An Explosion Through Twitter:
False rumour correction through social networks during the West, Texas fertilizer plant incident

By April Diver

Submitted to the Faculty of Extension
University of Alberta

In partial fulfillment of the requirements
for the degree of Master of Arts in
Communications and Technology

Aug 1, 2014

Table of Contents

"Digital networks have acted as a massive positive supply shock to the cost and spread of information, to the ease and range of public speech by citizens, and to the speed and scale of group coordination. […] These changes do not allow otherwise uncommitted groups to take effective action. They do, however, allow committed groups to play by new rules." – Clay Shirky (2011)

Acknowledgements

This study is dedicated to the first responders of the West, Texas incident many of whom perished while protecting their community on April 17, 2013.

This study would not have been possible without the support of my project supervisor Dr. Gordon Gow, my professors from the U of A Communications and Technology Graduate Program, and my cohort colleagues. Thanks also to Waldemar and Kevin who continue to amaze me with their ability to manipulate technology in the fight against Twitter for free research data. Thanks finally to my family and more specifically my husband who kept me on track right to the end.

Abstract

Social media is an essential communication network for the public during emergencies; however emergency managers have expressed significant concerns over the use of this form of communication by the public in propagating rumours and misinformation during emergency response and recovery. This study assesses false rumour correction in Peer-to-Peer networks during crisis by specifically exploring the context of Twitter use in the West, Texas Fertilizer Plant explosion on April 17, 2013. Three false rumours are examined as propagated by members of the public, media and other organizations through Twitter. This research examines the context and content of tweets to derive information about false rumour correction in digital, peer-to-peer networks. The aim of this study is to contribute to a better understanding of the conditions by which false rumours are corrected. These conditions could provide information about what actions emergency managers could take to encourage accuracy and collective correction in backchannel communication. The research design uses qualitative content analysis in two phases: tweets are categorized within three false rumour streams that occurred in the West, Texas explosion, and tweets which correct or question the information are coded further to identify additional characteristics. The design also explores the content of specific correcting tweets to identify any patterns or observations worth noting. The findings of this study suggest that misinformation in the early stages of a disaster does correct in some instances but not with sufficient frequency or spread to ensure that accurate emergency information is shared within the online community.

Introduction

Social media has become an essential communication network for the public during emergencies. Emergency managers and other official organizations involved in the emergency response should leverage and perhaps even participate in social media, however they have expressed significant concerns over the accuracy of these channels. When information goes viral and can no longer be managed by the authorities, emergency managers may be reliant on the public to pass along good information. Furthermore, emergency responders may have insufficient resources to actively and consistently participate in informal discussions and social media. However, when disaster information is more accurate, the burden on emergency responders and other authorities to correct misleading information can be reduced; emergency authorities who are not over-invested in correcting the message can instead focus on response and recovery to the emergency. Research is therefore required to understand the conditions that foster an accurate social media environment during a significant emergency.

This research assessed false rumour correction in Peer-to-Peer networks during crisis. Both the terms ‘misinformation’ and ‘false rumour’ are used in this study. Misinformation describes the content of a false rumour, where as ‘rumour’ implies the act where the misinformation is shared.

This research examined a number of potential contributing factors to the accuracy and self-correction of Twitter by examining the context of Twitter use in the West, Texas Fertilizer Plant explosion on April 17, 2013. At 7:29 PM Firefighters are dispatched to respond to a fire at the West Fertilizer Co. in West, Texas. At 7:53 PM a fertilizer tank

on scene holding ammonium nitrate explodes, killing fifteen people and injuring more than 300 (Dallas Morning News, n.d.).

This research intended to derive the factors that encourage people to correct information in Twitter. Three false rumours were explored as propagated by members of the public, media and other organizations through Twitter using qualitative content analysis. It is hoped that this research could then be used to identify methods and conditions that emergency managers could replicate to encourage accuracy and collective correction in social media networks, also referred to as the backchannel to officially released and vetted emergency communication. This research explored how social correction and accuracy occurred in Twitter. Specifically it considered if emergency responder organizations were invested in social media relations, and whether these subject matter experts played a substantial role in correcting errors. Alternatively the research also considered the role of traditional media to accurately report the facts in the many-to-many conversation of Twitter. Most importantly, it considered whether the informal Twitter community worked to resolve misunderstandings and rumours as a collective and under what conditions. Could information through Twitter be made more accurate without officials or experts directly participating in the conversation? These conditions may be valuable for emergency managers to replicate. Ideally the research might reveal resources and conditions (beyond the application of dedicated social media personnel) that could be invested into informal communication network management during future events to reduce or correct false rumours.

Research Questions

Specifically this research explored the following questions:

- Under what conditions can Twitter communication by non-experts self-correct?
- Do users leverage traditional media as a reference to correct rumours? (Do they perceive traditional media as a trusted source?)
- Will people in informal social networks work collectively to identify, share, and propagate corrected information?

Literature Review

To explore information correction behavior in peer-to-peer networks during disasters, it was important to first consider existing research and perspectives on the factors that may play a role in these research questions. This includes various rumour theories and models themes such as the interdependency and shift of traditional media towards new media, the role and content of peer to peer communication and specifically micro-blogs in disasters, the challenge of managing the message particularly in authority-to-public crisis communication, and community efficacy and collective information evaluation. Each of these themes were explored through historical and recent literature.

Communication Theory

Both the Social Exchange Theory and the Functional Theory may play a role in this project. Twitter is a communication channel in which individuals are motivated to communicate, but it also fosters group collaboration, decision-making, and action therefore both theories may be valuable. Scholars under the Social Exchange Theory believe that people communicate with each other through a context of rewards and costs. These contexts may include reciprocity, fairness, and negotiated rules. The rewards may include information approval, respect, power, group gain, and personal satisfaction (Littlejohn & Foss, 2011). The Functional Theory is alternatively a group decision-making theory, where group members share information, explore and identify errors in thinking, and persuade others. Scholars such as Dewey and Hirokawa state that faulty decisions stem from improper assessment, inappropriate goals and objectives, and inadequate information (Littlejohn & Foss, 2011). It will be important to consider how

individuals communicate with one another, as well as how the collective group decides what information is valid or needs to be corrected during an emergency situation.

Traditional and New Media: A New Relationship

The media, whether it is the traditional broadcast media of television and radio, which now includes websites of print and broadcast journalism (International Communication Association, 2011), or the more recent digital social media, plays a significant role in the experience and management of disasters. People use the media to learn about risks, compare their knowledge to the knowledge of those they trust, judge the accuracy of a message, and determine their involvement and what action to take (Vultee & Vultee, 2011). While traditional media such as television and newspapers continue to persist as disaster news sources, traditional media has typically been limited to provide one-way information dissemination. Alternatively, social media is built to enable two-way dialogue and interaction among organizations and individuals (Bortree & Seltzer, as cited by Fraustino, J. D., Liu, B. & Jin, Y., 2012).

Social media is defined as a two-way communication form where users may construct personal identities through self-presentation and dialogue, and create and collaborate in real-time. Social media platforms enable every user to be their own public relations manager through interactivity and information exchange (Smith, 2010). Interactive (social) media puts the audience into the news reporter's role. The audience's needs and habits now play a greater part in the communication cycle (Vultee & Vultee, 2011). Audiences can now follow both traditional media and citizen-reporters through an interactive medium, whereby they can participate, comment, accept, or reject openly the

information (International Communication Association, 2011). Where previously there were two different models of media -- public broadcasts by professionals, and private conversations between pairs of people – there is now a blended world whereby "the public and private media blend together, where professional and amateur production blur, and where voluntary public participation has moved from nonexistent to fundamental" (Shirky, 2010, para. 9).

A large portion of literature on social media and crises focuses on the increasing use of social media as an information source during disasters. Since 1995, the Internet has fostered a self-help network. It has played an increasing role in disseminating information about disasters; electronic networks provide access to maps, digitized photographs share a common picture of the damage, and friends and relatives can reach out to one another to provide status updates (Stephensen & Anderson, 1997). The first instance of Internet use in a disaster was during the Japan Kobe Earthquake in 1995, followed later that year by the Oklahoma City Bombings. Internet use during crisis has since evolved. Through a literature review, Fraustino, J. D., Liu, B. & Jin, Y. (2012) compared the use of social media across various events. During the September 11, 2001 terrorist attack in New York, U.S., television remained the dominant source of information and only 1% of people learned about the event through the Internet. Nearly four years later however, Hurricane Katrina in Southeast Louisana saw a dramatic increase in Internet and social media use. 73% of New Orleans residents visited online sites (media websites and municipal websites) for information, and more than half used the Internet to re-establish connections with their social network (Fraustino et al., 2012).

As social media gained popularity, traditional media have been criticized on several fronts. Firstly, they have been shown to contradict the messages provided by the experts, in this case the emergency managers. Mileti (1999) for example found that media weather predictions often conflicted with the warnings of official forecasters, and disaster details provided by traditional media were over-simplified and misleading specifically around the scale, location of damage, and disaster-related needs. More commonly however, the approach of traditional media outlets can intentionally or unintentionally introduce a bias into the message. Vultee & Vultee (2011) described the anecdotal framing as a method whereby the traditional media select and filter information for the consumer by focusing on stories that are likely to generate the most interest. This intentional bias is strategic, selecting the most interesting message and producing that message in ways that capture an audience.

According to Smith (2010), social media is more than a channel to communicate. He argues that user experience, risk, and motivations such as issue alliance and self-expression form a part of what is shared in new media, which influences what is shared by the citizen journalist. Citizen journalists, people without professional training, have the opportunity to broadcast using the modern tools of blogs and social media sites (Barnes, 2012). Along with the increased use of new media, there has been a notable interdependent relationship between traditional media and new media in crisis communication. Even traditional media outlets such as cable and broadcast news networks are more likely to use citizen generated digital content during the first stage of a crisis (Wigley & Fontenot, 2010). New media and new technologies are embedded with older technologies. Traditional media outlets have incorporated new media into their

business. New media complements the traditional ways of conceptualizing information sources. As both traditional and new media leverage authorities, peers, and each other, the lines between private and public broadcast become blurred (Barnes, 2012; Vultee & Vultee, 2011). *Huffington Post* is an example where citizen journalism (a user by the name of Adrianna Huffington and her friends) began by reporting mostly opinion. This outlet has now transformed into a mainstream source of entertainment, opinion and news (Barnes, 2012).

Emergency Management organizations continue to use a linear communication style. Emergency managers provide information during the initial stages of a disaster by using public warning systems, media announcements, and press conferences. This is described as linear – the message is developed by the experts (in this case the emergency managers), provided to a news media outlet, and then passed it along to the public. Using a linear communication style to disseminate crisis information is outmoded (Palen et al., 2007; International Communication Association, 2011). Social media now plays a more central role in all phases of a disaster, and is worth consideration by emergency management organizations.

Though the public continues to leverage traditional media reporting for their disaster information (and perhaps even validation), users may be moving away from a reliance on such sources. According to the International Communication Association (2011), even in back-channel communication many users still prefer to link back to traditional media sources, but a growing number of Twitter participants are alternatively linking to other social media sites as a source of information.

Micro-blogging in Disasters

Since 2006, micro-blogging has become an increasingly wide-used mode of communication on the internet (Bruns & Stieglitz, 2012). Micro-blog use during crises is also prevalent. In the Australian Floods of 2011, researchers found that several groups – affected locals, emergency services, mainstream media organizations – each participated to varying degrees in the micro-blog realm to provide and counter information (Bruns & Stieglitz, 2012).

