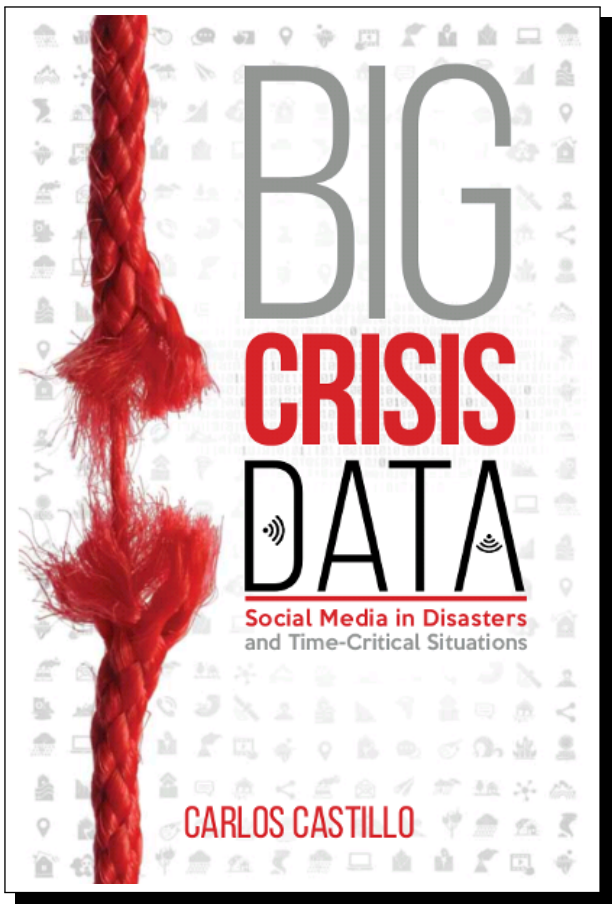
Big Crisis Data

Social Media in Disasters and Time Critical Situations

by Carlos Castillo



FREE PREVIEW
CHAPTER 8. VERACITY

Cite as: Carlos Castillo: Big Crisis Data. Cambridge University Press, 2016

BibTex:

```
@book{castillo2016big,
     title="Big Crisis Data",
     author="Castillo, Carlos",
     publisher="Cambridge University Press"
     year=2016
}
```

Contents of Chapter 8

8

Veracity: Misinformation and Credibility

In August 2014, CBS News published a story and a cellphone photo of a bizarre meteorological phenomenon. The reporter used a photo provided by a tugboat captain, who stated that he was not a meteorologist but described the image as a rare "sideways tornado." The phenomenon is actually more than rare: it does not exist. The reporter could have consulted with the TV station's meteorologist, who later easily identified the photo as a shelf cloud. The story was pulled off their website and then amended, but the embarrassment for the news network did not go away.[1]

Hoaxes in media are centuries old. Noted satirists such as Jonathan Swift in the seventeenth century and Mark Twain in the eighteenth were successful at spreading them well before the Internet (Walsh, 2006). Disaster-related media hoaxes predate the Internet by decades. A famous example was the 1938 radio adaptation of the alien-invasion novel by H. G. Wells, *The War of Worlds*, which at the time caused numerous calls to newspapers and the police, and created a significant media backslash for (unintentionally, according to its producers) "deceiving" the listeners.[2] Social media simply places the tools necessary to create and spread all kinds of information, including hoaxes, on the hands of many.

This chapter deals with concerns about the presence of false information in social media, which are frequently cited as one of the major obstacles to its adoption by humanitarian and emergency relief organizations (Hiltz et al., 2011; Hughes et al., 2014). Officers at these organizations have said that they often find themselves wondering if they can trust a given piece of information from social media, or not. Some of them believe that social media are more likely than other sources to contain bad, false, unverified, or inaccurate information (Bressler et al., 2012; Vieweg et al., 2014). Emergency managers, who may also want to integrate information provided by the community, have also expressed doubts about the accuracy and reliability of social media (Merrick and Duffy, 2013). While disaster response organizations are used to operate with "good enough" information during emergencies, they seem to hold higher, even "unreasonable" standards of accuracy for data from social media (Tapia and Moore, 2014).

In social media during emergencies, it is easier to find a message describing something that is true, than it is to find a message describing something that is false (Mendoza et al., 2010). However, even if false information constitutes a minority of what is posted in social media, the negative consequences – practical, economical, legal, political – of trusting false information can be potentially large. Techniques such as human computation and machine learning can be used to provide more elements for experts to make this decision.

We begin this chapter with a general discussion of trust issues in social media during disasters (§8.1), explain how those are usually addressed within a policy-based model of trust (§8.2), and the difference between misinformation and disinformation (§8.3). Then, manually intensive methods based on crowdsourced verification are described (§8.4), followed by methods based on automatic supervised learning and hybrid methods (§8.5).

[1] "CBS News Falls for Hoax, Reports on Nonexistent 'Sideways Tornado'." *Gawker*, August 2014. http://thevane.gawker.com/cbs-news-falls-for-hoax-reports-on-nonexistent-sidewa-1617162073

[2] "The Myth of the War of the Worlds Panic." Jefferson Pooley and Michael J. Socolow, *Slate*, October 2013. http://www.slate.com/articles/arts/history/2013/10/orson_welles_war_of_the_worlds_panic_myth_the_infamous_radio_broadcast_did.html.

