

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

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.

The increasingly collaborative, temporary, distributed, and mediated nature of knowledge work is obsoleting organizations built around self-interest, stability, co-location, and interpersonal relations. 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. 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. This project's findings have substantial implications for understanding the role of high tempo online knowledge collaboration.

Research Background

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) or emergency medicine (Faraj & Xiao, 2006). 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. 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). 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. 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, Weick, & Kramer, 1996). This leads to governance through networks of relations emphasizing reciprocity, socialization, and reputation rather than lines of authority (Brown & Eisenhardt, 1997; Jones, Hesterly, & Borgatti, 1997).

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 whom 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. Members of these temporary organizations use protocols and repertoires of proven interactions to manage task and role ambiguity, create plug-and-play teams with interchangeable roles, and support communities of practice to manage participation and organizing within specialties. 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.

However, organizations like disaster response teams or Wikipedia editorship lack the role clarity or group stability found in flight deck crews or emergency medical teams that enable

traditional forms of coordination. 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 than 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.

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). 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. 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. 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.

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. 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).

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. 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. To that end, this study 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.

Key research questions

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 events attract 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 milieu afterwards.

Research 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). Using the English Wikipedia's Application Programming Interface (API), the revision histories for every article in these categories and their subcategories were extracted in January 2012. 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. 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. 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.

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. Computing the difference between the date of the first edit and the date of the event itself, non-historical articles that had their first edit happening within 24 hours of the event itself were classified as breaking articles and the remainder were classified as

non-breaking articles. 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.

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, artifact-to-artifact citation networks, or user-to-user discussion or revert networks.

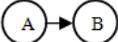
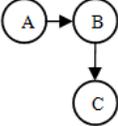
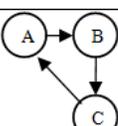
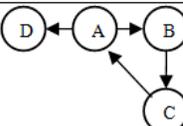
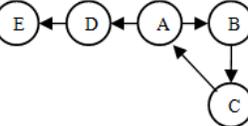
Revision number	Revision editor	Revision history chain graph
1	A	
2	B	
3	C	
4	A	
5	D	
6	E	

Table 1: Construction of a sociotechnical trajectory for an article from a revision history.

The 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. 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 $A \rightarrow B$ 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).

Primary results

Finding 1: Co-authorship structure changes over time

Figure 1 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, 2011; 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.

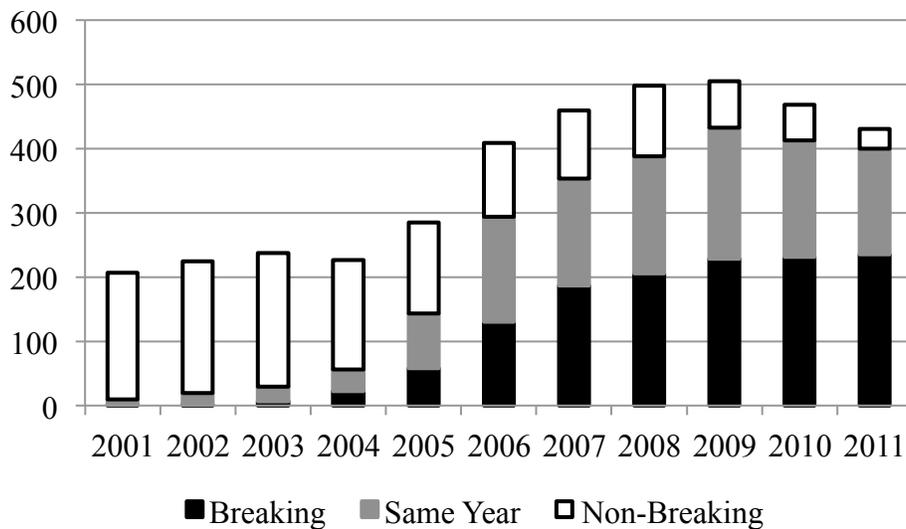


Figure 1: Frequency of article types per event year.

The bipartite (editor-article interaction) degree distributions for breaking articles (Figure 3) and their editors (Figure 2) reveal 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*. 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. Again, breaking articles exhibit remarkable similarity across the years while breaking articles' distributions are less centralized in early years than later years.