As one particular example of micro-blogging, Twitter has provided a valuable environment from which to cultivate data about content and users. Though it may be a potentially confusing medium during emergencies, it is also a rich source of microblog content and context (Vultee & Vultee, 2011). Twitter, as one of the most popular micro-blogging sites, has adopted a variety of methods to maintain efficiency and categorize exchanges such as @reply for specific responses and "retweets" for repeated messages from another user (Bruns & Stieglitz, 2012). Twitter specifically allows "context collapse" (Marwick and Boyd, 2010, p.122), allowing the user to filter and sensor what they present to others, and how much background about themselves and their information they will share. Twitter users can impede and fulfill public relations efforts with little stake in organizational goals (Smith, 2010). Like other forms of media, its legitimacy is publicly defined – the retweeting of each other's posts provides legitimacy and basic credibility to the message (Smith, 2010).

Vultee and Vultee (2011) explored micro-blogs to identify users and messages types during disasters of varying severity. They examined various user groups and identified several categories to describe the nature of messages: personal, commentary,

providing or seeking information, and promotional. They contrasted these categories against the severity of the emergency (5 point ordinal scale) and the phase of the emergency within which they occurred (warning, impact, recovery, and mitigation). Vultee and Vultee found that personal messages accounted for the highest portion of Twitter messages – more than two thirds in the warning (67.9%), impact (70.1%) and recovery (86.4%) of the six disasters examined. Government Twitter messages accounted for the lowest percentage of messages and were described by the researchers as "rare" (Vultee & Vultee, 2011, p.231). Most importantly they concluded that micro-blog discourse seemed to be largely conversational, and suggest that media and disaster professionals should anticipate that the audience of Twitter - now adapted to speed, concision, and timeliness - can be expected to punish carelessness of authorities and traditional media (2011).

Managing the message

Emergency managers have expressed concern about the accuracy of information in social networks. Though Hiltz and Gonzalez (2012) do not comment on the frequency or prevalence of inaccurate information, their study suggests that emergency managers perceive that social media networks will provide inaccurate, irrelevant and insignificant information, which could risk tying up response efforts. Citizen-generated information such as social media messages may not be easily confirmed, especially when accounts in social media are unverifiable (Crowe, 2011). Furthermore the traditional media's use of citizen-generated content has also raised concerns with emergency managers, as citizen information bypasses traditional gatekeepers and is then passed off through "official" channels (Wigley & Fontenot, 2010). Wigley and Fontenot's study on the Virginia Tech

Shooting revealed that broadcast media reporters were much more likely to use non-official sources then official sources (70% to 30%) to retrieve information about the unfolding event. The participation official or organizationally affiliated users is low; very few organizations use social media to provide top down communications. In a recent study, less than 13% of formal emergency response organizations in the US reported using social media for crisis communications (Security Director Report, 2009).

Trust also plays a significant role in determining which sources are selected and followed by the audience. Trust, as opposed to accuracy may influence which messages – the layperson's or the expert's - bear more importance to the audience (Wynne as cited by Haynes, Barclay and Pidgeon, 2008). Haynes, Barclay and Pidgeon (2008) examined the role of trust and its influence on risk communication during a volcanic crisis. They found that the public viewed friends and relatives as the most trusted source for information, and high trust allowed for competing messages to reinforce a belief that the risk to life and limb was less than as described by official reports.

Emergency managers also cite other potential information issues including the timing and appropriateness of the message. Even when information is accurate either through careful reporting or social correction (where the crowd corrects the outliers), highly sensitive, inappropriate or traumatizing information may be shared by the public, which would not have been shared until later by official authorities (Oh et al., 2013; Hiltz & Gonzalez, 2012). Where the information quality is not appropriate, such as including details that cause unnecessary distress such as the number of missing, emergency managers may be concerned that the more essential safety messages are not being clearly received by those who need them (Hiltz & Gonzalez, 2012).

In cases where the public has been left to participate without boundaries, and due to the viral nature of such information, it is important that false information is corrected before it jeopardizes public safety and event recovery (Palen, Hiltz & Liu, 2007; Security Director's Report, 2009). False rumours about looting for example have had public safety consequences in several events including Sept 11, 2001, and Hurricane Elena in Pinellas County, Florida, 1985 (O'Leary, 2004). In some instances people refused to evacuate, instead remaining in the hazardous area to protect their property, and in other instances overzealous police and security guards prevented the entry of legitimate disaster-response personnel into the disaster zone.

It is important to consider that not all emergency management organizations are heavily involved in social media. Crowe (2010) suggests that emergency management organizations tend to fall into one of three roles with social media communication: proactive – the organization will monitor, post, discuss issues; reactive – the organization will either monitor or post but not both, and will not actively engage the audience; or inactive – the organization has no role in social media.

In cases of community disasters, emergency responders need to make extra efforts to distribute reliable information and, at the same time, control collective anxiety in the community to suppress rumor spread (Oh, et al., 2013). However, the time, dedication, and level of involvement that is required for authorities to correct information may not be attainable. Emergency management agencies are often overwhelmed by the process and degree of dedication required to manage a robust and proactive social media presence (Hughes & Palen, 2012). Furthermore though some organizations are able to establish a full command and control structure whereby only information flows to the public from a

dedicated communications unit, many smaller emergency management organizations rely on volunteers, and dedicated communicator roles which are frequently under-resourced and over-tasked, creating a significant gap in the official message and the informal discussion (Crowe, 2011; Hughes & Palen, 2012; Eburns, n.d.). Palen, Hiltz, and Liu (2007) suggest that information generated from the bottom-up without validation cannot be assured to be accurate or appropriate for the audience therefore explicit attention should be paid to managing any false information before it impacts the emergency response.

Rumours: Models and Theories

Rumours are defined as unsubstantiated or unverified information aimed to provide a fact usually about issues, people or events of major significance to a group (Bordia & DiFonzo, 2004; Zhao, Cui, Qiu, Wang & Wang, 2013). For the purpose of this research, rumours were treated as a social phenomenon, a method to which information and misinformation was carried before it is accepted or denied. The causes of rumours in early research focus cognitive processing. Research in the 1920s and 1930s introduced three phases to rumour propagation – leveling, where the narrative becomes notably shorter and 70% of the details are lost; sharpening, where details are selected and filtered; and assimilation, where the rumour is distorted by our motivations (Bordia & DiFonzo, 2004). Not surprisingly this research emphasizes a reliance on cognitive memory, as it pre-dates computer-mediated communication where memory has since become less of an influence on rumour.

More recent pre-ICT (Information Communication Technology) research during the 1970's through to the 1990's focused largely on motivation. Groups who were introduced to uncertain situations requiring interpretation and understanding were motivated to pool knowledge to determine a solution through the rumour process (Bordia & DiFonzo, 2004). Other motivations that affect the transmission of rumours include social perception. For example, there is hesitancy to transmit an upsetting but implausible rumour as this may be associated with "crying wolf" (Rosnow as cited by Pezzo & Beckstead, 2006). On the other hand, an upsetting but plausible rumour may be transmitted because the sender desires to inform others, particularly friends and family, of impending danger and/or to elicit information from others (Pezzo & Beckstead, 2006).

Generally motivational and cognitive rumour research identified two main types of rumours: wish rumors - where hope is expressed; and dread rumours – those that express fearsome or undesirable events or consequences (Rosnow as cited Bordia & DiFonzo, 2004). These types are still considered in some post-ICT research. Using content analysis, Bordia and DiFonzo (2004) examined rumour transmission to find that the proportion of prudent statements (cautionary statements such as "I am not sure if this is true but…") in wish rumours was higher than in dread rumours. They suggested this may indicate that the penalties for spreading dread rumours may be more severe than the penalty for falsely raising hopes.

Rumours as fleeting, temporal phenomena were historically difficult to trace, however this has changed dramatically with the advent of widespread computer-mediated communication (Bordia & DiFonzo, 2004). Rumours, now recorded through text, picture and audio, can now be stored and retrieved. Furthermore ICTs have allowed anyone to

create, modify and share content, therefore information and rumours can spread faster and further than before (Zhao, Cui et al., 2013). Despite these changes, it is still valuable to consider previous models of rumour spread as a framework to understanding rumour propagation. Early models of rumours identified three main roles for participants: 'stiflers' (those who receive the rumour but choose not to spread it), 'spreaders' (those who spread the rumour), and 'ignorants' (those who are unaware of the rumour). With the introduction of ICTs and the large-scale networks, more recent models maintain these roles but also include ideologies of complex networks where participants in rumours may be 'susceptible', 'infected', and 'recovered' (Zhao, Cui et al., 2013).

The exploration of rumours as transmitted through social networks has become dramatically easier to map with digital technology. Zhang, Zhou, Zhang, Guan and Zhou (2013) found rumours evolve through the spreading process and may even become shorter, more concise, and more easily grasped as they move through a network. Rumours can infect the whole network, but this depends largely on the existence of connections among individuals as well as the individuals' strategies and motivations (Zhang, Zhou et al., 2013). In small networks, the majority of individuals are infected by multiple versions of the rumour. In a large network, the original rumour rather than the multiple versions, is predominant (Zhang, Zhou, et al., 2013). Tightly woven communities may also impose social pressure against verifying the accuracy of the message, discouraging the recipients from fact checking (Oh et al., 2013). However in particular, the Twitter community has been demonstrated to have weaker social ties and therefore Twitter users are less likely to accept dubious reports with ambiguous sources.

Users will openly express distrust and question the validity of information (Oh et al., 2013).

Based on the literature review on rumour theoretic, people are more apt to engage in rumours under circumstances with higher anxiety as rumours fill a void to resolve ambiguity about uncertainty (Oh et al, 2013). The 'Basic Law of Rumour' identified by Allport and Postman in 1947 recognized that rumour is a result of ambiguity and importance of the topic: $R = A \; X \; I$ (Bordia & DiFonzo, 2002). Pezzo and Beckstead (2006) identified that uncertainty is a good predictor of rumour emergence. Activities such as generating explanations, directing others to additional information, pointing out inconsistencies, identifying patterns and noting importance/unimportance of the topic are all part of resolving uncertainty during the rumour process (Anderson & Krull as cited by Bordia & DiFonzo, 2004). Competing messages often stemmed from scientific information, which conflicted with personal public observations (Barclay, 2008).

Bordia and DiFonzo (2004, with assistance from others such as Rosnow, Allport, Postman, and Buckner) identify four phases in the computer-mediated rumour mill:

1. Rumour is introduced – information providing and interrogatory statements (reactions and responses are solicited)
2. Rumour is explored (personal experience is shared, sense-making begins)
3. Rumour is evaluated (Sense-making dominates, discussion evaluates the rumour, evaluates details)
4. Problem Resolution (agreement is reached on what is accurate, or group looses interest).

Bordia and DiFonzo (2004) found that a higher percentage of interrogatory statements (questions seeking information) implied that people sought information in the face of uncertainty. Oh et al. (2013) also found that if unambiguous and localized situational information is not provided to the affected community in a timely manner,

their collective information processing is very likely to encourage rumors. Along this same theme, Bordia & DiFonzo (2002) found that as emotions stabilize and no longer drive the message, rumours will generally cease to grow.

For a final note on rumour theoretic, rumour research has often used disease-oriented terminology. Words such as “spread”, “spreader”, “infected”, “infection”, “injected” and cured” are evident throughout the literature (Zhang, Zhou, Zhang, Guan & Zhou, 2013; Bordia & DiFonzo, 2004; Pezzo & Beckstead, 2006; Thomas, 2007; Zhao, Cui et al., 2013). It is therefore not surprising that rumours hold a negative connotation to many including emergency managers. Rumours however play an important role in information seeking and information validation behavior in communication networks.