8.1 Emergencies, Media, and False Information

Criticism of media coverage of emergency situations is not new. A 1957 report published by the U.S. National Academy of Sciences raised serious issues about the way in which the media, specially the radio, informed and misinformed about disasters: "The methods of handling and reporting disasters by the mass media of communication have varied over a wide spectrum from the highly sensational, dramatic, false, and distorted at one end to the sane, verified, factual, and complete at the other" (Fritz and Mathewson, 1957).

The fact that reporters from mainstream media are increasingly relying on social media sources during crises, has opened a new front for criticism. Disclaimers indicating that information could not be independently verified, which tend to appear when social media photos or videos are used in news reporting, do not exempt media from being a target when those photos or videos convey the wrong information. For instance, in 2012 the BBC used an old photo from the war in Iraq to illustrate a more recent massacre in Syria.[3] In 2013, during the manhunt that followed the Boston Marathon bombings, the *New York Post* published on its front page the photos of two innocent bystanders who were mistakenly identified as suspects by the social media site Reddit. After the real culprits were located, U.S. President Barack Obama chastised the press: "In this age of instant reporting and tweets and blogs, there's a temptation to latch onto any bit of information, sometimes to jump to conclusions. But when a tragedy like this happens, with public safety at risk and the stakes so high, it's important that we do this right. That's why we have investigations. That's why we relentlessly gather the facts."[4] The Red Cross also advocates caution when dealing with unverified information: "Protection actors should take measures to minimize the risk of presenting a false or incomplete image of the issues they intend to address. In a crisis situation, a protection actor may feel under pressure to communicate findings that are not fully verified. When this happens, it is important to avoid hastily extrapolating firm conclusions, or being overly affirmative" (ICRC, 2013).

Social media provides a wealth of immediate information about developing situations, with multiple voices and perspectives that simply cannot be ignored by traditional news media and by humanitarian organizations. For these organizations, dealing with unverified information is often not a matter of choice, but a necessity. A degree of uncertainty is inevitable, and the extent to which information should be trusted is a matter that depends on weighting trade-offs that vary according to the situation (Tapia et al., 2013; Tapia and Moore, 2014). In other words, the risk of acting on incomplete information should not prevent an important operation from going ahead: "a lack of fully verified information is no reason for inaction when there are compelling reasons to suspect that violations have been committed, and might be repeated" (ICRC, 2013).

Timing is important for response and relief organizations: at the onset of a emergency response effort, when time presses but information is limited, there is more risk in ignoring social media information even if it has not been completely validated. As the situation evolves, the role of social media may be less central as other sources of information emerge.

Timing is also important for social media users. In time-critical situations, they have a preference for incomplete but early information over complete but late information. For instance, earthquake alerts from seismic sensors are usually validated carefully by experts before they are issued, but a study by Comunello et al. (2015) indicated that users strongly prefer an immediate message in social media about an earthquake with a provisional estimate of its magnitude (marked "provisional estimate"), instead of a message ten to twenty minutes later, once the magnitude has been confirmed.

Additionally, social media information, as other types of information, is affected by *information expiration*, that is, something that was valid before but is no longer valid now (Vieweg et al., 2014). Internet users may contribute to this problem. For instance, more people posted about a warning message concerning a shooting at a university, than about the notification that this warning had been lifted (Hui et al., 2012; Tyshchuk et al., 2012).

[3] "BBC News uses Iraq photo to illustrate Syrian massacre." Hannah Furness, *Telegraph*, May 2012. http://www.telegraph.co.uk/news/worldnews/middleeast/syria/9293620/BBC-News-uses-Iraq-photo-to-illustrate-Syrian-massacre.html

[4] "Statement by the President." *The White House, Office of the Press Secretary*, April 2013. https://www.whitehouse.gov/the-press-office/2013/04/19/statement-president.

8.2 Policy-Based Trust and Social Media

There are various ways in which we can approach the problem of trust, including: policy-based trust, reputation-based trust, and trust in information resources (Artz and Gil, 2007).

Policy-based trust means that only specific people and organizations are trusted for a given information category. Traditionally, governments, humanitarian, and emergency response organizations follow this model, which is to some extent mirrored by large Internet media dealing with emergency response. Information is trusted because it comes from an official source (Tapia and Moore, 2014). For instance, Twitter maintains a list of verified emergency-related accounts – only these accounts are authorized to push emergency notifications in case of a disaster.[5]

Reputation-based trust, or *source trust*, seeks to quantify to what extent we should trust on a certain source, based on the evidence we have collected so far. *Trust in information resources*, or *content trust*, seeks to quantify to what extent we should trust on a given piece of content, again based on the evidence we have collected so far. Source trust and content trust are inseparably linked: a trusted source produces and shares trusted content, and trusted content is produced and shared by trusted sources.

