Comparing Social Media and Traditional Surveys Around the Boston Marathon Bombing

Cody Buntain
Dept. of Computer Science
University of Maryland
College Park, Maryland 20742
cbuntain@cs.umd.edu

Erin McGrath
START Center
University of Maryland
College Park, Maryland 20742
ecmcgrath@umd.edu

Jennifer Golbeck
College of Information Studies
University of Maryland
College Park, Maryland 20742
golbeck@cs.umd.edu

Gary LaFree
START Center
University of Maryland
College Park, Maryland 20742
garylafree@gmail.com

ABSTRACT

Sociological surveys have been a key instrument in understanding social phenomena, but do the introduction and popularity of social media threaten to usurp the survey's place? The significant amount of data one can capture from social media sites like Twitter make such sources appealing. Limited work has tried to triangulate these sources pragmatically for research. This paper documents experiences in comparing analyses and results from a panel survey, a survey embedded within an experiment, and social media data surrounding the 2013 Boston Marathon Bombing. Our experience suggests the sources are complementary: social media provides better insight into behavior more rapidly and cheaper than surveys, but surveys can provide higher quality, targeted, and more relevant data.

CCS Concepts

•**Applied computing** → **Sociology;** •**Human-centered computing** → *Social networking sites;*

Keywords

social media, twitter, surveys, big data, boston marathon bombing

1. INTRODUCTION

Social science aims to understand or explain individual or collective behavior. One major tool for enhancing this understanding is to ask questions through surveys. Surveys have been an essential data collection instrument for scientists and policy makers for decades. Recently, the advent and wide-spread popularity of social media has introduced a new source of data and a different perspective from which to examine individual and collective behavior. While social media data's abundance and versatility has caused some to declare an end of life for traditional survey instruments, some in the social sciences have objected to social media's scientific value on the grounds of representativeness and validity. We suggest the truth lies between these two extremes, that social media and traditional survey work can instead complement each other to create a richer depiction of society more rapidly than we could before.

Through social media, scientists can examine actual behavior, albeit online, surrounding high-profile events, in contrast to survey subjects' self-reported suppositions or intents. Social media data can be collected more quickly and cheaply than traditional surveys. Traditional surveys, however, may provide greater insight into demographics and subjects' intentions while also yielding greater relevance to the research topic at hand.

Given these benefits, tradeoffs, and similarities, social media's utility in the social sciences is still being uncovered. This paper explores social media's advantages and costs with respect to traditional survey instruments, such as panel and cross-sectional surveys and survey experiments. We further ground this comparison by reviewing research on the 2013 Boston Marathon Bombing, a highly impactful terrorist attack that has been studied using both social media and surveys.

2. RELATED WORK

Since the introduction of Facebook and Twitter and the resulting explosion in popularity of social media, researchers have been finding new ways to leverage this data to answer sociological questions at a scale and pace previously unachievable. These works have led to powerful systems (e.g., identifying and warning others of earthquakes in Japan [33]), shed light on diffusion patterns for diseases (e.g., tracking the flu with Twitter [22, 23, 30]), and even claimed a role in regime change (the Arab Spring [13]).

As social media's use evolves in science, so too must the procedures surrounding its application and the details we must consider when drawing conclusions. It may be tempting to treat social media simply as another data source to triangulate, but sufficient caveats and new capabilities exist to warrant special consideration. Conversely, while the temptation to abandon old methods in favor of "cheap and easy" social media data is also present, we should consider how social media can augment existing techniques rather than replacing them.

Recent research efforts sought to address these methodological concerns. For instance, Bruns and Stieglitz presented systematic

Published as part of the #Microposts2016 Workshop proceedings, available online as CEUR Vol-1691 (http://ceur-ws.org/Vol-1691)

#Microposts2016, Apr 11th, 2016, Montréal, Canada.

methods for data collection from Twitter that focused on user and temporal metrics to describe Twitter conversations in a standard and replicable manner [2]. They also showed similarities among various types of events in Twitter, supprting the generalizability of Twitter activities across events. This standardization of metrics and methods for analyzing social media data continued with Kim et al., who described methodological concerns in gathering, storing, and analyzing Twitter data [19]. Like us, these authors published a retrospective of their prior work into cancer studies on social media and concluded with a set of helpful considerations and recommendations for researchers wishing to leverage Twitter data. Such considerations included demographic differences between Twitter users and national populations: Twitter is younger and more diverse than the population of Internet users in the United States, trends which are consistent with results from Italy's Twitter population from Vaccari et al.[37] and from Brazil's population found by Samuels and Zucco [34]. Recommendations included gaining familiarity with "big data" platforms like Apache's Hadoop to handle large data sets and standardizing metrics to account for population biases (e.g., ensuring one controls for population distributions to avoid correlating conditionally independent effects).

