Centralized, Parallel, and Distributed Information Processing during Collective Sensemaking

Peter Krafft*, **Kaitlyn Zhou**†, **Isabelle Edwards**†, **Kate Starbird**†, **Emma S. Spiro**†
*Massachusettts Institute of Technology, †University of Washington
Cambridge, MA; Seattle, WA
pkrafft@mit.edu, [katezhou,iare,kstarbi,espiro]@uw.edu

ABSTRACT

Widespread rumoring can hinder attempts to make sense of what is going on during disaster scenarios. Understanding how and why rumors spread in these contexts could assist in the design of systems that facilitate timely and accurate sensemaking. We address a basic question in this line: To what extent does rumor evolution occur (1) through reliance on a centralized information source, (2) in parallel information silos, or (3) through a web of complex informational interactions? We develop a conceptual model and associated analysis algorithms that allow us to distinguish between these possibilities. We analyze a case of rumoring on Twitter during the Boston Marathon Bombing. We find that rumor spreading was predominantly a parallel process in this case, which is consistent with a hypothesis that information silos may underlie the persistence of false rumors. Special attention towards detecting and resolving parallel information threads during collective sensemaking may hence be warranted.

ACM Classification Keywords

H.5.3 Information Interfaces & Presentation: Groups & Organization Interfaces: Collaborative computing, Computer-supported cooperative work; K.4.2 Social Issues; J.4 Social and Behavioral Sciences: Sociology

Author Keywords

Collective sensemaking; rumoring; rumor spreading; rumor evolution; collective intelligence; computational social science; computational modeling; social media; sociotechnical systems; Twitter

INTRODUCTION

Collective sensemaking is a central component of response to disaster scenarios. In these situations, accurate and efficient collective sensemaking can lead to life-saving information revelation, impacting both decision making and actions of affected populations. Communicating and interacting with others are key elements in the sensemaking process, where individuals and groups must collectively come to understand what is going on around them [30]. Today, collective sensemaking increasingly occurs in online settings [15]; social media platforms help facilitate crisis communication, but are also rife with misinformation [25, 38]. Building tools to improve collective sensemaking, particularly on social media, could therefore have a meaningful impact on peoples' lives. To build functional tools, we must better understand how the complex process of collective sensemaking occurs in these settings.

Rumors can often align with attempts to understand unfolding events when information is scarce, and as such rumoring behavior is often considered a form of collective sensemaking [30]. Rumors—defined as unverified information—often arise in situations characterized by extreme uncertainty and a lack of official information sources, as is the case in crisis contexts. Rumors provide explanations arising from the "collective sensibilities of the ... public" [6], lessening anxiety and restoring a sense of control in ambiguous conditions. Rumors are emblematic of collective sensemaking because in situations in which this process occurs, people often cannot or do not distinguish facts from false rumors.

Studies of rumoring behavior have a long history in the social sciences; prior work has tended to focus on the process of rumor spreading and the subsequent changes in content that result from serial transmission [3]. Although scholars proposed specific mechanisms leading to this phenomena (e.g. the tendency for rumors to retain specific content *(sharpening)* and lose other details *(leveling)*), empirical analysis of the collective sensemaking processes is lacking. Mapping how variations and permutations of rumor stories unfold and change over time could improve our understanding of collective sensemaking and our ability to facilitate efficient and accurate information gathering and decision making.

In this study, we focus on rumoring behavior immediately following the 2013 Boston Marathon Bombing. We map the dynamics of collective sensemaking by exploring variations of a particular rumor story over time. Occurring simultaneously, as well as in sequence, distinct threads of communication composed into a complex and widespread false rumor about a young girl who died while running the marathon in support of the Sandy Hook school shooting victims.

CHI 2017, May 6-11, 2017, Denver, CO, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-4655-9/17/05 ...$15.00.
http://dx.doi.org/10.1145/3025453.3026012

Feature	Meaning	Example Tweet
boy	Child was a boy, not a girl	Everyone is posting the little girl running saying 8 year old killed in Boston marathon... It was an 8 year old boy that was killed guys..
charity	Child was running for charity	So sad, this precious 8 year old was running in the Boston Marathon for charity. #RIPBabyGirl http://t.co/yldeNmxKYA
father	Child was waiting for father at the finish line	I thought it was an eight year old boy that was waiting for his father to cross the line that died in Boston not a girl running herself
sandy	Child was running for Sandy Hook victims	R.I.P. to this 8yr old girl who died in Bostons explosions, while running for the Sandy Hook kids. RT for respect. http://t.co/ZBhI4on7tN
spectator	Child was a spectator, not running in the race	Im pretty sure this lil girl wasnt running the marathon & died. Her tag says 5k on it. Stop posting this picture. http://t.co/EoFmmnQHgG
young	Child died	One of the casualties was an 8 year old girl running the marathon. The theme was 26 miles for 26 victims. The things people do make me sick.

Table 1: Descriptions and examples of the features identified in our tagging procedure.

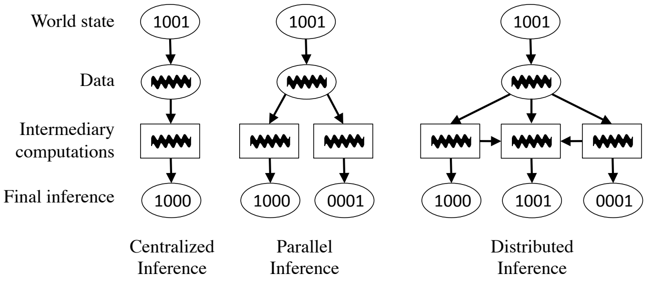


Figure 1: Inference algorithms take as input data that is assumed to be reflective of some underlying world state of interest. The algorithms then perform intermediary computations before arriving at final inferences.

Our approach is to map the information space of rumor-related discussion. From a computational point of view, it is natural to think of collective sensemaking as a form of inference from observations. Inference is the process of arriving at knowledge via the observation of data. For instance, one might wish to infer the parameters of a model, such as the value of the gravitational constant g in the Newtonian model of physics. In this illustrative hypothetical example we might attempt to find the gravitational constant that best explains trajectories of a set of falling objects we have observed. To find this best value of g, one common approach is to try a sequence of values and observe the explanatory power of each element in this sequence. However, such processes are often sensitive to the initial element of the sequence, or the sequence itself might involve random choices. Therefore running multiple sequences in parallel can improve the speed or accuracy of inference. In a more complex scenario, data may be distributed across different locations, and servers at each location must periodically communicate intermediate inferences with each other in a carefully designed distributed algorithm to ensure that final inferences are accurate. Figure 1 illustrates schematic diagrams of centralized, parallel, and distributed architectures for inference algorithms.

