Hot off the Wiki: Dynamics, Practices, and Structures in Wikipedia's Coverage of the Tōhoku Catastrophes

Brian Keegan, Darren Gergle, Noshir Contractor
School of Communication, Northwestern University
2240 Campus Drive, Evanston, IL, USA 60208
{bkeegan,dgergle,nosh}@northwestern.edu

ABSTRACT
Wikipedia editors are uniquely motivated to collaborate around current and breaking news events. However, the speed, urgency, and intensity with which these collaborations unfold also impose a substantial burden on editors' abilities to effectively coordinate tasks and process information. We analyze the patterns of activity on Wikipedia following the 2011 Tōhoku earthquake and tsunami to understand the dynamics of editor attention and participation, novel practices employed to collaborate on these articles, and the resulting coauthorship structures which emerge between editors and articles. Our findings have implications for supporting future coverage of breaking news articles, theorizing about motivations to participate in online community, and illuminating Wikipedia's potential role in storing cultural memories of catastrophe.

Categories and Subject Descriptors
K.4.3 [**Organizational Impacts**]: *computer-supported collaborative work*; H.5.3 [**Group and Organization Interfaces**]: *web-based interaction, collaborative computing*.

General Terms
Management, Measurement, Reliability, Human Factors

Keywords
Wikipedia, breaking news, current events, network analysis, bipartite network, emergent group, high tempo, collaboration

1. INTRODUCTION
As an encyclopedia that ~~"a~~nyone can edit", Wikipedia has attracted substantial scholarly interest in understanding the socio-technical processes that sustain motivated and coherent peer production of information. However, academic studies often presume that Wikipedia's collaborations are relatively stable and the motivations to contribute and ability to access reliable information do not change substantially over time. For the vast majority of articles on Wikipedia, this is a relatively sound assumption.

However, articles about current and breaking news events such as natural disasters, technological accidents, and political unrest do not readily fit into this mold. Coverage of breaking news events on Wikipedia saturates the lists of articles receiving the most users [40], revisions [42], and page views [24] in any given month going all the way back to its 2001 establishment. For example, the articles which attracted the most contributors in February 2011 included "~~2~~011 Egyptian revolution", "~~2~~011 Libyan uprising", "~~2~~010-2011 Middle East and North Africa protests", "~~H~~osni Mubarak", "~~B~~orn This Way (song)", "~~S~~uper Bowl XLV", and "~~2~~011 Christchurch earthquake" [40].

Given Wikipedia's prominence as a popular reference website, it is an obvious target for individuals to engage in information seeking following unexpected and highly salient events. Indeed, Wikipedia is increasingly framed by traditional journalists as a neutral and reliable information source [23]. However, in light of findings suggesting the growth of new editors, and articles has slowed [36] and indirect work increasingly occupies Wikipedians' attention [16], breaking news articles are a unique site to study how late-modern Wikipedians create and coordinate entirely novel content. In the wake of catastrophes, unique forms of social behavior and organization emerge to support information dissemination and disaster response. Projects like Wikipedia which empower users to alter and revise the content in light of new information are potentially well-suited to both supporting the emergent and temporary post-catastrophe organizations.

The complex and unpredictable ways in which breaking news unfolds which not only require complex forms of coordination but also sustaining attention over long periods of time and broad topical areas. At the outset, the temporary organizations which emerge to collaborate on breaking news topics are very unlikely to have previously interacted, possess the same information, have access to similar resources, or be willing to fulfill necessary roles. Despite this, Wikipedia articles are also perceived to be exemplars of timeliness, breadth, and reliability in the wake of disasters like the 2007 Virginia Tech massacre [6].

We examine the ways in which Wikipedia responded to the recent Tōhoku earthquake and tsunami to illuminate the processes which support its coverage of breaking news events. The Tōhoku earthquake and tsunami are certainly not representative of all breaking news events which themselves are not representative of most of the work on Wikipedia. However, this case serves as a boundary condition for evaluating how the Wikipedia community responds to unexpected events and creates content under urgent circumstances. These findings illuminate the practices and structures which can support other intensive forms of peer-production and also point to Wikipedia's role as a source of cultural memory and historiographical tool.

2. POST-CATASTROPHE ORGANIZING
2.1 Timeline of the Tōhoku catastrophes
On March 11, 2011 at 5:46 UTC, a 9.0-magnitude undersea megathrust earthquake occurred approximately 72 kilometers east of the northeastern Tōhoku region of Japan. The earthquake vertically displaced 5-8 meters of seabed which in turn created a

WikiSym'11, October 3-5, 2011, Mountain View, CA, USA.