Figures 4-A and 4-B plot the degree correlation of breaking articles which reflects the extent to which articles having many contributors also have contributors who have revised many other articles (Newman, 2002). 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. Because the concentration of editors on breaking articles does not differ significantly from non-breaking articles, this suggests collaborations involving breaking news articles do not coordinate this work by centralizing or decentralizing their work to a greater extent than non-breaking articles. However, 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.

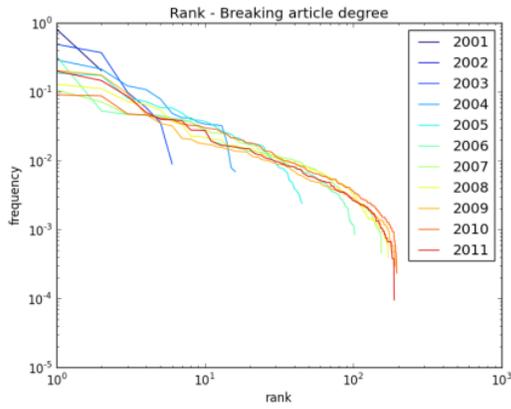


Figure 2-A: Breaking articles' bipartite degree distributions, by year

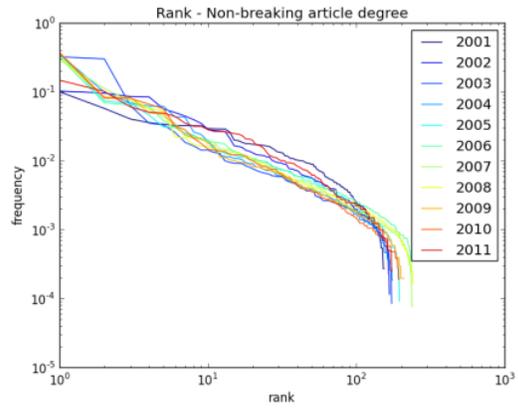


Figure 2-B: Non-breaking articles' bipartite degree distributions, by year

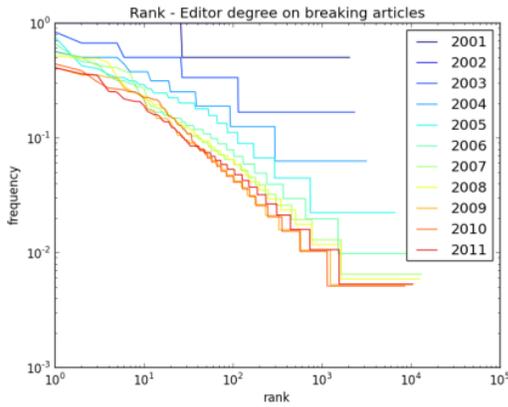


Figure 3-A: Editor bipartite degree distributions, by year for breaking articles.

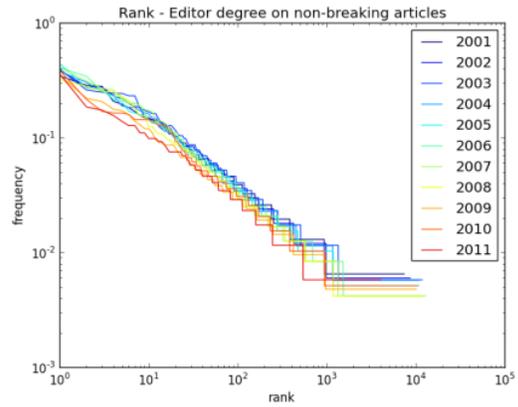


Figure 3-B: Editor bipartite degree distributions, by year for non-breaking articles.

