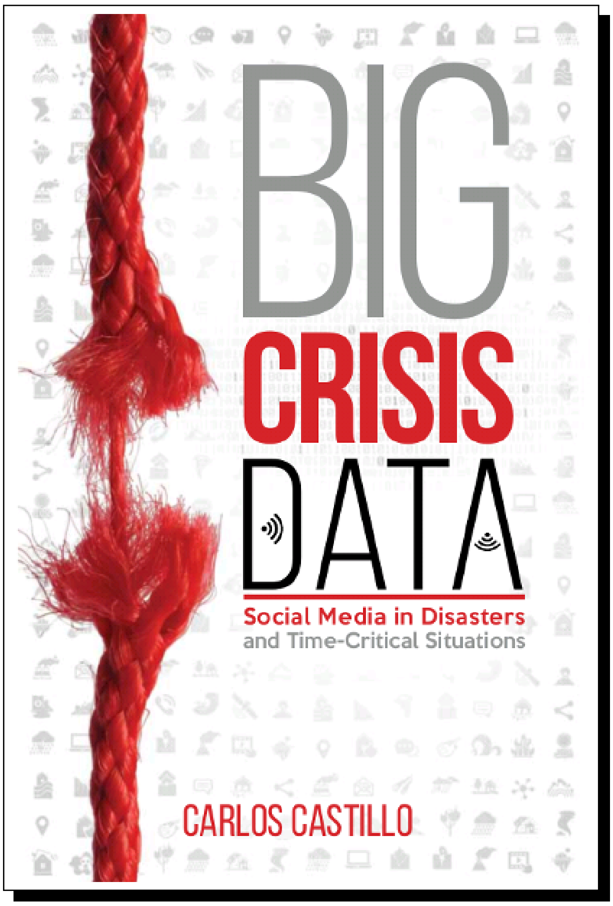
Big Crisis Data

Social Media in Disasters and Time Critical Situations

by Carlos Castillo



FREE PREVIEW
CHAPTER 9. VALIDITY

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 9

10

Validity: Biases and Pitfalls of Social Media Data

In 2008 Google launched Flu Trends, showing that the search volume of certain terms in a region was strongly correlated with levels of flu activity in that region.[1] They also found that the increase in the usage of flu-related terms happened days before health care authorities were able to report an increase in cases of flu. The reasons are twofold: there are delays in the official data collection done from hospitals, and people search for symptoms before visiting a doctor. Despite the success of Flu Trends, it was not beyond criticism. Lazer et al. (2014) highlighted a series of issues with its predictions, including a systemic bias that produced an overestimate in 100 out of the 108 weeks analyzed during a two-year period. As a more general criticism, Lazer et al. denounced this as an example of *big data hubris:* the assumption that a large dataset can be a substitute, rather than a supplement, to a traditional analysis method. The popular press has lambasted "big data fundamentalism," the idea that larger datasets imply more objective results.[2]

Researchers performing social science research have embraced and criticized, sometimes at the same time, the usage of large-scale datasets from social media. Social media, as a reflection of social interactions at large scale and in digitally accessible formats, provides a larger quantity of data at a much lower cost than alternative datasets, such as surveys or direct observations. However, the infamous "streetlight effect" may be at play here: scientists inclined to search for evidence where it is easier, instead of where better evidence is likely to be found.[3]

The representativeness of social media and other types of digital traces, and their lack of context, are often cited as key factors to distrust conclusions based solely on them. "Just because you see traces of data doesn't mean you always know the intention or cultural logic behind them. And just because you have a big N doesnt mean that its representative or generalizable."[4] For instance, methods to use trends found on Twitter data as direct predictions of political election results, have been to a large extent debunked (Gayo-Avello, 2012).

This chapter warns against a naïve interpretation of results obtained from social media data from emergencies. The quality of social media data for this purpose is affected by at least two types of factors. First, using social media as a data source introduces a series of biases that are difficult to evaluate and that can change rapidly over time. Second, automatic processing methods, particularly probabilistic ones, may introduce errors that might distort the results.

These concerns do not mean that we cannot answer relevant questions using this data, but instead that we should be cautious about the inferences drawn from the data, and the way those inferences are communicated to end users who make decisions at least partially based on them. It is important to continuously understand the biases being introduced and to validate results obtained through social media data against other sources of information. This chapter presents research that goes in that direction. First, example works comparing data obtained from social media with data from "offline" phenomena are presented (§10.1). Second, issues of sampling bias due to social (§10.2), contextual (§10.3) and technological (§10.4) problems are discussed, including methods for validating

[1] Google Flu Trends. http://www.google.org/flutrends/.
[2] "Why Big Data Is Not Truth." Quentin Hardy, *International New York Times*, June 2013. http://bits.blogs.nytimes.com/2013/06/01/why-big-data-is-not-truth/.
[3] See, for instance, "Why Scientific Studies Are So Often Wrong: The Streetlight Effect." David H. Freeman, *Discover Magazine*, December 2010. http://discovermagazine.com/2010/jul-aug/29-why-scientific-studies-often-wrong-streetlight-effect.
[4] "Big Data: Opportunities for computational and social sciences." Blog post by danah boyd, April 2010. http://www.zephoria.org/thoughts/archives/2010/04/17/big-data-opportunities- for-computational-and-social-sciences.html. "Big N" is an informal way of describing a large dataset, for instance, a large number of people or a large number of messages.

and contrasting this data with other sources. We then go deeper into spatial (§10.5) and temporal (§10.6) aspects of this validation.