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 [39]. 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 Great Tōhoku Earthquake and Tsunami the most expensive natural disaster ever recorded [40].

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.

2.2 Disaster sociology and emergent groups

Disaster sociologists have described the altruistic communities and organized behavior that emerge in the aftermath of disasters when the extent of the damage is uncertain and information is scarce [33]. 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 in parallel with the most acute phases responding to a disaster [35].

Online communities far removed from the risks and consequences of a disaster can also be sites for communication, sensemaking, and organizing [7, 28, 32, 34]. Citizens and responders appropriate and reconstruct the roles of technologies like mobile phones, text messaging, GPS, and microblogging to organize ad hoc responses, process and disseminate information, and provide social support [27]. While it is clear that information and communication technologies can support the emergence of collective intelligence in response to crises, these studies tend to emphasize the use of information communication technologies (ICTs) to generate and disseminate ephemeral information on a peer-to-peer or very local level [34]. These studies neither examine how less proximate organizations adapt their practices to seek and process information, nor develop lasting artifacts such as authoritative accounts of the event like those in Wikipedia. This raises several questions. Does Wikipedia's response to catastrophic breaking news events emulate the concentrated activity seen in offline communities despite its lack of propinquity? Are these responses also just as transient or does participation stabilize? How are practices and technologies on Wikipedia repurposed to support the demands of disaster response?

The notion that member-maintained communities require oversight and moderation to support high-quality collective effort implies that membership in and motivations to contribute to these communities are relatively stable [9]. However, communities and groups typically studied by organizational scholars rarely face the unpredictability, urgency, and reconfigurability demanded by groups in high-tempo situations such as disaster response [20], emergency medicine [13], aircraft carriers [8], or "spot" journalism [2]. High-tempo contexts are characterized by non-routine and extremely urgent work, abrupt consequences, and intense attention. Coordination in these volatile environments demands high levels of heedful and interrelated action, knowledge integration, and information processing [4]. Actors responding to high-tempo events are placed in a situation in which cognition is distributed, responsibility is shared, and profound differences in perspectives and norms exist which can lead to contentious interactions given the stakes of individual and joint reputation and interests [13].

Some organizations like emergency rooms and aircraft carrier crews respond to these demands by defining clear group membership, tasks, role or expertise [13, 39]. However, other groups and communities face contexts which wholly lack pre-existing organizational structures, the ability to routinize practices, or develop relevant expertise. These "emergent response groups" are unique because group members have diverse motivations, mixed perspectives, varied resources to contribute, and substantial volition to come and go as they please. Members are unlikely to have interacted previously and may not interact after the event. These factors contribute to unstable task definitions and the pursuit of multiple and potentially conflicting goals. The preconditions to facilitate efficient coordination such as known membership, expertise specialization and credibility, shared goals, and interaction routines are limited [20].

We argue that Wikipedia's response to breaking news events exhibits many of the features of theses emergent response groups. 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. The group of editors contributing to a breaking news article must assemble itself from Wikipedia editors' 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. However, articles about breaking news events may also lower the barriers separating reading and contributing to Wikipedia articles as forms of legitimate peripheral participation. New participants see the opportunities to not only fulfill niche roles but also contribute to a salient source of information [1, 5].

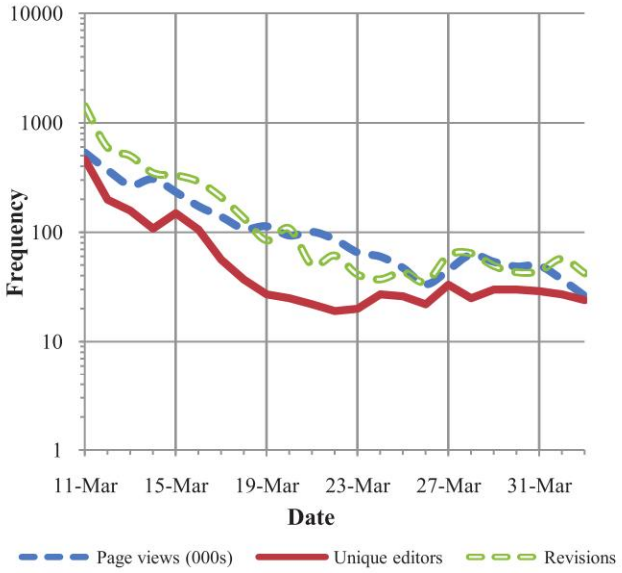

Figure 1: Page views, number of unique editors, and number of revisions per day for the "2011 Tōhoku earthquake and tsunami" English article's revision history. Logged y-axis.