Another pertinent work is Couper's comparison of social media and survey sciences [8] in his keynote at the 2013 European Survey Research Association. Couper sought to address hyperbolic claims that social media will push surveys into obsolescence and instead argued for a hybrid approach to help survey sciences evolve. Couper raised important issues that could adversely affect results from social media analysis and include user bias (not everyone is on Facebook even though Facebook has over one billion users), issues of access (data distribution rights and proprietary algorithms), and opportunities for mischief. Rather than discarding social media, Couper suggested exploring methods for integrating this data into existing survey sciences and developing methods and metrics to better understand quality and non-response issues. Our work follows a similar vein by identifying specific disadvantages in each method grounded in case study of the Boston Marathon Bombing and suggesting synergies between the two.

This push to combine social media and surveys has garnered interest with publications like that by Wells and Thorson, who used the Facebook platform to conduct a standard survey while simultaneously extracting (with consent) social media data from each respondent [39]. Wells and Thorson argued that their study of individuals through Facebook was enhanced with the individual's social context in a way that was impossible before, and this integration illuminated clearer channels of communication that would otherwise be difficult to untangle. The authors' experiences were not completely positive, however, in that Facebook's platform did not provide the promised results at several points throughout the study. Furthermore, recent changes to Facebook's privacy policy has made replicating the exact study with new participants impossible since social connections that used to be exposed via Facebook's platform are now no longer available. Together, these related works present two fields that are both evolving and trying to learn from each other. Our contributions are to support these past works and this evolution, describe the differences and synergies between the fields, and ground this discussion in a case study of the high-impact Boston Marathon Bombing.

3. COMPARING SOCIAL MEDIA AND SURVEYS

As illustrated above, a great deal of work has leveraged social media to answer sociological questions. These efforts have identified several synergies and divergences between the two methodologies. To explore these factors, we discuss several primary contrasting areas below. Each area presents the general differences one would expect to encounter in comparing any two experiments based from survey work and social media.

3.1 Observations and Inferences

A major difference between social media and surveys stems from the primary type of data each provides: social media mainly yields observations of online behavior and information about that behavior, whereas survey data yields more self-reported responses of subjects' attitudes or propensities toward both on- and off-line behavior. Because surveys are post-hoc, memory or recall error is just one of many non-sampling errors inherent in survey responses known to bias survey response [1]. This contrast does not suggest that social media is superior, since surveys yield invaluable information about individuals' perceptions, and many tools exist to control for these biases; it instead suggests the types of analysis one can perform on each data source differ substantially.

One such difference is clear in analyzing subjects' attitudes versus behavior. Surveys can directly ask respondents about their subjective experiences, perceptions, and attitudes about some entity or concept, and while the answers are self-reported (and therefore biased), the answer is at least directly observed. With social media data, however, most posts do not explicitly describe a user's attitude or subjective state; these attributes must be inferred (though some researchers have explored explicit mentions of phrases like "I am lonely" or similar to track observable emotions in social media [20]).

At the same time, traditional surveys can only ask the respondent about his or her behaviors with limited ability to observe their actual behaviors, and the connection between responses and actual behaviors is often tenuous. Social media, on the other hand, provides a wealth of information about user behavior since social media postings are made outside of the surveyed context; that is, social media data provides a record of a user's actual behavior. The relative value of behaviors versus attitudes is often study-specific, however, so the question of whether surveys or social media data provide better, more useful information is therefore likely to be study-specific as well.

As social media research matures, researchers are also making strides to address these deficiencies. Regarding subjective experience, researchers have explored methods for inferring such information. The field of natural language processing contains a great deal of literature on sentiment analysis, which infers positive, negative, or neutral feelings about a particular subject as extracted from textual or speech data [29]. As such, sentiment analysis has become a relatively mature area of research, capable of highly accurate results compared to humans. Researchers have had success in applying these sentiment analysis techniques to social media for a variety of public opinion mining tasks, including the 2013 work by Ceron and colleagues, which improved election forecasts using social media and sentiment analysis [6].

Survey instruments can surpass social media in observational data though, as we see in demographic information. Surveys often include demographic information as a matter of course, allowing analysts to reason about how socioeconomics, age, gender, and other traits affect responses. Social media, on the other hand, does not necessarily provide this information, and the Twitter platform has no direct way of obtaining a user's city/state/country of residence, gender, age, race, or other demographic characteristics. A significant amount of work has gone into inferring these demographic characteristics from users' posts [7, 16, 14, 5]. Many of these works presuppose access to significant portions of a user's social media stream, which may not always be available, but this triangulation can be performed with access to more data. Therefore, in instances where demographics are a priority, surveys provide primary data rather than the proxies one can obtain from social media.

3.2 Resource Costs

While social media and surveys are complementary with respect to observations and inferences, social media data has distinct advantages over survey data with respect to cost. Cost here means both financial cost and temporal cost.

Financial costs of social media data cover a wide range, from nearly free to a few thousand dollars per month. Though purchasing social media data can come with a high price tag, a significant collection of social media data from large populations can be obtained with relative ease and limited cost. Surveys, in contrast, yield fewer respondents and often require a financial incentive for respondents or financial resources to pay surveyors. Social media achieves this superiority by leveraging a service people are already incentivized to use rather than trying to motivate respondents to take a survey they might not ordinarily take. As a result, one can gather data from Twitter's 1% public sample stream, capture an average of 4.3 million tweets per day, and analysts need only pay for storage and processing power (both of which are available at little cost through cloud platforms).