The main problem with policy-based trust in social media, is that it forces a binary decision (to trust or not to trust) on an issue that is a matter of degree. Trustworthy information is acquired by collecting and contrasting information from multiple sources, and seldom yields perfectly reliable information: "When in doubt, the information should be tagged as unverified. Several levels of reliability may be used when deciding to tag the reliability of information obtained through open sources" (ICRC, 2013).

Intelligence analysts in the NATO military as well as those of Australia and New Zealand use a two-dimensional scale to grade intelligence. This includes a dimension of reliability of the source, and another of credibility of the information; both are evaluated independently. Source reliability is expressed on a scale from "A" (source completely reliable) to "E" (unreliable source), with "F" being unknown source reliability; information credibility is expressed on a scale from 1 (confirmed by other sources) to 4 (improbable), with 5 being unknown credibility (U.K. Ministry of Defence, 2014).

Internet users often trust social media as much as they trust traditional news media sources, even for contentious issues such as political elections (see, e.g., ORI Market Research, 2012). People do not express feelings of being misled, suspicious or mistrustful about social media information, as often as they express that they find that information in social media is useful (Taylor et al., 2012). However, they also show a strong propensity for referencing traditional sources including high-reputation news media outlets, official government sources, and well-established nongovernmental organizations (Thomson et al., 2012; Reuter et al., 2015b).

To some extent, Internet users also operate on a policy-based trust model, but their trust policies are more fluid than the ones applied by emergency and humanitarian organizations, and more dependent on the context. For instance, users in Pakistan perceive social media as more truthful during disasters than traditional media (Murthy and Longwell, 2013), while Japanese users believe the exact opposite, and rate social media as much less trustworthy than traditional media (Hokudai Earthquake Project, 2011).

8.3 Misinformation and Disinformation

One of the simplest ways of operationalizing trust – and by no means the only way – is to consider three elements: competence, benevolence, and integrity (Gefen, 2002).[6] Competence refers to the ability to give accurate information, benevolence to the willingness to do the effort, and integrity to honest behavior. Essentially, the main concern with respect to social media is the lack of any of these elements, that is, people not being in the position to provide precise information, not being willing to do so, or not being truthful about what they know.

Failures in the sense of competence are normally associated to *misinformation*, which is unintentionally spread; failures in integrity are normally associated to *disinformation*, which is deliberately spread (Stahl, 2006). Failures

[5] "Twitter Alerts: Critical information when you need it most." Gabriela Pena, *Twitter Blog*, September 2013. https://blog.twitter.com/2013/twitter-alerts-critical-information-when-you-need-it-most.

[6] Some authors add a fourth dimension, predictability (McKnight and Chervany, 2001), and others use just two factors: ability and intent (Thomson et al., 2012).

in benevolence may be related to obstacles or lack of incentives to provide information (bringing us back to the study of motivations of Section 7.3).

Misinformation tends to be more present than disinformation during natural disasters. Users may repost information that they have not verified yet, possibly weighing the cost of a false alarm versus the cost of not raising an alarm when they should.

Disinformation, while also present during disasters from natural hazards, tends to appear more often during human-induced crises. Lewandowsky et al. (2013) highlight that, given that perceptions of populations are becoming as important as the will of armies and governments to fight, controlling information is becoming a central aspect of modern warfare. In general, when there are two sides in conflict, or when there are legal liabilities or political consequences expected from an event, some people may have incentives to spread deceptive information.

In March 2015 it was uncovered that a group from Russia was using a number of "fake" social media accounts to propagate fake screenshots purporting to be of CNN reporting an explosion in Centerville, Louisiana, in the United States.[7] In June 2015 it was reported that this was an organization of about four hundred employees tasked with posting disinformation online.[8]

Some online disinformation campaigns involve a small set of identities interacting heavily with each other, but not with other, "regular" users. This type of campaign can be detected using graph-based methods, uncovering groups of users that are likely to be colluding (Ratkiewicz et al., 2011).

8.4 Verification Practices

Humanitarian organizations increasingly recognize that there is a trade-off between accuracy and speed (ICRC, 2013). Navigating this trade-off is an everyday activity for many journalists, who can provide valuable input to humanitarians regarding information verification practices. An ideal of professional journalism is that "being right is more important than being first," but implementing this ideal in reality is far from trivial.[9]

Over the years, a series of practices have emerged regarding verification of social media during emergencies. Many of these practices can be found in the "Verification Handbook" edited by Silverman (2014), which is a guide co-authored by journalists and other professionals of the media and emergency relief sectors. This handbook includes among other elements a taxonomy of types of fake disaster-related information online: (i) real photos from unrelated events; (ii) art, ads, film, and staged scenes; (iii) digitally edited images, (iv) social media accounts impersonating someone else (fake accounts); (v) altered social media reposts; (vi) fake social media screenshots; and (vii) fake websites.

The Verification Handbook also outlines a verification process or checklist for verifying content. First, determine the first occurence of this content online, to establish its *provenance*. Second, identify the author to establish whether the *source* of the information is reliable. Third, look at key aspects of the *content*, including its location and date. There are numerous tools that can be used during this process, including for instance reverse image search systems, that allow to search the Web for similar images to a given one.