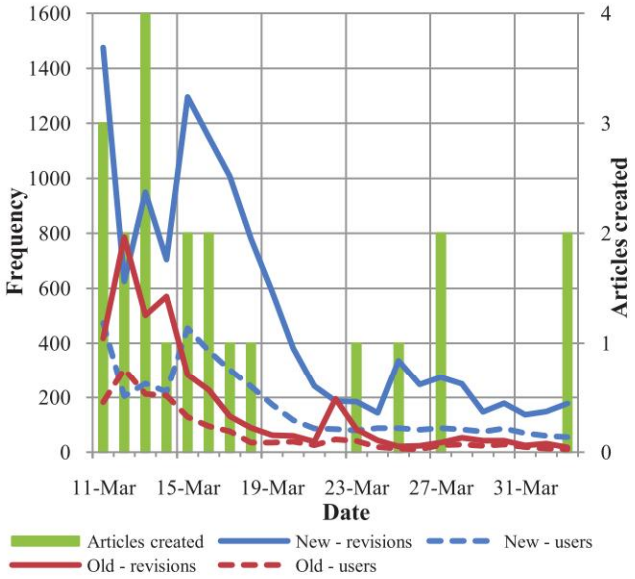

Figure 2: Revisions (solid lines) and unique editors (dashed lines) per day for the 62 "old" articles predating the earthquake and the 22 "new" articles created afterwards.

DATA AND METHODS

The data were extracted using a tool to download an article's revision history from Wikipedia's "Special:Export" interface. This Export interface outputted XML files containing 1,000 revisions including the article's name, article ID, revision comment, user ID, username, and timestamp. For articles with more than 1,000 revisions, these XML files were stitched together by article. The seed list of 84 articles were drawn from the "Tōhoku earthquake and tsunami" category on the English Wikipedia on April 3, 2011.

We assume that there is only one user per IP address. Articles were assigned binary "old/new" categories based upon their timestamps. Articles which existed prior to 05:30 UTC on March 11, 2001 were classified as "old" and articles created afterwards were classified as "new". Out of the 84 articles which were categorized by Wikipedians as affected by or related to the earthquake, tsunami, or nuclear incidents (including the articles about the events themselves), 62 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 22 articles which 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.

3. HIGH-TEMPO DYNAMICS

3.1 Shifting attention

Figure 1 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 84 articles affected by the earthquake, Figure 2 plots the changes in revisions and unique editors by day for "old" articles which 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

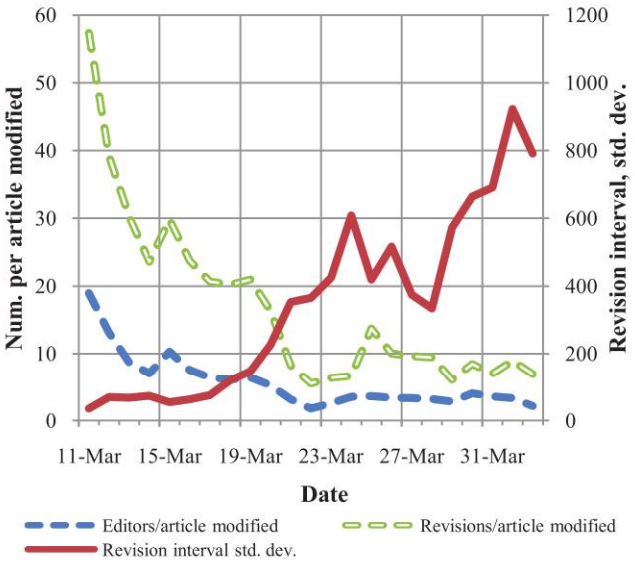

Figure 3: The variance in time between edits to any Tōhoku article increases dramatically in parallel with the dilution of editor attention and activity among actively-edited articles.

Figure 4: 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.

March 20. The shift in editor attention and activity from old articles to new articles remained stable after the first week. [1]

3.2 Article creation and spin-out

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 2 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 which can be modified in a given day increases. However, as Figure 2 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.

In Figure 3, the standard deviation of the interval of time that lapsed between edits to any of the 84 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" not necessarily a bad thing and may in fact reflect a shift from reactive tasks to systemic or maintenance tasks as described by [37]. 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 [17].

4. NOVEL PRACTICES

The intensity of collaboration in the most acute phase of the article imposes significant constraints on the ability for dedicated editors to substantially edit the article without other users interrupting with their own revisions. Likewise, the spin-out of articles described above increases the demands on editors to monitor and check multiple articles. The articles on the Tōhoku

[1] 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.

earthquake offer at least two examples of how Wikipedia editors adopt unique practices to deal with high tempo collaborations.