10.1 Studying the "Offline" World Using "Online" Data

The comparison done by Google Flu Trends of an "online" variable (searches for flu-related trends) with an "offline" variable (visits to hospitals for flu-related symptoms) provided a template for a series of works that performed similar comparisons. Researchers turned their attention to all kinds of user-generated data, including logs of online searches, blog postings, news articles, and social media messages, among other types of data. Correlations between time series extracted from these media and other variables were found for the stock market (Bollen et al., 2011), the market of jobs (Ettredge et al., 2005) and cars (Choi and Varian, 2012), the box office revenue of movies (Asur and Huberman, 2010), the sales of music and video games (Goel et al., 2010), and results in political elections (Tumasjan et al., 2010).

Scientists were quick to point out, however, that there are many limitations to these predictions. First, in a large group of time series, significant correlations among random pairs of series are likely to be observed. For instance, in the 2000–2009 period, the U.S. per-capita consumption of margarine and the divorce rate in the U.S. state of Maine were correlated at $r = 0.99$ (Vigen, 2015). Second, for media products (e.g., music), sometimes there are public data sources, such as Billboard Top 100 lists, that produce predictions of record sales that are similar or better than those obtained with large online datasets (Goel et al., 2010). Third, in the political arena, serious issues were observed regarding the generalizability of results obtained for one country in one type of election, to a different country, and/or a different type of election (Gayo-Avello, 2011).

De-biasing can be done to some extent, under certain circumstances (Zagheni and Weber, 2015), but in general no resampling methodology can be used to reproduce faithfully the results of a different study using a different survey methodology (e.g., phone surveys vs. social media). These problems are compounded when datasets over which this research is done are not publicly available, making it harder to disprove wrong conclusions (Lazer et al., 2014; Ruths and Pfeffer, 2014).

Data from online and social media data can be interpreted as a type of survey done over an opt-in "panel" of Internet users. However, there are many distortions in this data, including people who have multiple accounts and accounts used by multiple people, and the fact that online social connections are an inaccurate reflection of actual interpersonal ties (boyd and Crawford, 2012). More worryingly, the participation and topical coverage of this type of data is not only unrepresentative of the general population, it is also much more dynamic and less predictable than conventional survey panels: "In short, if online and social media data are to be treated as surveys, they must be treated as imperfect surveys indeed" (Diaz et al., 2014).

For many practitioners of humanitarian and emergency response, the distinction between "offline" and "online" may be relevant, but perhaps a more important distinction is between the "traditional" datasets they have used for years, such as weather reports, and "new" and emerging datasets, such as social media. A general problem of attempting to answer questions about society using online data, is that online data sets are big, but there is no guarantee that they are representative of the behavior of the general population. During a crisis, the presence of a social media message about a particular situation depends on many factors, including the person being willing and able to post information about the situation, his/her availability of a device able to post that message, the availability of network connectivity, and in the case of a geotagged message, the capability and intention of the device owner to include these geotags. Each of these factors introduces a new bias in the data.

The questions we explore next are basically about the extent to which datasets created from social media can be useful for research and practice on disaster response, about the systematic issues these datasets have, and the consequences of those issues.

10.2 The Digital Divide

The ICRC highlights two issues stemming from biases in crisis data. First, they can result in unintentional discrimination, leaving out some participants "owing to language ability, political affiliation, educational level, access to communication means when using crowdsourcing, or other factors." Second, even when there is no discrimination, "sampling bias hampers an accurate understanding of the situation and distorts the resultant protection response" (ICRC, 2013).

The *digital divide* is the enormous disparity that exists in terms of access to information and communication technologies among people living in different countries, as well as people living within the same country. A related concept is the *data-divide:* the lack of availability of high-quality data that affects low- and middle-income countries (Cinnamon and Schuurman, 2013). Worryingly, a large portion of research on social media for emergency management has been focused on populations that have widespread access to mobile social media. With some exceptions, there is little work on how access to these technologies impacts the results obtained by researchers (Hughes et al., 2014).

According to 2015 data, about 60% of the world has no access to the Internet. Furthermore, access is very uneven across the world. There is a huge gap between average levels of access in developed countries (78%), in comparison to developing countries (32%).[5] In many countries, many people do not have easy access to mobile phone and the Internet, they have a vulnerable structure that can be easily destroyed or interrupted, and they lack a reliable electricity supply (Whipkey and Verity, 2015).

Within countries there are also huge differences. For instance, in 2011 a Gallup poll of mobile phone users in 17 African countries revealed that mobile phone usage was significantly more common among the urban, affluent, and educated.[6] The same has been observed in China, where historically the Internet penetration in rural areas is less than one-third of that in urban areas (White and Fu, 2012 citing data from CNNIC[7]). Even in urban areas, within a city different neighborhoods may have vastly different levels of Internet access.[8]

The fact that huge differences exist in terms of access to social media and mobile devices, does not mean that these channels should not be used. The situation has some similarities with what happened when the system supporting the emergency phone number 911 was set up in 1968 in the United States. Many municipalities did not provide the service until the 1980s and it was not until the 1990s that it was widely adopted. Still, it is hard to dispute the utility of this service on the grounds that it was not universally accessible during its first two decades; never in history everyone has had access to the same communication technologies at the same time.[9]

In general, new technologies are less available for elderly people, less educated people, people with low income, people in rural areas, people belonging to a linguistic minority, and people with disabilities.[10] The latter group is particularly vulnerable during emergencies when information is communicated through social media, as Bricout and Baker (2010) note, because computing technologies are in general less adapted for people with disabilities than other forms of media, such as the television.