Common journalistic practices, such as contacting and interviewing sources, have also been applied by emergency managers to deal with social media content. Latonero and Shklovski (2011) interview the Public Information Officer of the Los Angeles Fire Department in the United States. In this interview, the officer recalls how he sent an e-mail to a person posting information, asking him to call back to his office at the fire department, and discussed on the phone the account of this eyewitness. "I felt I was able to add reasonable validation of what they were seeing, relayed that information to our responders in the field, and it turned out that there were people and property in danger in that area, things we couldn't see, that were over the horizon away from us."

Information verification platforms. Many verification practices can be codified in systems that assist users in evaluating the veracity of a claim posted in social media. Reuter et al. (2015a) describe a software where users can select and weight evaluation criteria from a predefined list, and use simple user interface elements such as sliders

[7] "Russia Is Hacking Your News Feed." Leonid Bershidsky, *Bloomberg View*, March 2015. http://www.bloombergview.com/articles/2015-03-11/russia-is-hacking-your-news-feed.

[8] "The Agency." Adrian Chen, *New York Times*, June 2015. http://mobile.nytimes.com/2015/06/07/magazine/the-agency.html.

[9] "Inside the BBC's Verification Hub." David Turner, *Nieman Reports*, July 2012. http://niemanreports.org/articles/inside-the-bbcs-verification-hub/.

to combine ratings across these criteria. This software also integrates automatically computed elements that can help evaluate veracity, such as the distance between the location of a user and the place where the crisis is taking place.

The verification practices used by journalists, analysts, and emergency responders cannot scale to a large volume of information, as they are limited by factors such as the size of the staff in these organizations. However, these practices can be scaled up with the help of volunteers through crowdsourcing. In a crowdsourced verification platform, volunteers are invited to look at claims posted by journalists or emergency managers (e.g., a photo showing the consequences of severe weather), and are asked to provide evidence in favor or against that photo. An example of such system is Verily,[10] introduced by Popoola et al. (2013).

Visible skepticism. Messages containing false information are frequently discredited or questioned in social media (Sutton, 2010; Castillo et al., 2013; Resnick et al., 2014). These self-correction mechanisms of social media are valuable, but they do not solve the problem of misinformation, because many people are exposed to false information but not to the messages correcting it (Resnick et al., 2014; Carton et al., 2015). However, this suggests a way of dealing with false information in social media: by encouraging users to discuss about dubious content in social media itself. For instance, users can be encouraged to post messages about incorrect information, as has been done in several emergencies by using the *#Mythbuster* hashtag.[11]

There is a temptation to try to remove or censor false information, but this may not be the best strategy in comparison with the practice of *visible skepticism*: "Repeatedly allowing a rumor to surface and be corrected is different from correcting it once and then blocking it from surfacing again. The latter stops the rumor's spread within the current platform, but the former may do more to quell its spread through other channels, by challenging it as it comes up again and again" (Dailey and Starbird, 2014). When users are exposed to the refutation of a message simultaneously to the message itself, they are significantly less likely to share dubious content, and repeated exposure to information countering a rumor is more likely to reduce users' propensity to repeat that rumor than a warning such as "this tweet may contain misinformation" (Ozturk et al., 2015).

8.5 Automatic Credibility Analysis

Aristotle's *Rethoric* offered a blueprint of persuasion based on three key elements, which are often succinctly described as credibility (*ethos*), emotions (*pathos*) and reason (*logos*). These elements can be used to describe methods for automatic credibility analysis. **Automatic reasoning.** Only a few initial steps have been done on automating the verification of information based on patterns of reasoning, that is, the *logos* aspect in the *Rethoric*. Dong et al. (2015) introduced Knowledge-Based Trust (KBT), which is computed by automatically extracting facts from a Web page (i.e., triples of the form <*subject, object, predicate*>), and then determining to what extent those facts agree with facts posted on other Web pages. The main technical difficulty is that with current technologies, the error rate of information extractors is actually higher than the rate at which people introduce factual errors in content. Dong et al. (2015) address this problem by incorporating the error rate of the information extraction into the probabilistic framework used to estimate how much trust to place in an extracted fact. **Sharing, refutation, and questioning.** While Mendoza et al. (2010) did not find a difference between how much truthful information and false rumors were retweeted in Twitter, they did note that information that turned out to be false was significantly more questioned and refuted. Questioning and refutation of information tends to appear at latter stages in the lifetime of a rumor. This was depicted by *The Guardian*'s study about the diffusion of news stories during the 2011 London riots.[12] A time slider can be used to interact with this visualization, making it evident that in many of these rumors, there is an initial period at which false messages "peak," followed by a period in which refutation messages are more prevalent, but in general in smaller quantities than the false messages that preceded it.

The fact that false rumors are more likely to be questioned and refuted can be used in automatic systems that

[10] Verily Crowdsourced Verification. http://veri.ly/.

[11] "Using #Mythbuster Tweets to Tackle Rumors During Disasters." Patrick Meier, *iRevolution*, January 2013. http://irevolution.net/2013/01/27/mythbuster-tweets/.

