that reciprocated ties occur at random. A model to test this process would then consist of two parameters: a parameter corresponding to the tendency for links to occur at random and a parameter corresponding to the additional tendency for links to be reciprocated. If there is a reason to suspect that reciprocation should occus more often in a given part of the network (women are more likely to reciprocate than men, for example) rather than occurring uniformly across the network, then this process needs to be modeled with yet another parameter (such as a gender attribute effect).

The observed network is only one possible realization of links among a given set of actors and the goal is to propose a model for the stochastic process which generated this network. The extent to which a particular process contributed to a network's structure can be ascertained by comparing the observed network to other networks generated by the hypothesized process. A statistical model thus assigns a probability to all possible networks such that networks exhibiting properties of interest are more likely to be used as a basis for comparison. In the friendship network example, if high levels of reciprocation are found in the observed network, the model should assign higher probabilities to the simulated graphs that also have high levels of reciprocation and use these as the basis for comparison.

However, the sample space permitted by networks' relational data is enormous because the addition or removal of a single link between any pair of actors necessarily results in a different network. These other networks may have substantively similar local and global metrics or very different structural properties. The whole sample space of a network includes the completely empty graph in which no connections are present between any nodes, the completely connected graph in which a connection is present between

every pair of nodes, and a huge number of permutations in between. The number of permutations for a network is a function of the number of nodes (n) it contains. For a directed graph, there are $2^{n(n-1)}$ possible permutations for the network while an un-directed graph has $2^{\frac{n(n-1)}{2}}$ possible permutations. Thus, a directed network of just 17 nodes has 2^{272} potential configurations that means there are more unique configurations for a 17-node network ($2^{272} \approx 7.58\text{x}10^{81}$) than there are atoms in the universe (~$1\text{x}10^{80}$). Because evaluating and assigning these probabilities to each distinct configuration of the network obviously becomes computationally intractable for graphs of even a seemingly trivial size like n=17, statistical models relying upon sampling and maximum-likelihood estimation are employed instead to generate parameter coefficients for p*/ERG models.

The estimation of parameters relies upon a maximum likelihood criterion to choose parameter values such that only the parameter that maximizes the structural tendency found in the observed network is selected for the model. If the parameter for reciprocity is equal to zero, then reciprocation occurs purely by chance. If this parameter is positive, then reciprocation occurs more often than one would expect by chance. If the parameter is negative, then reciprocation occurs less often than chance. Using a Markov chain Monte Carlo maximum likelihood estimation (MCMCMLE) approach, a distribution of random graphs is simulated based on a starting set of parameter values, these values are refined by comparing the distribution of the simulated graphs against the observed graph, and the process is iterated until the parameter estimates stabilize (Handcock, 2003; Snijders, 2002; Snijders, Pattison, Robins, & Handcock, 2006).

However, stabilization and convergence of parameters is not guaranteed. On one hand, a proposed model may be insufficiently specified such that the observed network is unlikely to occur even under the maximum likelihood coefficients. For example, the proposed model may vastly over- or under-estimate the amount of reciprocation or transitivity that is actually observed in the network. Using a goodness-of-fit approach to generate simulated networks and compare the resulting distribution of their network

metrics (i.e., degree distribution, diameter, embeddedness, etc.) against the observed network is a necessary confirmatory step for evaluating the validity of an estimated model.

On the other hand, estimation of models can also result in degeneracy when a model implies that every possible graph has a very low probability or the completely connected or unconnected graphs have a very high probability (M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, & M. Morris, 2008; Hunter, Goodreau, & Handcock, 2008). Alternatively for large networks, a model may fail to converge because it maximizes on a set of parameter values without sufficiently exploring the sample space for other maxima or because it lacks parameters which account for higher-order non-Markov processes in large networks (Goodreau, 2007; G. Robins, Snijders, Wang, Handcock, & Pattison, 2007). Similarly, over-specification of models with too many parameters results in networks that are substantively difficult to interpret, tremendously expensive to compute, and have statistically problematic estimates and standard errors given the loss of degrees-of-freedom. As a consequence, the estimation of networks using P*/ERGM approaches is as much an art as it is a science as it requires the inclusion of theoretically well-motivated parameters as well as careful model construction, iteration, and evaluation of model fit. The modeling approach and methods used in this analysis are outlined below.

Because the likelihood that a network tie is present or absent in a network is not independently and identically distributed (IID) from other network ties' likelihoods, a statistical model must account for these dependencies (Garry Robins et al., 2007). In a p*/ERGM, parameters allow the models to reflect dependencies on both endogenous tendencies for ties to exist because of other local structures in the network (e.g., popular articles continue to attract more editor attention) as well as exogenous attribute covariates (e.g., experienced editors prefer to work with other experienced editors). These methods allow tests of hypotheses about the tendency for ties to form as a result of the properties of the sending node, receiving node, as well as the presence and properties of other local ties and nodes.

The model produces a set of parameter estimates whereby estimates of zero indicate the modeled effect does not alter the likelihood from random chance, a positive parameter suggests the effect increases the likelihood of a tie, and a negative parameter implies the effect decreases the likelihood of a tie. Significance is tested using a *t*-ratio and concluded to be significant when the absolute value of this ratio exceeds a critical t-value of 1.96. Further details about the specification, estimation, and simulation of p*/ERGMs are beyond the scope of this chapter but can be found in (Faust, 1997; Garry Robins et al., 2007; G. Robins et al., 2007; Wang, Sharpe, Robins, & Pattison, 2009).

Data, Variables, and Methods

A variety of Wikipedia article genres like natural disasters, sporting events, and political scandals exhibit high tempo features that should require unique forms of coordination. This chapter returns to the corpus of commercial airline disasters for this study because these incidents occur with sufficient regularity to generate a large sample but are also "normal accidents" involving complex technology with prompt and serious consequences which make reliably notable events warranting coverage in Wikipedia. Articles for the corpus were drawn from the "List of accidents and incidents involving commercial aircraft."[54] The sample excludes hijackings and other instances of terrorism (such as the four flights involved in the September 11 attacks) because these incidents represent an archetype of catastrophe with distinct attention, salience, and causal attributions compared to "typical" accidents attributable to crew error, mechanical failure, or weather conditions. The resulting corpus includes 249 articles about commercial airline disasters occurring between January 1990 and through December 2010.

Based on the list of articles identified above, the revision histories for each article was extracted using Wikipedia's API. These data include revision-level data about the article name and ID, editor name and ID, timestamp, content, and comments. Registered users (*n*=6,462) identified by names and unregistered users (*n*=7,830) identified by IP addresses each have unique IDs. 14,292 unique user accounts made

[54] http://en.wikipedia.org/wiki/List_of_accidents_and_incidents_involving_commercial_aircraft

58,500 revisions to this corpus between September 16, 2001 and May 24, 2011. Each revision's article and editor ID were recorded as a duple representing an editor's modification of an article. Because p*/ERGMs estimate the binary presence or absence of a link rather than the weight or strength of a link, repeated editor-article duples were discarded creating a binary edgelist of 23,903 unique editor-article interactions. The edgelist was imported to the `statnet` statistical network analysis package in `R` for analysis using the `ergm` library (M. Handcock, D. Hunter, C. Butts, S. M. Goodreau, & M. Morris, 2008). The size of the resulting network required analysis on high-memory computing instances on Amazon Web Services' Elastic Cloud Compute (AWS EC2) infrastructure.

Bipartite Network Modeling

The Wikipedia revision data is a network consisting of a set of actors and a set of relationships among these actors. While traditional network analysis emphasizes unipartite or one-mode data where the relationships are between a single type of actor (i.e., people-to-people), two unique sets of actors actually exist in Wikipedia: editors and articles.

Because it is nonsensical for an article to edit another article or a user to edit another user, this interaction data is represented as a *bipartite graph* in which the nodes can be partitioned into exactly two mutually exclusive sets of actors such that ties only exist between sets and no ties exist within a set (S. P. Borgatti & M. G. Everett, 1997). Thus, a link exists between an editor node, *E*, and an article node, *A*, if *E* made a contribution to *A*, but neither *E*-*E* nor *A*-*A* links can exist. This bipartite structure is alternatively referred to as a "two-mode" or "affiliation" network. A bipartite network is a natural approach for modeling collaboration because it explicitly models the duality of persons and groups: a link between the social actors as one level of analysis and the groups to which they belong as another (Breiger, 1974). I employ p*/ERGM parameters designed specifically for bipartite networks (Faust, 1997; Wang et al., 2009).

Node Attribute Construction

A variety of article and editor variables were extracted from either the revision histories or article content to provide covariates for control, analysis, and modeling.