As noted by Castells (2001) among others, the digital divide is not simply a matter of Internet connections per person across different groups of people, but about the consequences of having or not having a connection. Analyzing annotations in Google Earth after the 2005 Hurricane Katrina, Crutcher and Zook (2009) noted a strong divide along racial lines: neighborhoods with large African American populations were less likely to have volunteered geographical annotations, even when those were as affected or more than other neighborhoods. Elwood (2008) also warns about volunteered geographical information systems as potential contributors to inequalities. For instance, census data undercounts people in places were there are more homeless or informal settlements, and a neighborhood perceived as vibrant by a minority group may be perceived as a dangerous neighborhood by those who do not belong to that minority. The way in which people contribute new information may contribute to increase or decrease power imbalances due to this data.

Coming back to the consequences of the digital divide, on the one hand, even if only few people in an affected area can send an SMS or post an update on social media, the fact that these messages reach relief workers can be potentially very helpful for the entire community. On the other hand, technologists must constantly remind them-

5 International Telecommunications Union (ITU) Statistics: http://www.itu.int/en/ITU-D/Statistics/.
6 "Mobile Phone Access Varies Widely in Sub-Saharan Africa." *Gallup Poll*, September 2011. http://www.gallup.com/poll/149519/mobile-phone-access-varies-widely-sub-saharan-africa.aspx.
7 China Internet Network Information Center: http://www1.cnnic.cn/IDR/ReportDownloads/.
8 "Map the iPhone Users In Any City, And You Know Where the Rich Live." Emily Badger, *CityLab*, June 2013. http://www.citylab.com/work/2013/06/map-iphone-users-any-city-and- you-know-where-rich-live/5961/.
9 "Big Data for Disaster Response: A List of Wrong Assumptions" Patrick Meier, *iRevolution*, June 2013. http://irevolution.net/2013/06/10/wrong-assumptions-big-data/.
10 Duke University Digital Divide Microsite: http://sites.duke.edu/digitaldivide/.

selves and users of their systems of all the gaps in their data, including the communities that are underrepresented or excluded (Crawford and Finn, 2014).

10.3 Content Production Issues

People selectively choose what to share online, and even the ones who seem to disclose "everything" online, and perhaps particularly them, in reality engage in a performance and presentation of a carefully edited self (Marwick and boyd, 2011; Marwick, 2013, ch. 5).

During a crisis, there are many reasons why someone may decide to post, or not to post, certain pieces of information online. For instance, in cases of armed conflict, some people may self-censor themselves to avoid becoming a target of violence (we come back to this in Section 11.2). In the same situation, others may engage in the spread of disinformation (as discussed in Section 8.3).

Depending on the specifics of a crisis, some types of information are rarely posted or absent from social media. Saleem et al. (2014) note that, for instance, lists of names of emergency-related casualties are in general not shared early during a crisis. Additionally, the information expiration problem discussed in Section 8.1 is a serious issue with respect to the information people share online during an emergency. People are more prone to indicate that a danger exists or that an item is needed, than that a danger no longer exists or that an item has already been supplied – contributing, for instance, to the "second disaster" of having to deal with donations that are no longer needed by those affected.[11]

The way in which a situation is described by the public also depends on each individual's capacity to communicate effectively. In an ideal case, their messages should use the same terms and expressions used by emergency management agencies – but there is no indication this is, indeed, the case. Semantic technologies can contribute to some extent to solve this mismatch (as discussed on Section 3.7).

Quantitative information is also scarce. Purohit et al. (2014) note that quantifiers in social media regarding emergency resources are extremely rare, for instance, basically no message about an emergency shelter in their data mentioned how many people could that shelter accommodate. Saleem et al. (2014) observed in the 2013 Alberta floods in Canada that most eyewitness reports of water level used figurative or vague language rather than standard units of length – however, they also note that photos depicting the water level were used extensively and could be interpreted (with manual labor) to estimate flood heights. Sellam and Alonso (2015) were able to find a few thousand tweets containing useful quantitative information in a corpus of millions, Crutcher and Zook (2009) cite volunteer-provided annotations about the 2005 Hurricane Katrina in Google Earth that speak about water levels of "5 inches" or reaching the "2nd story," and Earle et al. (2010) report that some people try to guess the magnitude of an earthquake, although frequently confusing intensity and magnitude (e.g., *"Minor earthquake. Maybe a 4"* and *"just felt an earthquake here . . . 5.5?"*).

10.4 Infrastructure and Technological Factors

Connectivity. After the 2010 earthquake in Haiti, cell phone towers were still operational enough for people to send and receive SMS messages, which has shown to be a quite resilient technology across disasters (Liu, 2014). However, disasters can certainly cause disruptions such as energy and communications blackouts, particularly in less-developed countries (Whipkey and Verity, 2015). In some crisis situations, mobile networks have shown to be less resilient than expected. Jennex (2012) studied the 2011 blackout in San Diego, observing that more than 90% of people either lost mobile Internet access or experienced a degradation of it, with only about 35% of mobile users retaining some sort of Internet access; 17% of those who attempted to update their Twitter status were able to do it without problems. Depending on the type of application, this may or may not be sufficient to obtain a general picture of the needs and concerns of those in the affected area.