Community Efficacy

Online informal communication networks may be more effective than government organizations at disseminating and exchanging information through their ability to easily form a community (Palen et al., 2007). Even after the response is formalized, unaffiliated citizens will continue to informally self-organize and communicate; online groups become virtual communities whereby members form policies and norms attempting to guide behavior. However the success of this varies (Palen et al., 2007).

Community efficacy is defined as the differential ability of groups to realize common values and maintain effective controls (Thomas, 2007). In his study, Thomas found that social cohesion (trust) and informal social control (willingness to follow informal rules of the group) are two essential factors for community efficacy. Community efficacy is built on “the link between cohesion—especially working trust—and shared

expectations for action" (Sampson as cited by Jaegar, Shneiderman, Fleischmann, Preece, Qu & Wu, 2007, p. 594). This has obvious effects on rumour transmission, as collective efficacy was found to foster the transmission of rumours (Thomas, 2007). According to Heverin and Zach (2012), Twitter provides a valuable platform for community efficacy by enabling sense-making during a crisis. Through the use of hashtags related to crisis events, users can request information, express emotions, and seek to identify reasons for the crisis. In this way, the purposeful use of hashtags enables a collective efficiency in the sense-making process.

On the other hand, large-scale emergencies are unpredictable and unplanned, circumventing the routine practices and norms that may govern social behavior, including communication and sense-making (Oh et. al, 2013). Social media services already have a reputation as rumor mills that aggregate and share misinformation. In an emergency environment, where the normal societal norms may be circumvented, communication may also no longer follow societal norms (Oh et al, 2013). Under emergency conditions, the social network may be an ideal environment for misinformation fanned by a collectively efficient spreading mechanism.

Trust remains a consistent theme in crisis communication. In previous disasters, the public viewed friends and relatives as the most trusted source for emergency information regardless of the channel used or the accuracy of the information (Barclay, 2008). However, governments generally focus disaster response on essential government services, technological infrastructure, and uncoordinated individuals who need urgent assistance; attempts to link residents and emergency response personnel in coordinated action are limited (Jaegar et al., 2007).

In traditional rumor research, anxiety is considered the most influential factor on the occurrence of rumours. However, in a digital communications environment, the most influential factor was ambiguity (Oh, et al., 2013). Oh, et al. (2013) specifically explored the social conditions under which rumours were created during three events: a shooting, a terrorist attack, and a product recall. They examined two types of ambiguity – content ambiguity and source ambiguity – and their impact on rumour propagation. They found that where the improvised virtual community of Twitter displayed cognitive distrust for ambiguous information, rumours were created to understand uncertain situations and reduce cognitive ambiguity. In particular while content ambiguity did not contribute to rumormongering, source ambiguity did significantly.

Alternatively, Barclay (2008) identified that some types of ambiguous content did contribute to rumours. In particular he found that emergency messages that contained precautionary attitudes ("the public should watch for tornado activity") and requested long-term evacuation ("residents are recommended to evacuate until further notice") lead to a perception of false alarms. This perception would then undermine the credibility of the risk resulting in increased rumours downplaying the severity of the event.

Though research on the conditions that encourage false rumour correction during emergencies is limited, emergency managers are already seeking approaches to control or eliminate rumours. Some emergency management groups have experimented with rumour hotlines but have been unsuccessful perhaps due to unpredictably of disaster conditions affecting power and phone services (Thomas, 2007). Other groups, such as the U.S. Federal Emergency Management Agency (FEMA) have provided websites specific to addressing the rumours that are generated from particular events (see

Oklahoma Tornado rumour control site: http://www.fema.gov/disaster/4117/updates/oklahoma-tornado-rumor-control).

Controlling rumours within the social network itself however may be a significant and unmanageable task for many emergency response groups who are frequently understaffed and already overwhelmed during the incident. Researchers suggest that Twitter may prevent sense-making due to information overload and dearth (Oh et al. 2013; Vultee & Vultee, 2011). This overload will certainly impact any methods that emergency managers might consider developing. Further understanding of how false rumours correct in social networks would be beneficial before strategies can be effectively developed.

The attitude, content, and form of rumours during emergency situations tend to be similar across the world (Bordia & DiFonzo, 2002). This adds some weight to the value of examining rumours in a few events in the hopes that correcting behaviours can be extrapolated to future events. It is also important to note that although emergency managers have expressed significant concern with rumours, some research models of rumour transmission suggest that rumour spreading at certain rates has a positive impact on the situation control, particularly when the facts regarding the situation are unknown as rumours may serve as a channel of information sharing helping to cope with unexpected disasters (Zhang & Zhang, 2009).

Social Network Misinformation in Disasters

Mendoza, Poblete, and Castillo (2010) examined Twitter use during the 2010 Chilean earthquake. Among their findings, they identified that tweets surrounding a false

rumour subject will deny the rumour in 50% of the sample. They also found that a significant number of tweets question the rumour. Finally they suggest that rumours might be detectable through an aggregate analysis (i.e. analysis that flags subjects that are highly questioned). Starbird, Maddock, Orand, Achterman & Mason (2014) examined more closely the relationship between misinformation and correction. They found that although misinformation and correction tend to rise and fall in tandem, they can occur in different magnitudes. They also found there is often a delay before misinformation is corrected and misinformation can persist beyond correction tweets, but generally at low levels.

Beyond Starbird et al. and Mendoza et al., there was limited literature that provided an in-depth exploration of misinformation in rumours, and more specifically how false rumours are corrected during disasters. Emergency managers, through their expressed concerns, have linked rumours to inaccuracy. However neither the rate of inaccuracy or the rate of self-correction (where participants in social media identify errors and propagate accurate information back through the network) has been thoroughly examined across a wide range of events and conditions.

Methodology

This research examined known false rumours, looking for instances of correction on Twitter. First and foremost it is accepted that rumours and more specifically false rumours (where the information was later identified to be incorrect) will occur in Twitter. This research sought to examine how and when social correction occurred. The design was split into two phases using qualitative content analysis: categorization of the tweets

within three false rumour streams, and codification of the content correcting and questioning tweets from Phase 1.

Dataset and Method of Collection

The Texas Fertilizer Plant Explosion was selected as the event to study because of the media attention it received both through mainstream (traditional) media and social media. The West, Texas event bears some similarity to the Boston Marathon Bombing, which has recently been the subject of a study on rumour transmission (see Starbird et al., 2014). This study provided a valuable point of comparison for this research as the research shares some similar methodology and events themselves shared similar characteristics. Both of these events occurred in relatively the same time period (2 days apart), and both occurred in the United States therefore both have a similar western 1st world influence. Furthermore both involved explosions where initially it was uncertain if it was accidental or intentional. Each event also involved similar authorities (first responder groups with similar command structure) and both received some degree of international media coverage and White House attention.

Some of the methodology used by Starbird et al. (2014) to identify three false rumours from the Boston Marathon Bombing was also adapted for this study. Starbird et al. (2014) identified and selected rumours that were later highlighted and criticized as false information in more traditional media networks (radio and television broadcasts) following the event. The first false rumour they examined claimed that an eight-year-old female runner was killed (Deneen, 2013). No children are allowed to run in the Boston Marathon and the actual 8 year-old victim, was male and died watching the race (Allen & Russell, 2013). Secondly they considered the false rumour where social media

participants hunted the wrong suspect. Sunil Tripathi became a centre of focus for social media as a primary suspect when a classmate of Tripathi posted a photo noting his resemblance to the unidentified suspects sought by the FBI. This information was then fuelled by two false rumours indicating that police were seeking Sunil in connection to the bombing, which was untrue (Starbird et al., 2014; Bidgood, 2013). His body would be discovered a few weeks after the bombing though had unfortunately died a few weeks before the bombing. This rumour was heavily criticized in broadcast media. The third false rumour captured by social and mainstream media was that the attack was carried out by the US government, otherwise known as a False Flag attack (Starbird et al., 2014).

Similar to the study by Starbird et al., three rumours from the 2013 West Texas Explosion were selected by completing a review of archived mainstream media reports. Using a non-tracking web search engine[1], the terms "West Texas Explosion" and "Rumour" were searched. Top results of this search revealed several articles published by recognized media outlets on various rumours. The first three rumours located in theses reports were used for this study.

Samples of tweets from each selected rumour were retrieved using searches with hashtag references and keywords. Most commonly the hashtag (e.g. #abc), a user-generated tagging mechanism in Twitter, provides an effective way to filter. Selected hashtags for instance can be retrieved and followed as a series of messages for research (Oh et al., 2013; Bruns & Stieglitz, 2012). Furthermore an examination of hashtags may

---

[1] *DuckDuckGo* was used to search mainstream media reports that were published following the event. This engine was selected as it is advertised as an engine that does not store search information or link searches to previous users' searches *(DuckDuckGo Privacy*, 2014). An engine that stores information about the user, though it may provide more relevant results in many cases, could have unnecessarily biased the search and the subsequent selection of rumours.

also provide some insight into the context of emotion or anxiety, e.g. “the tornado is coming towards us #terrified” (Burns & Stieglitz, 2012). The Content Filter *Topsy* was used to retrieve samples of each rumour selected. Content filters such as *Topsy* and *YourTwapperKeeper* may assist in sorting and filtering appropriate samples of the data (Oh et al., 2013; International Communication Association, 2011). Topsy stores data from Twitter beyond the seven-day archive, and indexes and ranks search results based upon the most influential conversations about each specific term, topic, page or domain queried *(Topsy.com,* 2014*).* Using Topsy’s Otter Application Programming Interface (*https://code.google.com/p/otterapi/* details code construction), the raw twitter data matching the Boolean search terms outlined below was retrieved and exported into spreadsheet format for categorization and manual coding.

The first false rumour in the West, Texas explosion that was explored in this research was the nuclear mushroom cloud that was shared across Twitter, mainstream media, and other sources as an image of the West explosion (see image 1). The mushroom cloud, distinctly recognized by the white bench in the lower part of the frame, is actually an old picture from a separate event in 2008, though it propagated both traditional and social media as an image of the West Texas explosion (Oremis, 2013). To retrieve a relatively focused discourse, this rumour was retrieved by searching any of the keywords “west texas”, #texasexplosion, #prayfortexas, #prayforwest, or “west explosion”[2] in association with “mushroom cloud”. It is likely that many tweets beyond

[2] The terms “west texas”, #texasexplosion, #prayfortexas, #prayforwest, and “west explosion” were selected as they generated that most results during initial searches using Topsy. It is unlikely that these filters would pull all results however, the sample size was

those tagged with “mushroom cloud” contain the false image. However as this research is more interested in the correction of the false rumour, it was anticipated that those correcting the image would be more likely to use a label of “mushroom cloud” to identify their subject, therefore this filter would still be useful to retrieve relevant tweets.

Figure 1 – Mushroom cloud posted as West Texas 2013 Explosion, originally from explosion near Waco Texas in 2008


(Oremis, 2013)

The second false rumour that was studied occurred throughout the first few days of the event. Though only 15 people died in the event, the number of deaths reported and shared fluctuated significantly. Specifically a rumour started that 60 people had died, when only 12 had at that time been confirmed as deceased (Associated Press, “KWTX reporting 60 to 70 dead at West, Texas fertilizer plant explosion”, 2013). This rumour

---

sufficient for the purposes of this project (209,398 tweets during the timeframe selected). These search terms were used for all three rumour sets Boolean searches.

was retrieved using a keyword search of any of "west texas", #texasexplosion, #prayfortexas, #prayforwest, or "west explosion" associated with "60 dead". This rumour was particularly important to this study because of the propensity for the number of victims in most crises to be misconstrued or falsely reported (Chan, T., Killeen, J., Griswold, W., & Lenert, L., 2004).