[12] "Reading the Riots." *The Guardian*, December 2011. This visualization uses propagation histories, described on Section 5.2. http://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter.

Table 8.1. *Partial list of features used in automatic information credibility systems for Twitter, from Castillo et al. (2011, 2013); Gupta and Kumaraguru (2012); Gupta et al. (2014).*

Type	Examples
Content	Features computed from the content of the message, such as: (1) Length of the tweet, (2) number of words, (3) number of unique characters, (4) number of hashtags, (5) number of retweets, (6) number of swear language words, (7) number of positive sentiment words, (8) number of negative sentiment words, (9) tweet is a retweet, (10) tweet is a reply, (11) number of special symbols [$, !], (12) number of emoticons [:-), :-(], (13) Number of @-mentions, (14) number of retweets, (15) time lapse since the query, (16) has URL, (17) number of URLs, (18) use of a URL shortener service
User	Features computed from the author of the message, such as: (1) registration age of the user, (2) number of messages posted, (3) number of followees, (4) number of followers, (5) ratio of followers to followees, (6) number of user-created lists in which it is included, (7) number of user-created lists created, (8) is a verified account, (9) length of self-description ("bio"), (10) length of screen name, (11) has a URL
Propagation	Features computed from a message's reposting history, including aggregates computed from content and user features, such as: (1) fraction of tweets that contain URLs, (2) fraction of tweets with hashtags, (3) fraction of sentiment words, positive and negative, (4) depth of reposting tree, (5) number of reposts

detect credibility, by grouping messages belonging to the same story and then performing a per-cluster analysis of the data (Castillo et al., 2011).

Information credibility. The majority of research on automatic methods to help people decide how much to trust on a given content, has been based on the notion of *information credibility*, which is more closely related to the concept of *ethos* in the *Rethoric*. *Ethos* is established through the credentials of an author, bringing us back to policy-based trust – however, it is also established through tone and style.

Fogg and Tseng (1999) describe information credibility as the quality of information being believable, and emphasize that this is a perceived quality that is made from multiple elements. Automatic systems that can emulate human perceptions of credibility, that is, predict whether a person would say that she believes in a given message, are well within grasp of current computational methods, particularly supervised learning. Two main elements are required: being able to collect ground truth from users, and being able to computationally model credible content. Ground truth for credibility is easier to obtain than from most other aspects of content trust. Determining if a piece of content is true or false may be extremely time consuming and often requires a great amount of background knowledge or context. In contrast, determining if a piece of content is believable is something we do every day on an intuitive basis, and hence tends to be a relatively fast operation.

Textual features. Basic methods to distinguish between credible and non-credible messages use a series of textual features computed from the text. These are similar to the stylometric features used by the NLP community for problems related to authorship attribution (Stamatatos, 2009).

To model the credibility of information on Twitter during emergencies, Castillo et al. (2011, 2013) use a series of content-based, user-based features, and propagation-based features which are used as input for a supervised classifier using a decision tree algorithm. As preprocessing, messages that are not deemed of interest by another automatic classifier are discarded from the set. Gupta and Kumaraguru (2012) use content- and user-based features as inputs to an SVM classifier. Their dataset consists of 14 crisis events, and their system is implemented as a browser plugin in Gupta et al. (2014). Some of the features used in these works are listed in Table 8.1.

These methods depend on human annotation for training, and as such, assume that people can to a certain degree agree on whether a message is credible or not. McCreadie et al. (2015) note that this is in general true, but the degree of agreement depends on the type of message and the way in which a question is framed. For instance, asking people whether a message contains controversial or disputed information tends to elicit less agreement than asking people whether a message states a fact.

With respect to the evaluation of methods for estimating information credibility, ideally this should be done on publicly accessible collections. *CREDBANK* (Mitra and Gilbert, 2015) is a reference collection of social media messages annotated with credibility assessments.

Topics and expertise. A different research direction attempts to model explicitly a form of source trust, that is, the expertise of authors on different topics, and assign more credibility to the messages they write on topics in which they are likely to be experts. Expertise modeling can be done, for instance, by studying the contents users produce

and how those contents propagate (Tang et al., 2009). These models of expertise have been used to identify and rank potentially credible or trustworthy users for a given topic (Canini et al., 2011; Zielinski et al., 2013).

With respect to crisis-related messages, Ito et al. (2015) describe a method for computing credibility using topics discovered by unsupervised methods. This is done by first running LDA to compute the topics of a user and the topics of a message. Two types of basic features can be derived from these, to measure the concentration of topics (e.g., whether the user repeatedly posts about the same topic) and to measure the correspondence of topics (e.g., whether the user's previous posts match the topic(s) of the current post).

Some people take advantage of the popularity of a certain hashtag, which means many people are searching for that hashtag, to write messages publicizing certain products (Earle et al., 2010). Spam detection methods for social media can be applied to remove these messages (Benevenuto et al., 2010), as well as bots (as described on Section 2.3). This is also related to methods for finding false product reviews, which is a well-studied subject that has many elements in common with the problem of determining whether a crisis-related message in social media should be trusted. Methods for detecting false product reviews are overviewed in Liu (2012, ch. 10).