- **Article attributes** – The date of the incident and timestamp of the first edit to the article were recorded. Based upon the difference between these times and Wikipedia's creation in January 2001, three possible types of articles exist: *breaking articles* about events which are written within 72 hours of the incident (*n*=93), *non-breaking articles* about events which happened while Wikipedia existed but were written well after the incident (*n*=50), and *historical articles* about events which predated Wikipedia's existence and are thus written well after the event (*n*=106).
- **Other article attributes** – The *word count* excluding markup syntax is recorded as a continuous variable. The modal *article quality* as evaluated by associated WikiProjects is coded as one of seven categorical attributes ranging from Stub to Featured Article-class.
- **Incident attributes** – The count of *fatalities* and *survivors* are recorded for each event as controls for the severity and thus salience of a particular incident. A dummy control variable for *OECD location* coded whether or not the crash occurred within or off the coast of one of the 34 developed member nations of the OECD.
- **Editor attributes** – The total number of revisions an editor made *within the corpus* over the 10 year span of time was recorded as an edit count. This count was binned into three categories to capture topical editing experience: 1 to 3 contributions (4 is the mean) was *unexperienced* (*n*=12,148), 4 to 42 contributions (mean plus one standard deviation) was an *apprentice* (*n*=1,992), and 43 or more contributions was *experienced* (*n*=152). Editors who had registered accounts (*n*=6,462) were dichotomized from non-registered users (*n*=7,830). Finally, editors were partitioned into three classes: *early contributors* who made their first contribution to the corpus before 2008 (*n*=5,366), *middle contributors* who made their first contribution to the corpus from 2008 to 2010 (*n*=6,224), and *late contributors* who made their first contribution to the corpus between 2010 and the present (*n*=2,702).

Results

A single large p*/ERGM which includes single level main effects and structural tendencies (summarized on Table 8-1) as well as multi-level attribute interaction parameters (summarized on Table 8-2) is specified to test the four hypotheses. This method reproduces similar "main effects" findings for the control variables as previous studies: the likelihood of editors revising an article increases with every

additional fatality associated with the incident, incidents occurring in or near OECD nations, higher quality, longer word counts, early contributors, and registered users.

Endogenous structural parameters which control for the latent tendency for links to be created by chance (*edges*), the network-level tendencies for articles or editors to become highly centralized (*article* and *editor degree distribution*), and the local-level tendencies for articles to accumulate multiple editors (*multi-editor article tendency*) as well as editors to contribute to multiple articles (*multi-article editor tendency*) are also specified. The estimates for these parameters are listed in Table 8-1 as "structural tendencies."

The estimate for the negative *edges* parameter reflects the log-likelihood of a network tie appearing entirely by chance and serves as the "intercept term" reflecting the density of the network if no other effects were present. This can also be interpreted as the "cost" of creating a tie which other structural tendencies, factors, and interactions will need to overcome. The positive *multi-article* and *multi-editor* structural parameters respectively reflect the latent tendency for articles to accumulate editors and for editors to edit many articles. The negative *article degree distribution* reflects the tendency for articles to avoid long-tailed degree distributions while the positive *editor degree distribution* captures the tendency for editors to have a very skewed distribution. These findings suggest a tie is most likely to form between articles with few co-authors and editors who have also edited many other articles.

This is evidence for H1 that breaking news articles are more likely to attract editors than non-breaking or historical articles. Both of these article types are less likely to have ties to other editors than breaking news articles. However, the p*/ERGM estimates testing H3 invert the expectation that experienced editors would be more likely to contribute to many articles than non-experienced editors. The negative estimates imply apprentice and experienced editors are much less likely than non-experienced editors to edit many articles. Experienced editors (within the corpus) make repeated contributions to a few articles

(again only within the corpus) rather than a few contributions to many articles. This specialization points toward rejecting H3.

Structural tendencies	
Edges	-2.626 (-217.0)***
Multi-editor article tendency	0.01321 (2.89)***
Multi-article editor tendency	0.08042 (36.9)***
Article degree distribution (α=2.5)	-4.476 (-16.8)***
Editor degree distribution (α=0.25)	4.616 (68.6)***
Article features	
Controls	
Fatalities	1.22E-03 (20.9)***
Survivors	1.61E-04 (1.40)
Location: OECD	0.145 (24.9)***
Quality – Start	0.158 (21.9)***
Quality – C	0.371 (56.1)***
Quality – B	0.366 (93.4)***
Quality – GA	0.309 (11.2)***
Quality – FA	0.127 (1.38)
Word Count	6.51E-05 (8.00)***
Main effects (Hypothesis 1)	
Temporal Type – Non-breaking	-0.549 (-33.1)***
Temporal Type – Historical	-0.519 (-45.3)***
Editor attributes	
Controls	
Registered	0.700 (86.3)***
Cohort – Middle contributor	-0.137 (-6.87)***
Cohort – Late contributor	-0.154 (-3.45)***
Main effects (Hypothesis 3)	
Experience – Apprentice	-4.14 (-525.2)***
Experience – Experienced	-2.00 (-66.3)***

Table 8-1: "Main effect" p*/ERGM estimates (t-statistics). Estimates are net of parameters in Table 2. Cells are shaded green for positive & significant estimates and red for negative & significant estimates.

Modeling Attribute and Structural Interaction Effects

p*/ERGM methods stand out from traditional regression approaches in their ability to model the interactions between local editor-article authorship structure, editor attributes, and article features with parsimonious and statistically-valid parameters. Bipartite p*/ERGM parameters like those visualized in Figure 8-1 are used to model the complex interactions between article features and editor attributes. These feature and attribute estimates are summarized in Table 8-2.

There are two broad classes of interactions reflecting the two possible explanations for editor-focused or article-focused processes to influence the collaboration structure. Editor experience and article temporal type are the interacting attributes I use in this analysis. “Article-focused interactions” capture the tendency for articles possessing a feature such as “breaking news article” to be edited by actors sharing similar or dissimilar attributes like experienced/apprentice/non-experienced editor. “Editor-focused interactions” capture the tendency for editors possessing an attribute like “experienced” to contribute articles sharing similar or dissimilar features like breaking/non-breaking/historical article. Figure 8-1 visualizes six structural signatures for article-focused interactions with editor attributes on the left and six structural signatures for editor-focused interactions with article features on the right.

Within “Article-focused interactions”, a general tendency is observed for significantly fewer interactions than would be expected by chance between non-experienced editors and apprentices (P2) and experienced editors (P4). Coauthorship among non-experienced editors (P1) is likewise rarer than random chance. Likewise, there appear to be strong disincentives for apprentice editors to work with each other (P3) on articles about contemporaneous incidents. Despite the expectation that experienced editors would fulfill crucial roles in high tempo collaborations around breaking news events by intensively collaborating together, after controlling for variables such as the severity of the event, experienced editors work together on high tempo collaborations significantly less often than expected by chance (P6). This finding is at odds with Chapter 5’s results emphasizing the rapid emergence of connected collaborations around breaking news articles, but can also be interpreted in light of the unusual editor trajectories discussed in

Chapter 7. The most experienced editors of breaking news articles participate in relatively few breaking news article collaborations and as a result, these editors are unlikely to interact across many distinct articles. However, apprentice editors are significantly more likely to contribute to most permutations of breaking, non-breaking, and historical articles than other types of editors. Reconciling this finding with the rapid emergence finding from Chapter 5 would suggest that it is these apprentice editors who make up the early responders that rapidly knit breaking article collaborations together rather than the experienced editors. Nevertheless, the other findings support H2 that the coordination demands of an article influence the tendency of editors with similar or dissimilar levels of experience to work together.

Within "Editor-focused interactions", editors of all levels of experience are unlikely to contribute to both breaking and historical articles (P10). The lack of shared coauthorship on breaking and historical articles suggests these are very distinct sub-genres with limited interaction between each groups' editors. However, apprentice editors have a tendency to edit diverse combinations of articles above and beyond the latent tendencies for editors to edit many articles. This offsets the main effect for apprentice editors to be unlikely to contribute to articles in general as well as the lack of a bridging role by experienced editors.

Experienced editors contribute to non-breaking articles (P9) at a rate much greater than chance and make contributions to different types of articles (P8, P10, P11) at rates much less (respectively) than expected by chance. Contrary to expectations, the effects of experienced editors' sustained contributions to only breaking (P7) or historical articles (P12) are weak and non-significant. Highly experienced editors are instead characterized by deep and sustained involvement in a few articles rather than stewardship of many articles. Instead, it is the apprentice editors who appear to play a crucial role not only contributing to many articles but also acting as crucial brokers providing bridges within breaking (P7), non-breaking (P9), and historical articles (P12) as well as between these article types (P8, P11). Again, these findings support H4 that an editor's level of experience will influence the tendency for them to edit similar or dissimilar types of articles.

The magnitude of the coefficients for these editor-focused main effects and interaction parameters are generally larger than the article-focused parameters. This suggests the attributes and structural interactions focused on editors play a stronger role in explaining the presence and absence of links between editors and articles than the features and structures focused on articles. Thus, editor attributes are more influential on the self-organization of high tempo Wikipedia collaborations than the features of articles.

The previous steps analyzed local-level processes but are these features sufficient to explain global network properties? The model's goodness-of-fit is assessed by simulating other networks based on this model and use the resulting distribution of networks to compare the properties of these generated networks to the observed network (Garry Robins et al., 2007). Using the `ergm` package's "gof" function, I simulate a sample of 10,000 networks based on the p*/ERGM in Tables 1 and 2 and measure fit using the degree distribution. Figure 7-2 plots the observed values (in blue circles) and distribution of simulated values (in yellow boxes) for the combined degree distribution for both modes of the network. The observed distributions are almost completely bounded by the distributions from simulated networks which reflects a good-fitting model.