The third false rumour in the West, Texas Explosion that was studied was the event's reported connection to the Boston Bombing a few days earlier. This rumor may have been fueled given the approximation of both events, and similarity of the hazard (an explosion). Tweets for this rumour were retrieved using "boston" associated with any of "west texas", #texasexplosion, #prayfortexas, #prayforwest, or "west explosion" in a keyword and hashtag search.

To refine the sample, all three rumour sets were retrieved within a 2-day time frame a few minutes after the event occurred. Where the Boolean search retrieved more than 5,000 tweets, only the first 5,000 tweets were used. Research suggests that rumours are most prevalent in the early stages of a disaster, therefore this time period is ideal to retrieve rumour threads. The dataset includes tweets from April 17 8:55PM CT to April 19 8:54PM CT.

Qualitative content analysis was selected as the preferred technique for this study. Content analysis is a research technique using a specialized procedure to inform practical action, which makes replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use (Krippendorff, 2013). This methodology was selected because it is unobtrusive and also eliminates any errors that occur when subjects are aware they being scientifically observed. Furthermore, any bias that may occur during a

direct subject and research interaction is prevented (Krippendorff, 2013). Though observation in real-time may have been valuable, a simulated event would not have had the same effect and may have been artificially transparent. This study used historical analysis. Content analysis of specific tweets leverages a vast amount of rich data already available to the general public. For this research, qualitative content analysis was performed using a deductive approach. A deductive approach allows the reviewer to begin data analysis with a coding template in mind, and organize the data according to an existing, though alterable, structure (Finfgeld-Connett, D., 2014). This flexibility is important as the purpose of a qualitative systematic review is "to test, adapt, expand, and in general, improve upon the relevance and validity of existing frameworks" (2014, p. 342).

This study has been divided into two phases. In Phase 1, the tweets from the three false rumours were categorized to identify and flag any correcting tweets. In Phase 2, tweets categorized as "correcting" are further examined and coded for "User Type", "Type of Communication", and "Referral Source". Each correcting tweet and were further examined through a qualitative exploration of the conversations and links within each thread. Specifically, the author sought to identify and retrieve patterns and/or specific conversations that provide insight into how correction occurs.

Mendoza, Poblete, & Castillo (2010) suggest that rumours in the form of tweets fall into one of several categories: 'affirms' – where the tweet propagates information confirming the information, 'denies' – where the tweet refutes the information item, or 'questions' – where the tweet asks about the information, and unknown/unrelated. Similarly, Starbird et al. (2014) adopted a categorization of 'misinformation',

'correction', and 'other' to code each distinct tweet within their rumour sets. As both methods are beneficial, this study used a blended approach using all of the categories from Starbird et. al (2014), as well as the 'questions' category from Mendoza (2010) for the first phase of the study. This served to specifically identify the correcting and questioning tweets for Phase 2. Additionally the author later (after initial coding) added the category of 'information' to account for tweets that seemed to contain valuable information about the event but did not question, confirm or deny the rumour, and 'unknown' for those tweets that could not clearly be placed into the rumour-related categories. All tweets within the three above identified rumour sets were coded for one of the five categories: information, misinformation, questioning, correction, and other/unknown. See below for the coding table used and Appendix A for full definitions of each category.

Table 1 - Coding Table for Phase 1

Variable Name	Level	Attributes	Code
1. Category of Tweet within Rumour Set	Nominal	Information	I
		Misinformation	M
		Questioning	Q
		Correcting	C
		Other/Unknown	U

The second phase of the research design dealt only with the tweets that fell into either the correcting or questioning category from Phase 1. Retweets in the correcting or questioning category (where the exact message has been passed along without substantial change) were also included in this phase. Each phase 2 tweet was manually coded for "User Type", "Type of Communication", and "Referral Source" (see table below and

definitions in Appendix A for full details). Other variables were also examined if certain variables arose during test coding to provide additional value to the study. The 'User Type' information used nominal coding to identify the user's background (e.g. government, private citizen), which may have indicated the user's level of trust within the community. As anticipated, there were some challenges in identifying the user, and as such the "anonymous/unknown" category was used. The 'Type of Communication' category identified whether the correcting or questioning tweet was an original post, retweet (indicated by RT), or direct reply (indicated by @). Lastly each tweet was coded for any source of referral it provided such as url. Where a url is included, the coding indicated the type of url source. Source types included traditional media – television, radio, print; an emergency management authority – any organization recognized as part of the response and recovery efforts; politicians – which included Mayors, Reeves, Senators, and any official members of the White House; or private citizens. Where a source could not be clearly identified, it was coded as uncertain.

Table 2 - Coding Table for Phase 2 (Correcting and Questioning Tweets Only)

Variable Name	Level	Attributes	Code
1. User Information	Nominal	Media Organization	1
		Official EM Organization	2
		Political/Leadership group	3
		Other Organization	4
		Private Citizen	5
		Anonymous	6
2. Type of communication	Nominal	Original source	O
		Retweet	R
		Direct Reply	DR
		Other/Unknown	99
3. Referral Source	Nominal	Media	Med
		EM/FR Authority	Au

		Politician	Pol
		Private	Priv
		Uncertain	Unc
		No source	n/a
		Unknown (link broken)	Ukn
		Other	Oth

Data Analysis Strategy

The analysis was quantitative initially to identify summary statistics on each of the coded variables within the rumour sets. This quantitative analysis was used primarily to narrow down the sample to the tweets which would be relevant for the second phase of the study. In the second phase of the study and after content coding was completed, the author used a qualitative approach to explore the conversations that surrounded the correcting tweets. This exploratory approach was used to highlight any patterns that emerged from a closer examination of the content (i.e. whether correcting or questioning tweets shared similar characteristics across all three rumour sets). Furthermore the frequency of retweets within each grouping of corrections was also considered in the hopes of understanding when and where correction was more successfully shared.

Validity (peer review, coding validity)

Trial coding was performed on a portion of tweets retrieved from different emergency event (#ABFlood) prior to the start of this research. This was used to refine the coding and ensure that the methodology would be appropriate and effective. The author and an assistant coded the first 20 tweets for this West study to ensure consistent coding results from different coders. Coding was compared to ensure a valid coding scheme, and clarify any differences in coding (Krippendorff, 2013). In response to a

number of questions from the second coder about distinctions between the tweet categories, the definitions found in Appendix A were refined to define the categories and coding frame more clearly.

To ensure a systematic and repeatable quantitative method, coding was also repeated on a portion of the sample at a later date to verify reasonable accuracy. Krippendorff (2013) suggests that intra-coding should yield a minimum of 70% reliability.

Potential Limitations of Methodology

Twitter recently changed its Application Interface preventing the use of data older than 7 days which limited the tools available to the researcher to explore the data; even services like Topsy are now limited due to this change (Kern, 2012). It is uncertain whether a full sample for the Boolean search was retrieved though Topsy or whether Topsy filtered by popular content. Many third party microblog storage providers such as Topsy focus on popular tweets and influential users while discarding the rest of the data. This was an accepted limitation for this research as other methods to retrieve data from the full stream of data were cost prohibitive.

Phase 1 Findings

This section includes both the findings from coding the data, including a quantitative summary, and a review of other emerging themes and discourse as revealed through the qualitative analysis. Although the qualitative exploration of tweet content provided a more detailed exploration of the patterns, it was valuable to first explore the

frequency (a quantitative summary) within the categories. This provided an indication of the patterns and established a basis for more directed qualitative analysis.

Table 3 indicates the frequency of the five categories used in the first phase of coding – 'information', 'misinformation', 'questioning', 'correcting', and 'unknown'. For reasons later described, 'unknown' tweets were removed from the samples where indicated, and not included in the comparisons of category frequencies.

Table 3 – Frequencies of Categories within Each Rumour Sets

Rumour Set	Category	Frequency Including 'unknown' category	Percentage Including 'unknown' category	Percentage Excluding 'unknown' category
60 Dead				
	Information	2	4.65%	4.76%
	Misinformation	33	76.74%	78.57%
	Questioning	5	11.63%	11.90%
	Correcting	2	4.65%	4.76%
	Unknown	1	2.33%	~
	TOTAL	**43**		
Mushroom Cloud				
	Information	5	12.20%	18.52%
	Misinformation	18	43.90%	66.67%
	Questioning	1	2.44%	3.70%
	Correcting	3	7.32%	11.11%
	Unknown	14	34.15%	~
	TOTAL	**41**		
Boston Connection				
	Information	205	6.37%	51.38%
	Misinformation	118	3.67%	56.57%
	Questioning	44	1.37%	11.03%
	Correcting	32	1.00%	8.02%
	Unknown	2817	87.59%	~
	TOTAL	**3216**		

Information Category Frequency

In the 'information' category, tweets provided accurate information or details about the event. These tweets went beyond more than exclamations of prayer or wishes for recovery (this type was categorized under 'unknown'), but attempted to provide some detail about the event. Generally these tweets seemed to have the intent to inform others. Tweets in the 'information' category also did not clearly present misinformation. Example:

> *Texas fertilizer plant blast injures many in WEST TEXAS HUGE EXPLOSION 'LOOKED like a MUSHROOM CLOUD' WITNESS REPORTS http://t.co/FeNdTjQ558*
>
> @hidefullofHoney; 8:44 PM - April 17, 2013

This category was also expanded to include tweets that provided commentary or evidence about the event's media coverage. In the researcher's view, these tweets were providing information about sources of event information including media transparency or efficacy. Example:

> *Chinese paper photoshops a cross into the West Texas mushroom cloud. Guess they're associating it with death/graves? http://t.co/aOe1LWiU9d*
>
> @ChinaHopeLive; 10:36 PM - April 18, 2013

> *@CNN not covering the TX explosion. They are still nattering about boston #if it didn't happen on a coast it didn't happen #prayfortexas*
>
> @talesin; 4:55 AM - April 18, 2013

Collectively there were 212 tweets in the information category. The information category was the most prevalent in the ‘Mushroom Cloud’ rumour set at 12.2% of all tweets within that group.

Misinformation Category Frequency

In the ‘misinformation’ category, tweets provided incorrect or inaccurate information to support or propel the rumour. Examples:

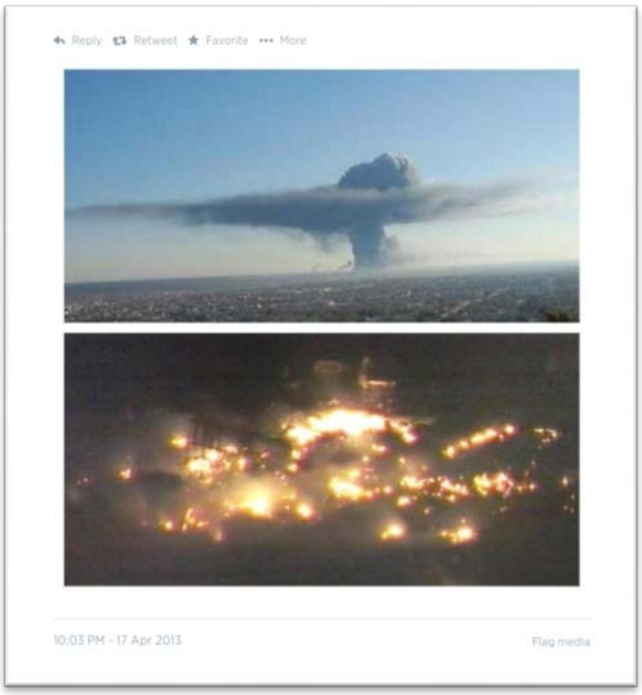
‘Mushroom Cloud’ rumour:

#Texas Fertilizer Plant mushroom cloud explosion. #prayfortexas

http://t.co/bQCx7H4RxH

@SetioDarmadi; 10:05 PM - April 17, 2013

Figure 2 – Photo from Tweet by @SetioDarmadi, 10:05 PM, April 17, 2013



‘Boston Connection’ rumour:

#prayfortexas how sick is our world we had first bombs at boston marthon n 2 days later now texas #sick

@xsportychic2395; 8:50 PM - April 17, 2013

’60 Dead’ rumour:

@cnnbrk is wrong I'm in TX right now. From the West Texas explosion there are 60 dead at least 100 injured. Love to all..