In other cases, such as the 2012 Hurricane Sandy in the United States, mobile phone voice service was interrupted in some areas, while Internet communications continued – causing people to flock to Internet-based

[11] "The 'Second Disaster': Making Well-Intentioned Donations Useful." Pam Fessler, *NPR*, January 2013. http://www.npr.org/2013/01/12/169198037/the-second-disaster-making-good- intentions-useful.

telephony to communicate with others about their situation (Whittaker, 31/10/2012). When Internet connectivity survives a disaster, it has the potential to provide crisis communications, particularly in places where other infrastructure is poor (White and Fu, 2012).

Even without Internet access due to a disaster or because of human intervention (e.g., an intentional Internet blackout mandated by the government during a demonstration), alternative network topologies may be used. "Mesh" networks in which devices connect to each other instead of to centralized servers, have been implemented in popular mobile apps that provide online messaging services independently of the availability of Internet.[12]

Damages to transport infrastructure can also affect the generation of content by mobile phone users. If the population is immobilized, for instance, during a flood situation where roads are dangerous, people do not have access to vantage points and instead will post from wherever they find themselves at that moment (Saleem et al., 2014).

Technology factors. Different platforms are used by different people for different purposes, and have different interaction modalities. For instance, some of them allow people to "dislike" or "downvote" content, while others do not. Different interaction modalities reinforce the fact that different platforms appeal to different demographics. In other words, the choice of a social media platform as a data source obviously affects the results that are obtained.

The data collection interfaces offered by social media platforms also introduce biases. Researchers do not know the exact sampling, ranking, and filtering methods used by the publicly available data consumption interfaces provided by social media (boyd and Crawford, 2012), and do not even know if those are stable or change over time. Undisclosed changes in sampling rates can be deceiving, for instance an increasing trend might be the result of an actual decrease in the data and a higher sampling rate. Morstatter et al. (2014) showed discrepancies between Twitter's public API and full datasets obtained from this service.

The usage of inference-based methods to enrich textual data, such as geocoding, is useful but may also introduce errors in the data. For instance, "someone might text the crisis maps phone number to report something they saw earlier, possibly texting from a shelter about a bridge that collapsed 10 miles down the road." (Gao et al., 2011).

10.5 The Geography of Events and Geotagged Social Media

Malik et al. (2015) studied whether geotagged tweets were spread in the United States following population patterns, in other words, whether they were representative of the locations of people. While at a very high level a map of geotagged tweets looks similar to a map of large metropolitan areas and cities, when looking in detail they observed that socioeconomical and demographic factors play a much bigger role than population densities in the density of geotagged tweets. Anecdotally, the largest density of geotagged tweets in the United States was observed from Las Vegas Strip, several large airports, and Walt Disney World; and several densely populated areas, such as prisons, had no geotagged tweets.

After this result, one could wonder if there is any chance that social media activity around a given place can be correlated with crisis or disaster events in that place. Actually, a number of researchers have documented that such correlations exist.

The U.S. Geological Service (USGS) has done a a multiyear, multidisciplinary effort to determine to what extent Twitter can be used as a social sensor for earthquakes (Earle et al., 2010; Guy et al., 2010). Their starting point is the empirical observation that conversations in social media including the word "earthquake" are rare, except in times and places where an earthquake has happened. Their studies have produced a number of increasingly more elaborate methods to interpret increases of social media activity.

Earthquakes have a typical distribution of intensities with an area of high intensity, and a radial pattern where intensities attenuate as the distance from this area increases. This distribution of intensities is usually matched by the activities observed in social media (e.g., by Evnine et al., 2014 for Facebook). This distribution can also be used to resolve ambiguities in the data. Robinson et al. (2015) present an example where they compare the distribution of social media messages posted in Australia and New Zealand in response to an earthquake in Melbourne (which was very concentrated around Melbourne) and the reaction in the same countries to an earthquake in Indonesia (which was much more spread).

[12] "Why a messaging app meant for festivals became massively popular during Hong Kong protests." Amar Toor, *The Verge*, October 2014. https://www.theverge.com/2014/10/16/6981127/firechat-messaging-app-accidental-protest-app-hong-kong.