Article-focused interactions (Hypothesis 2)			
	Breaking	*Non-breaking*	*Historical*
Only Non-Experienced (P1)	-0.0109 (-2.37)*	-0.0101 (-15.5)***	-0.0086 (-1.90)
Non-Experienced & Apprentice (P2)	-0.0153 (-3.33)***	-0.00122 (-2.59)**	-0.01449 (-3.02)**
Only Apprentices (P3)	-0.00317 (-0.68)	-0.03674 (-16.9)***	0.00166 (0.39)
Non-Experienced & Experienced (P4)	-0.01588 (-3.46)***	-0.02601 (-17.5)***	-0.01590 (-20.7)***
Apprentice & Experienced (P5)	0.00613 (1.10)	0.02869 (28.02)***	0.00415 (0.64)
Only Experienced (P6)	-0.03552 (-19.9)***	0.00578 (0.85)	0.00090 (0.23)
Editor-focused interactions (Hypothesis 4)			
	Experienced	*Apprentice*	*Non-Exp.*
Only Breaking (P7)	0.00552 (1.63)	0.11540 (28.3)***	-6.027 (-135.9)***
Breaking & Non-breaking (P8)	-0.06421 (-20.7)***	0.01308 (2.50)*	-5.977 (-52.7)***
Only Non-breaking (P9)	0.04338 (3.98)***	0.23140 (13.0)***	-5.116 (-32.7)***
Breaking & Historical (P10)	-0.07691 (-25.7)***	-0.01577 (-4.22)***	-6.136 (-81.7)***
Non-breaking & Historical (P11)	-0.05648 (-18.5)***	0.08355 (25.7)***	-5.517 (-54.8)***
Only Historical (P12)	0.00056 (0.12)	0.15120 (45.5)	-5.468 (-84.6)***

Table 8-2: Editor-article "interaction" p*/ERGM estimates (t-statistics) capturing the tendency for the row title to edit/be edited by the column title. Estimates are net of parameters in Table 1. Labels in parentheses next to attribute names are references for

Discussion

Returning to the bipartite network approach recognizes the agency of both articles and editors to influence the self-organization of collaboration. Editor attributes and article features are incorporated by modeling their interactions as a bipartite graph and using p*/ERGM methods. This expands on previous approaches by not only analyzing processes occurring at different levels of analysis but also by modeling the interactions between these levels in addition to controlling for potentially confounding processes. p*/ERGM statistical models disambiguated the article-focused and editor-focused interactions by specifying a model which simultaneously incorporates each of these potentially confounding processes to assess the relative contribution of each to the network structure. This approach revealed new insights regarding how Wikipedia editors and articles self-organize in relation to one another.

Prior scholarship has either developed editor-focused accounts examining how editor attributes (e.g., experience) influence collaboration patterns or article-focused accounts of why some articles features (e.g., task coordination demands) lead to higher quality or more contributions. This analysis simultaneously looked at both levels of analysis to better understand the relationship each has on the self-organization of collaborations involving extreme coordination demands and varied editor experience. These results suggest that editor experience and the features of articles in these editors' contribution history have a stronger influence on the self-organization of the collaboration than article features like coordination demands and the attributes of editors who contribute to these articles. This approach provides a more complete account of the processes that influence the structure of collaborations on Wikipedia than looking at descriptive statistics about the structure of network of editors or articles.

Applying this approach to the paradox of how breaking news articles exhibit high quality despite steep coordination costs and varied editor experience, this chapter unpacked how the attributes of editors have greater influence over this self-organization. These findings validated the hypotheses that not only are the coordination demands of articles matched with the number of editors who contribute to them (H1), but

that coordination demands of certain article types also lead editors to seek or avoid other types of editors *depending on the type of editor* (H2). However, as measured by both effect size and valence, the features of an article and its interactions with editor attributes play a secondary role in structuring the collaboration as compared to the attributes of editors and their interactions with article features. Although experienced editors exhibit a tendency toward concentrating their work in a few articles (H3), there is evidence that an editor's level of experience leads them to also work on or avoid certain articles *depending on the type of article* (H4).

While previous work examined how editors' varying level of expertise influenced how Wikipedia tools were used or other users were perceived and rewarded (Bryant et al., 2005; Kriplean et al., 2008), editors' patterns of contributions are mediated through their own intrinsic attributes, the coordination demands of an article, the kinds of articles they have contributed to in the past, and the types of editors who also contributed to those articles. Unwinding these intricate dependencies is crucial for understanding the processes that contribute to the formation, maintenance, dissolution, and re-emergence of social and technological interactions in Wikipedia and other online communities and distributed organizations.

Implications

Wikipedia's coverage of breaking news events suggests that peer production systems are capable of operating far from conditions of stable task demands and community membership (Keegan et al., 2011). These findings suggest that tasks that demand high tempo knowledge collaboration may benefit more from matching users to tasks based on their own experience level and history of contributions to similar tasks in the past rather than assembling a team solely by optimizing on the demands of the task or the experience of other members of a team. Recruiting members with diverse backgrounds and interests may increase group productivity (Chen et al., 2010), but these results suggest that special care should be paid to the particular configurations and combinations of interests rather than dimensionless indices of diversity.

Statistical models allow for a more parsimonious and theoretically coupled representation of dense and complex network structures by capturing the local-level interaction tendencies as well as the emergent macro-level structure. Moreover, complex dependencies in networks are difficult to make sense of with descriptive statistics, visualizations may not provide statistically valid inferences, and due to the differences in levels of analysis parameterizing and controlling for these complex dependencies is extremely difficult, if not impossible, with traditional OLS and even hierarchical regression (e.g., mixed model) techniques.

Like regression or other statistical approaches, p*/ERGM methods require specifying models with theoretically well-motivated parameters in addition to translating extant theoretical constructs into appropriate network parameters. Absent a theoretical rationale for model specification and appropriate controls, both types of models can recover spurious relationships. However, these approaches suggest scholars can pose more meaningful research questions about multi-level and multi-theoretical processes about self-organization in collective intelligence systems. p*/ERGMs also allow comparative network analysis by examining the similarities of the processes which structure networks of very different size, scale, and context (Faust & Skvoretz, 2002). This analysis only looked at one particular sub-genre of articles about airline crashes, but it would be possible to estimate models for other topics with a breaking news component such as earthquakes, hurricanes, or sporting events. p*/ERGMs for each of these could be estimated and a meta-analysis performed to compare the collaboration practices across topics or even other collaboration systems.

Limitations and Future Work

p*/ERGM methods are computationally intensive and become even more so as both the complexity of the model and the number of nodes in the network increases. Although new “peta-scale” computational infrastructures may address these bottlenecks, for the time being extending p*/ERGM methods to very large networks containing tens of millions of nodes like the entire Twitter, Facebook, or Wikipedia graphs

is impractical. However, well-motivated boundary specification and comparative analysis or sampling approaches combined with meta-analyses can make large-scale analysis more tractable (Handcock & Gile, 2010).

The p*/ERGM employed in this analysis assumed the data was cross-sectional and thus omitted potential temporal dependencies such as a tendency for an editor to contribute *after* another editor contributes. Longitudinal models of network change and dynamics can also be specified (Snijders, Van de Bunt, & Steglich, 2010). Although, editors' social roles play a role in coordinating work, it may also be the case that articles can fulfill "roles" socializing editors into particular collaboration norms or introducing them to effective coordination practices. This corpus also did not include editors' contributions to discussion pages and users' talk pages, which are major sites for coordination on these articles. Likewise, this analysis only employed a subset of many possible demographic and psychological attributes such as gender, location, education, disposition all of which can influence the depth and extent of editors collaborative behavior.

These results help to explain the processes that underpin the self-organization of these high tempo collaborations. This analysis set out to establish whether articles or editors exert greater influence over the creation of links in a bipartite network containing both. As a statistical model, this analysis moves beyond the analyses offered in previous chapters, which primarily offered descriptive accounts or tests of difference, to explain the emergence of this network as a result of multiple processes operating at multiple levels of analysis. While the findings demonstrate that the properties of articles and editors both play significant roles in the self-organization of these networks, editors patterns' of collaborating or avoiding certain types of articles are ultimately more influential on the structure of the network than articles' tendencies to attract or dissuade certain types of editors.

Chapter 9: Conclusions

This chapter will review the findings of the project in light of the theoretical framework, discuss their implications for designing socio-technical systems for high-tempo online knowledge collaboration, review the limitations of these methods, and propose a research agenda for on-going work in the domain. The project proposed an ecological feedback loop consisting of four components. First, high tempo online knowledge collaboration is only possible when there is a large pool of individuals with diverse capacities that can be brought to bear on tasks with diverse levels of complexity, modularity, and granularity. Second, there has to be an occasion for diverse users to self-assemble and focus their efforts outside of their typical contexts. Third, these individuals employ and exchange practices at multiple levels to coordinate their work reflecting a confluence of themes from high-tempo organization, distributed knowledge collaboration, and online community. Finally, these individuals return to their prior collaboration contexts with new skills and capacities to collaboration. Network theories and methods examining the structure of coauthorship and editing trajectories permitted an analysis of the diversity of editors' backgrounds, patterns of repeat participation, and changes in collaborations.