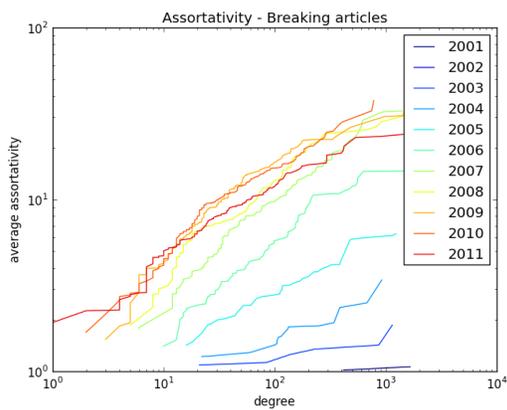


Figure 4-A: Breaking articles' bipartite degree correlation

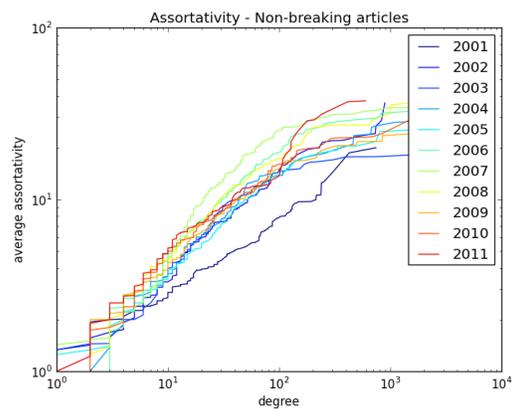


Figure 4-B: Non-breaking articles' bipartite degree correlation

These 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 occluded important differences in the structures of these coauthorship networks of both breaking and non-breaking articles as they changed over time. If breaking news articles have distinct coordination demands, this should manifest in the structural patterns of co-authorship as breaking articles exhibit more and denser activity earlier on in their “lives” than other types of Wikipedia articles. 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 Figures 5-A and 5-B.

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.

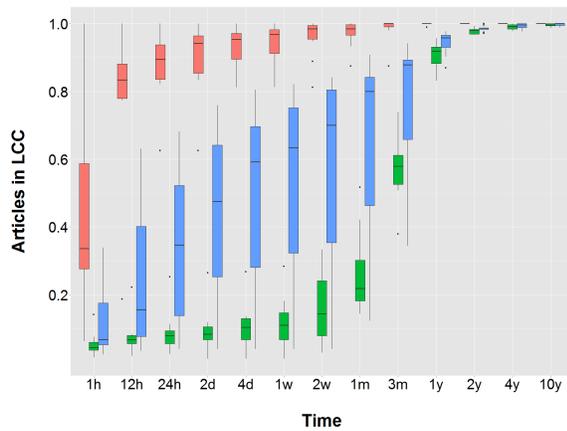


Figure 5-A: Articles in the largest connected component (LCC) for the cumulative network for breaking articles (red), non-breaking articles (blue), and historical articles (green).

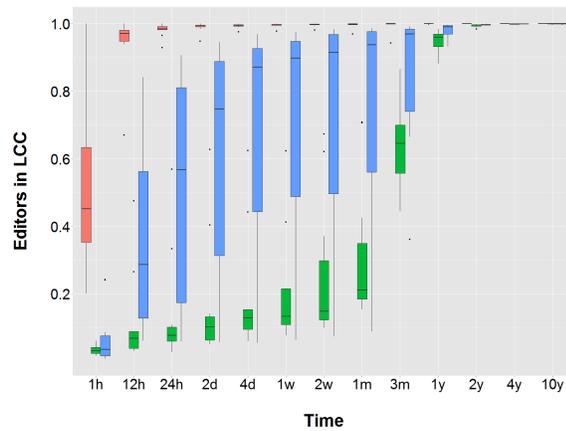
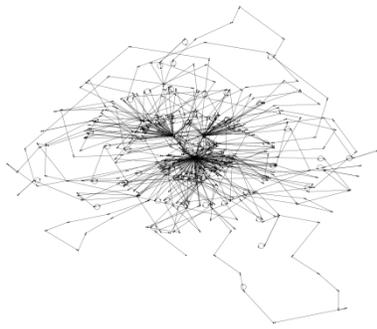


Figure 5-B: Editors in the LCC for the cumulative network for each article type and point in time.