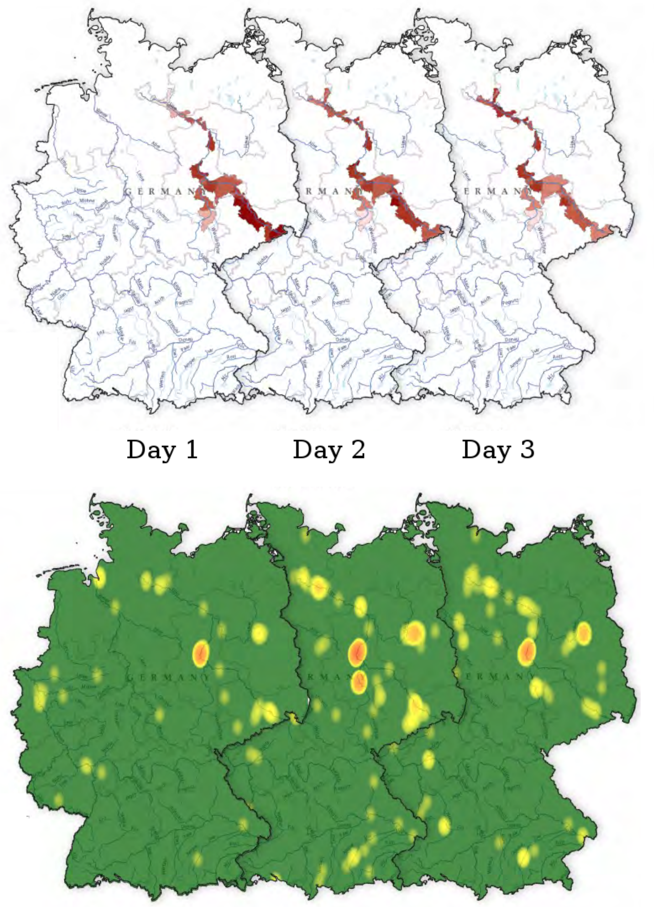


Figure 10.1. Comparison of changes in water height (top) and flood-related tweets (bottom), during floods in the river Elbe in Germany from June 8, 2013 to June 10, 2013. Adapted from Herfort et al. (2014), using World Hydro Reference as overlay map. Reproduced with permission from the authors.

As we noted on Section 2.1, earthquakes not only affect how users produce online information but also how they consume it. Bossu et al. (2008) show that the location of visitors to the most prominent website providing seismological information in Europe,[13] obtained from their IP address, can be used as a proxy to understand where an earthquake was felt.

Floods are other type of disaster where agreement with social media has been observed. Smith et al. (2015) compare water levels during 2012 inundations in different parts of Newcastle, England, against activities in Twitter measured through a relatively small sample of geocoded tweets, observing these are somewhat correlated. Dashti et al. (2014) observe a similar correlation between flood hazard zones determined using historical flood data in the county of Boulder in the United States, and tweets about the floods affecting the county in 2013.

Herfort et al. (2014); de Albuquerque et al. (2015) study floods during June 2013 in the river Elbe in Germany. They observe that flood-related messages in Twitter are indeed concentrated around cities close to the river, particularly Magdeburg. The similarity in geographical distributions is noticeable, and can be seen in Figure 10.1.

This bias toward cities is also observed by Sakaki et al. (2013) in their system for tracking the spatial movement of a phenomenon, such as a typhoon. They observe lower accuracy in determining geographical regions when the events happen on a sparsely populated area, such as an earthquake with an epicenter in the middle of the sea, or a typhoon happening far from the coast.

Additionally, geographical correlations have been observed in other nondisaster phenomena, including sightings of the Aurora Borealis which match well with geophysical data from geomagnetic storms (LaLone et al., 2015).

[13] European-Mediterranean Seismological Centre: http://www.emsc-csem.org/.

Data from social media can also be fused with other geographical data. For instance, the concentration of mobile phones in a given area, considered together with authoritative information about road networks, can be helpful during an evacuation for identifying bottlenecks or roadblocks (Oxendine et al., 2014).

Besides social media messages, data collected by mobile service providers has also been used to study geographical variables of disasters (Bagrow et al., 2011). Bengtsson et al. (2011) describe how information about the position of mobile phones, obtained from mobile service providers, was used to observe the movement of people out of Port-au-Prince in the aftermath of the 2010 earthquake in Haiti. Indeed, the estimations using mobile phone data were more accurate than those obtained by volunteers on the ground using surveys of buses and ships leaving the city, as confirmed by a detailed analysis done weeks after the disaster. **Intensity.** While determining the geography of an incident using social media data may be addressed to some extent with current approaches, determining the intensity of the phenomenon in different areas seems to be quite difficult. In general, the frequency of postings about an emergency situation in different areas cannot be used as a proxy for the severity of the situation in those areas. Negative results in this sense include Saleem et al. (2014), who observed that particularly dangerous areas may not be covered by tweets during an ongoing crisis, and Vieweg et al. (2014), who compared the number of tweets containing hurricane-related terms in different islands in the Philippines, finding only a weak correlation between the number of messages posted on each island, and variables such as number of people affected or number of houses destroyed.

Promising results were reported by Chen et al. (2014) in the prediction of the Air Quality Index (AQI) in four large Chinese cities, considering variables computed from term frequencies in messages in Sina Weibo. These terms are related to coughing, wearing a mask, having sore throat or cold symptoms, and can be combined into a prediction of the AQI using a function learnt from historical data. Mei et al. (2014) extended these results in various ways. First, they considered all the words in each message, learning a regression model of the AQI based on the presence or absence of each word. Second, they assumed an element of spatial continuity, in which nearby areas should have similar air quality at a given time. Third, they assumed temporal dependencies in the form of a Markov process, in which the air quality at a given time depends on the air quality in the time interval immediately preceding it.

10.6 Evaluation of Alerts Triggered from Social Media

In Section 6.4, two metrics for evaluating event-detection methods were described: sensitivity (the fraction of events that are detected), and false detection rate (the fraction of alerts that are false). These metrics have been used in the evaluation of several prototypes and deployed systems for generating crisis alerts from social media. There are few reference datasets for evaluating event detection methods (McMinn et al., 2013 is one exception), so in general evaluations of different methods are done on different collections, making them difficult to compare. In general, the results show that existing systems offer a range of trade-offs between sensitivity and false detection rate, and they tend to perform better for large-scale events that are clearly perceived by the affected populations.