The histories of encyclopedism and journalism from Chapter 3 illuminated how the identities of journalism and encyclopedism were challenged and transformed by profound social and technological changes leading both types of knowledge work to inhabit larger and more diverse social worlds. The evolution of these genres and slippages and transformations in their identities and boundaries expanded the scope and possibilities of encyclopedic and journalistic knowledge work. This history was instructive because the emergence and influence of Wikipedia's news work illuminated long historical processes of knowledge contestation and transformation. Chapter 4 demonstrated how the simultaneous imperatives for encyclopedic authority and journalistic relevance required Wikipedia editors to engage in boundary work to expand the possibilities of encyclopedic coverage to include and promote current events, exclude content and events which are insufficiently encyclopedic, and defend these choices by normalizing the

type and scope of content to be included through formal policies and regulations. Early Wikipedians, seeing the traffic and contributions for 9/11-related content, saw the potential to expand the scope of encyclopedic action. Managing the anomalies arising from the expansion of the encyclopedia into current events coverage required Wikipedians to converge on some shared meanings about the types of content which warrant inclusion, exclude other alternative interpretations, and legitimize these distinctions by institutionalizing and adopting policies enumerating appropriate types of content. These discussions reflected a need to develop consensus about shared norms reflected the diverse capacities and interests emerging among Wikipedia editors who began to encounter boundaries when engaging in collaborations outside of their familiar contexts.

Following this precedent for breaking news events to be included as Wikipedia articles, the number of articles expanded substantially. Using the revision histories logging the editors and timestamps for natural disasters, technological accidents, and violent conflict, Chapter 5 constructed a coauthorship network for all the articles within and between these genres. Comparing the centralization of editors and articles within breaking and non-breaking article collaborations, both types of collaborations exhibited substantial similarities in their structure and evidence that group members in high-tempo online collaborations differentially pattern their interactions during the most acute phases of article development as compared to non-breaking and historical articles. The immediate emergence of connected components in the early stages of breaking article collaborations and the stability of breaking article editors' interactions with each other over time points to the presence of role specialization in editing breaking articles. This section also made a methodological contribution demonstrating that cross-sectional analyses of breaking and non-breaking news articles' collaboration structures obscure significant differences in the dynamics and evolution of these collaborations over time.

Chapter 6 introduced a method to convert the sequence of edits made by editors within an article or within an editor into a directed graph. These trajectories made it possible to model which editors modify which

other editors revisions within a single article as well as repeated patterns of editor interactions within and across articles can be identified. Using this approach to model the differences between the trajectories of breaking and non-breaking articles, breaking articles' trajectories are initially characterized by highly clustered and centralized interactions with other editors but these features diminish over time and regress to structures found on non-breaking articles. Looking across different articles' trajectories, a well-connected latent collaboration structure of editors interacting with each other across articles emerged revealing potential communities of practice of editors engaged in repeat collaborations on breaking articles as well as articles about hurricanes.

Chapter 7 found that Wikipedia's collaborations on articles about current and breaking news events bring together a unique cast of characters with disparate backgrounds who fulfill distinct roles in these collaborations. This analysis suggests breaking news article collaborations rely to a great extent on functional roles of motivated editors self-selecting into these articles rather than structural roles such as news editors wholly dedicated to editing breaking news articles. While editors exhibited considerable variability in the structure of their editing trajectories reflecting their diverse backgrounds, trajectories within breaking articles follow regular structural patterns reflecting the presence of a highly centralized coordinators and substantial churn in contributor cohorts. The emergence and expansion of collaborative infrastructure on these breaking articles employs more improvisational features like disaster response rather than the regeneration of collaborative infrastructures like emergency room care.

Integrating these descriptive accounts, Chapter 8 developed a generative statistical network model to unpack how the attributes of editors and features of articles influence the self-organization of these collaborations. The results of the model demonstrate that editor experience and the features of articles in these editors' contribution history have a stronger influence on the self-organization of the collaboration than article features like coordination demands and the attributes of editors who contribute to these articles. This approach provides a more complete account of the processes that influence the structure of

collaborations on Wikipedia than looking at descriptive statistics about the structure of network of editors or articles. These findings validated the hypotheses that not only are the coordination demands of articles matched with the number of editors who contribute to them, but that coordination demands of certain article types also lead editors to seek or avoid other types of editors depending on the type of editor. Although experienced editors exhibit a tendency toward concentrating their work in a few articles, there is evidence that an editor's level of experience leads them to also work on or avoid certain articles depending on the type of article. Put another way, artifacts themselves play a role in the self-organization of these collaborations.

This project also has several limitations that warrant discussion and are opportunities for future research. First, the study examined only the English Wikipedia, which is by far the largest and most successful encyclopedia in the Wikimedia Foundation's stable of projects as well as being the most successful wiki. Cultural norms, knowledge claims, and collaboration practices vary widely across projects (Hara, Shachaf, & Hew, 2010; Hecht & Gergle, 2009, 2010). Second, the trajectories approach employed in Chapters 6 and 7 are locally coarse: just because editor B follows editor A does not mean that they are collaborating on the same. Using content-level features such as changes made at the paragraph or sentence level would capture editor-editor interactions jointly modifying the same sections of text (Halfaker et al., 2011; Priedhorsky et al., 2007; Rzeszotarski & Kittur, 2012). Third, as Chapter 5 demonstrated, the cross-sectional aggregation of network data obscured important and significant differences between different types of articles. Dynamic network models and longitudinal visualizations to understand the evolution of these collaborations and can understand assembly processes in greater details (Snijders et al., 2010). Finally, the analyses here used logged and digital traces of user behavior that potentially exclude crucial types of behavior (Howison, Crowston, & Wiggins, 2011). Editors on the Tōhoku earthquake and tsunami article relied on Internet Relay Chat (IRC) channels rather than the article discussion pages to coordinate tasks during the most intense phases of the collaboration. These

interactions are not logged or recorded and are thus omitted from the analysis omitting alternative explanations for the structure and dynamics of these collaborations. These analyses were likewise retrospective and do not include immersive or participatory accounts of the collaborations. Future work should interviews and participant observation methods to understand editors' backgrounds and motivations, unrecorded practices of improvisation, and processes of developing shared mind.

The existence of diverse genres of Wikipedia articles about breaking news events as well as collaborators who work across articles demonstrates the possibilities of peer production and open collaboration in online communities. Like other socio-technical systems like Twitter, Wikipedia capably self-organizes resources, activities, and tasks following unexpected events. The modularity of tasks and practices in these online communities leads to the retailoring of large pools of human capital to support rapid convergence and divergence in new domains. Ensuring the stability of the community of contributors and motivating sustained contributions over time is paramount to the success of many online communities, but participation in online communities does not always occur under conditions of stasis. In particular, this class of breaking news articles presents a novel lens for future work to re-evaluate extant conceptualizations of "communities of practice" as an integrating theory of socialization and coordination in online communities. Whereas communities of practice emphasize co-located and tightly knit groups practicing together long enough to develop mutual and shared understandings through sustained interactions and coordination (Bryant et al., 2005), breaking article coauthorship is characterized by temporary and distributed work on immaterial artifacts among unfamiliar collectivities of loosely-related individuals. A "collectivities" approach emphasizing how individuals' knowledge is exchanged and competencies are integrated in high tempo online collaborations may help reconcile how work proceeds when peripheral participation and deference to tenure are impracticable in contexts like those involving emergent response groups (Lindkvist, 2005).

Nevertheless, these findings have substantial implications for theorizing about the motivations to contribute to online communities. Wikipedia's breaking article collaborations highlight the flexibility to rapidly accommodate and socialize large influxes of participants attempting to make sense of unexpected events by balancing competing interests to support openness, flexibility, and autonomy against institutional needs for structure, norms, and socialization over very different time scales. These collaborations offer a compelling case to examine how online communities balance competing interests to support openness, flexibility, and autonomy against institutional needs for structure, norms, and socialization over very different time scales. This project contributed to our understanding of the structure and dynamics of collaboration in multidimensional networks within socio-technical systems. Despite the noisy behavioral patterns in these massive datasets extending over ten years, this project identified substantive social roles and sub-communities of shared interest. Wikipedians' commitments to the synthesis and dissemination of timely, neutral, and reliable information about current events provides a case to reflect on how new forms of organizing characterized by mass collaboration, peer production, and social media are giving rise to new forms of encyclopedism and journalism (Boczkowski, 2004; Deuze et al., 2007). This research has implications for designing and managing new communication and information technologies to respond resiliently to support high tempo, knowledge-intensive, and computer-mediated organizations that are increasingly prevalent for vital social actors like entrepreneurs, disaster responders, and political reformers.

References

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, 107-136.