8.6 Research Problems

Extending automatic verification methods. Content-based methods for verification are an interesting research direction, but they require multiple reports referring to the same situation, which may or may not be available at a given time. Logical inconsistencies in a message are one potential sign that the information is incorrect, but they are only one of many possible reasons in which information may be incorrect. More research is needed to understand to what extent these methods can contribute to verify claims done on social media during crises.

Creating new human verification methods. Rumors are a problem-solving strategy, a way of rapidly improvising an interpretation of an ambiguous situation (Shibutani, 1966); as such, they are inherent to sense-making. Understanding the verification practices used by professional journalists and/or emergency response organizations is key to be able to assist them. Verification practices for social media should also evolve as the technology matures.

Modeling the credibility of other contents generated by the public. Many types of emergency-related communications have to deal routinely with false information. For instance, every year there are literally millions of false calls to 911, the main emergency number used in the United States.[13] In the United Kingdom, the Metropolitan Police receives tens of thousands of misuse or hoax calls every year.[14] Computing methods developed for social media can be extended to deal with these types of false information during emergencies.

8.7 Further Reading

Artz and Gil (2007) present a brief conceptual framework for trust research in computer science. Sherchan et al. (2013) present a survey on trust in social media.

Hermida (2014, ch. 8), describes misinformation and disinformation in social media from the perspective of journalists and online content producers. Meier (2015, ch. 7 and 8) addresses verification of crisis data using human and machine intelligence.

[13] See, e.g., "City flooded with nearly 4 million inadvertent 911 calls on cell phones a year." Juan Gonzalez, *New York Daily News*, May 2012. https://www.nydailynews.com/new-york/city-flooded-4-million-inadvertent-911-calls-cell-phones-year-article-1.1074752 or "Fake 911 calls aren't cheap." Ed Boyle, *CBS News London*, June 2012. http://www.cbsnews.com/news/fake-911-calls-arent-cheap/.

[14] "999 'time wasters': Where do emergency services draw the line?" Alex Homer, *BBC News*, August 2014. http://www.bbc.com/news/uk-england-28562807.

Bibliography

Artz, Donovan, and Gil, Yolanda. 2007. A survey of trust in computer science and the semantic Web. *Web Semant.*, **5**(2), 58–71.

Benevenuto, Fabrıcio, Magno, Gabriel, Rodrigues, Tiago, and Almeida, Virgılio. 2010. Detecting spammers on Twitter. Pages 75–83 of: *Proceedings of 7th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference CEAS*, vol. 6. Redmond, Washington, USA: CEAS Conference.

Bressler, George H., Jennex, Murray E., and Frost, Eric G. 2012. Exercise24: Using social media for crisis response. *The World Financial Review*, Mar., 77–80.

Canini, Kevin Robert, Suh, Bongwon, and Pirolli, Peter L. 2011. Finding credible information sources in social networks based on content and social structure. Pages 1–8 of: *Proceedings of IEEE 3rd International Conference on Privacy, Security, Risk and Trust PASSAT*. Boston, Massachusetts, USA: IEEE.

Carton, Samuel, Park, Souneil, Zeffer, Nicole, Adar, Eytan, Mei, Qiaozhu, and Resnick, Paul. 2015. Audience analysis for competing memes in social media. Page In Press of: *Proceedings of 9th International AAAI Conference on Web and Social Media*. Oxford, UK: AAAI Press.

Castillo, Carlos, Mendoza, Marcelo, and Poblete, Barbara. 2011. Information credibility on Twitter. Pages 675–684 of: *Proceedings of 20th International Conference on World Wide Web Conference (WWW)*. Hyderabad, India: ACM.

Castillo, Carlos, Mendoza, Marcelo, and Poblete, Barbara. 2013. Predicting information credibility in time-sensitive social media. *Internet Research*, **23**(5), 560–588.

Comunello, Francesca, Mulargia, Simone, Polidoro, Piero, Casarotti, Emanuele, and Lauciani, Valentino. 2015. No misunderstandings during earthquakes: Elaborating and testing a standardized tweet structure for automatic earthquake detection information. In: *Proceedings of 12th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Kristiansand, Norway: ISCRAM.

Dailey, Dharma, and Starbird, Kate. 2014. Visible skepticism: Community vetting after Hurricane Irene. In: *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. University Park, Pennsylvania, USA: ISCRAM.

Dong, Xin L., Gabrilovich, Evgeniy, Murphy, Kevin, Dang, Van, Horn, Wilko, Lugaresi, Camillo, Sun, Shaohua, and Zhang, Wei. 2015. Knowledge-based trust: Estimating the trustworthiness of Web sources. Pages 938–949 of: *Proceedings of 41st International Conference on Very Larga Data Bases VLDB*. Kohala, Hawaii, USA: PVLDB.

Earle, Paul, Guy, Michelle, Buckmaster, Richard, Ostrum, Chris, Horvath, Scott, and Vaughan, Amy. 2010. OMG earthquake! Can Twitter improve earthquake response? *Seismological Research Letters*, **81**(2), 246–251.