In the case of collective sensemaking, individuals are observing and participating in discussion about a rumor on social media; they are forming beliefs or opinions based on their observations and interactions. Given that this process frequently occurs across geographically distributed participants, it is natural to ask whether the collective sensemaking we observe in these situations is consistent with processes of parallel or distributed inference. Abstracting this process may then allow us to draw on a variety of inference procedures that have been studied in the vast literature on optimization and machine learning to develop more precise mathematical models of collective sensemaking (e.g., [5, 16, 21, 29]).

BACKGROUND

Rumoring Behavior on Social Media

Early work on rumors studied the motivation for rumors, when and how they propagate, and how representations of information mutate as a rumor evolves through a population [2, 7, 27]. Recent work extends theories of rumoring to online environments, categorizing prominent attributes of rumors in social media platforms [6, 22, 38]. The relationship between attributes of this sort and the propensity for a rumor to spread has also been studied [20, 33]. Other works have looked at how information propagates through social networks without attending to how the representations of information change [4, 8, 10, 12, 37].

In closely related recent work, researchers examined how memes evolved through news articles [31] and how rumors evolved through the network of Facebook posts [1, 11]. This prior work focused on understanding the details of how rumors mutate individually, as well as how those mutations and the attributes of the resulting rumors were related to their propagation rates, but this prior work also identified points when distinct rumors or rumor variants merged together. Ultimately our results are largely consistent with the related results of this prior research, but our focus is on the information processing implications rather than the implications for genetic models of memetics.

To the best of our knowledge, none of this prior empirical work has attempted to infer the dependencies between rumor or conversation threads in order to measure the extent of parallel and distributed information processing occurring during collective sensemaking. Our work also contrasts with the closely related prior work on rumor evolution on Facebook in the richness of the rumor we study. Prior work has focused on studying quoted text and edits between quotes. This choice allowed for large-scale data analysis but limited the types of rumors that could be studied and the contextual detail that could be brought to bear in understanding this phenomenon. Here we use a semi-automated coding system using regular expressions to tag conceptual content of tweets associated with the rumor we study, and hence we can study the process of collective sensemaking at a different level of analysis. Compared to using quotes, we are able to observe conceptual changes rather than just textual changes, and our methods are applicable to any social media rumor rather than just copy-and-paste reposts.

Social Computation

Our work also connects to a growing literature that attempts to conceptualize human groups and social systems as information processing systems [9, 13, 17, 18, 19]. While rumoring behavior and collective sensemaking have been discussed informally in computational terms previously (e.g., [37]), to the best of our knowledge, ours is the first work that leverages an explicit information processing model in order to better understand observed rumoring behavior, and among the first to do so for collective sensemaking behavior (cf. [32]). Of course, many formal models of rumor behavior exist, including large bodies of literature on cultural diffusion [14] and on gossip algorithms [28], but the formal models in this area that are connected with data tend to lack coherent interpretations as systems that actually perform information processing, and the models that do have interpretations as information processing systems tend to be divorced from empirical data. The present work marries a formal information processing view of social systems with the empirical study of rumoring behavior and collective sensemaking.

ARCHITECTURES OF SOCIAL COMPUTATION

In designing software systems, one of the basic architectural decisions to be made first is whether the system will be implemented in a centralized, parallel, or distributed manner. This choice involves deciding not only whether to use multiple processing units but also how and when multiple processing units would communicate with each other during computation (typically either not at all or through shared memory in the parallel case, or through message passing in a prespecified network structure in the distributed case).

When viewing social phenomena through the lens of computation, it is also natural to ask in which of these capacities a phenomenon is operating. In the social context, the question is fundamentally about the complexity of the patterns of information flow between people. However, it is difficult to disassociate the terms parallel and distributed information processing from the algorithmic techniques typically associated with those areas in computer science. Our first step is therefore to achieve a level of conceptual clarity around what it means for an arbitrary information processing system to be operating in a centralized, parallel, or distributed capacity. This characterization obviously must be general enough so that it can apply to systems of people, who are clearly not arranged according to the typical architectures of centralized, parallel, or distributed software.

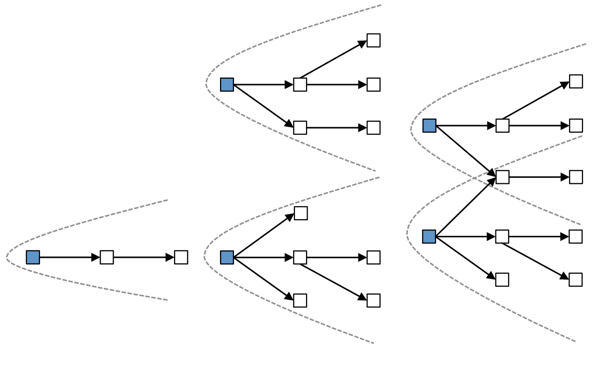

(a) Centralized (b) Parallel (c) Distributed

Figure 2: Informational dependency graphs associated with centralized, parallel, and distributed computation. The nodes represent computation blocks and the edges informational dependencies. The dotted lines show the downstream “influence cones” of the highlighted blue nodes—the set of nodes that their computation affects. Centralized computation is associated with sharp cones in an information pipeline. Parallel computation is associated with non-intersecting information silos. Distributed computation displays intersecting cones that form an information web.

Basic Framework

We first suppose that information processing, whether in software or in society, consists of compositions of “computation blocks”. In the case of social computation, a block might be a single person thinking about a single fact. In the context of our work, each computation block will be associated with a new inference event. These computation blocks are composed via informational dependencies, in which the output of one computation block is communicated as input into another block. In the social case, an informational dependency could arise from one person communicating an inference to another person. An informational dependency could also occur between one person thinking about a fact at one point in time and the same person rethinking that fact at a later point in time with the benefit of the inferences gleaned on the first occasion. The distinction between centralized, parallel, and distributed processing can then be understood in terms of the structure of the network of informational dependencies between computation blocks. These graph structures are illustrated in Figure 2.

Before describing each of these structures in turn, it is worth noting that unlike in software systems, where all engaged processing units play a functional role in the computation, social systems may also have people that perform actions that are not functionally relevant to the computation effectively embodied by the system. For example, a person may participate in a system by passively consuming the news. These tangential actors are informationally passive, and while they enrich the system, they will not affect its computational architecture.

