Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures

Jim Maddock†*, Kate Starbird†, Haneen Al-Hassani, Daniel E. Sandoval†, Mania Orand†, Robert M. Mason*
HCDE†, iSchool*, DUB
University of Washington, Seattle WA, 98195
{maddock, kstarbi, haneen1, dansand, orand, rmmason}@uw.edu

ABSTRACT
This study offers an in-depth analysis of four rumors that spread through Twitter after the 2013 Boston Marathon Bombings. Through qualitative and visual analysis, we describe each rumor's origins, changes over time, and relationships between different types of rumoring behavior. We identify several quantitative measures—including temporal progression, domain diversity, lexical diversity and geolocation features—that constitute a multi-dimensional *signature* for each rumor, and provide evidence supporting the existence of different rumor types. Ultimately these signatures enhance our understanding of how different kinds of rumors propagate online during crisis events. In constructing these signatures, this research demonstrates and documents an emerging method for deeply and recursively integrating qualitative and quantitative methods for analysis of social media trace data.

Author Keywords
Social computing; misinformation; rumoring; information diffusion; Twitter; crisis informatics;

ACM Classification Keywords
H.5.3 [Information Interfaces & Presentation]: Groups & Organization Interfaces - Collaborative computing, Computer-supported cooperative work; K.4.2 Social Issues

INTRODUCTION
Increasingly, social media platforms are becoming places where people affected by a crisis go to share information, offer support, and collectively make sense of the event [23]. Rumoring and its byproduct, misinformation, present a threat to the utility of these platforms, as noted by media [12] and emergency managers [13,14]. Distinguishing between misinformation and truth in online spaces is difficult, especially during the high-volume, fast-paced action of a crisis event as it unfolds. There is growing interest in technical solutions that could help detect misinformation [5,20,24] or preemptively reduce its ability to spread [4,7], and indeed these solutions would be quite useful in the crisis context. However, to generate the best strategies for identifying and reducing the impact of misinformation, we may first need to better understand how false rumors take shape and spread through online spaces.

Existing research on the spread of rumors online is primarily quantitative, including descriptive studies of trace data [9,15,26], theoretical research on network factors [4,7], and prescriptive studies that experiment with machine learning methods to classify rumors as true or false [5,15,24]. When qualitative methods are employed in this space, they typically consist of manual coding of large numbers of tweets, followed by quantitative methods—i.e. descriptive and statistical analyses—to infer meaning from patterns of codes [3,20,22]

Our research, a descriptive study intended to inform future predictive efforts, takes a different approach by profoundly and recursively integrating qualitative and quantitative methods to gain an in-depth understanding of how rumors develop and spread through social media after disaster events. This approach features the use of multi-dimensional *signatures*—or patterns of information propagation—to help understand, characterize, and communicate the diffusion of specific rumors. In this study, we examine four rumors that spread via Twitter after the 2013 Boston Marathon Bombings, establishing connections between quantitative measures and the qualitative "story" of rumors, and revealing differences among rumor types.

BACKGROUND

Definitions of Rumor and Rumoring
Rumor can be defined in a number of ways. In the social computing literature, Qazvinian et al. define a rumor as "a statement whose truth-value is unverifiable or deliberately false" [24, p. 1589], and Spiro et al. [26] offer a broader definition of rumor that includes any kind of informal information—i.e. not from "official" sources—without specifically considering its veracity.

Researchers of social psychology use a slightly different definition—for example, DiFonzo & Bordia define rumors as unverified statements that arise out of "danger or potential threat, and that function to help people make sense and manage risk" [8, p. 13]. Social psychologists often treat

CSCW '15, March 14 - 18 2015, Vancouver, BC, Canada
Copyright 2015 ACM 978-1-4503-2922-4/15/03…$15.00
http://dx.doi.org/10.1145/2675133.2675280

rumor as a process—i.e. *rumoring* [3,25]. Rumoring, in this perspective, is a collective activity that arises in conditions of uncertainty and ambiguity as groups attempt to make sense of the information they have [1,25]. Rumoring is also related to anxiety [2], and can be motivated at times by emotional needs—i.e. sharing information with others after an emotionally powerful event can be cathartic [10].

The Problem of Misinformation during Crisis

A byproduct of rumoring behavior is misinformation, which can be a problem in the context of disaster events [12,16]. The perception of online spaces being overrun by misinformation may limit the utility of these platforms during crisis situations, and some emergency responders are reluctant to include social media in their decision-making processes due to fear of misinformation [13,14]. There is a clear need for better understanding how and why misinformation spreads after disaster events.

Approaches for Studying Online Rumoring

Studying the diffusion of an online rumor presents an interesting methodological challenge. Important aspects of that rumor exist on two very different levels of analysis—i.e. both within patterns of information diffusion at the scale of thousands of posts, and within the specific content of individual tweets or links that help to shape, catalyze and propagate the rumor. Existing research has largely focused on the former unit of analysis and typically takes one of three approaches.

Quantitative Analysis

The first approach is purely quantitative and focuses on developing a high level understanding of the diffusion of rumors. This work includes descriptive, empirical studies [e.g. 9,15,26], as well theoretical research [e.g. 4,7]. Several of these studies analyze network features to understand the flow of misinformation [4,7,26]. Kwon et al. [15] include descriptive analysis of temporal characteristics, finding false rumors on Twitter have more spikes than true rumors. Friggeri et al. [9] examine the interaction between rumor corrections (via posted links to Snopes articles in comments) and the spread of rumors on Facebook, finding that large rumors keep propagating despite corrections.

Qualitative Coding with Quantitative Analysis

A second approach, which is largely descriptive, utilizes mixed methods. It first employs qualitative analysis to code large numbers of individual posts and then uses quantitative analysis to interpret patterns across those codes and over time [3,20,22]. The most influential of this research works to connect online rumoring to rumor theory from social psychology [1,25], using this knowledge to inform research questions and qualitative coding schemes. For example, studying rumor behavior in online discussion groups, Bordia & DiFonzo [3] identified five kinds of rumor statements, coded posts accordingly, and presented a model of rumor progression with four stages characterized by different proportions of each statement type. Oh et al. [22] coded tweets related to various rumors for six attributes, statistically analyzed how those codes related to each other, and found evidence that anxiety, personal involvement, and source ambiguity contribute to the spread of rumors online.

In research on Twitter use after the 2010 Chile Earthquake, Mendoza et al. [20] coded tweets related to several rumors as confirming or denying. They confirmed claims that the online crowd questions false rumors, and hypothesized that it might be possible to identify misinformation by automatically detecting those corrections.

Automatic Classification of Rumors as True or False

A third type of research focuses on developing machine learning algorithms for automatically detecting misinformation. In a follow up study to [20], researchers trained a machine classifier to determine the credibility of a news topic related to a set of tweets using corpus-level features that could relate to "questioning" behavior [5]. Qazvinian et al. [24] developed a machine-learning approach for classifying tweets as related to a false rumor, and as confirming, denying, or doubting a rumor. Their technique showed high precision and recall using content-related features in the tweet, and had reasonably high recall and precision for network and Twitter-specific features (i.e. hashtags and URLs). Kwon et al. [15] also explored solutions for automatically classifying rumors as either true or false by building off the Castillo et al. model [5], but introduced three types of additional features: temporal (looking at spikes over time), structural (looking at friend/following relationships) and linguistic (employing sentiment analysis). They found their feature set to be more effective than the features from Castillo et al [5].

Each of these machine learning studies measures the efficacy of their models—typically through accuracy, precision and recall scores—and provides some insight into a feature's relative predictive strength. Yet few studies provide significant insight into *how* and *why* rumors spread, and classification research has been limited to distinguishing between true and false information.

OUR APPROACH: CHARACTERIZING RUMORS THROUGH MULTI-DIMENSIONAL SIGNATURES

This research seeks to advance understanding of how online rumors form and spread, and to identify and distinguish between different types of rumors. We employ a unique methodological approach that relies upon a deep and recursive integration of qualitative and quantitative methods to construct multi-dimensional *signatures*—i.e. patterns of information flow over time and across other features [21]—that can be used to characterize rumors.