Allen, C., Nodelman, U., & Zalta, E. N. (2002). The Stanford Encyclopedia of Philosophy: A Developed Dynamic Reference Work. *Metaphilosophy, 33*(1-2), 210-228. doi: 10.1111/1467-9973.00225

Anderson, C. J., Wasserman, S., & Crouch, B. (1999). A p* primer: logit models for social networks. *Social Networks, 21*(1), 37-66. doi: Doi: 10.1016/s0378-8733(98)00012-4

Antin, J., & Cheshire, C. (2010). *Readers are not free-riders: reading as a form of participation on wikipedia*. Paper presented at the CSCW'10, Savannah, GA, USA.

Antin, J., Cheshire, C., & Nov, O. (2012). *Technology-mediated contributions: editing behaviors among new wikipedians*. Paper presented at the CSCW'12, Seattle, Washington, USA.

Aral, S., & Alstyne, M. V. (2011). The Diversity-Bandwidth Trade-off. *American journal of sociology, 117*(1), 90-171.

Argote, L., Turner, M. E., & Fichman, M. (1989). To centralize or not to centralize: The effects of uncertainty and threat on group structure and performance. *Organizational behavior and human decision processes, 43*(1), 58-74. doi: 10.1016/0749-5978(89)90058-7

Bakker, R. M. (2010). Taking stock of temporary organizational forms: A systematic review and research agenda. *International Journal of Management Reviews*.

Baldasty, G. J. (1992). *The commercialization of news in the nineteenth century*: Univ of Wisconsin Pr.

Bardram, J. E. (2000). Temporal Coordination–On Time and Coordination of Collaborative Activities at a Surgical Department. *Computer Supported Cooperative Work, 9*(2), 157-187. doi: 10.1023/a:1008748724225

Barzun, J. (1962). Notes on the Making of a World Encyclopedia. *American Behavioral Scientist, 6*(1), 7-14. doi: 10.1177/000276426200600102

Bechky, B. A. (2006). Gaffers, gofers, and grips: Role-based coordination in temporary organizations. *Organization Science, 17*(1), 3-21.

Bechky, B. A., & Okhuysen, G. A. (2011). Expecting the Unexpected? How Swat Officers and Film Crews Handle Surprises. *The Academy of Management Journal (AMJ), 54*(2), 239-261.

Bell, A. (1991). *The language of news media*: Blackwell Oxford.

Bell, B., & Kozlowski, S. (2002). A typology of virtual teams: Implications for effective leadership. *Group & Organization Management, 27*(1), 14.

Bellomi, F., & Bonato, R. (2005). *Network Analysis for Wikipedia*. Paper presented at the WikiMania, Frankfurt aum Main.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*: Yale University Press.

Berkowitz, D. (1992). Non-routine news and newswork: Exploring a what-a-story. *Journal of Communication, 42*(1), 82-94.

Berners-Lee, T., & Fischetti, M. (2000). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. New York, NY, USA: HarperInformation.

Birnholtz, J. P., Cohen, M. D., & Hoch, S. V. (2007). Organizational character: On the regeneration of Camp Poplar Grove. *Organization Science, 18*(2), 315-332.

Blumenstock, J. E. (2008). *Size matters: word count as a measure of quality on wikipedia*. Paper presented at the WWW'08, Beijing, China.

Boczkowski, P. (2004). *Digitizing the News: Innovation in Online Newspapers*. Cambridge: MIT Press.

Borgatti, S., & Everett, M. (1997). Network analysis of 2-mode data. *Social Networks, 19*(3), 243-269.

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks, 19*(3), 243-269.

Bosworth, S. L., & Kreps, G. A. (1986). Structure as process: Organization and role. *American Sociological Review*, 699-716.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: classification and its consequences*: The MIT Press.

Brandes, U., Kenis, P., Lerner, J., & van Raaij, D. (2009a). *Network analysis of collaboration structure in Wikipedia*.

Brandes, U., Kenis, P., Lerner, J., & van Raaij, D. (2009b). *Network analysis of collaboration structure in Wikipedia*. Paper presented at the WWW'09, Madrid, Spain.

Brass, D., Galaskiewicz, J., Greve, H., & Tsai, W. (2004). Taking stock of networks and organizations: A multilevel perspective. *Academy of Management Journal, 47*(6), 795-817.

Breiger, R. L. (1974). The duality of persons and groups. *Social Forces, 53*(2), 181.

Briggs, A., & Burke, P. (2009). *Social history of the media: from Gutenberg to the Internet*: Polity Pr.

Brown, A. (2011). *A Brief History of Encyclopedias: From Pliny to Wikipedia*: Hesperus Press.

Brown, J., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science, 12*(2), 198-213.

Brown, S., & Eisenhardt, K. (1997). The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly, 42*(1), 1-34.

Brown, T. M., & Miller, C. E. (2000). Communication Networks in Task-Performing Groups. *Small Group Research, 31*(2), 131-157. doi: 10.1177/104649640003100201

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage* (Vol. 45): Peter Lang Pub Inc.

Brunsson, N., & Jacobsson, B. (2002). *A world of standards*. Oxford: Oxford University Press.

Bryant, S. L., Forte, A., & Bruckman, A. (2005). *Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia*. Paper presented at the GROUP'05, Sanibel Island, FL.

Burke, P. (2000). *A Social History of Knowledge I: From Gutenberg to Diderot*: Polity Pr.

Burke, P. (2012). *A Social History of Knowledge II: From the Encyclopaedia to Wikipedia*: Polity.

Burt, R. S. (2004). Structural Holes and Good Ideas. *American journal of sociology, 110*(2), 349-399.

Bush, V. (1945, July 1945). As we may think. *Atlantic Monthly*, 101-108.

Butler, B., Joyce, E., & Pike, J. (2008). *Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia*. Paper presented at the CHI'08.

Butler, B., Sproull, L., Kiesler, S., & Kraut, R. (2002). Community Effort in Online Groups: Who Does the Work and Why? *Leadership at a Distance*.

Chen, J., Ren, Y., & Riedl, J. (2010). *The effects of diversity on group productivity and member withdrawal in online volunteer groups*. Paper presented at the CHI'10, Atlanta, GA, USA.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). *Using information scent to model user information needs and actions and the Web*. Paper presented at the CHI'01, Seattle, Washington, United States.

Cohen, N. (2007). The Latest on Virginia Tech, From Wikipedia. *The New York Times*.

Cohn, M., Mehl, M., & Pennebaker, J. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*(10), 687-693.

Collison, R. L. (1964). *Encyclopaedias: their history throughout the ages*: Hafner Publishing.

Conboy, M., & Steel, J. (2008). The future of newspapers. *Journalism Studies, 9*(5), 650-661.

Contractor, N., Monge, P., & Leonardi, P. M. (2011). Multidimensional Networks and the Dynamics of Sociomateriality: Bringing Technology Inside the Network. *International Journal of Communication, 5*(0).

Contractor, N., Wasserman, S., & Faust, K. (2006). Testing Multitheoretical, Multilevel Hypotheses about Organizational Networks. *Academy of Management Review, 31*(3), 681-703.

Contractor, N. S., Wasserman, S., & Faust, K. (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review, 31*(3), 681-703.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). *Social coding in GitHub: transparency and collaboration in an open software repository*. Paper presented at the CSCW'12, Seattle, Washington, USA.

Dahl, P. S., & Hansen, K. M. (2006). Diffusion of standards: the importance of size, region and external pressures in diffusion processes. *Public Administration, 84*(2), 441-459.

Darnton, R. (1984). Philosophers trim the tree of knowledge: the epistemological strategy of the Encyclopédie *The great cat massacre and other episodes in French cultural history* (pp. 191-215). New York, NY, USA: Basic Books.

Davenport, T. H. (1997). *Information ecology: Mastering the information and knowledge environment*: Oxford University Press, USA.

Davenport, T. H., Jarvenpaa, S. L., & Beers, M. C. (1996). Improving knowledge work processes. *Sloan Management Review, 37*, 53-66.

Dee, J. (2007, July 1, 2007). All the News That's Fit to Print Out, *The New York Times*.

Deuze, M. (2005). What is journalism?: Professional identity and ideology of journalists Reconsidered. *Journalism, 6*(4), 442.

Deuze, M., Bruns, A., & Neuberger, C. (2007). Preparing for an age of participatory news. *Journalism Practice, 1*(3), 322-338.

Dieckmann, H. (1961). The concept of knowledge in the Encyclopedie *Essays in Comparative Literature*. St. Louis, MO, USA: Washington University Studies.

DiMaggio, P., & Powell, W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 147-160.

Dunne, C., Riche, N. H., Lee, B., Metoyer, R. A., & Robertson, G. G. (2012). *GraphTrail: analyzing large multivariate and heterogeneous networks while supporting exploration history*. Paper presented at the CHI'12, Austin, Texas, United States.

Dynes, R. (1970). *Organized behavior in disaster*: Heath Lexington Books.

Eisenstein, E. L. (1980). *The printing press as an agent of change: Communications and cultural transformations in early-modern Europe* (Vol. 1): Cambridge Univ Pr.