Centralized Processing

Centralized processing is characterized by chain-structured informational dependency graphs. Centralized algorithms for inference typically involve a single processor taking data observations and gradually updating inferences based on that data (in a series of computation blocks). In the social context, one example of a centralized collective sensemaking phenomenon would be a crowd of people passively consuming news from a single source, perhaps one that updates its stories over time. The computation blocks in this case are found within the news source, and the passive observers are excluded from the information architecture.

Parallel Processing

Parallel processing is characterized by tree-structured informational dependency graphs. Parallel algorithms for inference typically involve multiple processors running isolated instances of an identical algorithm. While atypical for parallel inference algorithms, we also treat spawning child processes, e.g. that run replicated computation starting at the last observed parent state, as included within the scope of parallel architectures. In the social context, parallel processing could occur through multiple people actively engaging with news from a single source by modifying that news according to their own experiences. Parallel processing could also occur through distinct groups of people passively consuming news from single distinct sources. Or parallel processing could occur through people interpreting a situation using their own experiences without reference to information from others.

As is evident in these examples, parallel processing can be associated with rigid "information silos". From a technical point of view, in a parallel process, the descendants of two computation blocks with no path between them form distinct information silos. No two downstream computation blocks from each of those parent blocks will share new information between them.

Distributed Processing

Distributed processing is characterized by trellis-structured informational dependency graphs, i.e., directed acyclic graphs. Distributed algorithms for inference mainly involve periodically reciprocally sharing intermediary processed data or inferences between processors. In the social context, distributed processing could occur if people integrate information from multiple news sources or synthesize ideas heard from multiple people.

MODELING COLLECTIVE SENSEMAKING

The goal of our work is to understand the extent to which observed processes of collective sensemaking on social media are centralized versus parallel or distributed. If people mainly draw upon a single global source for their information about a particular scenario, and only copy, embellish, or refine that information privately, collective sensemaking would appear essentially like a centralized process. If individuals or small groups mostly replicate or refine distinct non-interacting threads, the process would appear essentially parallel. If multiple threads exist and continuously interact with each other, the process would be truly distributed. Each of these potential conclusions suggests different classes of algorithms to look towards in further work developing models of collective sensemaking as inference.

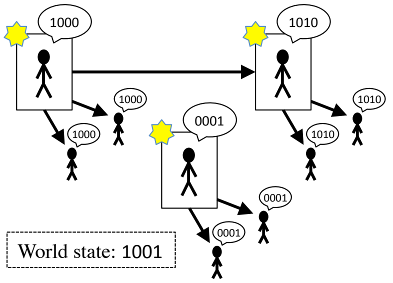


Figure 3: An illustration of our conceptual model of collective sensemaking/rumor spreading. Computation occurs at the boxed individuals who create new hypotheses. Each utterance shown in the figure would be an observation in our dataset.

To grapple with these questions, we suppose that the problem of collective sensemaking can be viewed as one of inferring an unknown present state of the world. We suppose that the state of the world can be represented by a binary vector of features. For example, in a disaster context, some features might be "there was a tornado", "there is a fire", "someone has died", et cetera. We also suppose that people communicate hypotheses about the world, again represented as binary vectors consisting of the features of the world state (e.g., using the features above, (1,1,0) indicates a hypothesis that there was a tornado and there is a fire, but nobody has died). Collective sensemaking involves the process of propagating, augmenting, refining, merging, supporting, and extinguishing these candidate hypotheses. We will define one thread in a process of collective sensemaking as the collected expressions of one candidate hypothesis (a particular combination of nonzero features).

To simplify the analysis, we assume that individual features or combinations of features are only discovered once, and all other occurrences of those features or combinations are copies or modifications of those first occurrences. While this simplifying assumption likely misses the complexity of real-world collective sensemaking, where truly independent sources could post essentially identical observations for example, the assumption allows us to begin to map a very complicated phenomena. Our conceptual model is illustrated in Figure 3.

DATA

The dataset we analyze consists of observations of collective sensemaking unfolding during the 2013 Boston Marathon Bombing. On April 15, 2013 two bombs were detonated near the finish line of the Boston Marathon. Hundreds of spectators and runners were injured by the blasts, and three individuals were killed. In the week that followed, police launched a significant manhunt for two suspects, during which time social media users collectively attempted to help apprehend the suspects. Several rumors propagated during this time, including the case study in this paper.

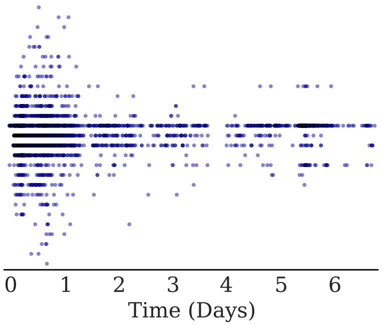


Figure 4: A plot of the temporal dynamics and frequency of the observed instances of each rumor thread. There is one row for each rumor thread, and one point for each tweet.

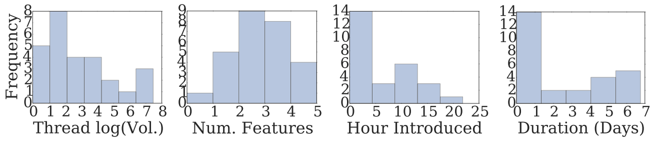
Collection

The data we analyze was collected in a prior study by using the Twitter Streaming API with the search terms “blast”, “boston”, “bomb”, “explosion”, and “marathon” [22]. Data collection began a few hours after the explosion of the first bomb and continued for seven days; some periods of observation have missing data due to rate limits of the Twitter API. Prior work has examined this data, identifying multiple rumor stories and qualitatively coding individual tweets based on their relevance to each rumor [22, 35, 38].

Tagging

Although the dataset we use was studied in prior work, we enrich the data by adding feature tags to 4,826 tweets associated with one larger rumor story. These tags correspond to “features” in our conceptual model. We focus on the “Girl Running” rumor, which contained claims and references to a young girl who was killed while running the marathon in honor of the victims of the 2012 shooting at the Sandy Hook elementary school in Newtown, CT. We focus on this rumor because it is well-defined while still containing interesting complexity in the number of distinct ideas present in the conversation. We use 4,826 distinct tweets identified as related to this rumor (via manual coding in prior work).