4.1 In-line casualty templates

Templates are Wikipedia pages which 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 which need to made to an article or to acknowledge contributions and achievements [14, 18, 19]. However, templates were adopted and implemented in unique ways by the Tōhoku earthquake editors to support their high-tempo collaboration across articles.

As was outlined in the previously, 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. Moreover, 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 4, 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 which 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 revisions on each page.

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. Despite its simplicity, 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 [18], this approach 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.

4.2 IRC backchannels

The rapid pace of changes on articles related to the catastrophe challenged the ability for expert editors to monitor and coordinate changes to the articles in real time. These expert users employ a variety of technologies to monitor recent and on-going changes and communicate via various back-channels such as Internet Relay Chat (IRC) channels and User talk pages coordinate their actions [31]. In particular, the integration of synchronous text-based chat tools like IRC into the repertoires of Wikipedia editing practices has not been previously examined or analyzed for even typical article collaboration. The role that IRC channels played in responding to breaking and current news events is likewise veiled because of a lack of activity logs and the inherent ephemerality of the medium. However, IRC certainly plays an even more important role in communicating and coordinating tasks for breaking news articles than they already do for standard articles.

Wikipedia and the Wikimedia Foundation operate a large and active array of IRC channels on the ―freenode" server network to support real-time discussions. These channels include forums for help, administrative disputes, administrator and mediator communication, and other language Wikipedias [41]. IRC lends itself to having many channels open and active in parallel and supporting persistent backchannel communication with other collaborators [22]. Furthermore, Wikipedia users have created ―bots" which report real-time changes made to a defined watchlist of Wikipedia articles back to the channel. Thus, an IRC channel can potentially serve as centralized ―command and control center" facilitating joint monitoring of a situation at a central location and synchronous communication with users spread across multiple persistent channels.

Among the many IRC channels which played a role in coordinating response to the Tōhoku catastrophes, the quasi-public channel owned by a user named ―Chzz" is emblematic of the affordances of these backchannels. Although he only ranks 154th among editors for the number of articles edited in the present corpus, Chzz's use of his own and other IRC channels provided crucial infrastructure to support a very complex collaboration.

―Chzz" is a British user on the English Wikipedia who previously taught English classes in Japan and has basic proficiency in the language. Unable to get in contact with his friends because of severed communication ties in Sendai, he logged onto the ―#wikipedia-ja" Japanese language general IRC channel and began reviewing the Japanese articles about the earthquake, power plants, and affected cities to find the latest information being introduced there. Immersed in this information and possessing a suddenly relevant skill set, he began using his familiarity with both the events and communication occurring on the Japanese language Wikipedia and IRC to monitor and introduce updated references in both languages on the English article.

Chzz was also a ―power user" of Wikipedia with over 80,000 contributions since his account registration in 2008. He began editing breaking news articles with the 2008 Sichuan earthquake as well as monitoring recent changes to the general Wikipedia corpus. During that time, his IRC channel became frequented by other power users engaged in similar tasks and a stable cohort of users joins and monitors his channel. This community of IRC-Wikipedians was highly proficient using the MediaWiki platform as well as navigating the Wikipedia community and include sysops/administrators as well as individuals affiliated with the Wikimedia Foundation itself. For example, Chzz's channel was one of the forums casualty template creator ―Dcoetzee" joined to