@FollowEWpriv; 9:17 PM - April 17, 2013

Collectively, there were 169 tweets in the misinformation category. The misinformation category was the most prevalent in the ‘60 Dead’ rumour subject at 76.74% (the ‘Mushroom Cloud’ set contained 43.9 % misinformation tweets, and the ‘Boston Connection’ set contained 3.7% misinformation tweets[3]).

Questioning Category Frequency

In the ‘questioning’ category, tweets challenged or questioned the rumour information. This category did not contain tweets that merely asked a question about the event unless the question could be reasonably related to the rumour subject. Examples that were included in the questioning category:

---

[3] When ‘unknown’ tweets were removed from the sample, the percentage of information tweets was 4.76%, 18.5%, and 51.4% for ‘60 Dead’, ‘Mushroom Cloud’, and ‘Boston Connection’ sets respectively.

Is the #texasexplosion a seperate incident to the boston ones or do they think they are connected?

@DeeVinyl; 12 :33 AM - April 18, 2013

60 DEAD?! RT @stevefullhart: KWTX is now reporting at least 60 people are dead in the West explosion, quoting local emergency officials.

@xtina1229; April 18, 2013

Questions unrelated to the rumour were placed in the information category. Collectively, there were 50 tweets in the questioning category. The questioning category was the most prevalent (11.63%) in the '60 Dead' rumour subject[4].

Correcting Category Frequency

In this category, tweets present information to correct or contradict the misinformation within the rumour subject. Examples of corrections:

'Boston Connection' rumour:

@samqb9 dog that just happened today. And besides, Boston was a planned attack, west Texas was just an accident as far as we know.

@BasedGoober; 9:43 PM - April 17, 2013

[4] When 'unknown' tweets were removed from the sample, the percentage of questioning tweets was 3.70%, 11.9%, and 11.0% for 'Mushroom Cloud', '60 Dead', and 'Boston Connection' respectively.

‘Mushroom Cloud’ rumour:

> *Local news cut off dad/kid West explosion video before kid starts pleading, but is using the debunked mushroom cloud pic... #KTLA*
>
> @csrbella; 10:08 PM - April 17, 2013

’60 Dead’ rumour:

> *Waco PD spox says he cannot confirm report of 60 dead in West explosion*
>
> @passantino; 12 :19 AM - April 18, 2013

There were 37 tweets in the correcting category across the three rumour sets. The correcting category was the most prevalent in the ‘Mushroom Cloud’ rumour subject (7.3%), followed by the ‘60 Dead’ rumour set (4.65%). The Boston connection rumour had the lowest correction rate (1.0%) but also contained the largest sample size[5].

Unknown Category Frequency

The ‘unknown’ category was used to identify tweets that were not suitable for any of the other categories. These tweets did not fit under the definition of an information tweet (provides details about the event) or any of the rumour specific categories - ‘correcting’, ‘questioning’, or ‘misinformation’. A significant number of ‘unknown’ tweets expressed sentiment similar to Tweets referred to by Vultee and Vultee (2011) as personal messages. Examples:

---

[5] When ‘unknown’ tweets were removed from the sample, the percentage of correcting tweets was 11.1%, 4.76%, and 8.0% for ‘Mushroom Cloud’, ‘60 Dead’, and ‘Boston Connection’ respectively.

All these events lately are seriously sick.. First Boston now Texas? Always be grateful. #PrayForTexas

@linnyhands; 9:15 PM - April 17, 2013

#prayfortexas United we stand United we fall!! Dont mess with Texas!!!! And also praying for everyone in Boston #bostonstrong

@liljazzy10; 9:27 PM - April 17, 2013

Explosions should stop. Many people are being affected. First in Boston and now in Texas. #prayfortexas

@Updates1D_x3; 8:33 PM - April 17, 2013

In some cases, personal messages were categorized as 'misinformation', but where they could not clearly be classified in either of these categories, such as the tweet examples above, they were placed in the unknown category.

In some cases, 'unknown' tweets may have also included questions about the event that were not trying to solicit information (rhetorical questions). In many cases rhetorical questions were difficult to categorize clearly and were therefore placed in the 'unknown' category. Examples below for instance could have been construed as a simple question, or might have implied an intentional event as part of the 'Boston Bombing Connection':

After Boston, Texas? When will this end? #PrayForTexas

@iamfrancisdiane; 10:20 PM - April 17, 2013

@Mikeaveli: this is crazy, first Boston now Texas, i hope my state isn't next? #prayfortexas #prayforboston

@Mikeaveli; 11:07 PM - April 17, 2013

The frequency of tweets that were categorized as 'unknown' was substantially higher in the 'Boston Connection' rumour set (2,817 tweets). This could have been a result of including the hashtags "prayforwest" which elicit emotional responses, however this tag was used across all three rumour samples to ensure continuity of results. Vultee and Vultee (2010) suggested that personal messages (which formed the majority of 'unknown' tweets in this study) will account for the highest portion of tweets. However this was only evident in the 'Boston Connection' rumour set (87.6%). Unknown category tweets occurred in the other two streams with much less frequency ('Mushroom Cloud' with 34.2% and '60 Dead' with 2.4%). In many cases where a personal message occurred in the '60 Dead' or 'Mushroom Cloud' rumour sets, the personal message was also accompanied by event details (information or misinformation), and for the purposes of this project, those tweets were classified in those respective categories. Examples:

Classified as misinformation:

Serious Explosion in Texas north of Waco! Numbers I've seen say 60 dead and probably climbing...sad sad week in our nation #prayfortexas

@SamiEarl; 9:42 PM - April 17, 2013

Classified as misinformation: *75 home destroyed in West Texas. Explosion caused a Mushroom Cloud. This is a mini atomic bomb. Wow*

@GSF_23; 10:14 PM - April 17 2013

Phase 2 Findings

Phase 2 focused on tweets classified as 'correcting' or 'questioning' only. Specifically, the author sought to identify patterns and/or specific conversations that provide insight into how correction occurs therefore these two categories became the most relevant for further and more detailed analysis. This phase included an examination of the participants (users), any external links or references that were included in the tweet, and other patterns or characteristics that emerged as part of the conversations that followed. An exploratory approach was used to highlight any patterns that emerged from a closer examination of the content. Retweets were also considered and explored within each group of correcting tweets to understand whether some corrections were more shared than others.

Participants

Users in questioning or correcting messages from all three rumour sets were primarily private users - those who by a quick review of their twitter profile appeared unaffiliated with any formal organization. Some private users also persisted in more than

one rumour subject. For example the user “Kolt”, corrected two different and unrelated rumour feeds:

> *And the picture of the "west explosion mushroom cloud" was actually a refinery in Georgia, 2008.*
>
> @DeltaNineSeven; 8:40 PM - April 17, 2013

> *OK, quit comparing the Boston bombings and the West explosion please. One was an act of terrorism, one was a chemical explosion, not a bomb.*
>
> @DeltaNineSeven; 8:21 PM - April 17, 2013

Some media users were also found in both the correcting and questioning tweets of the ’60 Dead’ and ‘Mushroom Cloud’ rumour sets, however no media users were found in the ‘Boston Connection’ questioning or correcting threads.

Table 4: Participant (User) Frequencies by Type

Type of Participant	Frequency in Correction Category	Frequency in Questioning Category
Private	Mushroom Cloud: 2 Boston Connection: 32 60 Dead: 1	Mushroom Cloud: 1 Boston Connection: 44 60 Dead: 3
Media	Mushroom Cloud: 1 Boston Connection: 0 60 Dead: 1	Mushroom Cloud: 0 Boston Connection: 0 60 Dead: 1

Similar to the findings of Vultee and Vultee (2010), the percentage of government and/or emergency management official messaging was very low. In the three rumour

sets, no government or emergency management official messages were found in any of the categories.

Retweet Frequency

Correction tweets were not heavily retweeted in any of the three defined rumour sets. 4 retweets was the highest number of retweets for one message in the 'Mushroom Cloud' thread. 5 retweets was the highest number of retweets for one message in the 'Boston Connection' thread. 10 retweets was the highest number of retweets in the '60 Dead' thread. Misinformation tweets on the other hand had higher retweet rates: there were 16 retweets in the 'Mushroom Cloud' rumour[6], 5 retweets in the 'Boston Connection' rumour, and 17 retweets in the '60 Dead' rumour.

Links and Tags

Links were not abundantly provided in either the 'questioning' category or the 'correcting' category. The majority of 'Boston Connection' tweets that were categorized as either correcting and questioning did not contain links. Though the sample sizes for both the 'Mushroom Cloud' and '60 Dead' threads were small, both the correcting and questioning categories (and therefore their retweets) included some tweets with links, however, in the correcting tweets, no private links were used in any of the three rumour sets.

---

[6] It is important to note that the author was able to find one example of the 2008 mushroom cloud picture which was retweeted more than 500 times, however this was not captured in the original rumour set as it did not include the Boolean search values for the set.

Table 5: Frequency of Link or Tag by Type

Type of Link/Tag	Frequency in Correction Category	Frequency in Questioning Category
Private	Mushroom Cloud: 0 Boston Connection: 0 60 Dead: 0	Mushroom Cloud: 0 Boston Connection: 2 60 Dead: 0
Media	Mushroom Cloud: 1 Boston Connection: 0 60 Dead: 0	Mushroom Cloud: 1 Boston Connection: 1 60 Dead: 0
No Link Provided	Mushroom Cloud: 2 Boston Connection: 32 60 Dead: 2	Mushroom Cloud: 0 Boston Connection: 41 60 Dead: 4

Other Findings

The sentiment score, defined by Topsy (2012) as how positively the subject is talked about compared to all other terms on Twitter that day, was relatively low across all samples. The sentiment score uses advanced natural language processing to analyze the sentiment — positive or negative — of each tweet compared to the sentiment of all topics discussed on Twitter at that time. Each tweet is assigned a score from 0 to 100, with 50 indicating neutral sentiment, and higher and lower scores representing positive and negative sentiment, respectively (*Observations on Twitter and the ACA - Health Reform & Public Programs Initiative - Programs - CHCF.org.*, 2014). The sentiment score in the West, Texas tweets retrieved under this study ranged from 3.35 to 10.15, and was an average sentiment score of 4.46. This was expected by the nature of the event and the type of messages that were sent regarding the West, Texas explosion.

Discussion

This section aims to draw connections between these findings, recent relevant research (in particular research by Mendoza, 2010 and Starbird et al. 2014), and key concepts from the literature review, including rumour theories, emergency communication trends and patterns, and past observations on collective correction through the social media during emergencies. The research study sought to examine the conditions under which Twitter communication by non-experts self-corrects. Do users work collectively to identify, correct, and propagate accurate information. It also examined whether users leverage other resources to correct information such as traditional media.