In an iterative and collaborative process, we compiled a list of distinct information features within the larger rumor story. Two authors of the present work took a random sample of 100 of these tweets, examining each tweet individually to identify and note new information that was being introduced into the rumor landscape. Using a list of keywords and phrases to identify each of these features, and after discussion among the research team, a consensus was reached about the set of appropriate features. Throughout this process, the level of detail of the features was refined. For example, two features that overlapped conceptually and empirically would likely be merged into a single feature; other cases were separated into multiple more specific features. For each feature, a regular expression-based query was developed to automatically retrieve and identify tweets from the dataset related to that particular rumor feature. Finally, some irrelevant and unreliable features were removed prior to analysis. The resulting set of features identified in this process, along with basic descriptions and example tweets, is shown in Table 1. We consider a single thread of discussion to be the set of tweets associated with a particular combination of these features.



(a) (b) (c) (d)

Figure 5: (a) Distribution of the logarithm of the volume (number of tweets) of each thread. (b) Distribution of the number of features associated with each thread. (c) Histogram of the times at which we first observe an instance of each thread. (d) Histogram of the duration of each thread.

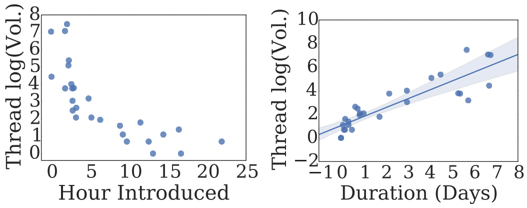


Figure 6: (Left) The time a thread first appeared versus its ultimate total tweet volume. (Right) The duration of time over which we observe a thread versus its total volume.

Descriptive Statistics

This dataset consists of 27 distinct rumor threads. In this section we present visualizations of basic descriptive statistics of these threads to provide context for our analysis. The time stamps of all tweets associated with the 27 threads are shown in Figure 4. Immediately this plot begins to hint at the structure of collective sensemaking in this case. We see that there are three threads that dominate in terms of the volume of tweets, and tweets associated with each of these three threads occur throughout our observation window. Figure 5 displays the distribution of tweet volume by thread as a histogram, again showing that most threads have low volume. We also observe that most threads have an intermediate number of features associated with them. Figure 5 also shows that most observed threads are created early, and last only a short time. Figure 6 shows that the most popular threads tend to be introduced earlier, and the length of time a thread exists appears to have an exponential relationship with the total volume of tweets associated with it. These results indicate the early-comers are likely to persist and enjoy widespread attention.

METHODS

Our main goal is to evaluate the degree to which rumoring behavior and by extension the collective sensemaking process observed in this case displays signatures consistent with centralized, parallel, or distributed information processing. We combine quantitative and qualitative assessment of the rumor threads in this analysis to give insight into how communication was unfolding over time. For the quantitative component of our analysis, we develop an automated method for classifying threads into different types of dependencies, associated with either centralized, parallel, or distributed computation.[1] For the qualitative component of our analysis, we look at example tweets from threads associated with different types of dependencies, drawing on expert knowledge of this particular case study. Our qualitative component adds richness and context for our observed results, increasing understanding as well as lending face validity to quantitative results.

Analysis Approach

To understand whether threads are formed in a centralized, parallel, or distributed manner we attempt to reconstruct the dependency structure between threads. Given our conceptual model we firstly suppose that any tweet that consists entirely of a previously observed combination of features belongs to the thread associated with that combination. When we see a tweet with a new combination of features, all of which are features that have been observed previously, we can deduce that this tweet must have been derived from some prior thread or threads. There are two possibilities in this case. If that specific combination of features has appeared as a subset of a combination of features that appeared in a single previous tweet, then the new thread could have been formed by a deletion event. If that specific combination of features has never appeared at all in any previous tweet, then the new thread must be from a merge of two prior threads. For example, if we observe the threads (1,1,0) and (1,0,1), and then later observe (1,1,1), this latter thread could have been due to a merge of these two earlier threads. Finally, when we find a new feature together with an otherwise old combination of features, we can think about this tweet as forming from a mutation event. Since all tweets we analyze were detected as associated with the Girl Running rumor, we treat threads that consist of completely novel features as mutations on the base thread $(0,0,\dots,0)$.

Each of these types of dependencies are associated with centralized, parallel, or distributed processing. Centralized and parallel processing would consist only of mutations and deletions events since in these forms of computation there is either only a single thread or multiple non-interacting threads. Distributed processing would also include merge events between threads since it consists of multiple interacting threads of computation.

Even with the deductions just described, there is still ambiguity in the dependency structure between threads. This ambiguity arises from cases where a thread could have been formed either by a deletion event or a merge event. For example, if a tweet with features $(1,1,1,1)$ exists already, as well as tweets with features $(1,0,0,0)$ and $(0,1,0,0)$, then a new tweet with features $(1,1,0,0)$ could have been a refinement of the first thread, or a merge of the second and third threads. We resolve these ambiguities in three different ways, detailed below. We use three different methods in order to provide a more robust analysis.

Algorithm 1 Merge Event Lower Bound (MELB)

```
for rumor thread $r$, sorted by first appearance time do
    $s := r$ with newly observed features stripped
    if $s$ has been observed previously then
        count $r$ as a mutation
    else
        if the features in $s$ are a subset of a prior thread's then
            count $r$ as a deletion
        else
            count $r$ as a merge
```

Algorithm 2 Merge Event Upper Bound (MEUB)

```
for rumor thread $r$, sorted by first appearance time do
    $s := r$ with newly observed features stripped
    if prior threads not strict supersets of $s$ could merge to $s$ then
        count $r$ as a merge
    else
        if $s$ has been observed previously then
            count $r$ as a mutation
        else
            count $r$ as a deletion
```

Counting Thread Types

Our first two methods of resolving ambiguous dependencies establish lower and upper bounds on the number of merge events that could have occurred between threads. Establishing a merge event lower bound (MELB) provides the most severe test within the confines of our conceptual model of whether distributed processing is occurring at all in the collective sensemaking event we study. The purpose of the lower bound is to avoid overestimation and a false positive declaration of distributed information processing. To establish a lower bound on the number of merge events, we simply attribute thread creation to deletion events whenever there is a case where a thread could have formed from either a deletion or a merge event. Therefore, with this method, merge events are likely undercounted. Obtaining a merge event upper bound (MEUB) allows us to avoid a false negative declaration of distributed information processing. To establish an upper bound, we attribute any thread creation to merge events whenever there is a case where a thread could have formed either from a deletion or a merge event. Pseudocode for these two methods is shown in Algorithms 1 and 2.