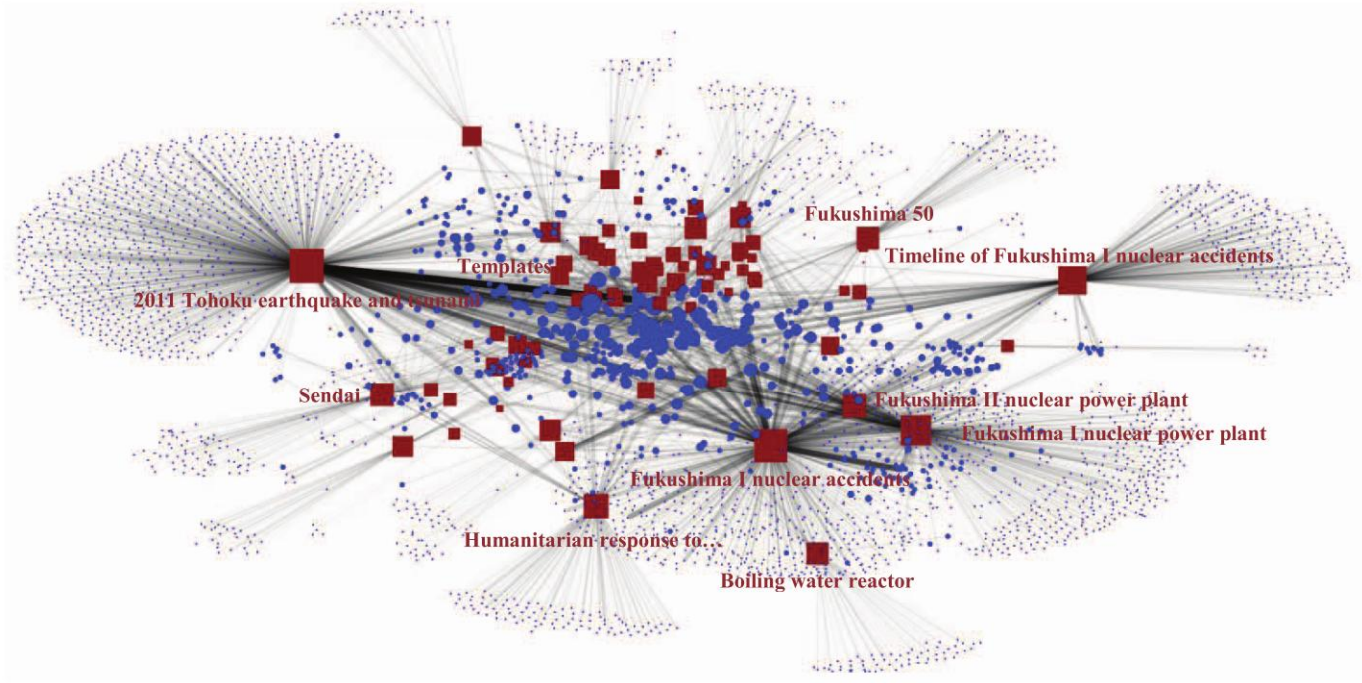

Figure 5: Two-mode network visualization of the 84 articles (red squares) and 3,217 editors (blue circles). Black links indicate an editor contributed to that article after March 11, 2011.

monitor revisions and protect Tohoku-related articles from vandalism and edit warring.

Unfortunately because the IRC channels are typically not logged, it is impossible to reconstruct which users were present and participating in which channels or the types of coordination in which users were engaged during the most acute phases of the collaboration. Future research should take the communication on these IRC channels as a point of departure for data collection to better understand the role they play in the dynamics, practices, and structures of deliberating and coordinating responses to breaking news articles.

5. EMERGENT STRUCTURES

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 [3]. 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.

We integrated all 15,379 revisions made after 05:30 UTC on March 11, 2001 to the 84 Tōhoku articles into a weighted bipartite coauthorship network. In this frame, 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 [3], these conversions result in the loss of information about editor-article edge weights and are not appropriate for our analysis.

5.1 Bipartite coauthorship network

The weighted bipartite network for the 84 articles, 3,217 editors, and 4,827 unique links between editors and articles is visualized in Figure 5. 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." As we discuss in section 6.2 below, this heterogeneous distribution of connections is characteristic of many complex networks.

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 both made many contributions to many articles. This strong central core of

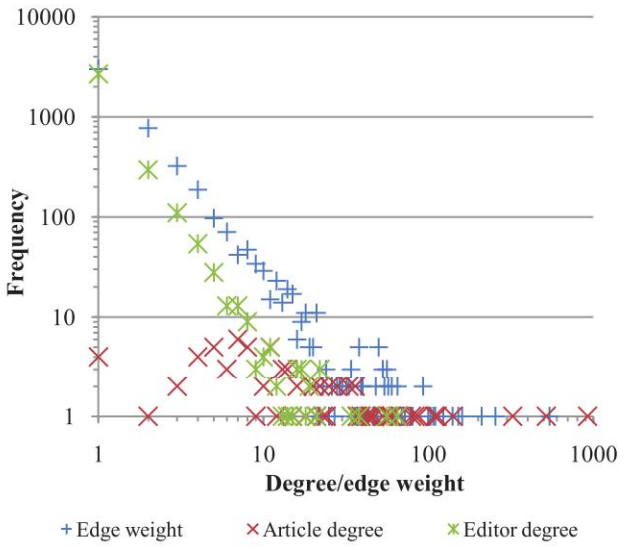

Figure 6: Degree and edge weight distributions for the bipartite network in Figure 5 on log-log scales. Power law distributions are fitted to the edge weight and editor degree distributions.

generalist editors consisted 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 edited across all editors in the corpus.

This approach is also limited by the fact that it does not capture the type or extent of the changes in content each editor's revision made. Thus, the central and prolific editors may simply be making relatively minor edits, rather than major substantive changes in content. Likewise, the analysis only captures the coauthorship structure related to the Tōhoku corpus instead of other instances of breaking news articles or Wikipedia more generally. Future research should extend analyses such as [29, 31] to examine which editors are authoring content which is ultimately viewed by readers and differences in the life-spans and activities of contributors to other breaking news or Wikipedia content.