Motivation and Rumour Spread

Some rumour theories suggest that an upsetting but plausible rumour may be transmitted because the sender desires to inform others, particularly friends and family, of impending danger and/or to elicit information from others (Pezzo & Beckstead, 2006). The rumours in the three threads examined seemed for the most part plausible. The rumoured number of fatalities was well within normal estimates for an explosion within a town's borders. The incorrect photo of the mushroom cloud, though not from the West, Texas explosion, was still from an area near West, Texas, and also from an event that occurred less than 5 years previously. The rumoured connection of the West, Texas explosion to the Boston bombing was also plausible given the close timing of both events, and past terrorism events (such as 9/11in 2001) which could have reasonably led participants to link two explosions in different states to a coordinated attack. Also in

support of Pezzo and Beckstead (2006), some rumour tweets attempted to elicit information from others as many tweets included questions about the event or attempted to provide information for validation.

To extract correcting tweets, the project focused on filters that would retrieve misinformation threads. It was therefore not surprising that when unknown tweets were removed from the sample, misinformation tweets (which still attempt to provide information) in all cases made up more than 50% of the sample. This remained the case when 'unknown' tweets were included in the sample for all cases except the 'Boston Connection' thread. This thread focused largely on expressing sentiment rather than providing or requesting information and which therefore did not align to the findings by Pezzo and Beckstead (2006).

Furthermore all three subjects explored through the rumour sets (and the West, Texas event itself) were high in anxiety. The 'Basic Law of Rumour' identified by Allport and Postman in 1947 (as cited by Bordia & DiFonzo, 2004, and findings by Pezzo and Beckstead, 2006) supports the high occurrence of rumours within subjects that cause anxiety.

As suggested by Smith (2010), retweeting may provide legitimization and perceptual credibility to the message. The rate of retweeting was higher in the 'misinformation' category than in the 'correcting' category. According to Smith, this would provide a false sense of legitimacy to the misinformation over the correction tweets. However the sample size for correction tweets was relatively small so this finding would require additional substantiation.

Specific motivations of participants could not be finitely determined in this study without additional information from the users (e.g. interview or survey).

The Role of Emergency Managers

In social media, Crowe (2010) suggests that emergency managers fit into one of three roles: proactive (managers or first responder groups monitor, post, and discuss issues), reactive (managers either monitor or post not both, and they do not actively engage their audience), and inactive (the managers have no role in social media). In the West, Texas explosion, emergency managers were inactive; they played no role in any of the three false rumour subjects. In this particular incident, the first responders consisted largely of volunteers. First responders came from West Volunteer Fire Department, Abbott Volunteer Fire Department, and Mills Volunteer Fire Department (n.d., Facebook "In Memory of West, TX Firefighters Last Alarm 4/17/2013"). These volunteer groups managed the response and many became victims (10 of the 15 later found dead were firefighters/EMS). It was therefore not surprising that the social media presence of this group only emerged after the event had ended, as members of the their families and the public paid memorial tribute to the loss of these first responders.

Emergency managers did however initially participate in traditional media broadcasts through on-screen appearances and press conferences. In particular, an EMS director on scene was quoted to report that he feared more than 60 people could be dead. The '60 Dead' rumour tweet emerged shortly after this statement. The ambiguous nature of the message and the media speculation that followed may have contributed to this

misinformation. The EMS Director's statement seems to have morphed into a confirmation by local media, then captured on Twitter by various users:

> *KTVT has graphic saying 60 dead in West explosion.*
>
> @TommyNoel; 8:33 PM - April 17, 2013

This rumour was rapidly supported and fueled in the social network backchannels where it propagated in numerous conversations well beyond the tweets retrieved by the '60 Dead' Boolean search used for this study. Misinformation about the victims of the event also showed up in other Twitter rumour sets:

> *First Boston Now Texas??? What's going on? 60+ Already dead many injuries.... #prayforwest #prayfortexas #prayforboston*
>
> @BuckyPhillips88; 9:33 PM - April 17, 2013

Oh et al. (2013) also found that if unambiguous and localized situational information is not provided to the affected community in a timely manner, their collective information processing is very likely to propagate rumors. It is possible that the absence of first responder and/or emergency manager participation may have contributed to the ambiguous conditions that fostered rumour development and prevented correction rates similar to those observed by Mendoza (2010).

Evolution of Rumours and Corrections

As expected, some rumours began to evolve within the backchannel network. Where a rumour may have begun as 60 dead, it later progressed to higher numbers:

> *60-70 Deaths, and 100 Injured. #PrayForTexas First Boston now this??? Scary world we live in...*
>
> @meglizjohn; 9:00 PM - April 17, 2013

> *#PrayForTexas this is awful. 186 people hurt in Boston, now 70-80 people dead on Texas. Saw the video and it gave me chills.*
>
> @bdonahue23; 9:44 PM - April 17, 2013

> *im sorry to those 100s who died today in Texas #prayfortexas :/ ill be praying for you and still wishing Boston the best*
>
> @LiSa_lOvEs_1D; 8:59 PM - April 17, 2013

This evolution may have also made it difficult for correcting tweets to gain legitimacy and traction. The actual number of deaths that was first announced as 12 and then later evolved to the final actual death toll of 15. The number of victims may have seemed implausible for an explosion within a town of 3,000 residents, which destroyed more than 100 homes, a nursing home and a school, and where mainstream media represented the damage with images such as that below:

Figure 3: A popular image used by media to represent the force of the explosion

(*A Neighborhood Nearly Obliterated by a Blast*, 2013)

Are concerns about information quality legitimate?

Emergency managers have cited potential information quality issues particularly when highly sensitive, inappropriate, or traumatizing information is inappropriately shared by the public (Oh et al., 2013; Hiltz & Gonzalez, 2012). The information quality was questionable as members of the public commented on the numbers of dead, long before this information was confirmed by official response organizations. Inaccurate information was clearly shared however it is not clear the impact of this misinformation. Did this misinformation cause harm or impact the response to the event? This project could not rule out that possibility, but the author was unable to find any evidence of harm from the rumours. This was in contrast to one of the false rumours in the Boston Bombing as studied by Starbird et al. (2014). One particular rumour was later reported to have caused harm. When a social media manhunt emerged for the wrong person, this misinformation caused harm to both the family and unintended victim's reputation and well-being (*Boston Bombings: Wrong Suspect Reveals Fears.* n.d.).

It is also worth discussing the role of links in Tweets. Links may help to provide legitimacy to the user's claim, particularly where an outside but reliable source is cited. The International Communication Association (2011) suggests that although many users prefer to link back to traditional media sources, a growing number of Twitter participants may alternatively be linking to other social media sites as a source of information. As mentioned earlier where participants included a link in their tweet, the link seemed to vary between media links and private links. Most notably however the correcting tweets contained no links of either type. The absence of private links (links to other social media sites and unaffiliated user pages) in correcting tweets was not surprising, as it makes sense that participants would not consider that a private link would provide a valuable and legitimate reference to support their claim. However the absence of traditional media links was unexpected. If traditional media sources are generally a trusted source of news information, and users wish to resolve ambiguity and be trusted by other users, then logically one might expect traditional media links to be used in correcting tweets to support the contradictory information. This was not the case in any of the three false rumour sets examined. Again this may in part be due to an absence of official or authoritative participation by emergency responders or other trusted experts, or a distrust of traditional media sources which was expressed in some of the streams (coded into the 'information' category).

How Does Correction Occur?

Similar to Starbird et al. (2014), this project found evidence of correction for each rumor but with considerably smaller proportions of correction than Mendoza et al.

(2010). In the three false rumour subjects, the rate of correction was lower than anticipated. Mendoza et al. identified an average of 53.8% correcting tweets in their rumour set, where as the highest portion found in this project in any of the three rumour sets was 11.1%.

Though the correcting tweets were infrequent in the three rumour sets, the data reveals that correction occurred in all three subjects in a similar way. Within the rumour subject, questioning occurred both before the correcting tweets emerged and afterward. Users also openly expressed distrust and question the validity of information (Oh et al., 2013):

> *first Boston gets attacked, now a fertilizer plant "catches fire"... things are getting sketchy.. #PrayersforBoston #PrayForTexas*
>
> @Chris_Gulli; 9:10 PM - April 17, 2013

More questioning tweets occurred after the correction tweet, which is congruent with dissonance theory. Dissonance theory suggests that when a participant is presented with conflicting information, the participant will desire to resolve that dissonance by filling the void with an explanation. This may reasonably explain why questioning is persistent once two or more versions of critical information are presented, as participants were attempting to resolve the contradiction.

> *Why is 9 News in Denver reporting 60 dead while CNN is reporting 34 injured? What the hell is happening in Texas? #texasexplosion*
>
> @IzzieB78; 8:58 PM - April 17, 2013

Several threads did demonstrate successful correction; replies and retweets supported the corrected message from the original tweeter. For example, the user below posted at 9:50 PM April 17th:

Figure 4: PicTweet of mushroom cloud

@sebp27; 9:50 PM - April 17, 2013

The above misinformation was corrected four hours later:

@sebp27 That's a 2008 TX chem plant explosion.

@castewar; 3:29 AM - April 18, 2013

However though some correcting and questioning tweets could have encouraged more accurate reporting if they had been shared and propagated, they instead faded out of the conversation. In some cases, misinformation even became more prominent after the

correction had been issued. The following was initiated by *140elect* at 9:07 PM, April 17th. It was retweeted 9 times, and the last retweet was at 4:26 AM on April 18th:

> *Credible reports place ~60 injured in #WestTX after fertilizer plant explosion. Unconfirmed reports 60 dead, hundreds injured. #prayfortexas*
>
> @140elect; 9:07 PM - April 17, 2013

This accurate information was then countered with false information within a few minutes:

> *@140elect @KHOU is reporting 70 confirmed dead including 5 firefighters and 1 police officer @tvdetective*
>
> @kamcc52; 9:09 PM - April 17, 2013

Private versus Traditional Media Participation

Before discussing the role of traditional media identified in the research, it is worth considering the prevalence of private users who participated in the 'correcting' feed. Private users make up the typical user base of Twitter (Fraustino, J. D. et al., 2012). The significant portion of private participators in this study's 'correcting' tweets may simply be a reflection of Twitter's typical user base or at least a reflection of those who participated in the original conversation. Alternatively it may be a reflection of the user type that is more likely to offer a diverse opinion and engage in information discourse (validation of misinformation through questioning, and correcting).

