Expressed Uncertainty and Denials as Signals of Online Rumoring

KATE STARBIRD, EMMA S. SPIRO, AHMER ARIF, FANG-JU CHOU, SINDHUJA NARASIMHAN, JIM MADDOCK, KELLEY SHANAHAN, AND JOHN ROBINSON UNIVERSITY OF WASHINGTON

1. INTRODUCTION

Social media platforms are important venues for communication during crisis events; both the public and emergency responders use online platforms to discover and curate event-related information. While many recognize the value of social media as an information source during crisis settings, there is also concern about misinformation [Sutton et al. 2008, Hughes and Palen 2012]. To ground our exploration of misinformation within online spaces, we leverage theories of rumoring behavior—a topic of longstanding interest in the social sciences [Caplow 1947, Allport and Postman 1947, Spiro et al. 2012]. Prior studies of rumoring, both during disasters and under more general circumstances, frame the "collective sensemaking" process that often occurs as a response to uncertainty within the information space. Such rumoring behavior often occurs during periods of collective stress, such as disasters, and in cases where official sources of news are unavailable (or insufficiently timely) [Shibutani 1966]. We hope that understanding the collective sensemaking process will inform automated methods of detecting rumors on social media platforms during crisis events.

1.1 Project Goals: Using Crowd Activity to Detect Misinformation

Our research aims to leverage the "collective intelligence" of the online crowd to identify rumors. Vieweg et al. [2008] and Starbird [2013] used the term "collective intelligence" to characterize the online crowd's work to identify, verify and amplify actionable information during crises. Starbird [2013] demonstrated how a network of social media users performed as a collective sensor for "good" information after a major earthquake. This research extends the view of collective intelligence in online crowds, examining how participation in conversations around crisis events—the collective sensemaking process that functions to both spread and correct rumors—can be analyzed at the network level to detect and classify rumors.

We propose using multi-dimensional "signals" or patterns of information propagation [Maddock et al. 2015] to detect and classify rumors in real-time [Starbird et al. 2014]. This builds upon insight from Mendoza et al. [2010] who proposed the possiblity of using linguistic markers of denials (where the crowd challenges or attacks misinformation) to identify rumors. We explore an additional component of rumoring behavior in online conversations—expressed uncertainty. Sociological studies of rumor show that sensemaking (after disaster events) is a response to uncertainty in the information space. Our research suggests that looking for uncertainty expressed by the crowd in relation to specific topics provides another mechanism (similar to identifying denials or corrections) for identifying rumors.

2. METHODS

We plan to analyze many different rumors across multiple events. Our initial study [Maddock et al. 2015] looked at six rumors spreading after the Boston Marathon Bombings. Here, we analyze and discuss two rumors from another crisis event—the December 2014 Syndey Siege.

2.1   Event Background and Data Collection

On 15 December 2014 at 9:45 AM AEDT a gunman took 17 people hostage in the Lindt Chocolate Cafe at Martin Place in central Sydney. The 16-hour standoff ended when police stormed the building. Tragically, two hostages and the gunman were killed. The crisis included many high-profile sub-events: local police cordoned off the surrounding area and attempted to negotiate, the gunman contacted media outlets with demands, five hostages managed to escape, and hostages were coerced into holding up a flag that was misidentified as belonging to the Islamic State of Iraq and the Levant (ISIL). We collected data on this event for the explicit purpose of examining rumoring behavior, using the Twitter Streaming API to track several event-related terms, including: `sydneysiege`, `martinplace`, `sydney`, `lindt`, and `chocolate shop`. We started the collection on 15 December at 11:06am AEDT and ended it two weeks later, capturing a total 5,429,345 tweets.

2.2   Identifying Rumors in the Tweet Corpus

We began the rumor identification process after the Siege ended, leveraging visual pattern analysis in conjunction with examination of external sources (i.e. media reports) to identify rumors. We then utilized an iterative process to develop and refine search strings to generate a comprehensive, low noise tweet corpus for each rumor. The Sydney Siege has yielded eight salient rumors that meet our criteria for coding. In this paper, we discuss the first two rumors that we have completely coded:

The Ray Hadley Rumor: This true rumor stated that an Australian radio host, Ray Hadley, spoke off-air to a hostage during the Sydney Siege. Hadley himself started this rumor.