Espinosa, A., Lerch, F. J., Kraut, R. E., Salas, E., & Fiore, S. M. (2004). Explicit vs. implicit coordination mechanisms and task dependencies: one size does not fit all. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Process and performance at the inter- and intra-individual level*. Washington, D.C., USA: APA.

Faraj, S., Jarvenpaa, S., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, orsc. 1100.0614 v1101.

Faraj, S., & Johnson, S. (2010). Network Exchange Patterns in Online Communities. *Organization Science*.

Faraj, S., & Sproull, L. (2000). Coordinating expertise in software development teams. *Management science*, 1554-1568.

Faraj, S., & Xiao, Y. (2006). Coordination in fast-response organizations. *Management science, 52*(8), 1155.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks, 19*(2), 157-191.

Faust, K., & Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology, 32*(1), 267-299.

Feld, S. L. (1981). The focused organization of social ties. *American Journal of Sociology, 86*(5), 1015-1035.

Feldstein, M. (2006). A Muckraking Model. *The Harvard International Journal of Press/Politics, 11*(2), 105.

Ferron, M., & Massa, P. (2011a). *Collective memory building in Wikipedia: the case of North African uprisings*. Paper presented at the 7th International Symposium on Wikis and Open Collaboration, Mountain View, CA.

Ferron, M., & Massa, P. (2011b). *Studying collective memories in Wikipedia*. Paper presented at the 3rd Digital Memories Conference, Prague, Czech Republic. 3rd Digital Memories Conference

Ferron, M., & Massa, P. (2011c). WikiRevolutions: Wikipedia as a Lens for Studying the Real-Time Formation of Collective Memories of Revolutions. *International Journal of Communication*, from http://www.ijoc.org/ojs/index.php/ijoc/article/view/1238

Fisher, D., Smith, M., & Welser, H. T. (2006). *You Are Who You Talk To: Detecting Roles in Usenet Newsgroups*. Paper presented at the HICSS'06, Honolulu, Hawaii, United States.

Foot, K., Warnick, B., & Schneider, S. M. (2005). Web‐Based Memorializing After September 11: Toward a Conceptual Framework. *Journal of Computer‐Mediated Communication, 11*(1), 72-96.

Foucault, M. (2002). Classifying *The order of things: An archaeology of the human sciences* (pp. 136-179). New York, NY, USA: Routledge.

Frank, O. (1981). A survey of statistical methods for graph analysis. *Sociological Methodology, 12*, 110-155.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81*(395), 832-842.

Fussell, S. R., Kiesler, S., Setlock, L. D., & Scupelli, P. (2004). *Effects of instant messaging on the management of multiple project trajectories*. Paper presented at the CHI'04, Vienna, Austria.

Gaio, L., Besten, M. d., Rossi, A., & Dalle, J.-M. (2009). *Wikibugs: using template messages in open content collections*. Paper presented at the WikiSym'09, Orlando, Florida.

Galtung, J., & Ruge, M. H. (1965). The structure of foreign news. *Journal of peace research, 2*(1), 64-90.

Gaved, M., Heath, T., & Eisenstadt, M. (2006). *Wikis of locality: insights from the open guides*. Paper presented at the WikiSym'06, Odense, Denmark.

Geels, F. W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research policy, 33*(6-7), 897-920.

Geiger, R. S. (2009). *The Social Roles of Bots and Assisted Editing Tools*. Paper presented at the WikiSym'09, Orlando, FL.

Geiger, R. S. (2011). The Lives of Bots. In G. Lovnik & N. Tkacz (Eds.), *Critical Point of View: A Wikipedia Reader*. Amsterdam, Netherlands: Institute of Network Cultures.

Geiger, R. S., & Ribes, D. (2010). *The work of sustaining order in wikipedia: the banning of a vandal*. Paper presented at the CSCW'10, Savannah, GA, USA.

Gieryn, T. F. (1983). Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review, 48*(6), 781-795.

Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009). *A conceptual and operational definition of 'social role' in online community*. Paper presented at the HICSS'42, Honolulu, HI.

Gleick, J. (2011a). Epilogue *The information: A history, a theory, a flood* (pp. 413-426). New York, NY, USA: Pantheon.

Gleick, J. (2011b). *The information: A history, a theory, a flood*: Pantheon.

Golder, S. A., & Donath, J. (2004). Social roles in electronic communities. *Internet Research, 5*, 19-22.

Goodreau, S. M. (2007). Advances in exponential random graph (p*) models applied to a large social network. *Social Networks, 29*(2), 231-248.

Gortner, E., & Pennebaker, J. (2003). The archival anatomy of a disaster: Media coverage and community-wide health effects of the Texas A&M bonfire tragedy. *Journal of Social and Clinical Psychology, 22*(5), 580-603.

Halfaker, A., Kittur, A., & Riedl, J. (2011). *Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work*. Paper presented at the WikiSym'11, Mountain View, California.

Hammer, E. M., & Zalta, E. N. (1997). A Solution to the Problem of Updating Encyclopedias. *Computers and the Humanities, 31*(1), 47-60.

Handcock, M., Hunter, D., Butts, C., Goodreau, S. M., & Morris, M. (2008). statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. . *Journal of Statistical Software, 24*(1).

Handcock, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley & P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis* (pp. 229–240). Washington, D.C.: National Academies Press.

Handcock, M. S., & Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics, 4*(1), 5-25.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software, 24*(1), 1548.

Hara, N., Shachaf, P., & Hew, K. F. (2010). Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology, 61*(10), 2097-2108. doi: 10.1002/asi.21373

Harcup, T., & O'Neill, D. (2001). What is news? Galtung and Ruge revisited. *Journalism Studies, 2*(2), 261-280.

Hartley, J. (2008). The supremacy of ignorance over instruction and of numbers over knowledge. *Journalism Studies, 9*(5), 679-691.

Hecht, B., & Gergle, D. (2009). *Measuring self-focus bias in community-maintained knowledge repositories*.

Hecht, B., & Gergle, D. (2010). *On the localness of user-generated content*. Paper presented at the CSCW'10, Savannah, GA, USA.

Heylighen, F. (2007). Why is Open Access Development so Successful? Stigmergic organization and the economics of information. In B. Lutterbeck, M. Barwolff & R. Gehring (Eds.), *Open Source Jahrbuch*.

Hill, B. (Forthcoming). Almost Wikipedia: Eight Collaborative Encyclopedia Projects and the Mechanisms of Collective Action.

Hill, W. C., & Hollan, J. D. (1994). History-enriched digital objects: Prototypes and policy issues. *The information society, 10*(2), 139-145. doi: 10.1080/01972243.1994.9960164

Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T. (1992). *Edit wear and read wear*. Paper presented at the CHI'92, Monterey, California, United States.

Howison, J., Crowston, K., & Wiggins, A. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal for the Association of Information Systems*.

Howison, J., Østerlund, C., Crowston, K., & Bolici, F. (2012). *Stigmergy and Implicit Coordination in Software Development*. Paper presented at the CSCW'12, Seattle, Washington, USA.

Hoyer, S. (2003). Newspapers without journalists. *Journalism Studies, 4*(4), 451-463.

Hu, M., Lim, E. P., Sun, A., Lauw, H. W., & Vuong, B. Q. (2007). *Measuring article quality in wikipedia: models and evaluation*. Paper presented at the CIKM'07, Lisbon, Portugal.

Hughes, A. (1951). Science in English encyclopædias, 1704–1875.—I. *Annals of Science, 7*(4), 340-370. doi: 10.1080/00033795100202461

Hughes, A. (1952). Science in English encyclopædias, 1704–1875.—II. *Annals of Science, 8*(4), 323-367. doi: 10.1080/00033795200200272

Hughes, A. (1953). Science in English encyclopædias, 1704–1875.—III. *Annals of Science, 9*(3), 233-264. doi: 10.1080/00033795300200173

Hughes, A. (1955). Science in English encyclopædias, 1704–1875.—IV. *Annals of Science, 11*(1), 74-92. doi: 10.1080/00033795500200065

Hundert, E. J. (1986). A Cognitive Ideal and Its Myth: Knowledge as Power in the Lexicon of the Enlightenment. *Social Research, 53*(1), 133-157.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association, 103*(481), 248-258.

Iba, T., Nemoto, K., Peters, B., & Gloor, P. A. (2010). Analyzing the Creative Editing Behavior of Wikipedia Editors: Through Dynamic Social Network Analysis. *Procedia - Social and Behavioral Sciences, 2*(4), 6441-6456. doi: 10.1016/j.sbspro.2010.04.054

Indratmo, & Vassileva, J. (2009). *Social Interaction History: A Framework for Supporting Exploration of Social Information Spaces*. Paper presented at the CSE'09.

Jones, C., Hesterly, W., & Borgatti, S. (1997). A general theory of network governance: Exchange conditions and social mechanisms. *Academy of management review, 22*(4), 911-945.

Jones, C., & Lichtenstein, B. (2008). Temporary inter-organizational projects: How temporal and social embeddedness enhance coordination and manage uncertainty *The Oxford Handbook of Inter-organizational Relations* (pp. 231-255). Oxford, UK: Oxford University Press.

Kafker, F. A. (1981). *Notable Encyclopedias of the 17th and 18th Centuries: Nine Predecessors of the encyclopedie*. Oxford: The Voltaire Foundation.