Preliminary research [27] describes the temporal signatures of three rumors across two codes (misinformation and correction). This study expands to include a more nuanced coding scheme—including speculation—and several more features that constitute dimensions of each rumor's signature. We use mixed-method analysis to tell the story of four distinct rumors spreading on Twitter after the 2013 Boston Marathon Bombings. Shifting repeatedly between quantitative and qualitative analysis, we describe each rumor's origins, propagation over time, relationships between rumor behavior types (e.g. sharing misinformation vs. correcting), prevailing URLs and domains, and lexical diversity, and work to connect each rumor's qualitative "story" to quantitative measures that reflect that narrative.

Event Background: 2013 Boston Marathon Bombings

On Monday, April 15 at 2:49pm, two bombs were detonated near the finish line of the Boston Marathon, killing three individuals and injuring 260 others. The instigators of the bombing fled the scene and a police investigation began immediately.

The FBI released photographs of two suspects on April 18 at 5:20pm. Three hours later, following an armed robbery near MIT, two men—who were later identified as bombing suspects—shot a police officer, carjacked an SUV, and engaged police in a violent firefight nearby Watertown, MA. One attacker was killed, but the other suspect escaped, setting off a manhunt through the Watertown neighborhood that concluded the next evening.

Between the bombings and the capture of suspects, the FBI requested photos or videos of the scene from the public. This call for information resulted in a massive volume of material, and social and mainstream media coverage of this information "resulted in a lot misinformation and false leads" [17]. During that time, social media users collectively and publicly searched through available photos and videos to find the perpetrators themselves. In some cases, innocent individuals were called out and accused, a behavior characterized as "digital vigilantism" [18].

Several other rumors took shape in social media spaces and began to spread. Like the errant search for suspects, some of these rumors stemmed from the online crowd's attempts to make sense of the situation, while others were simply falsehoods that targeted the emotions of an affected public.

Data Collection

For data collection, we used the Twitter Streaming API, tracking the following search terms: blast, boston, bomb, explosion, and marathon. We initiated this collection at 5:25pm[1] on April 15 and discontinued it at 5:09pm on April 22. Significantly, our collection did not cover the first few hours after the bombing and suffered several periods of data loss later in the event. It was also rate-limited (at ~50 tweets a second), causing data loss during the first six hours. It resulted in 10.6 million tweets contributed by 4.7M different authors. 56% were retweets and 47% contained a URL link, though another limitation of this data is that the data storage technique did not record metadata associated with a retweeted tweet or embedded URL, so these were calculated using textual searches, likely resulting in under-estimations for both measures.

Identifying rumors

In preliminary analysis of this Twitter data, we examined the most popular individual and co-occurring hashtags and noted a large number of viral stories. Further exploration of tweet text revealed certain hashtag groupings to be associated with specific rumors. From these we initially selected a subset of six rumors that were both diverse and highly visible during the event. Later, we eliminated two of these due to limitations in our collection that resulted in significant data loss (found to be >50%) for those which propagated during early periods of diffusion.

For each of the remaining rumors, we then identified a search string that returned from our total collection a comprehensive, low-noise set of tweets. For example, for Rumor #1, we used the following MySQL search string:

```
WHERE (text LIKE '%propos%' OR text LIKE
'%marry%') AND (text LIKE '%girl%' OR text
LIKE '%woman%')"
```

Table 1 lists the four rumors with total volume and number of distinct tweets in each rumor subset.

#	Rumor	Total	Distinct
1	Woman killed before proposal	3,371	1591
2	Girl killed while running	93,353	3275
3	Navy Seals as perpetrators	4,525	1996
4	Falsely accusing Sunil Tripathi	29,416	7445

Table 1. Rumor Subsets: Total Volume and Distinct Tweets

Tweet Coding

We coded each tweet within each rumor subset into one of seven distinct categories related to the rumor behavior type: *misinformation*, *speculation*, *correction*, *question*, *hedge*, *unrelated*, or *neutral/other*. These categories were developed through an iterative process during preliminary coding. We began with misinformation and correction, but as subtleties between rumor behavior types became salient, we added other codes. Table 2 provides examples for each of these codes (as they relate to Rumor #4, below).

Tweets coded as *misinformation* support the rumor without doubt, relaying the rumor as established fact and consequently spreading false information. *Speculation* tweets develop or support a growing rumor, and often introduce new information or commentary. The speculation tweet in Table 2 demonstrates this by inferring that the bomber 'may be' Sunil. In a *Hedge* tweet, an author passes along an existing rumor with some doubt about its veracity. These tweets show hesitancy, illustrating an unwillingness to declare the rumor as accurate, but continue to spread the

[1] All times are in EDT, local time in the affected region.

false information without a clear challenge. By contrast, a *question* tweet actively challenges or questions an existing rumor. Finally, *correction* tweets clearly negate the rumor.

Tweets coded as *neutral/other* are related to the rumor yet do not fit into one of the existing categories because the position is either neutral or unclear to researchers. Tweets that show no association to the rumor but exist within the subset due to search term noise are coded as *unrelated*.

Code	Example Tweet
Misinfo	`The 2 Boston suspects are mike muligeta and Sunil Tripathi. Sunil is still on the loose.`
Speculation	`This may be a shot in the dark but, I think the missing Ivy League student, Sunil Tripathi, is the guy responsible for the Boston bombings`
Hedge	`MIT suspects, per police scanners, are Mike Mulugeta and Sunil Tripathi. Probable, but not confirmed connection, with Boston Marathon Bombs`
Question	`Based simply on height, Boston suspect #2 doesn't look 6'2" Sunil Tripathi is. I just dont buy that its him.`
Correction	`Suspect #2 is NOT Sunil Tripathi! Both NBC and MSNBC confirmed this! #Boston`
Neutral/ Other	`Help us find Sunil Tripathi FB page was taken down within the 15 minutes... weird #Boston`
Unrelated	`Boston bombings, EQ and Sunil Narine!! The harlem shake. 3 disasters strike on the same day.`

Table 2. Coding Categories for Rumor-Related Tweets

Coding Process

Two researchers coded every distinct tweet (after removing retweets and close matches) in each rumor subset. Table 1 shows the number of distinct tweets per rumor. After the first round of coding, a third coder arbitrated disagreements by choosing the most appropriate code. Each rumor provided its own challenges, and initial agreement ranged from 70-90 percent. The most difficult distinctions occurred between *misinformation* and *speculation, speculation* and *hedge,* and for Rumor #3 in particular, *question* and *correction*. Agreement for collapsed categories (where misinformation includes speculation and hedge, and where correction includes question) was consistently much higher.

Analysis

This research utilizes a mixed-method approach, integrating quantitative, qualitative and visual analyses of tweets, to describe and characterize four rumors related to the Boston Marathon Bombings. These methods include the following:

Qualitative and visual analysis to understand the origins and evolution of each rumor: Using its temporal signature as a guide, we explore the progression of each rumor over time. The visual signatures allow us to identify significant moments in the rumor's propagation, and qualitative analysis enables us to understand the nature of the peaks, valleys, and interactions between rumor behavior codes.

Calculation of lexical diversity: To assess differences in the unique rumor behaviors present in tweet content, we calculate the lexical diversity—i.e. the number of different words that tweets in this corpus use. For each rumor behavior category, we create a dictionary containing every unique word mentioned by tweets with that code. Due to large imbalances in tweet volume for the different code categories, we normalize this calculation by selecting 100 random distinct tweets (excluding retweets) from each corpus.

Analysis of URL propagation and domain diversity: We explore the relationship between the rumoring behavior code of a tweet and URL links embedded in its text, as the latter represents a virtual link to content beyond the 140 character limit. Due to over-representation of a few domains across our rumor set, analysis at the domain unit is often more useful than focusing on individual URLs. We describe the domain diversity of each rumor—i.e. the distribution of embedded URLs across different domains—focusing predominantly on the top ten domains for each subset. Our data storage technique only retained the text of the current tweet (not upstream original tweets), which truncated a significant number of URLs.

Geolocation analysis: For the limited subset of tweets that include GPS data, we analyze tweet location to draw parallels between event proximity and rumor propagation.