Inferring Dependencies

Our third method resolves the same ambiguity in a different way. This method first defines a set of three atomic actions that can be composed to lead to new threads. The first action consists of merging all of the features of two threads, the second action consists of deleting one feature from one thread, and the third action consists of creating a new feature on one thread. These actions can be composed any number of times in the creation of a new thread, but this method assumes that newly observed threads were formed by the minimal number of actions upon previously existing threads. When this heuristic does not completely determine a single set of potential dependencies (for example, if a new thread could have been formed by a single deletion from an old thread, or a single merge of two old threads), then the dependency set with the highest mean tweet volume among the potential dependencies

[1] The code for our quantitative analysis is online at https://github.com/pkrafft/Centralized-Parallel-and-Distributed-Information-Processing-during-Collective-Sensemaking.

Type	Dependencies	MELB Method	MEUB Method
None	Single	1	1
Mutation	Single	6	6
Deletion	Single	10	3
Merge	Multiple	10	17

Table 2: Counts of the different thread dependency types. One thread (with zero features) represents discussion detected as relating to the rumor not displaying distinctive features.

with minimal edit distance is chosen. The algorithm we wrote for identifying the set of simplest dependencies for each new thread is a complex recursive algorithm, and hence we omit a description of its details due to space limitations. This method not only allows us to count the number of times each type of dependency occurred, it also allows us to identify which specific prior threads likely contributed to a new thread. Thus we can directly reconstruct and examine a specific plausible dependency structure of threads using this method.

RESULTS

The descriptive statistics we examined of our dataset already indicated that multiple prominent threads of discussion were occurring simultaneously during the propagation of the Girl Running rumor. This finding suggests that collective sensemaking in this case was either primarily parallel or primarily distributed. Distinguishing between these two possibilities requires examining how the main threads of discussion might have arisen. Beyond the major two or three threads of discussion, there were also a substantial number of threads with a lower volume of tweets associated with them. We are also interested in examining the common dependency patterns among these lower volume threads. The results from counting the number of instances of each possible type of dependency are shown in Table 2. A visualization of specific inferred dependencies between threads is shown in Figure 8.

Parallel Information Processing

The results of the MELB method, shown in Table 2, indicate that most threads we observe could have been formed either by mutations or deletions from old threads. Both the mutation and deletion actions only operate on a single existing thread, and hence are associated with centralized or parallel processing. Furthermore, as shown in Figure 7, threads with single dependencies tend to have higher total volume than threads with multiple dependencies (a difference that is statistically significant $p = 0.0437$ under a two-sided t-test with the MELB method and at $p < 0.001$ with the MEUB method). In fact, threads with either no dependencies or only a single dependency account for about 97% of the tweets we observe (including retweets) under the MELB method and still about 96% under the MEUB method. We further observe that a large portion of this percentage is attributable just to three threads, all of which have either zero or one possible dependency. Hence from our quantitative analysis, the process of collective sensemaking in this case appears to be predominantly parallel as opposed to distributed. Figure 8 indicates that our third method for inferring dependencies also supports this interpretation.

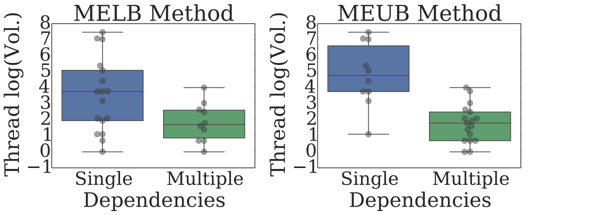


Figure 7: Threads with single dependencies tend to have higher log volume than threads with multiple dependencies. Each point in each plot is a single thread.

Examples

We look at three examples of threads with dependencies and dynamics that indicate parallel processing. As has been previously noted in investigations of the rumor studied here [22], and as confirmed by our quantitative analysis, this rumor consisted mainly of two threads. One thread involved reports that a young girl had died in the 2013 Boston Marathon Bombing, possibly while running in the marathon. In our data, this thread is associated with the "young" feature. Since this thread consists only of a single feature, the only possible ways in which it could have been created are through a deletion event of a more complicated thread, or through an innovation on top of the general discussion thread (that has no features). Our quantitative analysis reveals that "young" was indeed an innovation without prior dependencies beyond the general discussion around the event. Looking at the actual text of the tweets in our dataset, the first tweet we observe referencing this thread is the following: "RT @XXX: An eight year old girl who was doing an amazing thing running a marathon, was killed. I cant stand our world anymore".[2]

The main variant of this thread, which has been recognized in prior work [22], involved reports that the young girl who had died was running in memory of the victims of the Sandy Hook school shooting. In our data, this thread is associated with the "young,sandy" combination of features. Since this thread has two features, and since we can confirm that the "young" thread existed first within our tweets, the "young,sandy" thread could have been formed either by a merge of the "young" thread and a separate "sandy" thread, or it could have formed from a simplification of a more complex thread, or from a mutation of the "young" thread.

Our dependency inference method indicates that the "young,sandy" thread was in fact formed by a mutation of the "young" thread. That is, the first appearance of the "sandy" feature co-appeared with the "young" feature. The first tweet we observe with this combination was "RT @XXX: + Boston Marathon Bombing has me sobbing a fucking river. An 8 year old, a little girl who was running for Sandy Hooks,

[2] The first such tweet we observed was a retweet. This is either because the rumor was propagating earlier than the beginning of the data collection or because Twitter API rate limits led the data collection to miss the earliest tweet.

...” and we have no clear references to Sandy Hook before this tweet. However, we do observe earlier references to the little girl who died having been running for charity. The first such tweet we observe (classified as in the “young,charity” thread) was “So sad, this precious 8 year old was running in the Boston Marathon for charity. #RIP-BabyGirl http://t.co/yldeNmxKYA”. It could therefore be that “young,sandy” mutated from this thread rather than the “young” thread alone.

In either case, all of these threads have single dependencies among the existing threads of discussion, and hence are characteristic of parallel information processing. Participants in the online discussion appear to have seen prior reports in a single thread and introduced variants of the rumors that then propagated in parallel.

One final piece of evidence that supports the fact that these threads were indeed propagating in parallel highly isolated threads of discussion is that eventually (indeed actually rather early on) threads appeared that identified the rumor about the young girl as being false. For instance, this tweet appeared almost immediately after the first instance of the “young” thread that we observed, “Just wanna tell you that you have to qualify for the Boston Marathon so a 8 year old wouldnt be running!!!! that picture isnt of the girl!!!”, and a number of similar subsequent tweets also appeared. Yet the main rumor threads persisted despite these denials.