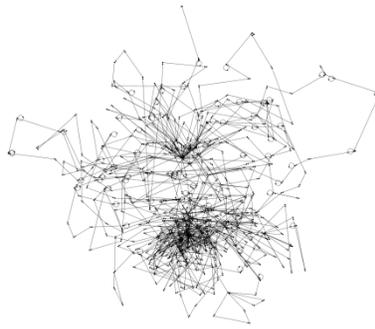
Finding 2: Sociotechnical trajectories capture complex structures and dynamics

Figure 6 depicts six representative article revision networks; three for breaking articles and three for non-breaking articles. By visual inspection, some breaking articles appear to have “tighter” patterns of revision activity, with fewer “chains” and more “loops”. However, other

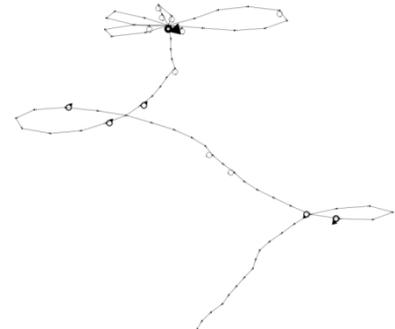
breaking articles like Kam Air Flight 904 exhibit the “loose”, chain-like structures I expect to find in non-breaking articles (see Figure 6c), and some non-breaking articles such as China Airlines Flight 611 exhibit the “tight” loop-like structures I expect to find in breaking articles (see Figure 6f).



**6-A: Adam Air Flight 574
(Breaking article)**



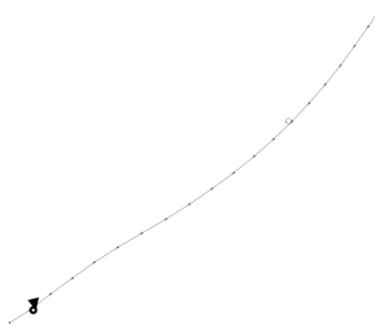
**6-B: Air France Flight 358
(Breaking article)**



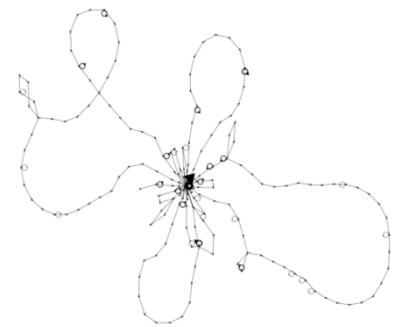
**6-C: Kam Air Flight 904
(Breaking article)**



**6-D: Lufthansa Flight 2904
(Non-breaking article)**



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



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

A statistical analysis of these networks reveals strong relationships between the immediacy with which an article is created after an event and how “tight” or “loose” the pattern of revision activity is in each network. 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. Articles with a “loose” revision history structure manifest greater final distance between editors’ revisions than articles with a “tight”

revision history structure. Motivated editors who repeatedly contribute on breaking news articles are make contributions after more diverse editors which in turn brings these otherwise peripheral editors closer together. 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-E) that exhibit a "loose" pattern of activity, 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-B) have low average betweenness since there are several editors connecting the entire network together providing multiple paths for interaction. Finally, articles about incidents that occurred farther in the past exhibit lower average clustering coefficients.

Trajectories in which editors make contributions modifying many other editors' contributions create denser networks with higher average clustering. The groups of editors whose revisions are modified by a single editor were less likely to mutually modify others' revisions. In contrast, the breaking news editors exhibited a greater tendency to revise in groups or clusters when engaging Wikipedia's editorial process.

Unlike prior work looking at co-authorship patterns *across* Wikipedia articles, 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. 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. 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.

Finding 3: Editors’ migrations across communities support collaboration

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, 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.

User Flodded edited the main article about the 2011 Tohoku earthquake and tsunami exclusively approximately 14 hours into the article's existence and continued to edit daily until March 23 making 542 of the article's 6202 revisions. 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.

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. User L.tak's contributions were also wide-ranging and varied, the most active editor on both the articles for the "Fukushima Daiichi power plant" article and talk page as well as the second most active editor on the "Fukushima Daiichi 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 respective topics.

User ACSE’s 41,778 revision editing history before the Japanese earthquake, tsunami, and nuclear disasters 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 also 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.

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. 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.

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.