FINDINGS

Rumor #1: Proposal Rumor

Rumor #1 claimed that a woman whose boyfriend was going to propose after the marathon died while running the race. This rumor may have developed from a small number of tweets sent prior to the bombings that mentioned a proposal taking place at the finish line. Soon after the bombings, the story appears to have morphed into a false rumor asserting that the woman had died. The first tweet in our dataset referencing this rumor is a retweet, which appears 55 times in our data set.

```
(April 15, 6:30pm): RT @TweeterA: a girl
who ran in the marathon was killed due to
the bombing & her boyfriend was gonna
propose to her afterwards
```

In our dataset, Rumor #1 consists of 3371 total tweets. It began to propagate a few hours after the event and had mostly run its course by about 11:00pm April 17, by which time 90% of volume had passed. The rumor reached a peak volume of 63 tweets per 10 minutes at 10:00pm April 15, but also had several subsequent local maximums. Figure 1 shows volume over time by rumor behavior code.

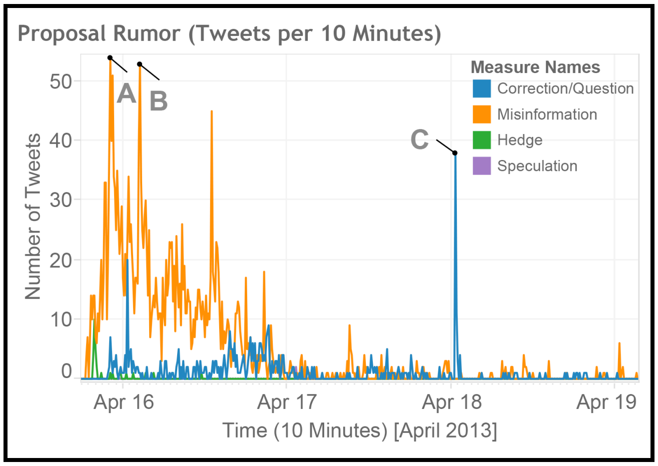

Figure 1. Rumor Codes Over Time for Proposal Rumor

An Added Photograph Sparks a Second Burst

The first peak largely consisted of text-only tweets. New variations of the rumor, where downstream authors restated the existing rumor in slightly different words, often led to subsequent increases in propagation and small spikes visible in Figure 1. The highest peaks (Figure 1, A & B) were generated by multiple variations of the rumor propagating at the same time.

Less than an hour after the rumor appeared, Twitter users began to introduce variations that included links to a photo:

```
(April 15, 7:11pm): this guy found his
girlfriend dead he was devastated he was
going to propose to her after the marathon
http://t.co/ZfIi7r4c3x
```

The above tweet has an embedded image showing a man at the scene of the blasts attending to an injured woman. This photo was originally published by the Boston Globe with a caption describing the scene as a man comforting a victim near the finish line. The URL links to a new copy of the photo posted within the pic.twitter.com domain without attribution. Manual analysis suggests that a large number of distinct tweets—i.e. different forms of the rumor—have this same photo embedded in them, an observation supported by URL and domain analysis. Figure 3 shows that while the first (and highest) peak in misinformation in Rumor #1 stems from mostly textual variations, the second major peak (Point B) is largely constituted by rumor variations that linked to a photo. These findings suggest the introduction of photos catalyzed additional bursts in the overall rumor.

Characterizing the Correction

13% of tweets in Rumor #1 were coded as corrections. Notably, the first correction of the photograph's context occurred hours before the peak volume of tweets containing the photo, indicating that the correction did little to stop the rumor's propagation. A much later burst of corrections, one that occurred days after the apex of misinformation (Figure 1, C), was a highly retweeted reprimand of rumor-spreading behavior more broadly:

```
(April 18, 12:30am) Some of these Boston
stories are so fake, like the one about
dude going to propose to the girl after the
race. That shit was made up
```

Differences and Interactions in Rumor Behavior Codes

Tweets coded misinformation dominated this rumor. Significantly, the rumor began as misinformation—i.e. very few tweets were coded as speculation. The crowd produced some corrections, but the overall ratio of correction to misinformation was small at about 3 to 16. The signal of misinformation for this rumor was therefore stronger than the signal of correction. Furthermore, corrections appear to get less amplification from the crowd; tweets coded as correction were somewhat less likely to be retweets than those coded as misinformation (51% vs. 65%). This difference is significant (Chi Square test, $p<0.01$).

Code	Total # Tweets	% RT	Lexical Diversity
Total	3576	54.9%	370
Misinformation	2586	65.2%	322
Speculation	2	50.0%	--
Hedge	34	70.6%	--
Question	22	27.2%	--
Correction	481	51.1%	379

Table 3. Tweets Per Code for Proposal Rumor

Lexical Diversity

Qualitative and visual analysis suggests that a large percentage of misinformation tweets use similar words. Subsequent quantitative analysis reveals that the combined textual content of misinformation tweets possessed a much lower lexical diversity than that of correction tweets. Taking a random sample of 100 distinct tweets from each coded category and averaging the results over 100 iterations results in a dictionary of 322 unique words for misinformation tweets compared to one of 379 words for correction. These differences in lexical diversity support the observation that corrective tweets tended to contain more original content than those that spread misinformation.

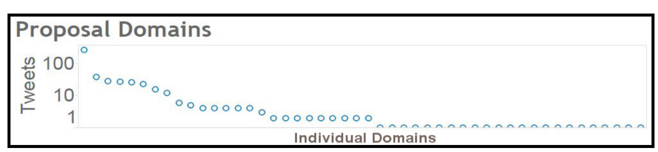

Figure 2. Domain Diversity, Proposal Rumor, Log Scale

Domain Diversity and Propagation Over Time

The distribution of tweets across the domains in the Proposal rumor is long-tailed (Figure 2). Twitter was the top domain (far left), linked to by 275 tweets; CNN.com, the second most sited domain was included only 39 times.

Of the top domains, each maps almost exclusively to a single code—misinformation or correction. Though our sample is too small to make a large generalization, here misinformation associated with social media domains—in this case Twitter and Instagram—while correction tweets that included URLs primarily linked to the CNN domain.

Examining the temporal relationships between URLs and rumor codes (Figure 3) indicates that certain domains seem to rise and fall in tandem with peaks of misinformation and correction. Both the Twitter and CNN domain groupings correspond to—and indeed match the shape of—a second peak of misinformation or correction, respectively. Because our URL counts are artificially low, it is likely that these domain signals are even stronger, and therefore fit the overall plots even more closely. Recall that the Twitter domain was associated with the photo that helped catalyze the rumor's spread; though we cannot claim that external content provided by URLs drives social media conversation, it does appear to facilitate both propagation and correction of rumors.

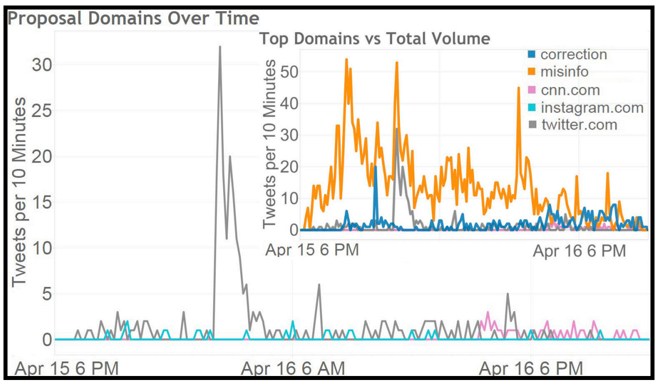

Figure 3. Domains Over Time for Proposal Rumor

Rumor #2: Girl Running Rumor

Rumor #2 stated that an eight-year-old girl died while running the Boston Marathon. This rumor may have originated from sense-making activities of the crowd as it attempted to process official information that an eight-year-old child had been killed at the bombings. An early tweet in our collection lends evidence to this hypothesis:

```
(April 15 6:09pm): What if one of the
people who died's baby girl was running her
first marathon ever today? #PrayForBoston
```

It's unclear whether this tweet is indeed connected to the rumor, but an hour later the rumor transformed into distinct misinformation:

> **@tweeterB** `(April 15 7:17pm): An eight year old girl who was doing an amazing thing running a marathon, was killed. I cant stand our world anymore`

This rumor has many similarities to the proposal rumor, and began to propagate only about 45 minutes later. However, Rumor #2 had a much higher tweet volume than Rumor #1, consisting of 93,353 tweets within our dataset. It began to spread rapidly at about 8:15pm on April 15 and reached its peak rate (2,287 tweets in ten minutes) at 9:50pm. It started to fade rapidly by midnight and 90% of its volume completed by 10:45am the next morning. Figure 4 shows volume over time by rumor behavior code.