Fogg, BJ., and Tseng, Hsiang. 1999 (May). The elements of computer credibility. Pages 80–87 of: *Proceedings of Conference on Human Factors in Computing Systems (CHI)*. ACM, Pittsburgh, Pennsylvania, USA.

Fritz, Charles E., and Mathewson, John H. 1957. *Convergence behavior in disasters: A problem in social control: a special report prepared for the committee on disaster studies*. National Academy of Sciences National Research Council.

Gefen, David. 2002. Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM Sigmis Database*, **33**(3), 38–53.

Gupta, Aditi, and Kumaraguru, Ponnurangam. 2012. Credibility ranking of tweets during high impact events. In: *Proceedings of 1st Workshop on Privacy and Security in Online Social Media PSOSM, at WWW 2012*. Lyon, France: ACM.

Gupta, Aditi, Kumaraguru, Ponnurangam, Castillo, Carlos, and Meier, Patrick. 2014. Tweetcred: Real-time credibility assessment of content on Twitter. Pages 228–243 of: *Proceedings of 1st Workshop on Privacy and Security in Online Social Media SocInfo*. Barcelona, Spain: Springer.

Hermida, Alfred. 2014. *Tell everyone: Why we share and why it matters*. Doubleday Canada.

Hiltz, Starr Roxanne, Diaz, Paloma, and Mark, Gloria. 2011. Introduction: Social media and collaborative systems for crisis management. *ACM Transactions on Computer-Human Interaction (TOCHI)*, **18**(4), Article 18, 6 pages.

Hokudai Earthquake Project. 2011. *General consumer survey*. Press release.

Hughes, Amanda L., Peterson, Steve, and Palen, Leysia. 2014. Social media in emergency management. Chap. 11, pages 349–392 of: Trainor, J. E., and Subbio, T. (eds), *Issues in Disaster Science and Management: a Critical Dialogue between Scientists and Emergency Managers*. FEMA in Higher Education Program.

Hui, Cindy, Tyshchuk, Yulia, Wallace, William A., Magdon-Ismail, Malik, and Goldberg, Mark. 2012. Information cascades in social media in response to a crisis: A preliminary model and a case study. Pages 653–656 of: *Proceedings of 21st International Conference on World Wide Web*. Lyon, France: ACM.

ICRC. 2013. Managing sensitive protection information. Chap. 6, pages 77–102 of: *Professional Standards for Protection Work*. International Committee of the Red Cross.

Ito, Jun, Song, Jing, Toda, Hiroyuki, Koike, Yoshimasa, and Oyama, Satoshi. 2015. Assessment of tweet credibility with LDA features. Pages 953–958 of: *Proceedings of 24th International Conference on World Wide Web Companion*. Florence, Italy: ACM, for International World Wide Web Conferences Steering Committee.

Latonero, Mark, and Shklovski, Irina. 2011. Emergency management, Twitter, and social media evangelism. *International Journal of Information Systems for Crisis Response and Management*, **3**(4), 67–86.

Lewandowsky, Stephan, Stritzke, Werner GK., Freund, Alexandra M., Oberauer, Klaus, and Krueger, Joachim I. 2013. Misinformation, disinformation, and violent conflict: From Iraq and the "war on terror" to future threats to peace. *American Psychologist*, **68**(7), 487–501.

Liu, Bing. 2012. *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool.

McCreadie, Richard, Macdonald, Craig, and Ounis, Iadh. 2015. Crowdsourced rumour identification during emergencies. Pages 965–970 of: *Companion Proceedings of Crowdsourced Rumour Identification During Emergencies RDSM at WWW 2015*. Florence, Italy: ACM, for International World Wide Web Conferences Steering Committee.

McKnight, D. Harrison, and Chervany, Norman L. 2001. Trust and distrust definitions: One bite at a time. Pages 27–54 of: *Trust in Cyber-societies*. Springer.

Meier, Patrick. 2015. *Digital humanitarians*. CRC Press.

Mendoza, Marcelo, Poblete, Barbara, and Castillo, Carlos. 2010. Twitter under crisis: Can we trust what we RT? Pages 71–79 of: *Proceedings of 1st Workshop on Social Media Analytics SOMA*. Washington DC, USA: ACM, for ACM.

Merrick, D., and Duffy, Tom. 2013. Utilizing community volunteered information to enhance disaster situational awareness. In: *Proceedings of 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Baden Baden, Germany: ISCRAM.

Mitra, Tanushree, and Gilbert, Eric. 2015. Credbank: A large-scale social media corpus with associated credibility annotations. In: *Proceedings of 9th International AAAI Conference on Web and Social Media*.

Murthy, Dhiraj, and Longwell, Scott A. 2013. Twitter and disasters. *Information, Communication, and Society*, **16**(6), 837–855.

ORI Market Research. 2012. *Social media election survey report*. Press release.

Ozturk, Pinar, Li, Huaye, and Sakamoto, Yasuaki. 2015. Combating rumor spread on social media: The effectiveness of refutation and warning. Pages 2406–2414 of: *Proceedings of 48th Annual Hawaii International Conference on System Sciences HICSS*. Kauai, Hawaii, USA: IEEE.