The Lakemba Raids Rumor: This false rumor asserted that the Australian Federal Police carried out home raids in Lakemba, a Sydney suburb, in parallel with the siege. These claims were denied by authorities. The rumor may have originated from the sighting of twenty officers conducting a pre-arranged tour of the Lakemba Mosque as part of a police induction day.

2.3   Coding Scheme and Process

To capture aspects of rumoring activity, we code tweets associated with each rumor along two dimensions. The first, designed to identify crowd corrections, consists of five mutually exclusive categories: *Affirm, Deny, Neutral, Unrelated,* and *Uncodable.* The second coding dimension captures another property of rumoring and sensemaking behavior—expressed uncertainty. Tweets with *Uncertainty* suggest in some way that the veracity of the rumor is not completely established:

```
Anyone in Lakemba that can confirm police raids in the area? #Australia
#Sydney #Sydneysiege
```

We have three trained coders manually code every distinct tweet (removing retweets and very close matches) in each rumor corpus. We use a "majority rules" process for adjudication where agreement by two or more coders determines the final code. Table 1 shows the number of tweets in each rumor assigned to *Affirm*, *Deny* and *Uncertainty* codes.

Rumor	Total	Affirm	Deny	Other	Uncertainty
Ray Hadley talks to Hostages	2683	2576	30	77	234
Lakemba Raids	1382	514	820	48	6

Table 1. Number of tweets in each rumor with each code

3. PRELIMINARY FINDINGS

Figure 1 shows the temporal signature (as in Starbird et al. 2014) — the volume of tweets over time, by code — for each rumor. Examining this data illuminates how the crowd engaged differently with the two rumors.
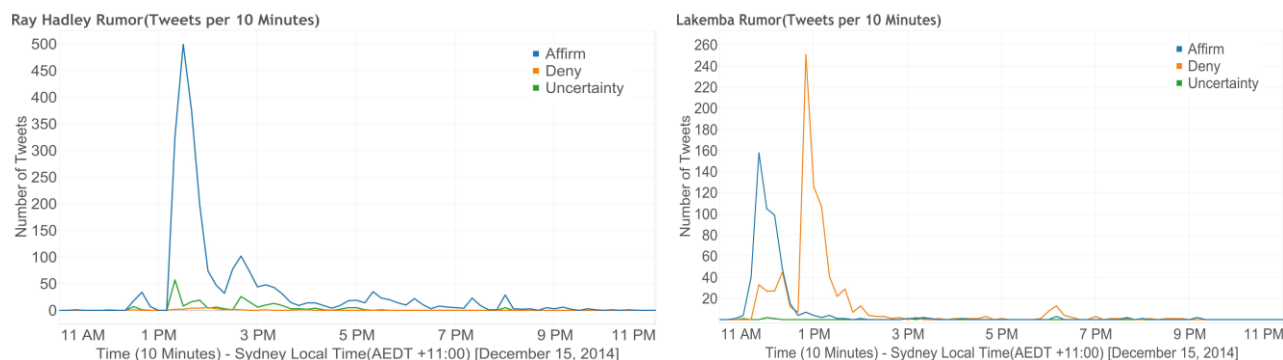

Fig. 1. Tweet volume over time for Hadley Hostage Rumor (left) and Lakemba Raids Rumor (right).

In the **Ray Hadley** rumor, affirmations peak about an hour after this rumor begins. Whilst there are very few denials of this true rumor, the crowd did express uncertainty about the information (at a rate of about 1 uncertainty per 10 affirmations). This uncertainty often took the form of skepticism related to Hadley's reputation as an attention-seeking, "shock jock" radio host:

```
@<mention> There's a bit going on with Hadley's alleged contact with a
hostage. Just don't know how reliable it is
```

The **Lakemba Raids** shows an initial wave of affirmations followed by a strong denial. This is the first time in our research where the volume of denials exceed affirmations of the rumor. Interestingly, though an initial denial wave accompanied the original rumor (between 11am and 1pm), the stronger denial arose after the rumor had faded, representative in this case of public reprimand for rumor spreading. The crowd expressed very little uncertainty about this rumor.