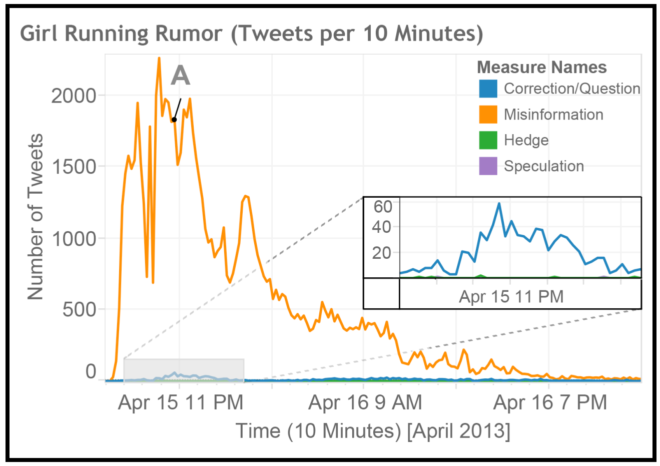

Figure 4. Rumor Codes Over Time for Girl Running Rumor

Similar Signatures: Propagation through an Image

The tweet generated by @tweeterB was the first instance of the Girl Running rumor, but it only received two retweets. The next version, which was retweeted much more widely (555 times), appeared four minutes later. It included a key addition—a link to an image of a young girl running a race.

```
(April 15 7:21pm): The 8 year old girl that
sadly died in the Boston bombings while
running the marathon. Rest in peace
beautiful x <link>
```

This photo was clearly an intentional false addition to the rumor, at least initially. As various other versions of the rumor began to spread in the following hours, most included that photo. Shortly after 11:20pm, the volume of misinformation increased dramatically, and at least 82% of tweets related to the rumor during that period included the image. By 9:50pm misinformation reached a peak volume of 2266 tweets per ten minutes.

Characterizing the Correction

Corrections only constitute about 2% of the total volume of the rumor. Until 9:50pm on April 15—at about the time that misinformation was reaching its highest tweet rate—they remained relatively rare. Corrections peaked less than an hour later at 56 tweets per ten minutes. Figure 4 shows the alignment and stark difference in amplitude between correction and misinformation signals.

The set contained a few different forms of correction. The first simply refuted the rumor outright:

```
(April 15, 11:42pm): Whoever started this
8yo girl dying while running the Boston
```

```
marathon for Sandy Hook needs to be dealt
with swiftly. #hoax
```

Another correction contradicted the rumor, offering a logical explanation for why it could not be true:

```
(April 15 11:46pm): An eight year old girl
did not die today in Boston while running
the marathon. You have to be 18+ to be in
it. Come on, people!
```

Finally, a third type attacked the rumor, refuting it with a factual explanation:

```
(April 16, 10:35pm): The 8-year-old victim
today was a boy: <link> Please stop RTing
nonsense about it being a girl running for
a cause.
```

It is impossible using our methods to establish a causal connection between the rise in corrections and the decline of misinformation, yet the temporal graph might indicate an interaction. Misinformation volume dropped dramatically, though temporarily, at 10:40pm on April 15 (Figure 4, A), while correction volume peaked at 56 tweets per 10 minutes during that same interval. However, by about an hour later misinformation had peaked again, before beginning a steady decline, interrupted by one more burst.

Code	Total # Tweets	% RT	Lexical Diversity
All Codes	93,353	96.0%	362
Misinformation	90,841	97.6%	260
Speculation	4	25.0%	NA
Hedge	8	37.5%	NA
Question	112	27.7%	NA
Correction	1931	48.2%	414

Table 4. Tweets Per Code for Girl Running Rumor

Differences and Interactions in Rumor Behavior Codes

Similar to the Proposal rumor, the Girl Running rumor contained much more misinformation than correction (48:1) and little speculation. In contrast to Rumor #1, here the percentage of retweets within tweets coded as misinformation is high, at 97.6%, demonstrating that the retweet mechanism played a major role in the overall signal of misinformation for the Girl Running rumor.

Domain Diversity and Propagation Over Time

Quantitative analysis of domains reveals similar patterns between the Proposal and Girl Running rumors. Again, the distribution of tweets across domains was long-tailed, with a sharp drop off between the top domain (twitter.com at 17,356) and the next domain (instagram.com at 288). CNN was the fourth most-tweeted domain at 202 mentions. Domains appear to map primarily to a single code—i.e. either misinformation or correction but not both. The Facebook domain is one exception, balancing relatively evenly between misinformation (14 tweets) and correction (11 tweets). Social media constitute much of the top ten domains in this rumor, and all strongly associate with misinformation. Again, CNN is the one domain primarily linked to by correction tweets (201 times).

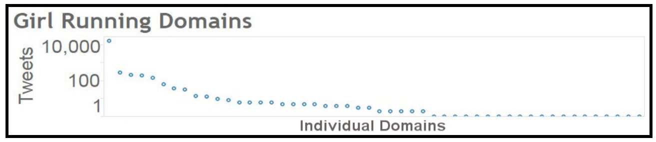

Figure 5. Domain Diversity, Girl Running Rumor, Log Scale

The shape of the temporal signatures of the Twitter domain—in this case photos on pic.twitter.com—again matches the shape of the misinformation signal overall (see Figure 6). For Rumor #2, it appears that the Twitter domain played a role in the propagation of misinformation from a very early stage, which aligns with our qualitative findings. Additionally, the CNN domain peaks at about the same time as corrective tweets, indicating some interaction between external sources and the propagation of correction as well.

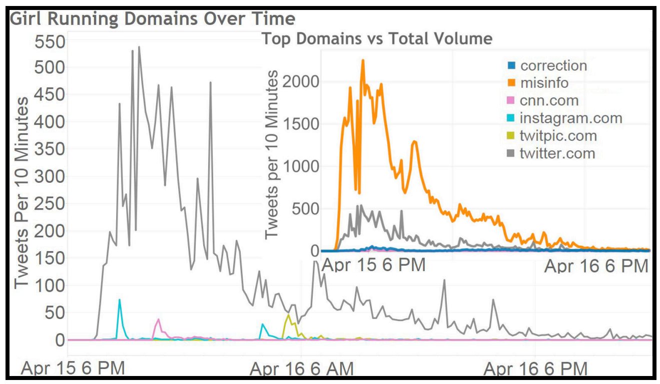

Figure 6. Domains Over Time for Girl Running Rumor

Rumor #3: Navy Seals Rumor

Rumor #3 speculated that either the Navy Seals or Craft Security facilitated the Boston Marathon Bombings. This rumor was associated with a popular #falseflag hashtag in our dataset. A "false flag" is a term used to describe an attack designed to appear as if carried out by someone other than its perpetrators.

At 6:44pm, April 15 the first tweet related to this rumor appeared in our data set:

```
RT @TweeterD: BOSTON BOMBINGS HAVE CIA
BLACK OPS WRITTEN ALL OVER IT! BALL
BEARINGS IN BOMBS - PLACEMENT OF BOMBS -
PATRIOTS DAY - JUST SAYIN...
```

Similar tweets appeared sporadically over the subsequent 24 hours. Initially, the rumor propagated at extremely low volumes, never exceeding two tweets an hour.