Given an average of nearly 13 messages per user per year on Twitter, these archives also contain messages from many different users. Many of these messages, however, might also be spam or unrelated to the analyst's questions. Herein lies the financial tradeoff between social media and surveys: Surveys potentially provide higher-quality responses at a higher cost, whereas social media provides a huge number of possibly low-quality data points at much lower cost.

A further advantage of these large data collections is reduction in cost from re-use. Since collecting data sets from these public sources is undirected, they can be used repeatedly as research questions are answered and new hypotheses are generated. Advantages of this reusability for regression become clear when analysts identify new research questions that prior survey work may not have covered. Rather than running new surveys and introducing confounding factors like delay and additional costs, analysts can revisit the original social media data and run new analyses directly.

Social media is not only financially cheaper, its real-time nature makes it temporally cheaper as well. That is, one can acquire and analyze social media data much more rapidly than surveys can be designed, implemented, and analyzed. Since social media streams can be captured in real time, one can evaluate public responses online and get an immediate sense of events on the ground. Indeed, this area of real-time social media analysis has spawned a significant sub-field in computer science [35, 15, 27, 38, 32, 11]. Because it is real-time data of primary online behavior, social media mitigates recall bias inherent in survey respondents' recollections of events, instead illuminating social media users' immediate reactions. To assess such quasi-experimental research problems with surveys, the surveyors, in lieu of predicting future events, must get lucky in the timing of their survey.

Social media data's availability and low cost are its primary attractions, but its low cost also provides a useful mechanism for combining it with survey work. That is, it is relatively inexpensive to sample data from social media like Twitter's public stream continuously and use this data as a foundation for deeper investigations, either in social media or with surveys.

3.3 Relevance

As alluded to in previous sections, social media data can suffer from quality issues since users' posts may be unrelated to the target research questions, because of spam, rumor, or similar events going on elsewhere in the world. One way these quality issues have been conceptualized is through noise, and the content or users discussing information relevant to research questions is the signal. Social media can then be described as very noisy, or having low signal-to-noise ratio. Surveys, if properly designed and implemented, provide better quality controls and allow for more targeted questions and responses, thereby increasing the signal to noise ratio.

While spam is an important source of noise in social media, social media platforms are already taking steps to reduce spam, and surveys have limited susceptibility to spam. We therefore focus on the more pertinent question of identifying signal or relevance in these media. Relevance can be measured across three axes: temporal relevance, topical relevance, and geographical relevance.

Temporal relevance can be measured by proximity to an event of interest. If a researcher is interested in public response to an event, social media data can have stronger temporal relations to the specific event by virtue of its timeliness. That is, since one can capture social media immediately before, during, and after an event, a researcher can minimize the opportunity for additional bias to affect an individual's response. For example, a survey can ask a question about an individual's willingness to work with police following a terrorist attack (thereby ensuring high topical relevance), but if this question was asked after a cultural backlash against police, as occurred in August of 2014 with mass protests against police in Ferguson, Missouri, responses can be skewed. Since surveys can take a non-trivial amount of time to design and distribute to respondents, this risk of confounding factors is larger compared to social media, though with computer-aided survey tools, the survey field is reducing this lag. Similar to surveys, however, as one moves farther away from the event of interest, social media's amnesiac characteristics suggests it likely becomes less reliable as users move on to the next big trend. Social media has a short memory, with interest in major events like crises returning to pre-event levels within a few days to weeks (as shown by Olteanu et al. [26] and Buntain et al. [4]), so effects of interest may only be discernible for a short period. The prompting in surveys can illuminate these effects for longer periods.

Topical relevance measures whether individuals are discussing content related to an event of interest. As part of the survey design process, surveyors can be as explicit as they desire about the topic or event under consideration. In social media, however, it is more difficult to identify posts related to a particular topic unless the author has explicitly and intentionally tagged the content. This issue is further complicated by the peculiar, abbreviated, and colloquial language social media users often employ to circumvent length restrictions in posts (e.g., Twitter's 140-character limit). Hashtags, or tokens with a "#" symbol prepended to them, are often used in social media to connect posts to topics, but large amounts of relevant content omit these markers, as seen in Kim et al.'s analysis of message about the Affordable Care Act [19]. Careful selection of topically relevant keywords can help identify higher quality or more relevant social media messages as well, also discussed by Kim et al. [19]. Researchers from natural language processing, machine learning, and other fields are working to facilitate this topic identification, however, by publishing methods for systematizing the hashtag collection process (e.g., Bruns and Stieglitz [2]) and by making topic modeling in social media easier.