Earle et al. (2011) compared simple keyword-based approaches to measure the number of crisis-related messages in Twitter over time (e.g., counting how many messages contain “earthquake” and related terms), with data from seismological sensors. They found that although many earthquakes were not tweeted about, when people tweeted about an event, detections took in most cases less than two minutes, “considerably faster than seismographic detections in poorly instrumented regions of the world.” During a five-month period and when compared to validated seismological data, the sensitivity of their system was nominally low: it detected 48 out of 5,175 earthquakes. However, most of the events it did not detect were of smaller magnitude and not actually felt by human populations, which means they had less potential to cause damage. Additionally, they false discovery rate was about 4%: only 2 out of the 48 detections were spurious.

Avvenuti et al. (2014) compared their system, *EARS*, against data from the Italian National Institute of Geophysics and Vulcanology (INGV). Similarly to Earle et al. (2011), they also reported that earthquakes of low magnitude (lower than 3.0) are very hard to detect using social media because they are basically detected only by physical sensors, but not felt by people. For earthquakes of magnitude 3.5 and above (20 in their sample of 70 days), their sensitivity is around 80% and their false detection rate, around 15%.

Robinson et al. (2013) evaluated an earthquake detection system for Australia and New Zealand based on Twitter data. They report a sensitivity of 81% (17 alarms on 21 events over a 12-months period), and a false detection rate of 55% (a total of 31 alarms, out of which 17 were correct). They used a simple heuristic of imposing a minimum support of three tweets before triggering an alarm, which further reduced the false detection rate.

Power et al. (2013) study an alert system for fires based on automatic classification of Twitter messages. The system generated 42 fire alerts over a four-months period, of which 20 were related to actual fires (false discovery rate of 52%), which is a smaller number of false alarms than another system they describe, which was based entirely on keywords.

Merging social media data with other sources. *LITMUS* (Musaev et al., 2014) is an advanced system for detecting landslides that integrates data from multiple sources, including physical data measuring seismic activity and rainfall, and social media data from Twitter, YouTube, and Instagram. First, the world is divided into geographic cells covering 2.5 minutes of latitude and longitude each. Then, signals from each data source in each cell are aggregated using a weighted sum, in which the weights of different signals are calculated using a separate training set, and are related to the accuracy of each source independently when detecting a landslide. The authors show that the false discovery rate of this method is lower than one relying exclusively on physical sensors, at the expense of a loss of sensitivity of around 20%.

Merging data sources is a promising approach, but it also presents some challenges. Data integration problems may arise, which require systems that can interoperate. This is one more argument for the use of semantic technologies, and for the development of appropriate data standards that allow unambiguous data fusion.

10.7 Research Problems

Measuring the impact of bias. In crisis situations, some activities require high-quality data, whereas in other cases "good enough" data suffices (Zook et al., 2010; Tapia and Moore, 2014). Organizations used to take decisions with incomplete information, rarely come back to retrospectively evaluate to what extent their decisions were affected by incomplete or erroneous information, and what were the costs of those decisions.

The responsibility of measuring and countering the bias introduced by a new source of information lies with those who provide the information. Evaluating the extent to which bias affects decisions is necessary to enable the creation of reliable decision-support systems.

Estimating reliability. Similarly, more research is needed to be able to estimate the reliability of different predictions done with this data. For instance, systems that trace the geographical boundaries of a phenomenon using data from social media can provide a more accurate picture by also displaying the uncertainty of those boundaries. Systems that generate automatic alerts can benefit its users by exposing what is the estimated reliability of each of the alerts.

Understanding the value of disclaimers. Communicating the reliability of an inference, or reminding users of the lack of those reliability estimates, are also important to make the output of computational tools more valuable as decision-making elements. Crisis maps produced by the Standby Task Force during Typhoon Ruby carry the following footnote: "Social media is not necessarily representative or verified. Please keep this in mind when interpreting the crisis map."[14] This follows recommendations by organizations including the ICRC, which states that "Any external report should mention the reliability of its contents in general terms. Incidents that are not yet fully established can be included, as long as the level of reliability is clearly disclosed" (ICRC, 2013).

However a disclaimer that some information cannot be independently verified may not be effective at stopping an image from being widely circulated as "true," as we saw for the case of the photo of Iraq being used to illustrate an event in Syria by the BBC (Section 8.1). More research is needed to understand how journalists, emergency managers, and the public, interpret and act upon different types of disclaimer.

10.8 Further Reading

Crawford and Finn (2014) examine many limitations, both in terms of biases and ethical questions, of using social

[14] MicroMappers map for Typhoon Hagupit, 2014. http://maps.micromappers.org/2014/hagupit/tweets/.

media data about disasters. Tufekci (2014) exposes several methodological issues and interpretative pitfalls of research using social media data. boyd and Crawford (2012) identify general issues in current "big data" research, particularly the one that uses social media and online data.

Mejova et al. (2015) address several applications of Twitter data analysis to problems in the domains of health care, political opinion, city sensing, socioeconomic indicators, and disaster response. Their main focus is how messages on Twitter correlate with variables on these domains. Zagheni and Weber (2015) describe methods for reducing the bias introduced by using nonrepresentative populations extracted from the Internet. Reliable external information can be used to calibrate the measurements, and in some cases even without external information, if all one cares about are trends, even biased data can be used to some extent.

Hughes et al. (2014) includes a practitioner's view of social media during emergencies, including several issues related to data reliability connected to the topics of this chapter.

Bibliography

De Albuquerque, João Porto, Herfort, Benjamin, Brenning, Alexander, and Zipf, Alexander. 2015. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, **29**(4), 667–689.

Asur, Sitaram, and Huberman, Bernardo A. 2010. Predicting the future with social media. Pages 492–499 of: *Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology WI-IAT*, vol. 1. Los Alamitos, California, USA: IEEE Computer Society.