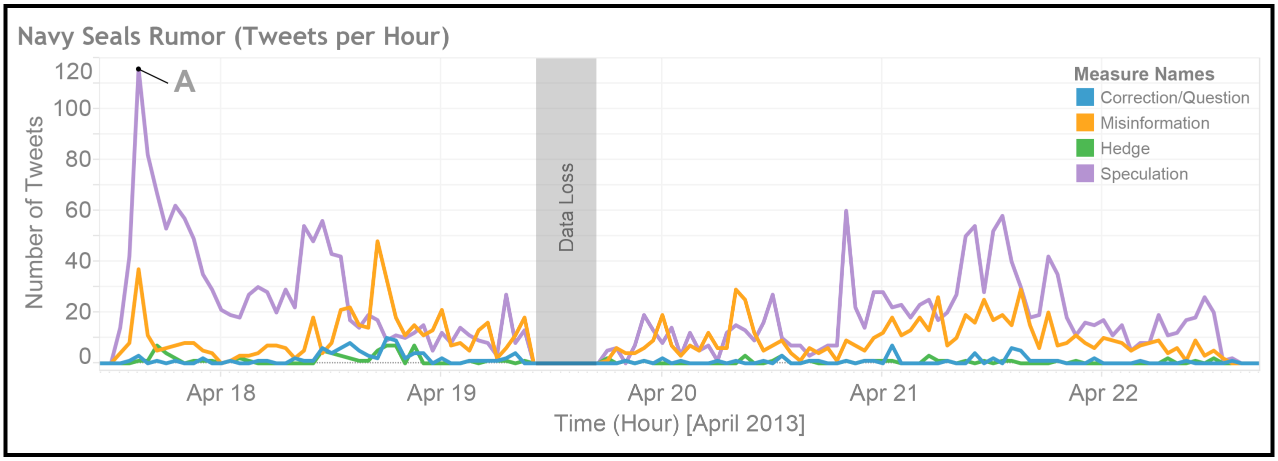

Figure 7. Rumor Codes Over Time for Navy Seals Rumor

That changed at about 1:40pm on April 17 when 35 tweets referencing the rumor were shared within one hour. Almost all included a link to an InfoWars.com article arguing that Navy Seals were behind the bombings:

```
(April 17, 1:38pm) Truth has been revealed:
Boston Bombing culprits found as NAVY SEALS
UNDERCOVER http://t.co/PFahtn28qy
THIS IS UNBELIEVABLE.
```

InfoWars is an independent media outlet run by Alex Jones, who also has a popular radio show. Links to the InfoWars website were the first links to appear in any tweets related to this rumor. Their precipitation to the first peak in misinformation further demonstrates how external content can help Twitter rumors evolve and spread.

Code	Total # Tweets	% RT	Lexical Diversity
All Codes	4151	30.1%	503
Misinformation	1171	35.4%	489/310*
Speculation	2551	28.5%	437/273*
Hedge	102	24.5%	NA/NA
Question	54	20.1%	NA/345*
Correction	87	18.4%	NA/368*

Table 5. Tweets Per Code for Seals/Craft Rumor
*** Due to low volume of distinct tweets for each code, the 2nd number here was calculated using only 50 random tweets.**

The total volume of tweets related to the Seals Rumor is 4525, similar to the Proposal rumor. This rumor, however, demonstrates a very different pattern over time than the first two with a much longer temporal range. Rumor #3 first appeared in significant volumes on April 17 and was still propagating at a consistent rate at the end of our collection period on April 22. The peak volume of 163 tweets per hour occurred between 3pm and 4pm on April 17. Though the volume did drop significantly after the initial peak, it continued to fluctuate between 20 and 70 tweets per hour for the remainder of the collection period.

This rumor notably contained two distinct variations, accusing either the Navy Seals or private agents from Craft Security. These different versions seem to arise from new interpretations of images and explanations from "experts" within the online crowd's sensemaking activities.

Speculative Signatures

In contrast to the first two, the Seals Rumor was largely driven by speculation, which accounted for 61% of the total volume. Additionally, it began with speculative tweets, and speculation largely constituted the first major peak (Figure 7, A)—which occurred from 3pm to 5pm, April 17. Multiple identical versions of the following tweet made up the majority of this initial peak, appearing 125 times in that period (and 699 times overall):

```
(April 17 2:44pm) Navy SEALs Spotted at
Boston Marathon Wearing Suspicious
Backpacks - http://t.co/4iIEKLyZtd
```

These tweets propagated not as retweets, but as original tweets with the same words, though often with different shortened links. All pointed to the same webpage, another article on the InfoWars site, and were likely generated by a tweet button on that page.

A Weak Correction

Correction and question tweets occurred in similar volumes for this rumor and were often hard to differentiate. However, even taken together, both categories make up only 3.4% of the rumor. This rumor therefore spread without much correction. When tweets containing corrections or questions did appear, they were less likely to be retweeted than speculation or misinformation tweets.

Lexical Diversity

Qualitative analysis suggests that tweets containing speculation and misinformation tended to vary more textually than they did in the previous two rumors. Where variations on a small group of initial tweets defined both the Girl Running and Proposal rumors, from the Seals rumor multiple themes emerged over the course of several days.

(April 18 5:50pm): RT @TweeterC: 4 #BostonMarathon #Bombing Images show Operatives of Craft Inter, a Blackwater-style private military/security Inc <link>

(April 20 12:12pm): FBI Boston Bombing Video Altered To Hide Fact Bomber Was Black Ops Mercenary? http://t.co/viUcjRs92I via @BeforeItsNews

Measures of lexical diversity support these observations, showing the Seals Rumor to be much more diverse overall—503 unique words in 100 random tweets compared to 362 for Rumor #1 and 370 for Rumor #2. Again, corrections show more lexical diversity than misinformation. This pattern is not only consistent across the correction code for all four rumors, but also seems to extend to other similar categories—i.e. correction and question both have higher lexical diversity than misinformation and speculation.

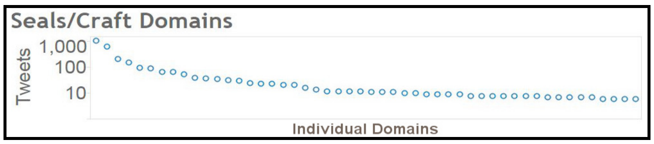

Figure 8. Domain Diversity, Seals Rumor, Log Scale

Domain Diversity and Propagation Over Time

The Seals rumor had a much higher level of domain diversity than Rumors #1 and #2—i.e. tweets in this rumor linked to many different domains. Figure 8 shows the signal of domain diversity, demonstrating a less steep slope than the previous rumors.

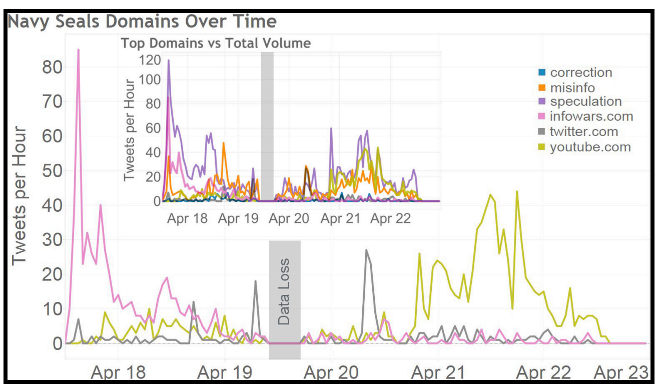

Figure 9. Domains Over Time for Navy Seals Rumor

None of the top domains were primarily associated with tweets coded as correction. Unlike the first two rumors, Twitter was not the predominant domain (219 tweets) and fell significantly below both Youtube.com (1083 tweets) and InfoWars.com (663 tweets), incidating that a different kind of external content shaped this rumor. Other alterative news sites appear in the top ten most tweeted domains, including beforeitsnews.com (160 tweets) and secretsofthefed.com (54 tweets). Conversation was therefore aided by outside media sources, but it was never addressed by traditional, "mainstream" outlets.

The temporal signature created by domains (Figure 9) again tends to align with spikes of misinformation. Similar to the previous rumors, volume increases among the top domains appear to correspond to spikes of misinformation and speculation. The initial maximum correlates to the appearance of the InfoWars domain, while the somewhat consistent spikes later in the window rise in tandem with YouTube.com, almost all linking to a video of a retired U.S. Army officer asserting that the attack was a false flag.