Geographic relevance, like temporal and topical relevance, addresses a researcher's desire to sample individuals who are located in a particular area or near a given event. Similar to topical relevance, surveys are better equipped for targeting specific geographic areas than social media data from platforms like Twitter. Since social media data come from all over the world, it is difficult to limit results to particular or small geographic areas. While some social media user profiles contain location information, existing research shows this data is unreliable (e.g., users stating their locations as “Earth” or “Mars”) [7]. Instead, we can sometimes rely on geolocation information embedded in messages in the form of GPS coordinates. Unfortunately, while Twitter’s 1% sample stream produces many messages (an average of over $3,000$ tweets per minute), only a small percentage of those messages include this geolocation information (between $1-3\%$, or about 40 messages per minute). Researchers have tried to address this problem by inferring user locations from their Twitter streams and interactions with other users [7]. Furthermore, with the popularity of check-in capabilities in applications like Swarm, Yelp, and Facebook, users are even more likely to indicate location information on their feeds. When analyzing Twitter’s 1% sample stream, however, these techniques often break down since a single user appears in this stream relatively rarely (an average 13 times per year and a median of 3 times). By contracting with social media platforms or data resellers, researchers can increase their accuracy in geolocating users (Twitter now provides this capability through a subscription service), but this step can increase costs and does not address reliability issues of self-reported locations.

Regarding relevance as a whole, it seems social media is most useful for rapid assessments and getting direct insight into or in reaction to a particular event. Surveys can ensure more relevant responses but at additional cost in time, effort, and data volume.

3.4 Validity

External validity is the extent to which the findings could be generalizable outside the sample analyzed within the study. Random samples are purposively random to avoid selection of a sample with a bias, or of respondents that already hold characteristics within the population, that may be determining the outcome. In the same way, independent variables are tested along with control variables to determine a causal effect of the independent on the dependent variable by ruling out any confounding factors. In a truly experimental design, external validity is the highest because researchers have a control group, and a treatment group. These groups are both random samples that are exactly the same except for the fact that one sample has received the treatment which researchers are hypothesizing causes the outcome they are trying to determine, or not.

Such samples are not only impossible for studies in which respondents have interaction effects on each other, such as in social media, they are undesirable. Studies of social network systems seek to bound the system in question, rather than derive a random sample, because the respondents affect each other by nature of the network structure. A random sample would only give researchers a piece of the system in which interaction effects may or may not be present at the strengths with which they actually occur. Wells and Thorson presented a similar point in their work on combining large data sets from Facebook with survey data and suggested the need to avoid interactions and “pluck” individuals from a random sample was no longer as necessary given the types of data now available [39].

Furthermore, researchers are increasingly demonstrating generalizability of social media results across events and platforms. Research by Olteanu et al. showed similarities in public and organizational response to crises on Twitter across 26 different events of varying type, duration, and severity [26]. Similarly, Bruns and Stieglitz also demonstrated how various classes of events (TV broadcasts versus crises/protests) exhibited similar characteristics in social media postings and were well-separable by these features [2].

4. STUDYING THE BOSTON MARATHON BOMBING

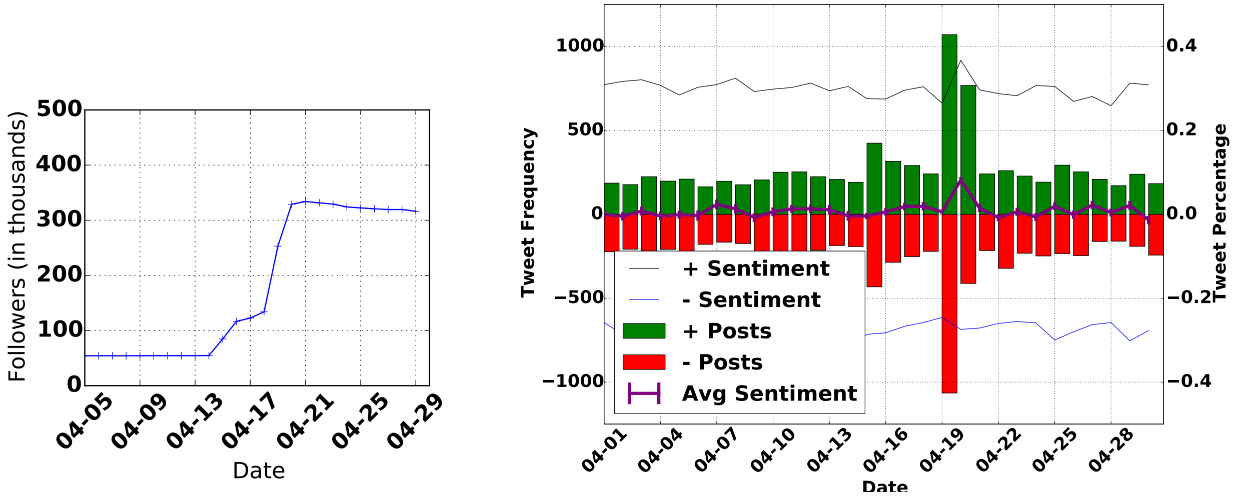
On 15 April 2013, at 14:49 EDT/18:49 UTC, two improvised explosive devices were detonated near the Boston Marathon’s finish line, killing four and injuring approximately 260 people [12]. Over the next four days, local and federal law enforcement agencies engaged in an unprecedented investigation and manhunt, culminating in a car chase and shootout between police on the evening of the 18th and door-to-door search in a Boston suburb on the 19th. At the conclusion, one suspect, Tamerlan Tsarnaev, was dead, and the second suspect, Dzhokhar Tsarnaev, was badly injured and in police custody. These events shocked the United States, paralyzed the city of Boston for several days, and was covered almost exclusively by nearly all major news media and social media. Social media played a major role in this event, with a quarter of Americans following the events via social media [31], and law enforcement organizations using social media to keep the community calm and well-informed (the Boston Police Department even was lauded for its use of social media [9]).