Distributed Information Processing

While the few threads that include the vast majority of tweets we observe appear to be associated with parallel information processing, we also find signatures of distributed information processing in the tail of the discussion activity. As shown in Table 2 we identify at least ten threads, a substantial proportion of the total number of threads we observe, that cannot have been formed by deletions from or mutations of prior threads. This 40% or so of threads do not constitute a large volume of tweets, but they may shed light on how some people access the information available during collective sensemaking, and they highlight the complexity in the details of the collective sensemaking process.

Examples

We examine two examples of inferred merge events to check whether these cases appear to be legitimate rather than artifacts of the tagging procedure or of our inference procedure. The first example occurs between the threads “young,boy” and “young,sandy”, which appear to merge into “young,sandy,boy”. Here are the tweets associated with “boy” that appear before the first appearance of “young,sandy,boy”:

- “Everyone is posting the little girl running saying 8 year old killed in Boston marathon... It was an 8 year old boy that was killed guys..”
- “RT @XXX: For those seeing photo of 8 yr old girl running with claims that she died: CNN has reported it was an 8 yr old BOY wh ...”

The next tweet with a “boy” feature is then in the “young,sandy,boy” thread: “HES A BOY @XXX: R.I.P. to the 8 year-old girl who died in Bostons explosions, while running for the Sandy Hook kids. #prayforb...”

Clearly this tweet has integrated information about the boy who died while spectating with an observation of the Sandy Hook thread. The “young,boy” thread also continues in parallel with tweets such as the following:

- “Just sayin it was an 8 year old boy that died, not a girl, for all those who posted a picture of a little girl running #GodBless #Boston”
- “RT @XXX: Is this true? CNN says it was a boy RT @XXX: Girl, 8, died in #Boston while running ...”

And eventually we see another “young,sandy,boy” instance: “An 8 year old girl died, and an 8 year old boy in Boston today. And they both were running in honor of sandy hook. Like really?”

Another merge event we observe occurs between the “young,boy” and “spectator” threads. Here we find a reference to the fact that the girl people had been talking about could not have been running in the marathon because of her age, as well as a reference to the report that a young boy had died, not a girl: “It was a boy that died not the 8 year old girl.. plus she wasnt running you need to be 18 to run the marathon. #prayforboston”

These threads, though comparatively low in volume, indicate that some people are paying attention to multiple parallel threads that are occurring, and integrating information from across them into new threads. Again, however, these merge events appear primarily to create new parallel threads rather than extinguishing old threads.

DISCUSSION

The results presented above explore collective sensemaking during the 2013 Boston Marathon Bombing. The evidence we examined suggested that collective sensemaking in this case contained both parallel and distributed information processing. These results have implications for understanding collective sensemaking in the context of crisis events.

Sources of Parallel and Distributed Processing

Given the observed high degree of parallelism in this case study of collective sensemaking, as well as some degree of distributed information processing, it is important to consider how these components (and their combination) may arise in social media systems; collective sensemaking in online environments may well be constrained by the features of the platforms used. In this case, the structure of collective sensemaking during extreme events is driven primarily by the affordances and limitations of the platform interface itself. Twitter, for example, allows users to search for distinct content of interest using keywords or hashtags. Search results then display tweets associated with distinct subsets of the larger discussion surrounding this rumor, and the event itself. Users can also receive content through pre-existing or newly formed social relationships (i.e. follower ties). Alternate mechanisms of exposure might be third-party platforms, or even word-of-mouth.

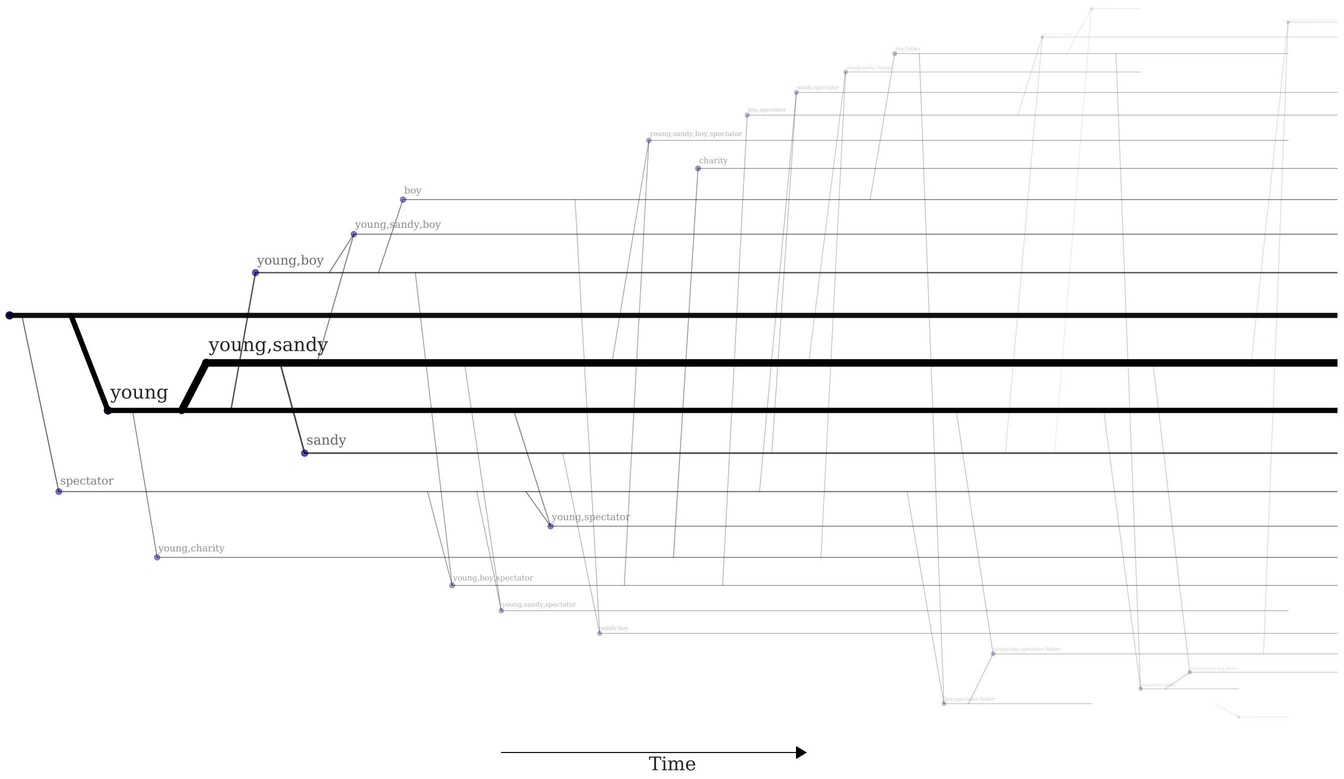


Figure 8: A reconstruction of the dependencies between threads of the rumor we study. Each horizontal line is a thread, with the thickness of the line being monotonically related to the volume of tweets associated with the thread.