Rumor #4: Falsely Accused Rumor

Rumor #4 asserted that missing Brown University student Sunil Trapathi was a suspect, partially based on photos published by the FBI shortly after 5pm on April 18. After the actual suspects were identified, this rumor became a symbol for the negative impacts of online misinformation during the event.

Our dataset indicates that more than one person seeded this rumor, and that it likely existed elsewhere online prior to appearing on Twitter (e.g. Reddit). The first two tweets in our collection, sent about two hours after the photos were released, appear to be independent observations alleging similarities between Trapathi and images of a Boston Marathon suspect:

(April 18 7:38pm) Sunil Tripathi - Some might think he looks like the kid in Boston. But the FBI photos are too grainy to say for sure. http://t.co/9y5CivjlCX

(April 18 7:47pm) Sunil Tripathi. One of the Boston bombers. I'm calling it

The total volume of the Falsely Accused rumor was 29,416 tweets. Its lifespan was relatively short, beginning on April 18 at 7:38pm and tapering significantly by April 19 at 10pm. Peak volume was 4675 tweets per ten minutes at 3:00am on April 19.

Building Speculation Becomes Misinformation

The first few hours of the rumor were characterized by persistent, low-level speculation (520 of the first 725 tweets); a large number of these linked to a specific sub-Reddit thread. This trend changed drastically at 2:50am April 19, when the following tweet was shared:

(April 19, 2:50am): BPD scanner identified the names: Suspect 1: Mike Mulugeta Suspect 2: Sunil Tripathi. #Boston #MIT

Notably the reference to the police scanner added a layer of complexity to the Falsely Accused rumor; however, the scanner never actually claimed that Trapathi was a suspect. Yet after that point, due to a large number of retweets and references to the scanner claim, the signal of misinformation began to dominate.

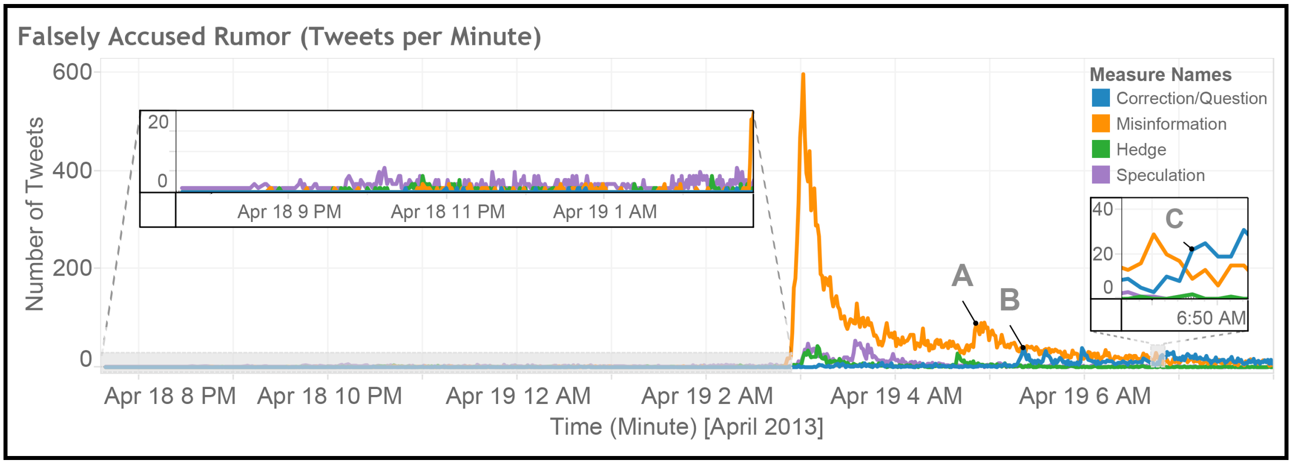

Figure 10. Rumor Codes Over Time for Falsely Accused Rumor

Interactions between Rumor Codes: The Crowd Corrects

Among the four, the Falsely Accused rumor is the only one where correction volume eventually surpassed misinformation. Early on, the combined signal of correction and question flickered at a few tweets per hour, including:

```
(April 19, 2:40am): Come on, you've been
watching way too many thrillers. Sunil
Tripathi doesn't look at all like one of
the Marathon bombing suspects. (Question)
```

```
(April 19, 3:20am): No, the scanner DID NOT
say suspect missing was Sunil Tripathi,
they said repeatedly light skinned, white.
#MITshooting #Boston (Correction)
```

After a second rolling maximum of misinformation before 5am (Figure 10, A), largely driven by a tweet claiming to show Tripathi in the same photo as the young boy who was killed in the blasts, correction volume began to approach misinformation volume, reaching about 40 tweets a minute on several occasions. These pulses never reached the volume of the initial peaks in misinformation, but they did approach (5:21am) and eventually surpass (6:48am) the volume of misinformation at those times. The first of those points (Figure 10, Point B) corresponds to a rise in tweets quoting an NBC statement that Tripathi was not one of the suspects. The second (Figure 10, Point C) occurred as Twitter users begin to widely share similar information from the AP that named Tsarnaev as the actual suspect.

After its peak, correction volume did not rapidly decline as with other rumors, but remained greater than the volume of misinformation for the remainder of the rumor's lifespan. Much of this persistent activity focused on public reprimand of the crowd for falsely accusing Tripathi.

Lexical Diversity

Lexical diversity for tweets related to the Falsely Accused rumor was much higher overall—at 555 unique words per 100 distinct tweets—than the Proposal (370) and Girl Running (362) rumors, and slightly higher than the Seals rumor (503). Consistent with patterns seen in other rumors, speculation and misinformation tweets had less lexical diversity than correction and question tweets.

Code	Total # Tweets	% RT	Lexical Diversity
All Codes	27,522	77.0%	555
Misinformation	19,024	80.0%	486
Speculation	2620	73.2%	442
Hedge	1222	70.9%	463
Question	546	49.8%	575
Correction	4110	70.6%	554

Table 6. Tweets Per Code for Falsely Accused Rumor

Domain Diversity and Propagation Over Time

Domain diversity was high for the Falsely Accused rumor as well. Though the rumor contained roughly a third of the total tweet volume compared to the Girl Running rumor, the number of domains with relatively high representation was significantly larger. Interestingly, many of the domains in the top ten most tweeted were social media sites, including Twitter, Reddit, Imgur, rt.com, YouTube, Twitchy, and the Tumblr account of celebrity Perez Hilton. This domain distribution differs sharply in slope from the first two rumors, and drastically in content from the third.

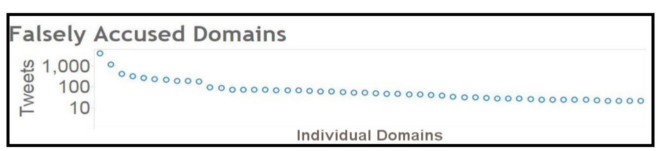

Figure 11. Domain Diversity, Falsely Accused Rumor

Domains were again strongly associated with either misinformation or correction, and Facebook.com was again a notable exception to this with a misinformation to correction ratio of about 2:1. YouTube.com had a somewhat balanced ratio at 5:1. NDTV.com, an Indian news site, represented the only correction-dominated domain in the top 10.

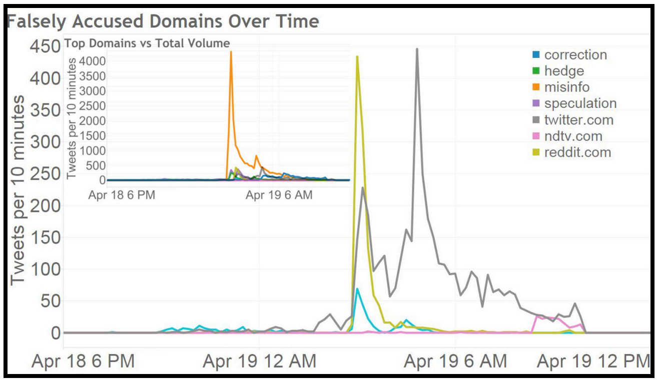

Figure 12. Domains Over Time for False Accused Rumor

Temporal analysis (see Figure 12) shows relatively similar parallels between domains and total tweet volume, though as with top domain analysis, the temporal shape differs to reflect the rumors' dissimilar origins. The Reddit.com domain group increases sharply first and is only later replaced by Twitter as the conversation migrated across platforms. Tweets that fall within the Twitter domain group again rise with misinformation, but notably less intensely after peak total misinformation. Aligned with patterns for correction-heavy domains in two previous rumors, the NDTV domain rises and falls with correction and peaks at nearly the same time as maximum correction.