The body of work surrounding social media’s utility to the social sciences has been growing rapidly, both in new applications and comparisons with old techniques. Relatively few of these investigations have had the opportunity to explore differences and complements between social media and traditional survey work in the midst of and in response to a major crisis event. The Boston Marathon Bombing in April of 2013 presents an important case study in this regard. Since the bombing, it has been studied from several different angles: public response on social media [4]; a cross-sectional/panel survey of public willingness to report activity to and perceptions of law enforcement and the US government administered before and after the bombing [21]; and a survey to discern information seeking and searching patterns that included a sample of those exposed to the Boston attack [17]. Given these three distinct perspectives of the same highly-followed event, we explore below how the theoretical differences described above manifest themselves.

4.1 Twitter Versus Facebook

The social media studies on which we focus make exclusive use of the Twitter microblogging platform. While Facebook is a much larger community (comprising 72% of the online adult community in the United States versus Twitter’s 23% [10]), acquiring data from Twitter is far easier than Facebook. Twitter provides a free and publicly accessible service through which any user can programmatically search and stream a random sample of 1% of all tweets being published to Twitter at a given moment[1], and Facebook has no analogous public service. While Twitter has known issues with respect to representation of the population (Twitter is often younger and more male than the average population [19, 37, 34]), research has shown consistency with Twitter populations and offline phenomena like election results [37, 34] and census data [3]. Results are further limited by our reliance on Twitter’s 1% public sample stream, which has known biases against low-volume topics (i.e., topics discussed by few users), but for a major event like the Boston Marathon Bombing, existing research suggests this bias should be limited, especially regarding network structure [25].

[1] https://dev.twitter.com/streaming/overview



(a) Boston Police Department Followers

(b) Sentiment Towards Police

Figure 1: General Twitter Activity During Events

5. THE BOMBING THROUGH SURVEYS AND SOCIAL MEDIA

In the following sections, we present differences encountered while studying public response to the Boston Marathon Bombing through the surveys previously described and social media.

5.1 Attitude and Behavior

In LaFree and Adamczyk's 2015 work, the authors leveraged longitudinal data from a national survey on public perception of law enforcement to investigate the American population's willingness to support police before and after the Boston Marathon Bombing [21]. Their regression model of cross-sectional and panel survey data indicated certain segments of the population were more willing to report terrorism-related suspicious behavior to the police following the bombing. While we can **infer** an increase in willingness will support and increased interaction with law enforcement, this data does not allow us to examine whether this increased willingness actually translated into changes in behavior away from the survey.

In contrast, social media data from the Twitter microblogging service surrounding the Boston Marathon Bombing yielded observations of online behavior. These observations showed a significant increase in references to and followership of police, especially the Boston Police Department. For example, the number of users following the Boston Police Department (BPD) Twitter account increased by a factor of 5 (54K to 264K followers) in response to the bombing and the ensuing manhunt, as shown in Figure 1a, an increase two orders of magnitude above the average increase experienced by accounts during the event. This analysis of social media data supported the hypothesis that users are more likely to interact online with law enforcement following such an event; however, we cannot assess the demographics of users who began following the BPD or why. It's worth noting that online behavior may differ from an individual's offline behavior given the anonymity afforded by social media; this anonymity may provide protections against social desirability bias since the user is disconnected from the individual, though more research is needed here.

Straddling the line between observation and inferential data, however, is sentiment analysis. Survey results show respondents reported higher willingness to work with police, and analysis of social media shows users began seeking information from law enforcement en masse in response to the bombing; both of these results suggest the public opinion toward law enforcement became more positive. To explore this possible connection, we employed TextBlob's[2] sentiment analysis framework, which includes a state-of-the-art sentiment scoring system. For each tweet posted in April of 2013 and posted from the United States (according to provided geolocation information), we scored the tweet in the range $[-1, +1]$, where -1 indicates very negative sentiment, and +1 indicates very positive sentiment [3]. Figure 1b illustrates this process by showing the daily number of positive and negative posts mentioning police as well as the average positive, negative, and total sentiment. This sentiment analysis showed a significant increase in positive sentiment on 19 April, during the manhunt for the Tsarnaev brothers, with this increase returning to pre-event levels within a few days. The connection between "willingness to work with" and positive sentiment is unclear though. Willingness is, by definition, a subjective attitude about a propensity toward a specific behavior, and positive sentiment toward police in social media may proxy willingness to support law enforcement.