Kafker, F. A., & Loveland, J. (2009). *The Early Britannica: The growth of an outstanding encyclopedia*. Oxford, England: Voltaire Foundation.

Kane, G., Majchrzak, A., Johnson, J., & Chen, G. (2009). *A lifecycle model of perspective making and perspective taking in fluid online collectives*. Paper presented at the International Conference on Information Systems, Phoenix, AZ, USA.

Kane, G. C., & Alavi, M. (2008). Casting the net: A multimodal network perspective on user-system interactions. *Information Systems Research, 19*(3), 253-272.

Keegan, B. (2011). *Breaking News, Breaking Planes, and Breaking Hearts: Psycholinguistic Dynamics and Sensemaking in Collaborative Accounts of Catastrophe*. Paper presented at the International Communication Association, Boston, MA.

Keegan, B., & Gergle, D. (2010). *Egalitarians at the Gate: One-Sided Gatekeeping Practices in Participatory Social Media*. Paper presented at the Computer Supported Collaborative Work, Savannah, GA.

Keegan, B., Gergle, D., & Contractor, N. (2011). *Hot off the Wiki: Dynamics, Practices, and Structures in Wikipedia's Coverage of the Tōhoku Catastrophes*. Paper presented at the WikiSym'11, Mountain View, CA.

Keegan, B., Gergle, D., & Contractor, N. (2012). *Do Editors or Articles Drive Collaboration?: Multilevel Statistical Network Analysis of Wikipedia Coauthorship*. Paper presented at the CSCW'11, Seattle, WA, USA.

Kenny, N. (1991). *The palace of secrets: Béroalde de Verville and Renaissance conceptions of knowledge*: Oxford University Press, USA.

Kittur, A., Chi, E., Pendleton, B., Suh, B., & Mytkowicz, T. (2007). *Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie*. Paper presented at the CHI'07, San Jose, CA.

Kittur, A., & Kraut, R. (2008). *Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination.* Paper presented at the CSCW 2008, San Diego.

Kittur, A., & Kraut, R. (2008). *Harnessing the wisdom of crowds in wikipedia: quality through coordination*. Paper presented at the CSCW'08, San Diego, CA, USA.

Kittur, A., Lee, B., & Kraut, R. (2009). *Coordination in collective intelligence: the role of team structure and task interdependence*. Paper presented at the CHI'09, Boston, MA, USA.

Klinenberg, E. (2005). Convergence: News production in a digital age. *The Annals of the American Academy of Political and Social Science, 597*(1), 48-64.

Kogan, H. (1958). *The Great EB: The Story of The Encyclopaedia Britannica*: University of Chicago Press.

Kovach, B., & Rosenstiel, T. (2001). *The elements of journalism*: Crown Publishers.

Kreps, G. A., & Bosworth, S. L. (1993). Disaster, organizing, and role enactment: A structural approach. *American journal of sociology*, 428-463.

Kriplean, T., Beschastnikh, I., & McDonald, D. W. (2008). *Articulations of WikiWork: Uncovering Valued Work in Wikipedia through Barnstars*. Paper presented at the CSCW'08, San Diego, CA, USA.

Lamont, M., & Molnar, V. (2002). The study of boundaries in the social sciences. *Annual review of sociology*, 167-195.

Lampe, C., Wash, R., Velasquez, A., & Ozkaya, E. (2010). *Motivations to participate in online communities*. Paper presented at the CHI, Atlanta, GA.

Lampland, M., & Star, S. L. (2009). *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life*: Cornell Univ Pr.

Lanier, J. (2011). *You are not a gadget: A manifesto*: Vintage.

Latapy, M., Magnien, C., & Vecchio, N. D. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks, 30*(1), 31-48. doi: 10.1016/j.socnet.2007.04.006

Lindkvist, L. (2005). Knowledge Communities and Knowledge Collectivities: A Typology of Knowledge Work in Groups*. *Journal of Management Studies, 42*(6), 1189-1210.

Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2010). *Exploring iterative and parallel human computation processes*.

Loubser, M. A. X., & Den Besten, M. (2008). *Wikipedia Admins and Templates: The Organizational Capabilities of a Peer Production Effort*. Paper presented at the 2008 European Academy of Management Conference.

Loveland, J. (2006). Unifying Knowledge and Dividing Disciplines: The Development of Treatises in the Encyclopaedia Britannica. *Book History, 9*(1), 57-87.

Loveland, J. (2012). Why Encyclopedias Got Bigger... and Smaller. *Information & Culture: A Journal of History, 47*(2), 233-254.

Lundin, R. A., & Söderholm, A. (1995). A theory of the temporary organization. *Scandinavian Journal of management, 11*(4), 437-455.

Majchrzak, A., Jarvenpaa, S., & Hollingshead, A. (2007). Coordinating expertise among emergent groups responding to disasters. *Organization Science, 18*(1), 147.

Malone, T., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys (CSUR), 26*(1), 119.

McLure-Wasko, M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *Management Information Systems Quarterly, 29*(1), 4.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-444.

Melanson, M. (2010). Why Wikipedia Should Be Trusted As A Breaking News Source *ReadWriteWeb* (Vol. 2012).

Messner, M., & South, J. (2011). Legitimizing Wikipedia: How US national newspapers frame and use the online encyclopedia in their coverage. *Journalism Practice, 5*(2), 145 - 160.

Meyerson, D., Weick, K., & Kramer, R. (1996). Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research, 166*, 195.

Mitchelstein, E., & Boczkowski, P. J. (2009). Between Tradition and Change: A Review of Recent Research on Online News Production. *Journalism: Theory, Practice & Criticism*.

Mituzas, D. (2011). Wikipedia article traffic statistics, from http://stats.grok.se/

Monge, P., & Contractor, N. (2003). *Theories of communication networks*: Oxford University Press, USA.

Monge, P. R., & Contractor, N. S. (2003). *Theories of Communication Networks*. New York: Oxford University Press.

Montgomery, D. (2011). Minutes after Virginia earthquake, it was on Wikipedia, *The Washington Post*. Retrieved from http://www.washingtonpost.com/lifestyle/style/minutes-after-virginia-earthquake-it-was-on-wikipedia/2011/08/24/gIQAQqQMcJ_story.html

Murray, F., & O'Mahony, S. (2007). Exploring the foundations of cumulative innovation: Implications for organization science. *Organization Science, 18*(6), 1006-1021.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of management review*, 242-266.

Nerone, J., & Barnhurst, K. G. (2003). US newspaper types, the newsroom, and the division of labor, 1750–2000. *Journalism Studies, 4*(4), 435-449.

Newman, M. (2002). Assortative mixing in networks. *Physical Review Letters, 89*(20).

Orlikowski, W., & Hoffman, D. (1997). An Improvisational Model of Change Management: The Case of Groupware Technologies. *Sloan Management Review, 38*(2), 11-21.

Pachucki, M. A., Pendergrass, S., & Lamont, M. (2007). Boundary processes: Recent theoretical developments and new contributions. *Poetics, 35*(6), 331-351.

Palen, L., & Vieweg, S. (2008). *The emergence of online widescale interaction in unexpected events: assistance, alliance & retreat*. Paper presented at the CSCW'08, San Diego, CA, USA.

Panciera, K., Halfaker, A., & Terveen, L. (2009). *Wikipedians are born, not made: a study of power editors on Wikipedia*. Paper presented at the GROUP'09, Sanibel Island, FL.

Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology, 52*(2), 169-193.

Pennebaker, J., & Harber, K. (1993). A social stage model of collective coping: The Loma Prieta earthquake and the Persian Gulf War. *Journal of Social Issues, 49*(4), 125-145.

Pentland, B. T., Hærem, T., & Hillison, D. (2011). The (N) Ever-Changing World: Stability and Change in Organizational Routines. *Organization Science, 22*(6).

Pentzold, C. (2009). Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies, 2*(2), 255.

Petit, M. J. (2003). Loose-Leaf Publications. In M. Drake (Ed.), *Encyclopedia of Library and Information Science* (Vol. 4, pp. 1700-1711). New York: Marccel Dekker.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review, 106*(4), 643.

Popper, K. (1994). *Knowledge and the Body-Mind Problem: In defence of interaction*: Routledge.

Priedhorsky, R., Chen, J., Lam, S. K., Panciera, K., Terveen, L., & Riedl, J. (2007). *Creating, destroying, and restoring value in wikipedia*. Paper presented at the Proceedings of the 2007 international ACM conference on Supporting group work, Sanibel Island, Florida, USA.

Protess, D. L. (1992). *The journalism of outrage: Investigative reporting and agenda building in America*: The Guilford Press.

Quarantelli, E., & Dynes, R. (1977). Response to social crisis and disaster. *Annual review of sociology, 3*, 23-49.

Quinn, A. J., & Bederson, B. B. (2011). *Human computation: a survey and taxonomy of a growing field*.

Rafaeli, S., & Ariel, Y. (2008). Online motivational factors: Incentives for participation and contribution in Wikipedia. In A. Barak (Ed.), *Psychological Aspects of Cyberspace: Theory, Research, Applications*, (pp. 243–267). Cambridge, UK: Cambridge University Press.