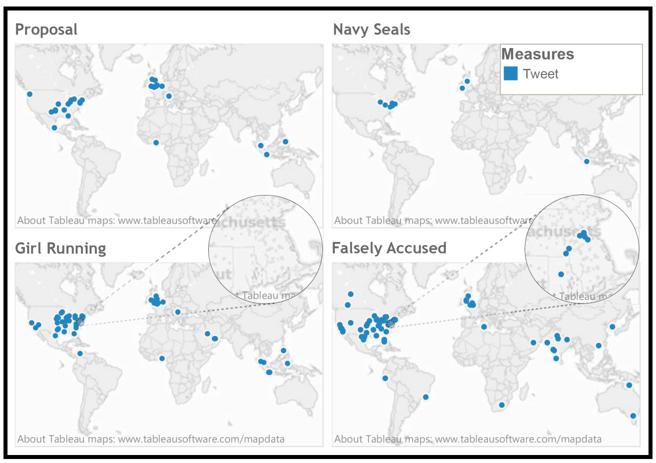

Figure 13. GPS Locations for Geo-located Tweets by Rumor

Geolocation and Rumor Propagation

The Falsely Accused rumor also stands out from the other rumors when we examine the GPS coordinates in geo-located tweets within the set, though the percentages of geo-located tweets are very small (<1% for each rumor). Figure 13 shows the GPS location of each geo-located tweet in each rumor. Interestingly, none of the rumors contain more than a single tweet posted from the Boston area (within 50-mile radius of the city's central point) *except* for the Falsely Accused rumor (8 tweets). The difference between Falsely Accused and Rumors #1 and #2 in the distribution of local vs not-local tweets is statistically significant (Chi Square, $p<0.05$). Though the geo-data is too sparse to draw strong conclusions, this analysis suggests that Rumor #4 was locally relevant in a way that the other rumors were not. Additionally, this rumor is also the only one with tweets geolocated to India, likely due to the fact that Tripathi was of Indian heritage.

DISCUSSION: MULTI-DIMENSIONAL SIGNATURES OF ONLINE RUMORS

This paper presents an in-depth examination of rumoring behavior, providing several empirical insights into how misinformation spreads via social media after disaster events and demonstrating how certain visual and quantitative measures reflect the qualitative "story" of a rumor. This work also documents a novel method for doing mixed-method research on information diffusion through social media, using multi-dimensional *signatures* to characterize large volumes of tweets related to a topic.

The concept of signatures has been introduced briefly in previous work [21,27], but here we explore the concept more deeply and discuss how it can be operationalized.

Signatures are a conceptual metaphor for understanding information diffusion online. A signature is a distinctive *representation* of information diffusion—in this case, a rumor spreading through Twitter—that allows us to quantitatively describe phenomena that are complex, multi-dimensional, and (we argue) largely qualitative.

Identifying Signature Dimensions for Characterizing the Spread of Rumors

Signatures consist of multiple *dimensions* that help to describe and represent salient features of the data. In this work, we explored five dimensions of rumor propagation: temporal progression of rumor spreading behavior (within individual tweets), URL domain diversity, domains over time, lexical diversity, and geolocation information.

Temporal graphs of tweet volumes over time clearly demonstrate (visually) the concept of a *signature*. Kwon et al. [15] connect temporal properties of rumors to rumor veracity, though they claim false rumors contain more repeating spikes over time, a pattern that is not consistent in the crisis-specific rumors analyzed in this paper. Examining temporal interactions across rumor behavior codes in additional to volume, as we do here, allows for additional insight into a rumor's origin and propagation.

Domain volume and diversity are markers of the influence of outside sources on rumor propagation. Like temporal signatures by code, peaks in a domain's temporal signature potentially demonstrate critical moments in rumor development when news, social media, or other external sources alter the content or course of the rumor. The shape of the domain diversity curve indicates interaction between tweets and outside sources.

Lexical diversity in textual communication has been identified as a marker of truthfulness [29]. Our research suggests that lexical diversity is also an important feature of rumor propagation, providing a measure of how much a rumor changed over time, and possibly reflecting the number of unique messages and voices that participated in its spread. Lexical diversity appears to correlate with different kinds of rumor spreading behavior—e.g. speculation has a higher lexical diversity than misinformation. Combining these observations, rumors with a low lexical diversity spread virally without substantial content variation, while rumors with high lexical diversity are more "conversational" and seem to more fully engage the crowd in collaborative sensemaking.

The geographic patterns of tweets may be another important feature in a rumor's signature. Though our geo-location data was sparse, we could see patterns suggesting Rumor #4 was more "locally relevant" than the other rumors. Using other techniques to increase the proportion of geo-located data [e.g. 6] could increase the utility of this feature.

Using Signatures for Comparing Rumors

Like a hand-written signature, a data signature is distinctive and recognizable. Signatures can therefore be used as tools for representing, communicating and identifying something that is fairly complex, such as patterns over large volumes of data. Signatures can also be operationalized as a tool for comparative studies. For example, some rumors may have similar signatures that can belong to specific rumor types. This research shows strong support for at least one rumor type—the *Internet Meme* rumor.

Internet Meme Signatures

The Girl Running and the Proposal rumors developed and propagated in very similar ways, and their signatures have many similarities. Both began as misinformation, and started to diffuse widely with the addition of a photograph. Both had very weak corrections when compared with the overall signal of misinformation. They also had relatively low measures of lexical diversity; they were spread by many authors, but through very similar content. We believe that these two rumors show a possible common "type" of rumor propagating through social media.

A Conspiracy Theory Signature?

The other two rumors were quite distinct from the first two and from each other. The Navy Seals rumor was largely speculative, beginning with and persisting primarily as speculation. Like the "false" rumor pattern in the Kwon et al. study [15], the Navy Seals rumor's signal pulsed repeatedly, often in tandem with links to a specific domain that was promoting this theory of the event. Much of the information related to this rumor came from outside of Twitter. Both domain diversity and lexical diversity were much higher than the Internet meme rumors, as participants in the conversation added their own evidence and explanation. This rumor may be indicative of a conspiracy theory type rumor, but analysis of more rumors will be needed to establish this as a type.

A Collective Sensemaking Signature?

Though similar to the Navy Seals rumor in some ways, the Falsely Accused rumor likely represents another type of rumor signature. This rumor also began with speculation, as members of the crowd attempted to make sense of available information and identify the bombers. Later, the rumor shifted from speculation to misinformation, as downstream individuals passed along early theories as fact. Eventually, for this rumor (and only this rumor), correction caught up and surpassed misinformation, demonstrating the concept of the self-correcting crowd [11,20]. Lexical and domain diversity were both high, indicating wide participation as opposed to simple amplification. Notably this was the only rumor in the set, according to geo-location data in the tweets, that resonated among accounts in the Boston area. This rumor may be part of a larger category of collective sensemaking signatures, though again future research is needed to confirm this type.

Using Multi-Dimensional Signatures as a Method for Analyzing information Diffusion

In this research, we also explore how multi-dimensional signatures can be used as a structuring construct to guide a method of analysis that deeply integrates qualitative and quantitative methods.

Here, we utilize the concept of signatures as a tool for recursive data analysis. These signatures are central components of a repeating process of data representation, observation and reflection, employed to iteratively make sense of the complexity of high-volume information diffusion. For each rumor, we generated a signature (feature by feature), often creating a visualization or some other quantitative representation. These representations were not final results, but opportunities to return to the data with new questions. For example, visualizations of tweets over time revealed peaks and troughs that often aligned with critical moments in rumor development. Subsequent qualitative analysis at those "points of interest" provided insight into how a rumor formed, evolved, and spread.