Sentiment analysis's attraction primarily comes from its ability to be automated and process the millions of tweets we extracted very rapidly. It is also possible to use crowdsourcing systems like Amazon's Mechanical Turk or CrowdFlower to acquire manual codings for this data from humans at close to the same scale (but at higher cost). Existing work has explored these avenues with good results in coding types of content, user gender, and sentiment [24, 26, 19, 3].

We also considered capturing differing emotional states of social media users in response to the Boston Marathon Bombing. Existing work by Pang, Cameron, and Jin has modeled how emotional responses drive communication behaviors in a crisis, in which the authors identified and coded various public emotions from newswire stories in response to several types of crises [18, 28]. For crises specific to bombings and terrorist attacks, Pang et al. posited the

[2] https://textblob.readthedocs.org

three primary emotional responses from the public were fear, anger, and anxiety. Related work examined what demographics experience these emotions and these subjects' propensity toward specific online and offline behaviors [17]. These works seek to infer the public's emotional responses from secondary reports in articles in traditional media, whereas, with social media, we can directly investigate a subset of primary responses and analyze their language for evidence of different emotions.

Figure 2 illustrates these emotional responses as identified in Twitter surrounding the Boston Marathon Bombing using Mohammad's 2013 word-emotion association lexicon [24]. Results from this figure are consistent with Pang et al. with respect to the increase in references to fear on April 19th, the final day of the manhunt for the Tsarnaev brothers. While we can capture and measure the public's collective emotion and intensity, social media data provides little insight into individual experiences of emotions and their link to offline behaviors.

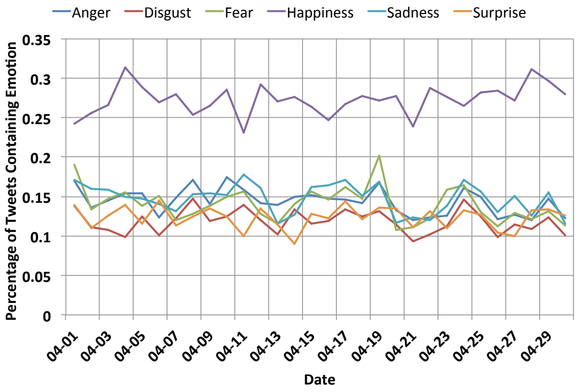


Figure 2: Twitter Emotions

These examples suggest data from social media platforms like Twitter were better at supporting analyses of online, observed behaviors and collective sentiments, whereas traditional surveys illuminated individual-level attitudes and propensities toward certain behaviors and their relation to demographics. More succinctly, social media was better as primary evidence of collective behavior, while surveys provided publics' self-reported attitudes and propensities to act.

5.2 Financial Costs

The volume and availability of data from social media, especially Twitter, are two of its most attractive features, and here we outline the costs for capturing this data. As in similar work (e.g., Buntain et al. [4]), we leveraged an existing corpus of tweets gathered from Twitter's 1% public sample stream, built with the twitter-tools library[3] developed for evaluations at the NIST Text Retrieval Conferences (TRECs). In collecting from Twitter's public sample stream, we connected to the Twitter API endpoint (provide **no filters**), and retrieved a sampling of 1% of all public tweets, yielding approximately 4,000 tweets per minute. In total, this corpus contained 3.5 billion tweets from 1 April 2013 to 31 May 2015. To investigate effects specific to the Boston Marathon Bombing, however, we concentrated on the month of April surrounding the bombing, which contained 134,245,610 tweets.

To perform social media analysis at scale (an issue raised by Kim et al. in 2013 [19]), significant computation resources are often necessary, and we used the Apache Spark distributed processing platform, much of the code for which is available at the author's GitHub repository[4].

Previously, we alluded to using cloud platforms to reduce costs for data collection and analysis. Without additional contracts with data resellers, this infrastructure expense is the primary driver of costs in this research. If we were to use Amazon's S3 cloud storage facilities, it would cost approximately $300 per month to store our entire 3.5-billion-tweet data set or about $11 per month for the Boston-specific data. Additional costs for a virtual system in Amazon's Elastic Cloud to run the actual data collection and analysis can range from $5 per month up to around $300 per month for a more powerful system, depending on researcher needs. Contracting rates with a Twitter-authorized data reseller like Gnip can cost an additional $2,000-$4,000 per month.

We then compare these social media costs with running surveys. The survey research we discuss was performed over the span of several years from 2012 to 2015, using Knowledge Networks (now known as GfK Custom Research). Knowledge Networks is a polling company that holds a U.S. patent for its selection methodology that ensures reliable U.S. representativeness. The first wave of the three-wave panel survey was implemented in November 2012, the second wave was in February of 2014, and the survey was completed in August of 2014. This three-wave survey cost approximately $45,000. The same company, GfK, also performed the second survey experiment on crisis emotions and social media behaviors in May of 2013, with a cost of approximately $95,000.

Assuming one were to contract with a data reseller and run a relatively powerful system on Amazon's Elastic Cloud, for a monthly cost of around $3,300, one could run many experiments on social media data for more than a year for the cost of a high-quality survey from a company like GfK.