Rainie, L., Internet, P., & Project, A. L. (2002). *One year later: September 11 and the Internet*: Pew Internet & American Life Project.

Rantanen, T. (2007). The cosmopolitanization of news. *Journalism Studies, 8*(6), 843-861.

Ratcheva, V., & Simpson, M. (2011). Temporary Organisations: Implications for Knowledge Work. *SSRN*.

Reagle Jr, J. (2010). Good Faith Collaboration: The Culture of Wikipedia: MIT Press.

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies, 28*(3), 377.

Rieh, S. Y., & Hilligoss, B. (2007). College students' credibility judgments in the information-seeking process. *The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning*, 49-71.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 173-191. doi: DOI: 10.1016/j.socnet.2006.08.002

Robins, G., Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika, 64*(3), 371-394.

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 192-215.

Rzeszotarski, J., & Kittur, A. (2012). *Learning from history: predicting reverted work at the word level in wikipedia*. Paper presented at the CSCW'12, Seattle, Washington, USA.

Salcetti, M. (1995). The emergence of the reporter: Mechanization and the devaluation of editorial workers. In H. Hardt & B. Brennen (Eds.), *Newsworkers: Toward a history of the rank and file* (pp. 48-74). Minneapolis, MN, USA: University of Minneapolis Press.

Saunders, C. S., & Ahuja, M. K. (2006). Are all distributed teams the same? Differentiating between temporary and ongoing distributed teams. *Small Group Research, 37*(6), 662.

Schlesinger, P. (1987). *Putting" reality" together: BBC news*: Routledge.

Schmidt, K., & Simonee, C. (1996). Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW), 5*(2), 155-200. doi: 10.1007/bf00133655

Schwab, A., & Miner, A. S. (2008). Learning in hybrid-project systems: The effects of project performance on repeated collaboration. *The Academy of Management Journal ARCHIVE, 51*(6), 1117-1149.

Scott, W. R., & Meyer, J. W. (1994). *Institutional environments and organizations: Structural complexity and individualism*: Sage Publications, Inc.

Seward, Z. (2009). Here's the AP document we've been writing about. Retrieved from http://www.niemanlab.org/2009/08/heres-the-ap-document-weve-been-writing-about/

Shirky, C. (2008). *Here comes everybody: the power of organizing without organizations*: Penguin Press HC, The.

Shklovski, I., Palen, L., & Sutton, J. (2008). *Finding community through information and communication technology in disaster response*. Paper presented at the CSCW'08, San Diego, CA, USA.

Siegel, R. (2007). How Wikipedia Breaks News, and Adjusts to It, *NPR*. Retrieved from http://www.npr.org/templates/story/story.php?storyId=7320255

Smith, M. A., & Fiore, A. T. (2001). *Visualization components for persistent conversations*. Paper presented at the CHI'01, Seattle, Washington, United States.

Snijders, T. A. B. (2002). Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure, 3*(2).

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specificiations for exponential random graph models. *Sociological Methodology, 36*(1), 99-153.

Snijders, T. A. B., Van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*(1), 44-60.

Solnit, R. (2009). *A Paradise Built in Hell: The Extraordinary Communities That Arise in Disaster*. New York: Viking Penguin.

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420. doi: 10.1177/030631289019003001

Stecchini, L. C. (1962). On Encyclopedias in Time and Space. *American Behavioral Scientist, 6*(1), 3-6. doi: 10.1177/000276426200600101

Steensen, S. (2010). Online Journalism and the Promises of New Technology. *Journalism Studies, 12*(3), 311-327. doi: 10.1080/1461670x.2010.501151

Stone, L., & Pennebaker, J. (2002). Trauma in real time: Talking and avoiding online conversations about the death of Princess Diana. *Basic and applied social psychology, 24*(3), 173-183.

Stover, C. F. (1962). Change and Rationality in Encyclopedism. *American Behavioral Scientist, 6*(1), 35-38. doi: 10.1177/000276426200600108

Suh, B., Chi, E., Pendleton, B., & Kittur, A. (2007). *Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations*. Paper presented at the VAST'07, Sacramento, CA, USA.

Suh, B., Convertino, G., Chi, E., & Pirolli, P. (2009). *The singularity is not near: slowing growth of Wikipedia*. Paper presented at the WikiSym'09, Orlando, Florida.

Sullivan, L. E. (1990). Circumscribing Knowledge: Encyclopedias in Historical Perspective. *The Journal of Religion, 70*(3), 315-339.

Susi, T., & Ziemke, T. (2001). Social cognition, artefacts, and stigmergy: A comparative analysis of theoretical frameworks for the understanding of artefact-mediated collaborative activity. *Cognitive Systems Research, 2*(4), 273-290. doi: 10.1016/s1389-0417(01)00053-5

Timmermans, S., & Epstein, S. (2010). A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization*. *Annual review of Sociology, 36*, 69-89.

Tuchman, G. (1972). Objectivity as strategic ritual: An examination of newsmen's notions of objectivity. *American Journal of sociology*, 660-679.

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *The American Journal of Sociology, 79*(1), 110-131.

Van de Ven, A., Delbecq, A., & Koenig Jr, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, 322-338.

Van Dijck, J., & Nieborg, D. (2009). Wikinomics and its discontents: a critical analysis of Web 2.0 business manifestos. *New Media & Society, 11*(5), 855-874. doi: 10.1177/1461444809105356

van Tuyll, D. R. (2010). The Past Is Prologue: How Nineteenth-Century Journalism Might Just Save Twenty-First-Century Newspapers. *Journalism Studies, 11*(4), 477-486.

Von Hippel, E., & Von Krogh, G. (2003). Open source software and the" private-collective" innovation model: Issues for organization science. *Organization Science, 14*(2), 209-223.

Von Krogh, G., & von Hippel, E. (2006). The promise of research on open source software. *Management Science, 52*(7), 975-983.

Waisbord, S. (2011). *Can NGOs Change the News?*

Wang, P., Sharpe, K., Robins, G. L., & Pattison, P. E. (2009). Exponential random graph (p*) models for affiliation networks. *Social Networks, 31*(1), 12-25.

Wasserman, S., & Faust, K. (1994). Centrality & Prestige *Social Network Analysis: Methods and Applications* (pp. 169-219). New York: Cambridge University Press.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and applications*. New York: Cambridge University Press.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika, 61*(3), 401-425.

Wasserman, S., & Robins, G. (2005). An introduction to random graphs, dependence graphs, and p*. In P. J. Carrington, J. Scott & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 148–161). New York: Cambridge University Press.

Wattenberg, M., Viégas, F., & Hollenbach, K. (2007). Visualizing activity on wikipedia with chromograms. *Human-Computer Interaction–INTERACT 2007*, 272-287.

Wegner, D. (1987). Transactive memory: A contemporary analysis of the group mind. *Theories of group behavior*, 185-208.

Weick, K. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly, 38*(4).

Weick, K., & Roberts, K. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 357-381.

Weick, K., & Sutcliffe, K. (2007). *Managing the unexpected: resilient performance in an age of uncertainty*: Jossey-Bass Inc Pub.

Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). *Finding social roles in Wikipedia*. Paper presented at the iConference'11, Seattle, WA.

Wikipedia. (2011, April 2010). Wikipedia Statistics: English - Zeitgeist, 2010, from http://stats.wikimedia.org/EN/TablesWikipediaEN.htm#zeitgeist

Wilkinson, D. M. (2008). *Strong regularities in online peer production*.

Wood, C. (2011). Wikirage: What's hot now on Wikipedia, from http://www.wikirage.com/

Wootton, C. W., & Wolk, C. M. (2000). The Evolution and Acceptance of the Loose-Leaf Accounting System, 1885-1935. *Technology and Culture, 41*(1), 80-98.

Yates, D., Wagner, C., & Majchrzak, A. (2010). Factors affecting shapers of organizational wikis. *Journal of the American Society for Information Science and Technology, 61*(3), 543-554.

Yeo, R. (1991). Reading Encyclopedias: Science and the Organization of Knowledge in British Dictionaries of Arts and Sciences, 1730-1850. *Isis, 82*(1), 24-49.

Yeo, R. (2000). Encyclopaedic knowledge. In M. Frasca-Spada & N. Jardine (Eds.), *Books And The Sciences In History* (pp. 207-218). Cambridge, UK: Cambridge University Press.

Yeo, R. (2001). Introduction: The encyclopaedic tradition *Encyclopaedic visions: Scientific dictionaries and enlightenment culture* (pp. 1-34). Cambridge, England: Cambridge Univ Pr.

Zimmer, M. (2009). Renvois of the past, present and future: hyperlinks and the structuring of knowledge from the Encyclopédie to Web 2.0. *New Media & Society, 11*(1-2), 95-113. doi: 10.1177/1461444808099573

Zurcher, L. A., Jr. (1970). The "Friendly" Poker Game: A Study of an Ephemeral Role. *Social Forces, 49*(2), 173-186.

Appendix A – Data acquisition and analysis code