Throughout this process, qualitative findings provided insight into which features to measure and what representations to create. For example, initial qualitative coding led us to believe that lexical diversity would be an important feature for distinguishing between corrections and misinformation, and possibly for characterizing different types of rumors. This approach aligns with Starbird and Palen's [28] work that used in-depth qualitative analysis to identify important quantitative features of information diffusion.

Each quantitative representation is therefore both an outcome of qualitative research (to identify the dimensions) and an instrument for qualitative research. As the features are identified and the dimensions represented, the signature then becomes a tool for conceptualizing patterns across large amounts of data—for both understanding and communicating how the rumor formed and spread.

Demonstrating and Documenting an Emerging Method

Qualitative analysis of large amounts of Twitter data is hard to perform and perhaps even harder to communicate. The process is largely iterative, moving from data exploration; to sampling; to iterative rounds of qualitative analysis (including open coding of small data samples and mass coding of large swaths of data); to creating quantitative and visual representations of this coding; to reflecting on those representations; and perhaps returning to data exploration with new questions. This paper demonstrates how multi-dimensional data signatures can be tools for structuring and communicating this emerging method of analysis.

FUTURE WORK

This study is part of a larger research effort that aims to improve techniques for automatically identifying and classifying rumors propagating on Twitter. The current study identifies several features that may be useful in distinguishing between true and false rumors. In future work, we intend to leverage this concept of signatures, in combination with our growing understanding of how and why rumors propagate, to develop methods of categorizing different types of rumors. The three rumor types identified here are likely part of a larger set, and future research will expand and refine this list of prototypical signatures as we converge on the most useful set of categorizations.

CONCLUSION

In this paper, we integrate qualitative, quantitative, and visual analysis to examine the spread of four rumors propagating via Twitter after the 2013 Boston Marathon bombings. We identify five quantitative features that reflect salient aspects of the qualitative story of each rumor, including temporal progression of rumor behavior codes, URL domain diversity, domain propagation over time, lexical diversity, and geolocation information. These features constitute multi-dimensional *signatures* that can be used to characterize different types of rumors. In constructing these signatures, this research demonstrates and documents an emerging method for deeply and recursively integrating qualitative and quantitative methods for analysis of social media trace data.

ACKNOWLEDGMENTS

This research is a collaboration between the emCOMP Lab and SoMe Lab at UW and was partially supported by NSF Grants 1342252 and 1243170. The authors thank Shawn Walker for his efforts to transform the original Twitter data into datasets that were more easily accessed and analyzed.

REFERENCES

1. Allport, G. W. & Postman, L. An analysis of rumor. *Public Opinion Quarterly 10*, 4 (1946), 501-517.
2. Berger, J. Arousal increases social transmission of information. *Psychological Science 22*, 7 (2011), 891-893.
3. Bordia, P., & DiFonzo, N. Problem solving in social interactions on the Internet: Rumor as social cognition. *Social Psychology Quarterly*, *67*, 1 (2004), 33-49.
4. Budak, C., Agrawal, D., & El Abbadi, A. Limiting the spread of misinformation in social networks. In *Proc. of WWW 2011*, ACM, 665-674.
5. Castillo, C., Mendoza, M., & Poblete, B. Information credibility on twitter. *WWW 2011*, ACM, 675-684.
6. Cheng, Z., Caverlee, J., & Lee, K. You are where you tweet: A content-based approach to geo-locating twitter users. *Proc. of Conference on Information and Knowledge Management, CIKM 2010*, 759-768. ACM.
7. Chierichetti, F., Lattanzi, S., & Panconesi, A. Rumor spreading in social networks. In 36th *Intl. Colloquium on Automata, Languages and Programming (ICALP*), pp. 375–386. Springer, 2009.
8. DiFonzo, N, & Bordia, P. *Rumor psychology: Social and organizational approaches*. American Psychological Association, 2007.
9. Friggeri, A., Adamic, L. A., Eckles, D., & Cheng, J. Rumor Cascades. *AAAI, 2014*.
10. Gantz, W., & Trenholm, S. Why people pass on news: Motivations for diffusion. *Journalism & Mass Communication Quarterly 56*, 2 (1979), 365-370.
11. Herrman, J. Twitter is a Truth Machine. Blog in *Buzzfeed*. (October, 2012). Available at: http://gofwd.tumblr.com/post/34623466723/twitter-is-a-truth-machine
12. Hill, K. Hurricane Sandy, @ComfortablySmug, and the Flood of Social Media Misinformation. *Forbes.com.* (October 30, 2012). Available at: http://www.forbes .com/sites/kashmirhill/2012/10/30/hurricane-sandy-and-the-flood-of-social-media-misinformation/
13. Hiltz, S. R., Kushma, J., & Plotnick, L. Use of social media by US public sector emergency managers: Barriers and wish lists. *Proc. of ISCRAM, 2014*.
14. Hughes, A. L., & Palen, L. The evolving role of the public information officer: An examination of social media in emergency management. *J. of Homeland Security & Emergency Management 9*, 1 (2012), 1-20.
15. Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. Prominent Features of Rumor Propagation in Online Social Media. In *IEEE 13th International Conference on Data Mining*, *ICDM 2013*, 1103-1108.
16. Leberecht, T. Twitter Grows Up in Aftermath of Haiti Earthquake, *CNET News*, (January 19, 2010) http://www.cnet.com/news/twitter-grows-up-in-aftermath-of-haiti-earthquake/

17. Leonard, H. B., Cole, C. M., Howitt, A. M., & Heymann, P. B. Why was Boston strong? Lessons from the Boston Marathon Bombing. Program on Crisis Leadership and Program in Criminal Justice Policy and Management, Harvard Kennedy School, April 3, 2014.
18. Madrigal, A. Hey Reddit, Enough Boston Bombing Vigilantism. The Atlantic. (April 17, 2013). http://www.theatlantic.com/technology/archive/2013/04/hey-reddit-enough-boston-bombing-vigilantism/275062/
19. Madrigal, A. #BostonBombing: The Anatomy of a Misinformation Disaster. *The Atlantic* (2013, April 19) http://www.theatlantic.com/technology/archive/2013/04/-bostonbombing-the-anatomy-of-a-misinformation-disaster/275155/
20. Mendoza, M., Poblete, B., & Castillo, C. Twitter under crisis: can we trust what we RT?. *Proc. of 1st Workshop on Social Media Analytics 2010*, ACM, 71-79.
21. Nahon, K., Hemsley, J., Mason, R. M., Walker, S., & Eckert, J. Information flows in events of political unrest. *Proc. of iConference 2013*, 480-485.
22. Oh, O., Agrawal , M., & Rao , H. (2013). Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. *MIS Quarterly, 37*, 2 (2013), 407-426.
23. Palen, L, Anderson, K.M, Mark, G, Martin, J, Sicker, D, Palmer, M., and Grunwald, D. A Vision for Technology-Mediated Support for Public Participation & Assistance in Mass Emergencies & Disasters, in *Proc. of ACM-BCS Visions of Computer Science* (2010), 1-12.
24. Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q. Rumor has it: Identifying misinformation in microblogs. In *Proc. of the Conference on Empirical Methods in Natural Language Processing*, (2011), 1589-1599.
25. Shibutani, T. *Improvised news: A sociological study of rumor*. Ardent Media, 1966.
26. Spiro, E. S., Fitzhugh, S., Sutton, J., Pierski, N., Greczek, M., & Butts, C. T. Rumoring during extreme events: A case study of Deepwater Horizon 2010. *Proc. of ACM Web Science Conference 2012*, ACM, 275-283
27. Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, RM. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. *iConference 2014*, 654-662.
28. Starbird, K. & Palen, L. (How) Will the Revolution be Retweeted? Information Propagation in the 2011 Egyptian Uprising. In *Proc. of CSCW 2012*, 6-17.
29. Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision & Negotiation*, 13(1), (2004), 81-106.