Avvenuti, Marco, Cresci, Stefano, Marchetti, Andrea, Meletti, Carlo, and Tesconi, Maurizio. 2014 (Aug.). EARS (Earthquake Alert and Report System): A real time decision support system for earthquake crisis management. Pages 1749–1758 of: *Proceedings of 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, USA.

Bagrow, James P., Wang, Dashun, and Barabasi, Albert-Laszlo. 2011. Collective response of human populations to large-scale emergencies. *PloS One*, **6**(3), e17680.

Bengtsson, Linus, Lu, Xin, Thorson, Anna, Garfield, Richard, and Von Schreeb, Johan. 2011. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine*, **8**(8), e1001083.

Bollen, Johan, Mao, Huina, and Zeng, Xiaojun. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, **2**(1), 1–8.

Bossu, Rémy, Mazet-Roux, Gilles, Douet, Vincent, Rives, Sergio, Marin, Sylvie, and Aupetit, Michael. 2008. Internet users as seismic sensors for improved earthquake response. *Eos, Transactions American Geophysical Union*, **89**(25), 225–226.

boyd, danah, and Crawford, Kate. 2012. Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, **15**(5), 662–679.

Bricout, John C., and Baker, Paul MA. 2010. Leveraging online social networks for people with disabilities in emergency communications and recovery. *International Journal of Emergency Management*, **7**(1), 59–74.

Castells, Manuel. 2001. *The internet galaxy: Reflections on the internet, business, and society*. Oxford University Press.

Chen, Jiaoyan, Chen, Huajun, Zheng, Guozhou, Pan, Jeff Z, Wu, Honghan, and Zhang, Ningyu. 2014. Big smog meets Web science: Smog disaster analysis based on social media and device data on the Web. Pages 505–510 of: *Web Science Track, Companion: Proceedings of the 23rd International Conference on World Wide Web (WWW)*. Seoul, Korea: ACM, for International World Wide Web Conferences Steering Committee.

Choi, Hyunyoung, and Varian, Hal. 2012. Predicting the present with Google trends. *Economic Record*, **88**, 2–9.

Cinnamon, Jonathan, and Schuurman, Nadine. 2013. Confronting the data-divide in a time of spatial turns and volunteered geographic information. *GeoJournal*, **78**(4), 657–674.

Crawford, Kate, and Finn, Megan. 2014. The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal*, 1–12.

Crutcher, Michael, and Zook, Matthew. 2009. Placemarks and waterlines: Racialized cyberscapes in post-Katrina Google Earth. *Geoforum*, **40**(4), 523–534.

Dashti, Shideh, Palen, Leysia, Heris, Mehdi P., Anderson, Kenneth M., and Anderson, Scott. 2014. Supporting disaster reconnaissance with social media data: A design-oriented case study of the 2013 Colorado Floods. In: *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. University Park, Pennsylvania, USA: ISCRAM.

Diaz, Fernando, Gamon, Michael, Hofman, Jake, Kiciman, Emre, and Rothschild, David. 2014 (May). *Online and social media data as a flawed continuous panel survey*. Tech. rept. Microsoft Research.

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.

Earle, Paul S., Bowden, Daniel C., and Guy, Michelle. 2011. Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics*, **54**(6), 708–715.

Elwood, Sarah. 2008. Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, **72**(3-4), 173–183.

Ettredge, Michael, Gerdes, John, and Karuga, Gilbert. 2005. Using Web-based search data to predict macroeconomic statistics. *Communications of ACM*, **48**(11), 87–92.

Evnine, Ariel, Gros, Andreas, and Hofleitner, Aude. 2014 (Aug.). *On Facebook when the Earth shakes...* Tech. rept. Facebook Data Science.

Gao, Huiji, Barbier, Geoffrey, and Goolsby, Rebecca. 2011. Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, **26**(3), 10–14.

Gayo-Avello, Daniel. 2011. Don't turn social media into another "Literary Digest" poll. *Communications of ACM*, **54**(10), 121–128.

Gayo-Avello, Daniel. 2012. No, you cannot predict elections with Twitter. *Internet Computing, IEEE*, **16**(6), 91–94.

Goel, Sharad, Hofman, Jake M., Lahaie, Sbastien, Pennock, David M., and Watts, Duncan J. 2010. Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, **107**(41), 17486–17490.

Guy, Michelle, Earle, Paul, Ostrum, Chris, Gruchalla, Kenny, and Horvath, Scott. 2010. Integration and dissemination of citizen reported and seismically derived earthquake information via social network technologies. Pages 42–53 of: *Advances in Intelligent Data Analysis IX*. Springer.

Herfort, Benjamin, de Albuquerque, João Porto, Schelhorn, Svend-Jonas, and Zipf, Alexander. 2014. Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the river Elbe flood in June 2013. In: *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. University Park, Pennsylvania, USA: ISCRAM.

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.

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

Jennex, Murray E. 2012. Social media–truly viable for crisis response. In: *Proceedings of 9th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Vancouver, Canada: ISCRAM.

LaLone, Nick, Tapia, Andrea H, Case, Nathan A, MacDonald, EA, Hall, M, and Heavner, Matt. 2015 (May). Hybrid community participation in crowdsourced early warning systems. In: *Proceedings of 12th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*.

Lazer, David, Kennedy, Ryan, King, Gary, and Vespignani, Alessandro. 2014. The parable of Google flu: Traps in big data analysis. *Science*, **343**, 1203–1205.