5.2 Degree and weight distributions

Complex networks often have a highly heterogeneous distribution of links among the nodes in the network [25]. In the context of the weighted bipartite graph of Figure 5, there are a minimum of three different distributions of activity which can be measured: the distribution articles' connections to distinct editors (article degree), the distribution of editors' connections to distinct articles (editor degree), and the distribution of the number of edits editors made to a single article (edge weight).

Figure 6 plots these distributions on a log-log scale. The distribution of editors' edge weights (in blue) is highly heterogeneous: there are 3,009 editors who made a single contribution to an article (upper-left corner of the distribution) while one editor was recorded making 542 contributions to a single article (center-bottom). Editors' degrees (in green) are similarly heterogeneous; 2,685 editors contributed to only one article while one editor contributed to 65 articles. Although the distribution of articles' connections (in red) has some significant outliers, it is much more homogenous and linear that the previous two distributions.

Long tail degree distributions in a network are often indicative of a latent process of preferential attachment whereby outliers' connectivity or activity is reinforced via a positive feedback loop [25, 26]. Thus, editors who intensively edited a single article will continue to intensively edit that article. Similarly, editors who have edited many articles and potentially fulfill a "generalist" role will continue to expand their repertoire and edit a diverse set of articles. The relatively linear distribution of unique editors across the articles is surprising. It suggests the processes of information seeking resist the positive feedback loops which contribute to the long tail distributions of the other distributions. In other words, the attention paid by editors to popular articles does not result in these "rich" articles becoming "richer."

Although further research is necessary, two processes could be driving this finding. First, the "deputization" of many editors into contributing to specialized thematic subgroups (e.g., nuclear plant accident related articles) flattens the curve such that more articles receive attention from more editors than would be expected under conditions of preferential attachment. Alternatively, the frequency and intensity of editing activity on the most prominent and popular articles effectively shuts out the editors who lack the patience or skill to deal with repeated edit conflicts when they attempt to make a contribution. As a result, these users avoid the popular articles and shift their attention toward other related articles in an attempt to make a contribution. This leaves the expert and committed editors who are already likely to have a hand in editing many articles to edit the prominent articles.

6. DISCUSSION AND IMPLICATIONS

Our study provided a view into the self-organized and distributed collaboration which emerges in the English Wikipedia community the aftermath of a catastrophe. Despite the considerable organizational impedances associated with emergent groups collaborating in a high tempo context and the lack of consequential proximity to the event, the 84 Wikipedia articles affected by the Tohoku earthquake exhibited a substantial concentration of activity from thousands of users over several weeks. We identified several dynamic features of this collaboration including the intense interest and attention focused on the topic in the immediate aftermath and the subsequent decline and dispersion of attention among newly-created articles. Editors on the primary Tohoku catastrophe articles likewise adopted novel practices like using transcluded in-line templates to update casualty numbers and topical communication backchannels like IRC in reaction to the intense editing activity and complex coordination requirements. The structure of the collaboration also suggested that sub-communities of editors specialized in intensively editing thematically similar content.

6.1 Implications

Our findings have substantial implications for theorizing about the motivations to contribute to online communities. 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 described by Palen, Shklovski, et al. [34], but also contexts with less immediate consequences, more poorly defined organizational

factors, and less rich interpersonal social interactions. Scholars should thus examine how short-lived online communities can actually be instances of successful collaboration.

Online communities emphasizing group identity in particular need to have the flexibility to rapidly accommodate and socialize large influxes of participants attempting to make sense of unexpected events and engage in diverse forms of collective action in response to them. Instances of fluid, close coordination which occur under conditions of minimal self disclosure and limited consensus are simultaneously high-risk but also potentially ecstatic experiences for individuals [12]. Breaking news articles on Wikipedia offer a compelling case to examine how online communities balance the competing interests to support openness, flexibility, and autonomy against institutional needs for structure, norms, and socialization over very different time scales.

This analysis of Wikipedia's coverage of a breaking news event likewise highlights interesting parallels between the goals of Wikipedia and traditional journalism. Both share fundamental motivations to collect, synthesize, and disseminate neutral and reliable information to a general audience in a timely manner [10, 21]. Yet because of its low participation costs, autonomous tasks, and algorithmic integration of users' contributions, Wikipedia editors are active mediators in the social co-construction of knowledge and meaning-making in a way that traditional journalism could never be [11, 15]. Nevertheless, the extent to which Wikipedia and journalists rely upon significantly different organizational forms to accomplish similar goals should highlight not only the affordances and constraints of their respective organizational forms but also expand the boundaries of what each is possible of achieving.