Given the similarity between the language used in affirming and denying tweets (e.g. copy and pasting an affirming tweet and adding a denial of the content at the beginning or end), and given how persistent the dominant threads were even in the face of conflicting information, the limited display of information in the interface seems unlikely to represent the whole story. It is likely extremely difficult for users to have a complete and accurate picture of all rumor-related conversation in the total information space; users instead see a very specific, potentially biased, subset of crisis communication. It is unsurprising then, given these features of the communication channel, that multiple parallel discussions may exist.

Another source of parallelism is suggested when examining the text of the example tweets, as well as referencing prior work about rumoring during crisis events, specifically on social media platforms. Social media posts persist, though perhaps not visibly, and can be “re-discovered” at much later time points. Users who may have missed detailed rumor discussion, and even resolution of uncertainty within the rumor, can re-introduce rumors by re-posting outdated content. Indeed, prior work has observed cases of rumor resurgence [4, 22].

Moreover, rumor-related messages often contain references to external sources, such as mass media outlets. In the case reviewed here, for example, a large portion of the conversation surrounding denials of the Girl Running rumor explicitly referenced content from CNN which noted it was actually a young boy who was killed while spectating. Each individual thread of discussion we examined was centralized by definition since they consisted of the same information content. It is possible that distinct threads obtain their coherency from some centralized, external information source such as CNN. Some amount of parallelism could therefore be due to different people attending to difference news outlets.

Finally, there has been continued interest in the ability of the distributed crowd on social media platforms to engage in crisis response and recovery activities, coordinating aid activities, situational awareness, crisis mapping, donations, etc. [23, 24, 36, 34]. Distributed information processing could arise from the coordinated actions of populations that have prior experience with crowdwork and crisis participation, and who are tweeting about various different discussion threads.

Relation to Theories of Rumoring

It is also interesting to consider how our information processing perspective relates to traditional characterizations of rumoring processes such sharpening and leveling, as well as other social phenomena such as information silos and echo chambers. Deletion events can likely be viewed as sharpening or leveling. Merge events, mutations, and innovations are probably related to “assimilation”. Echo chambers and the spread of false news could potentially be exacerbated by parallel rumoring while potentially being mitigated by centralized or distributed rumoring. Future work could study these connections in more depth.

Implications for Design

Research that improves our understanding of collective sensemaking, and the types of information processing that accompany it, has many implications for the design of systems to help facilitate these processes. For example, making observed parallel processing visible to users could help online crowds transition from parallel to centralized or distributed systems for collective sensmaking, perhaps increasing efficiency and allowing human resources to be directed at multiple tasks rather than multiple versions of the same task.

Prior work has often focused on building systems to detect rumors (e.g. [26]). Less attention has been paid to building systems that help facilitate collective sensemaking and rumoring behavior itself. Such work could aid in crisis response, allowing online crowds to reach consensus or at least collective understanding of appropriate response actions. These systems could also assist in decision making by directly facilitating communication and interaction that builds understanding and awareness of crisis situations and how events unfold. Our study represents just one step in this direction.

Methodological Limitations

One limitation of our study is in the raw data itself. The dataset was collected using Twitter's Streaming API, and using particular keyword searches. Hence, it is likely the case we only have a sample (albeit a large sample) of the full set of tweets discussing this rumor. If multiple threads were introduced close together in time, lacking comprehensive coverage of all tweets could disrupt the dependencies we infer.

Our tagging procedure also introduces a potential source of error. Features were manually curated through an extensive and resource intensive procedure, but ultimately the features associated with specific tweets were determined automatically via regular expressions and are therefore noisy. Tagging is susceptible to both false positives and false negatives; errors could disrupt our inferred dependencies for the same reason that missing tweets do.

We also need to consider possible error due to our modeling assumptions. The main modeling assumption we leverage is that features are only innovated once. This assumption is almost certainly not met, but it may be approximately satisfied if most people get their information from looking at what others say rather than investigating the event independently, or if each feature is innovated by a single source external to the platform (such as a news source). If this assumption is not satisfied, then we may inappropriately infer merge events when none have actually occurred. Without this assumption, then, apparent merge events between threads could also be attributable to a thread rediscovering a feature contained in a separate thread, rather than having an actual interaction with that thread.

While strong assumptions simplify models and limit their immediate application in real-world settings, they also provide necessary parsimony when exploring complex phenomena. Moreover, the qualitative analysis used to augment our quantitative findings helps to evaluate the impact of all of these limitations. By carefully checking the inferences from our quantitative analysis against prior work and against the text of the tweets in the raw data, we can gain confidence in our conclusions. Future work could examine probabilistic models that account for gaps in our data, for unreliable features, and for varying probabilities of different features being discovered independently in order to relax our modeling assumptions and assign probabilities to merge events.

Finally, we must be careful not to generalize too readily from the single Girl Running rumor that we study. The amount of parallel and distributed information processing that we observe could be tied to the type of rumor we chose to study (a clearly false rumor that was likely intentionally started by a person who knew it to be untrue). Further, many of the affirming permutations around this rumor were likely purposeful misinformation, so the sensemaking process was mostly around questioning/correcting the rumor. Future work could conduct the type of analysis we have performed here with a broader set of rumors.

CONCLUSIONS

In this work we present a new way to analyze and characterize individual rumoring scenarios. Our conceptual model opens the door to richer mathematical models of rumor spreading than currently exist. Our analysis methods and conceptual model also lead to inferential capabilities and attendant visualizations that could be leveraged immediately in a system for collective sensemaking.

We illustrate these techniques using data from a rumor that has been extensively analyzed previously using mixed-method, interpretivist research methods, including detailed qualitative analysis in conjunction with temporal signatures [22, 35, 38]. This prior work serves as a comparison for the current work, allowing us to demonstrate how our new methodological approach leads to additional insights about how the rumor changed over time, and providing insight into why it kept propagating (widely) after it had been corrected. This prior work also demonstrates that this rumor is similar in temporal signature shape to the majority of Twitter rumors, and highly representative of one of three Twitter rumor types.