Liu, Sophia B. 2014. Crisis crowdsourcing framework: Designing strategic configurations of crowdsourcing for the emergency management domain. *Computer Supported Cooperative Work (CSCW)*, **23**(4-6), 389–443.

Malik, Momin M, Lamba, Hemank, Nakos, Constantine, and Pfeffer, Jürgen. 2015 (May). Population bias in geotagged tweets. In: *Proceedings of 9th International AAAI Conference on Web and Social Media (ICWSM)*.

Marwick, Alice, and boyd, danah. 2011. To see and be seen: Celebrity practice on Twitter. *Convergence: The International Journal of Research into New Media Technologies*, **17**(2), 139–158.

Marwick, Alice E. 2013. *Status update: Celebrity, publicity, and branding in the social media age*. Yale University Press.

McMinn, Andrew J., Moshfeghi, Yashar, and Jose, Joemon M. 2013. Building a large-scale corpus for evaluating event detection on Twitter. Pages 409–418 of: *Proceedings of Conference on Information and Knowledge Management (CIKM)*. San Francisco, California, USA: ACM.

Mei, Shike, Li, Han, Fan, Jing, Zhu, Xiaojin, and Dyer, Charles R. 2014. Inferring air pollution by sniffing social media. Pages 534–539 of: *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE.

Mejova, Yelena, Weber, Ingmar, and Macy, Michael (eds). 2015. *Twitter: A digital socioscope*. Cambridge University Press.

Morstatter, Fred, Pfeffer, J"urgen, and Liu, Huan. 2014. When is it biased?: Assessing the representativeness of Twitter's streaming api. Pages 555–556 of: *Web Science Track, Companion: Proceedings of 23rd International Conference on World Wide Web (WWW)*. Seoul, Korea: ACM.

Musaev, Aibek, Wang, De, and Pu, Calton. 2014. LITMUS: Landslide Detection by Integrating Multiple Sources. In: *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. University Park, Pennsylvania, USA: ISCRAM.

Oxendine, Christopher E., Schnebele, Emily, Cervone, Guido, and Waters, Nigel. 2014. Fusing non-authoritative data to improve situational awareness in emergencies. In: *Proceedings of 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. University Park, Pennsylvania, USA: ISCRAM.

Power, Robert, Robinson, Bella, and Ratcliffe, David. 2013. Finding fires with Twitter. Pages 80–89 of: *Proceedings of the Australasian Language Technology Association (ALTA) Workshop*.

Purohit, Hemant, Castillo, Carlos, Diaz, Fernando, Sheth, Amit, and Meier, Patrick. 2014. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday Online Journal*, **19**(1).

Robinson, Bella, Power, Robert, and Cameron, Mark. 2013. A sensitive Twitter earthquake detector. Pages 999–1002 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.

Robinson, Bella, Power, Robert, and Cameron, Mark. 2015. Disaster monitoring. Chap. 6 of: Mejova, Yelena, Weber, Ingmar, and Macy, Michael (eds), *Twitter: a Digital Socioscope*. Cambridge University Press.

Ruths, Derek, and Pfeffer, Jürgen. 2014. Social media for large studies of behavior. *Science*, **346**(6213), 1063–1064.

Sakaki, Takeshi, Okazaki, Masahide, and Matsuo, Yoshikazu. 2013. Tweet analysis for real-time event detection and earthquake reporting system development. *Knowledge and Data Engineering, IEEE Transactions on*, **25**(4), 919–931.

Saleem, Haji M., Xu, Yishi, and Ruths, Derek. 2014. Novel situational information in mass emergencies: What does Twitter provide? *Procedia Engineering*, **78**, 155–164.

Sellam, Thibault, and Alonso, Omar. 2015. Raimond: Quantitative data extraction from Twitter to describe events. Pages 251–268 of: Cimiano, Philipp, Frasincar, Flavius, Houben, Geert-Jan, and Schwabe, Daniel (eds), *Engineering the Web in the Big Data Era*. Lecture Notes in Computer Science, vol. 9114. Springer International Publishing.

Smith, Luke, Liang, Qiuhua, James, Phil, and Lin, Wen. 2015. Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*, Jan.

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.

Tufekci, Zeynep. 2014 (June). Big questions for social media Big Data: Representativeness, validity and other methodological pitfalls. In: *Proceedings of 8th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. Pages 178–185 of: *Proceedings of 4th International AAAI Conference on Weblogs and Social Media*. Washington DC, USA: AAAI Press.

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.

Vigen, Tyler. 2015. *Spurious correlations*. Hachette Books.

Whipkey, Katie, and Verity, Andrej. 2015 (Sept.). *Guidance for incorporating Big Data into humanitarian operations*. Tech. rept. Digital Humanitarian Network.

White, James D., and Fu, King-Wa. 2012. Who do you trust? comparing people-centered communications in disaster situations in the United States and China. *Journal of Comparative Policy Analysis: Research and Practice*, **14**(2), 126–142.

Whittaker, Zack. 31/10/2012. Internet usage rocketed on the East Coast during Sandy: Report. *ZDNet Online*, Oct.

Zagheni, Emilio, and Weber, Ingmar. 2015. Demographic research with non-representative internet data. *International Journal of Manpower*, **36**(1), 13–25.

Zook, Matthew, Graham, Mark, Shelton, Taylor, and Gorman, Sean. 2010. Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, **2**(2), 7–33.

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