6.2 Limitations and future work

Any scholar wishing to disentangle the intricate threads which support a project as large and complex as Wikipedia must balance the competing imperatives to develop narrow but grounded accounts or general but coarse descriptions. The multi-method approach employed herein reflects an attempt to bridge the divide between these empirical divides. Nevertheless, it still reflects the dynamics, practices, and structures associated with a single catastrophe with its own unique salience, historical importance, and socio-cultural implications. The recency of the event also necessarily occludes on-going processes and future developments in which the observed dynamics, practices, or structures are potentially altered. Thus, further work is obviously necessary to establish the observed dynamics, practices, and structures generalize to other large scale catastrophes, other genres of breaking news articles, and Wikipedias in other languages.

The quantitative approaches employed herein are likewise descriptive in nature and fail to provide a statistical model which can be used to simulate or confirm how breaking news articles evolve. Moreover, these single-level analyses omit the possibility that multiple processes occurring at multiple levels of analysis interact to generate the particular dynamics and structures observed. Methods like agent-based simulation, exponential random graph models, and actor-based stochastic models could be leveraged to understand the multi-level, multi-theoretical processes of social exchange, closure, reciprocity, homophily, diffusion, and collective which govern network evolution [8].

This analysis omitted any consideration of the dynamics, practices, or structures associated with changes in content. However, these content, their changes over time, and their record in the revision history point to the potential for collective intelligence platforms such as Wikipedia to be employed as forms of authoritative knowledge and cultural memory [30]. As historiographical artifacts, breaking news articles are potentially the *sine qua none* for capturing the prevailing interpretations of events as they occurred. The nature of Wikipedia's open peer-production model coupled with its archival history of all revisions suggest the possibility of scholars being able to engage in "information archeology" of collective intelligence databases to not only capture forgotten and supplanted interpretations of events but who changed them, what replaced them, and how they changed and stabilized over time.

7. ACKNOWLEDGEMENTS

We would like to thank the members of the Laboratory for Collaborative Technology ("Collablab") for their support and feedback as well as Isaac Wilson, David Hartglass, and Steven Howard for their contributions to the data scraping tool. We would also like to thank Derrick Coetzee and other members of the ##chzz IRC channel for their feedback.

8. REFERENCES

[1] Antin, J. and Cheshire, C. Readers are not free-riders: reading as a form of participation on Wikipedia. In *Proc. CSCW'10* (Savannah, GA, USA, 2010). ACM.

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

[3] Borgatti, S. and Everett, M. Network analysis of 2-mode data. *Social Networks*, 19, 3 1997, 243-269.

[4] Brown, S. and Eisenhardt, K. The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly*, 42, 1 1997, 1-34.

[5] Bryant, S. L., Forte, A. and Bruckman, A. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proc. GROUP'05* (Sanibel Island, FL, USA 2005). ACM,

[6] Cohen, N. The Latest on Virginia Tech, From Wikipedia. *The New York Times.* April 23, 2007.

[7] Cohn, M., Mehl, M. and Pennebaker, J. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15, 10 2004, 687-693.

[8] Contractor, N., Wasserman, S. and Faust, K. Testing Multitheoretical, Multilevel Hypotheses about Organizational Networks. *Academy of Management Review*, 31, 3 2006, 681-703.

[9] Cosley, D., Frankowski, D., Kiesler, S., Terveen, L. and Riedl, J. How oversight improves member-maintained communities. In *Proc. CHI'05* (Portland, Oregon, USA, 2005) ACM.

[10] Deuze, M. What is journalism?: Professional identity and ideology of journalists Reconsidered. *Journalism*, 6, 4 2005, 442.

[11] Deuze, M., Bruns, A. and Neuberger, C. Preparing for an age of participatory news. *Journalism Practice*, 1, 3 2007, 322-338.

[12] Eisenberg, E. Jamming: Transcendence through organizing. *Communication Research*, 17, 2 1990, 139.

[13] Faraj, S. and Xiao, Y. Coordination in fast-response organizations. *Management Science*, 52, 8 2006, 1155.

[14] Gaio, L., Besten, M. d., Rossi, A. and Dalle, J.-M. Wikibugs: using template messages in open content collections. In *Proc. WikiSym'05* (Orlando, Florida, USA, 2009). ACM.

[15] Glaser, M., Globe, B., Great, O., Post-Standard, S. and Foundation, W. Collaborative Conundrum: Do Wikis Have a Place in the Newsroom? *Onlineartikel in Online Journalism Review OJR,* September 2004.