Popoola, Abdulfatai, Krasnoshtan, Dmytro, Toth, Attila-Peter, Naroditskiy, Victor, Castillo, Carlos, Meier, Patrick, and Rahwan, Iyad. 2013. Information verification during natural disasters. Pages 1029–1032 of: *Social Web for Disaster Management (SWDM), Companion: Proceedings of 22nd International Conference on World Wide Web Conference (WWW)*. Rio de Janeiro, Brazil: ACM, for IW3C2.

Ratkiewicz, Jacob, Conover, Michael, Meiss, Mark, Gonçalves, Bruno, Patil, Snehal, Flammini, Alessandro, and Menczer, Filippo. 2011. Truthy: Mapping the spread of astroturf in microblog streams. Pages 249–252 of: *Proceedings of 20th International Conference Companion on World Wide Web*. Hyderabad, India: ACM, for ACM.

Resnick, Paul, Carton, Samuel, Park, Souneil, Shen, Yuncheng, and Zeffer, Nicole. 2014. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In: *Proceedings of Computational Journalism Conference*. New York, USA: Brown Institute for Media Innovation, Columbia University.

Reuter, Christian, Ludwig, Thomas, Ritzkatis, Michael, and Pipek, Volkmar. 2015a. Social-QAS: Tailorable quality assessment service for social media content. Pages 156–170 of: *End-User Development*. Springer.

Reuter, Christian, Ludwig, Thomas, Kaufhold, Marc-André, and Pipek, Volkmar. 2015b. XHELP: Design of a cross-platform social-media application to support volunteer moderators in disasters. Pages 4093–4102 of: *Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI)*. Seoul, Korea: ACM Press.

Sherchan, Wanita, Nepal, Surya, and Paris, Cecile. 2013. A survey of trust in social networks. *ACM Computing Surveys*, **45**(4), 47:1–47:33.

Shibutani, Tamotsu. 1966. *Improvised news: A sociological study of rumor*. Ardent Media.

Silverman, Craig (ed). 2014. *Verification handbook*. European Journalism Centre.

Stahl, Bernd Carsten. 2006. On the difference or equality of information, misinformation, and disinformation: A critical research perspective. *Informing Science: International Journal of an Emerging Transdiscipline*, **9**, 83–96.

Stamatatos, Efstathios. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, **60**(3), 538–556.

Sutton, Jeannette N. 2010. Twittering Tennessee: Distributed networks and collaboration following a technological disaster. In: *Proceedings of 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Seattle, Washington, USA: ISCRAM.

Tang, Jie, Sun, Jimeng, Wang, Chi, and Yang, Zi. 2009. Social influence analysis in large-scale networks. Pages 807–816 of: *Proceedings of 15th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. Paris, France: ACM, for ACM.

Tapia, Andrea H., and Moore, Kathleen. 2014. Good enough is good enough: Overcoming disaster response organizations' slow social media data adoption. *Computer Supported Cooperative Work (CSCW)*, **23**(4-6), 483–512.

Tapia, Andrea H., Moore, Kathleen A., and Johnson, Nicolas. 2013. Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations. Pages 770–778 of: *Proceedings of 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Baden Baden, Germany: ISCRAM.

Taylor, Mel, Wells, Garrett, Howell, Gwyneth, and Raphael, Beverley. 2012. The role of social media as psychological first aid as a support to community resilience building. *Australian Journal of Emergency Management, The*, **27**(1), 20–26.

Thomson, Robert, Ito, Naoya, Suda, Hinako, Lin, Fangyu, Liu, Yafei, Hayasaka, Ryo, Isochi, Ryuzo, and Wang, Zian. 2012. Trusting tweets: The Fukushima disaster and information source credibility on Twitter. Page 10 of: *Proceedings of 9th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Vancouver, Canada: ISCRAM.

Tyshchuk, Yulia, Hui, Cindy, Grabowski, Martha, and Wallace, William. 2012. Social media and warning response impacts in extreme events: Results from a naturally occurring experiment. Pages 818–827 of: *System Science (HICSS), 2012 45th Hawaii International Conference on*. IEEE.

U.K. Ministry of Defence. 2014. *Understanding and intelligence support to joint operations*. 3rd edn. Joint Doctrine Publication (JDP). Development, Concepts and Doctrine Centre (DCDC), Ministry of Defence, UK.

Vieweg, Sarah, Castillo, Carlos, and Imran, Muhammad. 2014. Integrating social media communications into the rapid assessment of sudden onset disasters. Pages 444–461 of: *Proceedings of International Conference on Social Informatics SocInfo*. Barcelona, Spain: Springer.

Walsh, Lynda. 2006. *Sins against science: The scientific media hoaxes of Poe, Twain, and others*. SUNY Press.

Zielinski, Andrea, Middleton, S., Tokarchuk, L., and Wang, Xinyue. 2013. Social media text mining and network analysis for decision support in natural crisis management. Pages 840–845 of: *Proceedings of 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Baden Baden, Germany: ISCRAM.

If you enjoyed this free preview of Big Crisis Data,
get the book at http://bigcrisisdata.org/