Although these preliminary findings are not necessarily generalizable across rumor types and events, they both extend the scope of online rumoring and sense-making behavior, and suggest new avenues for automatic detection and classification:

(1) True vs. False Rumoring. Previous research predominantly investigates false stories, which are generated intentionally or from the information scarcity that accompanies disasters. As the Hadley rumor demonstrates, similar characteristics of rumoring behavior e.g. affirmations, denials, and uncertainty, could also apply to true narratives, but may generate different signatures.

(2) Affirmations vs. Denials. For the first time, we identify a false rumor in which the total volume of denials exceeds the volume of affirmations. This finding corroborates previous research by Mendoza et al. [2010], and warrants further analysis to help explain its unique signatures.

(3) Low vs. High Uncertainty. Development of our new coding scheme specifically prioritizes uncertainty as an important dimension in sensemaking behavior. Notably, uncertainty can be present in both affirmations and denials, as well as true and false rumors. Uncertainty signals may align with specific types of rumors and make help distinguish between them—e.g. a purposefully fake news story (likely low) and a conspiracy theory (likely high uncertainty).

REFERENCES

Gordon W. Allport and Leo Postman. 1947. The Psychology of Rumor. New York, NY: Henry Holt.

Theodore Caplow. 1947. Rumors in War. *Social Forces*. 25, 3 (1947), 298-302. DOI:http://dx.doi.org/10.2307/3005668

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web (WWW'11)*. ACM, 675-684. DOI:http://dx.doi.org/10.1145/1963405.1963500

Amanda L. Hughes, and Leysia Palen. 2012. The evolving role of the public information officer: An examination of social media in emergency management. *Journal of Homeland Security and Emergency Management*. 9, 1 (2012), 22. DOI:http://dx.doi.org/10.1515/1547-7355.1976

Jim Maddock, Kate Starbird, Haneen Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. 2015. Characterizing online rumoring behavior using multi-dimensional signatures. To appear in *Proceedings of the 2015 Conference on Computer Supported Cooperative Work (CSCW'15)*. ACM, Vancouver, BC. DOI:http://dx.doi.org/10.1145/2675133.2675280

Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under crisis: Can we trust what we RT? In *Proceedings of 1st Workshop on Social Media Analytics (SOMA'10)*. ACM, New York, NY, 71-79. DOI:http://dx.doi.org/10.1145/1964858.1964869

Onook Oh, Manish Agarwal, and Raghav Rao. 2013. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*. 37, 2 (2013), 407-426.

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

Tamotsu Shibutani. 1966. Improvised news: A sociological study of rumor. Indianapolis, IN: Bobbs-Merrill.

Emma S. Spiro, Jeannette Sutton, Matt Greczek, Sean Fitzhugh, Nicole Pierski, and Carter T. Butts. 2012. Rumoring during extreme events: A case study of Deepwater Horizon 2010. In *Proceedings of the Web Science 2012 Conference (WebSci'12)*. ACM, Evanston, IL, 275-283. DOI:http://dx.doi.org/10.1145/2380718.2380754

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* (CHI'13). ACM, 801-810. DOI:http://dx.doi.org/10.1145/2470654.2470769

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. In *Proceedings of iConference 2014*. DOI:http://dx.doi.org/10.9776/14308

Jeannette Sutton, Leysia Palen, and Irina Shklovski. 2008. Backchannels on the front lines: Emergent use of social media in the 2007 Southern California wildfires. In *Proceedings of the 5th International Information Systems for Crisis Response and Management Conference (ISCRAM'08)*. Washington, DC, 624-632.

Sarah Vieweg, Leysia Palen, Sophia B. Liu, Amanda L. Hughes, and Jeannette Sutton. 2008. Collective intelligence in disaster: Examination of the phenomenon in the aftermath of the 2007 Virginia Tech shootings. In *Proceedings of the 5th International Information Systems for Crisis Response and Management Conference (ISCRAM'08)*. Washington, DC, 44-54.