This ability to run many experiments on the same social media data set is also valuable and further reduces costs by allowing researchers to re-use the data. We encountered this reusability issue while investigating public perceptions of law enforcement surrounding the Boston Marathon Bombing: researchers posed a new question regarding the public's primary information sources online (our analysis showed it was a mixture of the BPD, the Boston Globe newspaper, and several national news organizations [4]). Answering this question with the existing survey data would have been difficult since the surveys did not include specific questions for such an analysis. With social media data, however, we were able to run a completely new experiment on the existing data set. Similarly, we were also able test sentiment towards a completely new entity (United States sentiment towards Muslims) without needing to collect new data.

In these ways, collecting data on social media is unsurprisingly financially cheaper than collecting survey responses.

5.3 Temporal Costs and Relevance

In our experiments, social media was also faster to collect than survey data. The three-wave panel survey had a significant gap between the Boston Marathon Bombing and the second wave (nearly 10 months), and the crisis emotions experiment had a delay of about one month. In contrast, since we were collecting social media data already, we not only had immediate access to the data, we also had data before, during, and after the event. Even if we had not been collecting data, one could purchase the desired time frame

[3] https://github.com/lintool/twitter-tools

[4] https://github.com/cbuntain/TweetAnalysisWithSpark

from a data reseller without issue. Therefore, similar to a quasi-experimental treatment, social media can provide insights about public behavior both before and after a significant event like the Boston Marathon Bombings more easily since data can be collected with little investment in the design of the data collection instrument, and analyses may be performed later.

Social media is not wholly temporally superior, however, as users tend to move on to the next big trending topic fairly rapidly. In our investigations of the Boston Marathon bombing and the work by Olteanu et al., conversation around major crisis events returned to pre-event levels within a few days or weeks of the event [26, 4]. As a result, measuring significant effects of an event like the Boston Marathon bombing several months after the event is extremely difficult in social media given its undirected nature. Our survey work, on the other hand, was still able to identify these significant effects almost a year later [21].

5.4 Topical Relevance and Noise

As we began cleaning our social media data and tried to focus on topical content about the bombing, the utility of surveys in focusing individuals' response became clear. One of our research questions concerned how public perceptions of police changed in response to the bombing, but it was surprisingly difficult to filter out irrelevant content. Part of this difficulty comes from social media's global nature; on 20 April 2013, residents of New Delhi staged a mass protest in response to local law enforcement's poor handling of the kidnap and rape of a five-year-old girl [36]. This protest had significant impact on social media with widely circulated messages featuring the hashtag "#delhirape," which challenged the validity of our sentiment analysis, as these posts expressed anger and outrage toward police.

Several approaches exist to separate these topically divergent #delhirape tweets from target topic. One could simply discard any post mentioning #delhirape, but as Kim et al. found, many posts relevant to #delhirape but without the hashtag would be not be removed [19]. The approach we used was to focus our investigation on only those posts originating in the United States, which (as we discuss in the next section) significantly reduces the amount of data we are able to analyze. Surveys solve this problem trivially since they were exclusively given to U.S. residents, and dealt very clearly with terrorism, social media's global nature introduced confounding factors in the data.

5.5 Geographic Relevance

As hinted above, one can address issues of topical relevance by constraining social media data to a specific location. Since user postings can include geolocation information, on the surface, this approach seems straightforward. Digging further down, however, difficulties become apparent as so few posts actually contain this geographic information (only 1-3%), which severely restricts our analysis capabilities and population when dealing with Twitter's 1% public sample stream.

A good example of this issue was our attempt at comparing sentiment towards police in New England (the area in which the bombing occurred) to the rest of the United States. There simply were not enough social media messages in our Twitter dataset that were both relevant to law enforcement and posted from New England to make a significant comparison to the rest of the country. Here again we see an issue where more data can solve the problem; existing research has shown one can infer user location in social media with sufficient data, but the Twitter sample stream is not adequate, so one would need to contract with data resellers. From surveys, we could trivially test whether willingness to work with law enforcement was stronger closer to the event and attenuated with distance.

This issue is also related to issues of demographics in social media: since many social media accounts, especially those in Twitter, provide very little in the way of demographic information, it is difficult to segment the data set's population into bins (geographic, gender, or other) that would be clear from survey data.

6. CONCLUSIONS

This paper documents our experiences in triangulating analyses and results from survey instruments and social media data surrounding the 2013 Boston Marathon Bombing. These observations suggest social media's primary datasets of online behavior provide insights more rapidly and cheaply than surveys, but surveys can provide higher quality, targeted, and more relevant data, albeit at a higher cost in terms of resources and time. In our study, and in others for which post hoc data are gathered from the 1% Twitter Stream, the findings are not generalizable in the classic sense of external validity. We argue that in complement with studies that use traditional social scientific design, like these survey studies, the study of a sample of Twitter users gives a more complete picture of how public attitudes are impacted by events like terrorist bombings because they give us insight into social interactions and the effects individuals have on one another in ways that traditional experimental designs in surveys explicitly seek to avoid. Rather, science is more likely to benefit by combining both modes of data to understand and explain changes in individual and collective behavior surrounding impactful events.