Finding 4: Editor and article features both influence the co-authorship structure

Coauthorship patterns on Wikipedia thus foreground the duality of persons and material artifacts (Breiger, 1974; Contractor, Monge, & Leonardi, 2011): collaborations occur around articles exhibiting particular features but articles also emerge from the contributions of editors with distinct traits. The material properties 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.

Using 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, 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 provide a more complete account of self-organization on Wikipedia. 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 (Contractor, Wasserman, & Faust, 2006; Monge & Contractor, 2003). A statistical approach to network analysis called p*/exponential random graph modeling 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 (Robins, Pattison, Kalish, & Lusher, 2007; Wang, Sharpe, Robins, & Pattison, 2009).

“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. 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, the results of the analysis reveal that experienced editors work together on high tempo collaborations significantly less often than expected by chance. “Editor-focused interactions” capture the tendency for editors experienced in making many contributions to articles to contribute articles that share the same features such as being breaking articles. Contrary to expectations, the effects of experienced editors’ sustained contributions to only breaking or historical articles 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 editors with moderate amounts of experience who appear to play a crucial role not only contributing to many articles but also acting as crucial brokers providing bridges within breaking, non-breaking, and historical article types. The analysis suggests 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.

Conclusions

Wikipedia's breaking articles provide exhaustive large-scale and longitudinal data logs of user behavior and interaction with digital artifacts that allow analysis into 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, evidence suggests 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.

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 (1991) 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 result 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).

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, Forte, & Bruckman, 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 high tempo knowledge work in 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 contributes to our understanding of the structure and dynamics

of knowledge collaboration in multidimensional networks within large and complex 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, Bruns, & Neuberger, 2007). This research has implications for designing new communication and information technologies as well as managing new organizational form to respond resiliently to support high tempo, knowledge-intensive, and computer-mediated organizations that are becoming increasingly prevalent for vital social actors like entrepreneurs, disaster responders, and political reformers.

References

- Baker, W. E., & Faulkner, R. R. (1991). Role as resource in the Hollywood film industry. *American Journal of Sociology*, 97(2), 279-309.
- Bakker, R. M. (2010). Taking stock of temporary organizational forms: A systematic review and research agenda. *International Journal of Management Reviews*.
- 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.
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*: Yale University Press.
- Boczkowski, P. (2004). *Digitizing the News: Innovation in Online Newspapers*. Cambridge: MIT Press.
- Breiger, R. L. (1974). The duality of persons and groups. *Social Forces*, 53(2), 181.
- 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.
- 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.

- Contractor, N., Monge, P., & Leonardi, P. (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: An analytic framework and empirical example. *Academy of Management Review*, 31(3), 681-703.
- Deuze, M., Bruns, A., & Neuberger, C. (2007). Preparing for an age of participatory news. *Journalism Practice*, 1(3), 322-338.
- Faraj, S., & Xiao, Y. (2006). Coordination in fast-response organizations. *Management science*, 52(8), 1155.
- Ferron, M., & Massa, P. (2011). 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>
- Goodman, R. A., & Goodman, L. P. (1976). Some management issues in temporary systems: A study of professional development and manpower-the theater case. *Administrative Science Quarterly*, 494-501.
- 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.
- Lindkvist, L. (2005). Knowledge Communities and Knowledge Collectivities: A Typology of Knowledge Work in Groups*. *Journal of Management Studies*, 42(6), 1189-1210.
- Majchrzak, A., Jarvenpaa, S., & Hollingshead, A. (2007). Coordinating expertise among emergent groups responding to disasters. *Organization Science*, 18(1), 147.
- Meyerson, D., Weick, K., & Kramer, R. (1996). Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research*, 166, 195.
- Monge, P., & Contractor, N. (2003). *Theories of Communication Networks*. New York: Oxford University Press.
- Newman, M. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20).
- Pentzold, C. (2009). Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies*, 2(2), 255.
- 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.
- Wang, P., Sharpe, K., Robins, G., & Pattison, P. (2009). Exponential random graph (p*) models for affiliation networks. *Social Networks*, 31(1), 12-25.
- Weick, K., & Roberts, K. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 357-381.