[16] Kittur, A., Chi, E., Pendleton, B. A., Suh, B. and Mytkowicz, T. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the

Rise of the Bourgeoisie. In *Proc. CHI'07 (San Jose, CA, USA 2007).* ACM.
[17] Kittur, A., Lee, B. and Kraut, R. Coordination in collective intelligence: the role of team structure and task interdependence. In *Proc. CHI'09* (Boston, MA, USA 2009). ACM.
[18] Kriplean, T., Beschastnikh, I. and McDonald, D. W. Articulations of WikiWork: Uncovering Valued Work in Wikipedia through Barnstars. In *Proc. CSCW'08* (San Diego, CA, USA 2008). ACM.
[19] Loubser, M. A. X. and Den Besten, M. Wikipedia Admins and Templates: The Organizational Capabilities of a Peer Production Effort. SSRN, 2008.
[20] Majchrzak, A., Jarvenpaa, S. and Hollingshead, A. Coordinating expertise among emergent groups responding to disasters. *Organization Science*, 18, 1 2007, 147.
[21] Matheson, D. Weblogs and the Epistemology of the News: Some Trends in Online Journalism. *New Media & Society*, 6, 4 August 1, 2004, 443-468.
[22] McNely, B. Backchannel persistence and collaborative meaning-making. In *Proc. SIGDOC'09* (Bloomington, IN, USA, 2009). ACM.
[23] Messner, M. and South, J. Legitimizing Wikipedia: How US national newspapers frame and use the online encyclopedia in their coverage. *Journalism Practice*, 5, 2 2011, 145 - 160.
[24] Mituzas, D. *Wikipedia article traffic statistics*., 2011. http://stats.grok.se
[25] Newman, M. The structure and function of complex networks. *SIAM Review*, 45, 2 2003, 167-256.
[26] Newman, M. E. J. Clustering and preferential attachment in growing networks. *Physical Review E*, 64, 2 2001, 025102.
[27] Palen, L. and Liu, S. Citizen communications in crisis: anticipating a future of ICT-supported public participation. In *Proc. CHI'07* (San Jose, CA, USA, 2007). ACM.
[28] Palen, L. and Vieweg, S. The emergence of online widescale interaction in unexpected events: assistance, alliance & retreat. In *Proc. CSCW'08* (San Diego, CA, USA, 2008). ACM.
[29] Panciera, K., Halfaker, A. and Terveen, L. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proc. GROUP'05* (Sanibel Island, FL, USA, 2009). ACM.
[30] Pentzold, C. Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies*, 2, 2 2009, 255.
[31] Priedhorsky, R., Chen, J., Lam, S. K., Panciera, K., Terveen, L. and Riedl, J. Creating, destroying, and restoring value in Wikipedia. In *Proc. GROUP'07* (Sanibel Island, FL, USA, 2007). ACM.
[32] Qu, Y., Huang, C., Zhang, P. and Zhang, J. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In *Proc. CSCW'11* (Hangzhou, China, 2011). ACM.
[33] Quarantelli, E. and Dynes, R. Response to social crisis and disaster. *Annual Review of Sociology*, 31977, 23-49.
[34] Shklovski, I., Palen, L. and Sutton, J. Finding community through information and communication technology in disaster response. In *Proc. CSCW'08* (San Diego, CA, USA 2008). ACM.
[35] Solnit, R. *A Paradise Built in Hell: The Extraordinary Communities That Arise in Disaster*. Viking Penguin, New York, 2009.
[36] Suh, B., Convertino, G., Chi, E. H. and Pirolli, P. The singularity is not near: slowing growth of Wikipedia. In *Proc. WikiSym'09* (Orlando, FL, USA 2009). ACM.
[37] Wattenberg, M., Viégas, F. and Hollenbach, K. Visualizing activity on wikipedia with chromograms. *INTERACT 2007,* 2007, 272-287.
[38] Weick, K. and Roberts, K. Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 38, 1993, 357-381.
[39] Wikipedia *2011 Tohoku earthquake and tsunami*. Wikimedia Foundation, 2011. http://en.wikipedia.org/wiki/2011_Tohoku_earthquake_and_tsunami
[40] Wikipedia *Wikipedia Statistics: English - Zeitgeist*. Wikimedia Foundation, 2011. http://stats.wikimedia.org/EN/TablesWikipediaEN.htm#zeitgeist
[41] Wikipedia *Wikipedia:IRC*. Wikimedia Foundation, 2011. http://en.wikipedia.org/wiki/Wikipedia:IRC
[42] Wood, C. *Wikirage: What's hot now on Wikipedia*. 2011. http://www.wikirage.com/