To better understand the relationship of traditional media and informal social networks, the author explored the rumour of ‘60 Dead’. According to the discussions on Twitter, the misinformation in this particular thread originated from several traditional media outlets:

> *@KTLA omg! 60 dead? :-(very very sad.. #prayfortexas*
>
> @sToRmY_Ali; 9:15 PM - April 17, 2013

> *KWTX is now reporting at least 60 people are dead in the West explosion, quoting local emergency officials.*
>
> @xtina; April 17, 2013

The mushroom cloud rumour may also have originated from a traditional media source, but that could not be confirmed. It was however clearly perpetuated by local and national media reports that used the stock 2008 photo in their coverage:

> *Local news cut off dad/kid West explosion video before kid starts pleading, but is using the debunked mushroom cloud pic... #KTLA*
>
> @CSSRMolly; 10:08 PM - April 17, 2013

In both cases, many Twitter users openly criticized official media outlets for generating and/or supporting rumours and misinformation. Additionally the absence of traditional media links in ‘correcting’ tweets showed that users did not leverage

traditional media as a reference to correct rumours. Social media was only criticized by one user as a potential source of misinformation:

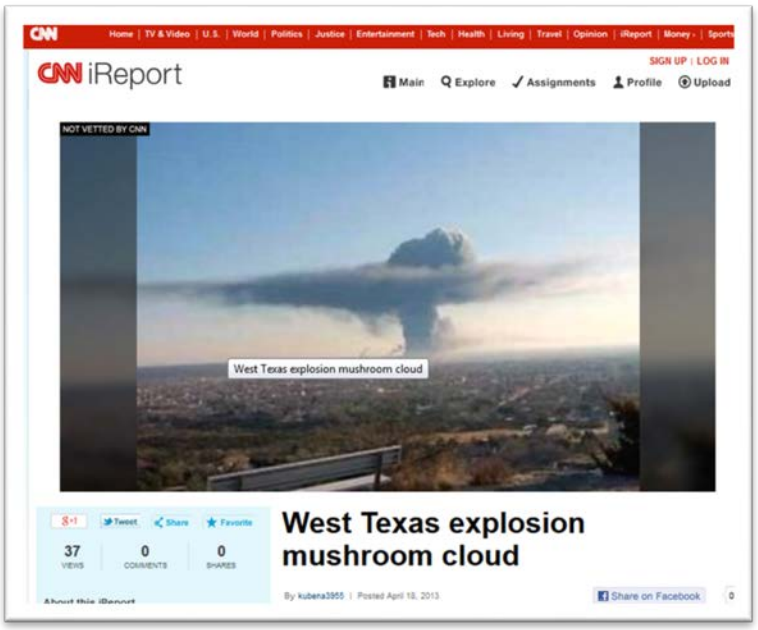
> *That picture of the enormous mushroom cloud from West Texas fertilizer explosion? Yeah, likely photoshopped. Twitter is almost as bad as CNN*
>
> @getnickwright; 8:20 PM - April 17, 2013

However the relationship between traditional media and Twitter could not be clearly defined. As suggested by Wigley & Fontenot (2010), traditional media outlets such as broadcast news networks leveraged citizen generated digital content during the first stage of a crisis. References to the Huffington Post were found in the samples, and specifically demonstrated the potential complexity of citizen participation. As also suggested by the literature, it was unclear if this source was perceived as a traditional media source (trusted by the public to be factual and vetted) or if it was understood as a source influenced heavily by public participation. One tweet replied directly with "@HuffingtonPost over 60 people killed in west texas blast! Happening now" (Dan Seago, 2013, http://twitter.com/djs0824/status/324728428697878531) in attempt to provide more information for the Huffington's news report. It is unknown if that may have contributed to the inaccuracies later reported by the Huffington Post in this particular incident.

The "citizen journalist" was clearly leveraged in the West, Texas explosion by major media outlets. Outlets such as CNN created unique spaces whereby citizens could share event details as they unfolded (e.g. CNN iReport). These unique spaces, visited by

a smaller network (in the case below 37 users) may have contributed to misinformation by creating pockets of discussion. Here, misinformation persisted as it was left uncorrected and isolated from 'correcting' tweets in the larger discussion. This may have created an additional delay in a correction as the correction was not stored on the original site. The misinformation however likely continued to infiltrate the larger audience through 'like' and 'share' links.

Figure 5 – iReport user reporting website provided by CNN



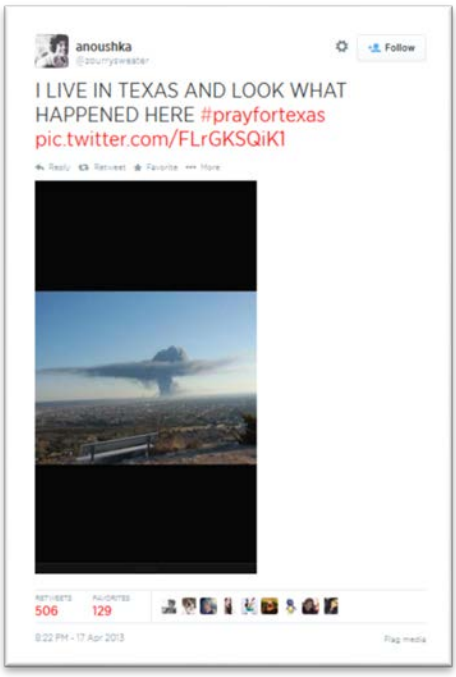
(kubena3955, 2013, http://ireport.cnn.com/docs/DOC-959388?ref=feeds/latest&utm_source=twitterfeed&utm_medium=twitter)

Although new media complements the traditional ways of conceptualizing information sources – sharing through authorities, peers, and broadcast media – Vultee

and Vultee suggest (2011), new media may confuse the audience when citizen-generated content is perceived as legitimate public journalism. Postings such as the example above, may have also worked against corrective behavior by appearing to be provided by a legitimate media outlet (trusted user) despite warnings on the site that the stories and photos had not been vetted by CNN.

Like other forms of media, Twitter's legitimacy is publicly defined – the retweeting of each other's posts, provides legitimacy and basic credibility to the message (Smith, 2010). The private citizen who can send a rumour through his own network (if he has a significant number of ties) may also complicate correction by increasing the number of ties a correction must now penetrate. For example, the tweet below was retweeted over 500 times:

Figure 6: TwitPic containing false mushroom cloud image



Users who retweeted the image from figure 6 above had anywhere from 171 to over 7,000 followers. This is presents a different problem for correction than the isolated audience of the CNN iReport discussed earlier. Rumours may have spread more quickly by the user 'anoushka', but further research would be required to reveal whether the same might hold true for a 'correcting' tweet received by this user.

Community Efficacy

The new digital age has been presented as a method to harness and leverage collective behavior. In the context of an emergency, an optimistic view of estranged participants suggests that they may temporarily work together towards a common goal; e.g. participants in an informal social network may successfully work together towards eliciting change. Thomas (2007) suggests that community efficacy may also foster the transmission of rumours. It could not be clearly demonstrated that community efficacy fostered the transmission of rumours and their correction. Examples within the three rumour streams suggest some potential for collective action but no evidence was found of collective action towards correcting misinformation:

> *Fertilzer is not GUNS, however, to have a Plant in middle of A Town! #Texas Answers @GovernorPerry also #georgeWbush FailureX2*
>
> @ScrewFoxNews; 4:21 AM - April 18, 2013

Correctors (those who provided 'correcting' tweets) may have been deterred by unexpected group behavior. The thread of tweets below demonstrates significant

criticism a user received over validating and correcting false information. In this example, users criticized a user's accurate reporting as an attempt to discredit the event and insult the victims of the explosion:

Original Post:

> *That picture of the enormous mushroom cloud from West Texas fertilizer explosion? Yeah, likely photoshopped. Twitter is almost as bad as CNN*

First Objection:

> *@getnickwright May want to delete that tweet. Looks like people actually got hurt.*

Second Objection:

> *@getnickwright I'd delete that tweet before someone gets upset...*

Response by Original Poster:

> *why would I want to delete the tweet? I said nothing about the seriousness of the explosion, just noted that viral photo was fake*

Third Objection:

> *If I were there in Boston during the bombing, and someone were to say the explosion looked photoshopped I'd be offended.*
>
> (complete feed from @getnickwright; 8:20 PM to 8:28 PM -April 17, 2013)

According to Rosnow (as cited by Pezzo & Beckstead, 2006) there is hesitancy to transmit an upsetting but implausible rumour as this may be associated with "crying wolf". Though this was suggested as a demotivator for rumour spread, the author found

that it may have been a motivation that prevented correction as demonstrated in the above example.

The type of event may also have had an influencing effect on the potential for group formation and collective action (in this case the spread of corrective information). The West Explosion was a hort duration event, as opposed to an emergency event with long term response and recovery phases. In events where response and recovery are drawn out over several weeks or months, there may be more opportunity for informal Twitter users to form groups whereby collective behaviours emerge.

Future Considerations

The results of this project were preliminary and based on a relatively small sample. Future research should expand the sample size to further validate any findings. It would also be valuable to compare rates and patterns of correction with events where emergency managers have played a more significant role in the social media sphere in the early stages of an event. Some users – "super correctors" - persisted in several samples to provide corrections and may be valuable to explore further to identify if and how they could be leveraged to help maintain accuracy in informal networks particularly in cases where emergency managers are not able to participate. It might be fruitful to explore the "cry wolf" scenario further to identify whether other events may contain similar rejection behavior towards those who present corrected information during a crisis. Lastly, it would be interesting to explore the effectiveness of other resources that emergency mangers have begun to use to control false rumours. In particular the rumour website, started by FEMA during Hurricane Sandy was designed for the specific purpose of

combating rumours in online communities (Ribeiro, 2012). If successful, this may be a less onerous resource for emergency managers to manage.

Conclusion

This research explored false rumour correction in Twitter networks, during the West, Texas Fertilizer Plant explosion on April 17, 2013. Three false rumours were explored with the hopes of deriving factors that could encourage people to correct information in Twitter. Various user types influenced the results of this research. In this incident, emergency managers and other first responder groups appear to have not been invested in social media relations, therefore correction did not stem from this particular group. Twitter users did not typically leverage traditional media as a reference to correct West, Texas rumours. On the contrary, traditional media played a significant role in creating/supporting misinformation, which then perpetuated into the Twitter network. Despite these limitations, some members of the informal Twitter community did work to resolve misinformation. It was however not possible to clearly identify specific conditions which encouraged correction. This might be better achieved by applying this methodology to another event to then provide a comparison of the effectiveness of some of the influential factors (role of media and emergency managers, effect of 'crying wolf', and the influence of 'super correctors'). Further research is also needed to determine what strategies and resources could be applied to encourage correction in informal communication during future emergency incidents.

The findings of this study suggest that misinformation in the early stages of a disaster does correct in some instances but not with sufficient frequency or spread to

ensure that accurate emergency information is shared within the online community. Emergency managers are therefore rightful to be concerned about the accuracy of information in Twitter, but would be hard pressed to rely on the online community to correct this information without intervention. The accuracy of information is also shared and broadcast by the traditional media, who did not appear to issue corrections to false information where needed. Though it is likely unreasonable to expect emergency managers to invest significant time in the social media conversation, particularly in the early stages of an emergency response, there may be alternate methods to encourage online accuracy and correction. More time-effective and resource friendly methods might include managing a single rumour website, to help highlight common misperceptions in the emergency.

It has been suggested that digital networks have enabled collective action (and in this case correction) due to "positive supply shock to the cost and spread of information, to the ease and range of public speech by citizens, and to the speed and scale of group coordination" (Shirky, 2011, para. 3). However, Shirky made careful note that "these changes do not allow otherwise uncommitted groups to take effective political action. They do, however, allow committed groups to play by new rules." This study may suggest that informal groups in the early stages of a disaster are not yet committed and therefore cannot effectively and collectively correct false rumours. The persistence of rumours in the early response stages of an emergency therefore remains a legitimate concern. There is however some correction that occurs even when emergency managers and other formal response organizations are unable to participate directly in the social networks, but this correction is limited.

Appendix A – Code Book

Definitions of Variables and Attributes:

Phase 1 Variable 1 – Category of Tweet Within Rumour Set: The intended purpose of tweet with each rumour set (is it the rumour, a question or challenge about the information, a correction/statement of alternative information to the rumour, or neither)

Phase 1 Variable 1 Attributes:

a) Information: Provide information about the event but does not question, correct or support the rumour
b) Misinformation: Provides/supports the rumour
c) Questioning: Questions and/or challenges the information (e.g. "is this true??", "I don't think this is true")
d) Correction: States the falsehood of the information and/or presents the truth.
e) Other/Unknown: cannot be clearly placed in the above categories, does not relate to the event at all or is unclear. Statements of hope or prayer will also fall into this category.

Phase 2 Variable 1 – User Information: the Twitter user's affiliation whether it is a personal account or they are speaking on behalf of an established entity.