Our substantive finding that rumoring in the case we studied was largely centralized and parallel may be central to explaining why rumors resurface even after they have been corrected, but is also interesting in its own right. Given that the rumoring process could easily have been highly distributed, some underlying forces must be favoring centralized and parallel rumor evolution. Future work could identify these underlying forces. Finally, future work could compare how the prominence of different dependency types and the underlying forces leading to them vary across different types of rumors or different social media and interaction contexts.

ACKNOWLEDGEMENTS

This work was supported in part by the NSF GRFP under grant #1122374, by NSF grants 1420255, 1541688, and 1640940 and by ARO W911NF-15-1-0270. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

REFERENCES

1. Lada A Adamic, Thomas M Lento, Eytan Adar, and Pauline C Ng. 2016. Information evolution in social networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 473–482.

2. Gordon W Allport and Leo Postman. 1946. An analysis of rumor. *Public Opinion Quarterly* 10, 4 (1946), 501–517.

3. Gordon W Allport and Leo Postman. 1947. *The Psychology of Rumor*. Henry Holt.

4. Ahmer Arif, Kelley Shanahan, Fang-Ju Chou, Yoanna Dosouto, Kate Starbird, and Emma S Spiro. 2016. How Information Snowballs: Exploring the Role of Exposure in Online Rumor Propagation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 466–477.

5. Dimitri P Bertsekas and John N Tsitsiklis. 1989. *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall.

6. Prashant Bordia and Nicholas DiFonzo. 2004. Problem solving in social interactions on the Internet: Rumor as social cognition. *Social Psychology Quarterly* 67, 1 (2004), 33–49.

7. Theodore Caplow. 1947. Rumors in war. *Social Forces* (1947), 298–302.

8. Justin Cheng, Lada A Adamic, Jon M Kleinberg, and Jure Leskovec. 2016. Do Cascades Recur?. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 671–681.

9. Simon DeDeo. 2014. Group Minds and the Case of Wikipedia. *Human Computation* (2014).

10. P Alex Dow, Lada A Adamic, and Adrien Friggeri. 2013. The Anatomy of Large Facebook Cascades. In *Seventh International AAAI Conference on Weblogs and Social Media*.

11. Adrien Friggeri, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor cascades. In *Eighth International AAAI Conference on Weblogs and Social Media*.

12. Sharad Goel, Duncan J Watts, and Daniel G Goldstein. 2012. The structure of online diffusion networks. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 623–638.

13. Rob Goldstone and Georg Theiner. 2017. The Multiple, Interacting Levels of Cognitive Systems (MILCS) Perspective on Group Cognition. *Philosophical Psychology* (2017).

14. Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. 2013. Information diffusion in online social networks: A survey. *ACM SIGMOD Record* 42, 2 (2013), 17–28.

15. Thomas Heverin and Lisl Zach. 2012. Use of microblogging for collective sense-making during violent crises: A study of three campus shootings. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 34–47.

16. Eshcar Hillel, Zohar S Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. 2013. Distributed exploration in multi-armed bandits. In *Advances in Neural Information Processing Systems*. 854–862.

17. Edwin Hutchins. 1995. *Cognition in the Wild*. MIT Press.

18. Michael Kearns, Siddharth Suri, and Nick Montfort. 2006. An experimental study of the coloring problem on human subject networks. *Science* 313, 5788 (2006), 824–827.

19. Peter M Krafft, Julia Zheng, Wei Pan, Nicolás Della Penna, Yaniv Altshuler, Erez Shmueli, Joshua B Tenenbaum, and Alex Pentland. 2016. Human collective intelligence as distributed Bayesian inference. *arXiv preprint arXiv:1608.01987* (2016).

20. Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*. IEEE, 1103–1108.

21. Nancy A Lynch. 1996. *Distributed Algorithms*. Morgan Kaufmann.

22. Jim Maddock, Kate Starbird, Haneen J Al-Hassani, Daniel E Sandoval, Mania Orand, and Robert M Mason. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 228–241.

23. Andrew Mao, Winter Mason, Siddharth Suri, and Duncan J Watts. 2016. An experimental study of team size and performance on a complex task. *PloS one* 11, 4 (2016), e0153048.

24. Patrick Meier. 2012. Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time. *Journal of Map & Geography Libraries* 8, 2 (2012), 89–100.

25. Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1589–1599.

26. Paul Resnick, Samuel Carton, Souneil Park, Yuncheng Shen, and Nicole Zeffer. 2014. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In *Proc. Computational Journalism Conference*.

27. Stanley Schachter and Harvey Burdick. 1955. A field experiment on rumor transmission and distortion. *The Journal of Abnormal and Social Psychology* 50, 3 (1955), 363.

28. Devavrat Shah. 2009. *Gossip Algorithms*. Now Publishers Inc.

29. Ohad Shamir. 2014. Fundamental limits of online and distributed algorithms for statistical learning and estimation. In *Advances in Neural Information Processing Systems*. 163–171.

30. Tamotsu Shibutani. 1966. *Improvised News: A Sociological Study of Rumor*. Ardent Media.

31. Matthew P Simmons, Lada A Adamic, and Eytan Adar. 2011. Memes Online: Extracted, Subtracted, Injected, and Recollected. *International AAAI Conference on Weblogs and Social Media* 11 (2011), 17–21.

32. Paul R. Smart and Katia P. Sycara. 2013. Collective Sensemaking and Military Coalitions. *IEEE Intelligent Systems* 28 (2013), 50–56.

33. Emma S Spiro, Sean Fitzhugh, Jeannette Sutton, Nicole Pierski, Matt Greczek, and Carter T Butts. 2012. Rumoring during extreme events: A case study of Deepwater Horizon 2010. In *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 275–283.

34. Kate Starbird. 2013. Delivering patients to sacré coeur: Collective intelligence in digital volunteer communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 801–810.

35. Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert M Mason. 2014. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *iConference 2014 Proceedings* (2014).

36. Kate Starbird, Grace Muzny, and Leysia Palen. 2012. Learning from the crowd: Collaborative filtering techniques for identifying on-the-ground Twitterers during mass disruptions. In *Proceedings of 9th International Conference on Information Systems for Crisis Response and Management, ISCRAM*.

37. Kate Starbird and Leysia Palen. 2012. (How) will the revolution be retweeted?: Information diffusion and the 2011 Egyptian uprising. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. ACM, 7–16.

38. Kate Starbird, Emma Spiro, Isabelle Edwards, Kaitlyn Zhou, Jim Maddock, and Sindhuja Narasimhan. 2016. Could This Be True?: I Think So! Expressed Uncertainty in Online Rumoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 360–371.