Phase 2 Variable 1 Attributes:

a) Media – The user is a recognized member of, or is speaking as a media organization (TV, radio, print)
b) Official – The user is a recognized member of, or is speaking as an official emergency management organization, first responder organization or recognized non-profit who may assist in the event's emergency response.
c) Politician – The user is a recognized member of, or is speaking as a political representative (includes Mayor, Reeve, Senator, Marathon Official, Marathon Organizer, and any official representation of the White House).
d) Other organization - The user is a recognized member of, or is speaking as a representative of an established group or organization not recognized above.
e) Private citizen – The user is a private citizen, no recognizable group as listed above.
f) Anonymous – The user is unknown, could be a private citizen or a whistle blower but this cannot be confirmed (no details are available on the user beyond his handle).

Phase 2 Variable 2 - Type of Communication: the intended purpose of the communication whether it is simply to inform, to correct, to question or just to share with further participants.

Phase 2 Variable 2 Attributes:

a) Original source/new communication – the first instance of the rumour/information, will not be a retweet unless new information is significantly altered.
b) Retweet - Retweet and/or repeated information for the purpose of sharing, no significant change in content
c) Direct Reply – replies directly to a user (will generally contain an @ symbol).
d) Other – none of the above or it is unclear

Phase 2 Variable 3 – Referenced Source of Information: Is there an external reference source for the information included (is there an indication of where the original or shared information came from?). This may be indicated by a URL, hashtag or by @replies.

Phase 2 Variable 3 Attributes:

a) Media – A reference is included to a traditional media outlet (television, radio, print) tweet, article or report.
b) EM/FR Authority – A reference is included to an official organization's webpage, report or social media account. This organization must be recognized as part of the response and recovery efforts and may include First Responders and/or Emergency Management organizations. Reports through media should be coded as Media.
c) Politician – A reference is included to a political speech, report or quote from any political leadership group (includes Mayor, Reeve, Senator, and any official member of the White House) from their own report, organization's webpage, or social media account. Reports through media should be coded as Media.
d) Private citizen – A reference is included to another Twitter user who is not a member of a recognizable authority or media outlet.
e) Anonymous – A reference is included to an anonymous source.
f) No source indicated - A reference is not included. Tweet may also be the source.

References

A Neighborhood Nearly Obliterated by a Blast (2013). New York Times.com. Retrieved July 11, 2014, from http://www.nytimes.com/interactive/2013/05/10/us/damage-from-west-texas-explosion.html?_r=0

Allen, E. & Russell, J. (2013). Martin Richard, 8-year-old boy killed in Boston Marathon bombings, mourned by Dorchester neighbors - The Boston Globe. *BostonGlobe.com*. Retrieved March 22, 2014, from https://www.bostonglobe.com/2013/04/16/dorchester-neighbors-mourn-martin-richard-year-old-boy-killed-boston-marathon-bombings/keiXi55ZFf2YGityy16EuN/story.html

Associated Press (2013, April 17). KWTX reporting 60 to 70 dead at West, Texas fertilizer plant explosion - The Daily Texan. Retrieved March 22, 2014, from http://www.dailytexanonline.com/news/2013/04/17/kwtx-reporting-60-to-70-dead-at-west-texas-fertilizer-plant-explosion

Barclay, J. (2008). The issue of trust and its influence on risk communication during a volcanic crisis. *Bulletin of Volcanology, 70*(5), 605–621.

Barnes, C. (2012). Citizen Journalism vs. Traditional Journalism. *Caribbean Quarterly, 58*(2/3), 16–27.

Bidgood, J. (2013, April 25). Sunil Tripathi, Student at Brown, Is Found Dead. *The New York Times*. Retrieved from http://www.nytimes.com/2013/04/26/us/sunil-tripathi-student-at-brown-is-found-dead.htmlBordia, P., & DiFonzo, N. (2002). When social psychology became less social: Prasad and the history of rumor research. *Asian Journal of Social Psychology, 5*(1), 49.

Bordia, P., & DiFonzo, N. (2002). When social psychology became less social: Prasad and the history of rumor research. *Asian Journal of Social Psychology*, *5*(1), 49.

Bordia, P., & DiFonzo, N. (2004). Problem Solving in Social Interactions on the Internet: Rumor As Social Cognition. (2004). *Social Psychology Quarterly*, *67*(1), 33–49.

Boston Bombings: Wrong Suspect Reveals Fears. (n.d.). *Sky News*. Retrieved June 15, 2014, from http://news.sky.com/story/1080260

Bruns, A., & Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, *17*(4), 3–3.

Bruns, A., & Stieglitz, S. (2012). Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services*, *30*(3/4), 160–185. doi:10.1080/15228835.2012.744249

Chan, T. C., Killeen, J., Griswold, W., & Lenert, L. (2004). Information technology and emergency medical care during disasters. *Academic Emergency Medicine*, *11*(11), 1229–1236. doi:10.1197/j.aem.2004.08.018

Crowe, A. (2011). The social media manifesto: A comprehensive review of the impact of social media on emergency management. *Journal of Business Continuity & Emergency Planning*, *5*(1), 409–420.

Deneen, S. (2013). Rumor: Girl, 8, running Boston Marathon died in bombing. Retrieved March 22, 2014, from http://news.msn.com/rumors/rumor-girl-8-running-boston-marathon-died-in-bombing

DuckDuckGo Privacy. (n.d.). *DuckDuckGo*. Retrieved July 6, 2014, from https://duckduckgo.com/privacy

Eburn, M. (n.d.). Media access to emergencies: Command, control or co-ordination? *Australian Journal of Emergency Management, The, 25*(1), 13.

Finfgeld-Connett, D. (2014). Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. *Qualitative Research, 14*(3), 341-352. doi:10.1177/1468794113481790

Fraustino, J. D., Liu, B. & Jin, Y. (2012). Social media use during disasters: A review of the knowledge base and gaps. *Final Report to Human Factors/Behavioral Sciences Division, Science and Technology Directorate, U.S. Department of Homeland Security.* College Park, MD: START, 2012.

Haynes, K., Barclay, J., & Pidgeon, N. (2008). The issue of trust and its influence on risk communication during a volcanic crisis. *Bulletin Of Volcanology*, (5), 605.

Heverin, T., & Zach, L. (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),* 34–47. doi:10.1002/asi.21685

Hiltz, S. R., & Gonzalez, J. J. (2012). Assessing and improving the trustworthiness of social media for emergency management: A literature review. *Norwegian Information Security Conference / Norsk Informasjonssikkerhetskonferanse*, 6.

How social media are changing crisis communications - for better and worse. (2009). *Security Director's Report*, *9*(11), 2-5.

Hughes, A. L., & Palen, L. (2012). The evolving role of the public information officer: An examination of social media in emergency management. *Journal of Homeland Security & Emergency Management*, *9*(1), 1–20.

International Communication Association. (2011). Twitter and Disasters: the uses of Twitter during the 2010 Pakistan floods. *Conference Papers*, 1–30, Presented at 2011 Annual Meeting.

Jaeger, P. T., Shneiderman, B., Fleischmann, K. R., Preece, J., Qu, Y., & Wu, P. F. (2007). Community response grids: E-Government, social networks, and effective emergency management. *Telecommunications Policy, 31*(10-11), 592–604.

Kern, E. (2012, November 29). As the firehose matures, Twitter tightens grip on valuable asset. *Gigaom*. Retrieved from http://gigaom.com/2012/11/29/as-the-firehose-matures-twitter-tightens-grip-on-valuable-asset/

Krippendorff, K. (2013). *Content analysis: an introduction to its methodology* (3rd ed.). Los Angeles; London: SAGE.

Mannix, L. (2011). *Queensland floods: media's looting reports put lives at risk.* Retrieved from http://www.crikey.com.au/2011/01/14/queensland-floods-medias-looting-reports-put-lives-at-risk/

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114-133. doi:10.1177/1461444810365313

Mendoza, M., Poblete, B., & Castillo, C. (2010). Twitter under crisis. *International Conference on Knowledge Discovery & Data Mining*, 71.

Mileti, D. S. (1999). Disasters by design: A reassessment of natural hazards in the United States. Washington, D.C.: Joseph Henry Press, 1999.

Observations on Twitter and the ACA - Health Reform & Public Programs Initiative - Programs - CHCF.org. (2014). Retrieved July 11, 2014, from http://www.chcf.org/programs/healthreform/aca140

Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*, *37*(2), 407–A7.

O'Leary M (2004). The first 72 hours: A community approach to disaster preparedness. Lincoln (Nebraska), iUniverse Publishing. http://www.iuniverse.com/bookstore/book_detail.asp?&isbn=0-595-31084-2

Oremus, W. (2013, April 18). This mushroom cloud is not from the fertilizer-plant explosion. *Slate*. Retrieved from http://www.slate.com/blogs/future_tense/2013/04/18/is_this_texas_mushroom_cloud_photo_real_yes_but_it_s_from_2008.html

Palen, L., Hiltz, S. R., & Liu, S. B. (2007). Online forums supporting grassroots participation in emergency preparedness and response. *Communications of the ACM*, *50*(3), 54–58.

Pezzo, M. W. & Beckstead, J. (2006). A multilevel analysis of rumor transmission: Effects of anxiety and belief in two field experiments. *Basic & Applied Social Psychology*, *28*(1), 91-100.

Ribeiro, J. (2012). FEMA battles online misinformation with rumor control. TechHive. Retrieved August 8, 2014, from http://www.techhive.com/article/2013516/fema-battles-online-misinformation-with-rumor-control.html.

Shirky, C. (2010). Tapping the Cognitive Surplus. *Futurist*, *44*(6), 21-22.

Shirky, C. (2011). From innovation to revolution: Do social media make protests possible?: Shirky replies. *Foreign Affairs*, *90*(2), 153-154.

Smith, B. G. (2010). Socially distributing public relations: Twitter, Haiti, and interactivity in social media. *Public Relations Review*, *36*(4), 329–335. doi:10.1016/j.pubrev.2010.08.005

Starbird, K., Maddock, J., Orand, M., Achterman, P., Mason, R. M. (2014). Rumors, false flags and digital vigilantes: Misinformation on Twitter after the 2013 Boston marathon bombing. *iConference 2014*. Retrieved March 20, 2014 from http://faculty.washington.edu/kstarbi/Starbird_iConference2014-final.pdf

Stephenson R. & Anderson P.S. (1997). Disasters and the information technology revolution. *Disasters*, *21*(4), 305–334.

Thomas, S. A. (2007). Lies, Damn Lies, and Rumors: An Analysis of Collective Efficacy, Rumors, and Fear in the Wake of Katrina. *Sociological Spectrum*, *27*(6), 679–703. doi:10.1080/02732170701534200

Vultee, F., & Vultee, D. M. (2011). What we tweet about when we tweet about disasters: The nature and sources of microblog comments during emergencies. *International Journal Of Mass Emergencies & Disasters*, *29*(3), 221-242.

Wigley, S., & Fontenot, M. (2010). Crisis managers losing control of the message: A pilot study of the Virginia Tech shooting. *Public Relations Review*, *36*(2), 187–189. doi:10.1016/j.pubrev.2010.01.003

Zhang, Z., & Zhang, Z. (2009). An interplay model for rumour spreading and emergency development. *Physica A*, *388*(19), 4159–4166. doi:10.1016/j.physa.2009.06.020

Zhang, Y., Zhou, S., Zhang, Z., Guan, J. & Zhou, S. (2013). Rumor evolution in social networks. *Physical Review E: Statistical, Nonlinear & Soft Matter Physics*, *87*(3-A), 1-5. doi:10.1103/PhysRevE.87.032133

Zhao, L., Cui, H., Qiu, X., Wang, X., & Wang, J. (2013). SIR rumor spreading model in the new media age. *Physica A*, *392*(4), 995–1003. doi:10.1016/j.physa.2012.09.030