7. REFERENCES

[1] H. Assael and J. Keon. Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing*, 46(2):114–123, 1982.

[2] A. Bruns and S. Stieglitz. Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology*, 16(2):91–108, 2013.

[3] C. Buntain, J. Golbeck, and G. LaFree. Powers and Problems of Integrating Social Media Data with Public Health and Safety. In *Bloomberg Data for Good Exchange*, New York, NY, USA, 2015. Bloomberg.

[4] C. Buntain, E. McGrath, J. A. Golbeck, and G. LaFree. Evaluating Public Response to the Boston Marathon Bombing and Other Acts of Terrorism through Twitter. *in press*, 2016.

[5] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella. Discriminating Gender on Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 146 of *EMNLP '11*, pages 1301–1309, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[6] a. Ceron, L. Curini, S. M. Iacus, and G. Porro. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 16(2):1–19, 2013.

[7] R. Compton, D. Jurgens, and D. Allen. Geotagging one hundred million twitter accounts with total variation minimization. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 393–401. IEEE, 2014.

[8] M. Couper. Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods*, 7(3):145–156, 2013.

[9] E. F. Davis Iii, A. A. Alves, and D. A. Sklansky. Social Media and Police Leadership: Lessons From Boston. In *New Perspectives in Policing Bulletin*. Washington, DC: U.S. Department of Justice, National Institute of Justice, NCJ 244760., 2014.

[10] M. Duggan. The Demographics of Social Media Users. Technical report, Pew Research Center, 2015.

[11] A. Gupta and P. Kumaraguru. Twitter explodes with activity in mumbai blasts! a lifeline or an unmonitored daemon in the lurking? 2012.

[12] History.com Staff. Boston Marathon Bombings, 2014.

[13] P. N. Howard, A. Duffy, D. Freelon, M. M. Hussain, W. Mari, and M. M. Mazaid. Opening closed regimes: what was the role of social media during the Arab Spring? *Available at SSRN 2595096*, pages 1–30, 2011.

[14] W. Huang, I. Weber, and S. Vieweg. Inferring Nationalities of Twitter Users and Studying Inter-national Linking. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 237–242, New York, NY, USA, 2014. ACM.

[15] A. L. Hughes and L. Palen. Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3/4):248, 2009.

[16] J. Ito, T. Hoshide, H. Toda, T. Uchiyama, and K. Nishida. What is He/She Like?: Estimating Twitter User Attributes from Contents and Social Neighbors. *Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International*, (ii):1448–1450, 2013.

[17] Y. Jin, J. D. Fraustino, and B. Liu. The scared, the outraged, and the anxious: How crisis emotions, involvement, and demographics predict publics' conative coping. In *the Annual Convention of the International Communication Association*, San Juan, Puerto Rico, 2015.

[18] Y. Jin, A. Pang, and G. T. Cameron. Toward a Publics-Driven, Emotion-Based Conceptualization in Crisis Communication: Unearthing Dominant Emotions in Multi-Staged Testing of the Integrated Crisis Mapping (ICM) Model. *Journal of Public Relations Research*, 24(3):266–298, 2012.

[19] A. E. Kim, H. M. Hansen, J. Murphy, A. K. Richards, J. Duke, and J. A. Allen. Methodological considerations in analyzing twitter data. *Journal of the National Cancer Institute - Monographs*, (47):140–146, 2013.

[20] F. Kivran-Swaine, J. Ting, J. J. Brubaker, R. Teodoro, and M. Naaman. Understanding Loneliness in Social Awareness Streams: Expressions and Responses. In *International AAAI Conference on Web and Social Media*, pages 256–265, 2014.

[21] G. LaFree and A. Adamczyk. Change and Stability In Attitudes Toward Terrorism: the Impact of the Boston Marathon Bombings. Preprint, START Center, University of Maryland, jun 2015.

[22] V. Lampos and N. Cristianini. Tracking the flu pandemic by monitoring the social web. *2010 2nd International Workshop on Cognitive Information Processing, CIP2010*, pages 411–416, 2010.

[23] V. Lampos, T. De Bie, and N. Cristianini. Flu detector - Tracking epidemics on Twitter. In J. Balcázar, F. Bonchi, A. Gionis, and M. Sebag, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 599–602. Springer Berlin Heidelberg, 2010.

[24] S. M. Mohammad and P. D. Turney. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3):436–465, 2013.

[25] F. Morstatter, J. Pfeffer, H. Liu, and K. Carley. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. *Proceedings of ICWSM*, pages 400–408, 2013.

[26] A. Olteanu, S. Vieweg, and C. Castillo. What to Expect When the Unexpected Happens: Social Media Communications Across Crises. In *In Proc. of 18th ACM Computer Supported Cooperative Work and Social Computing (CSCWâĂŹ15)*, number EPFL-CONF-203562, 2015.

[27] M. Osborne, S. Moran, R. McCreadie, A. Von Lunen, M. Sykora, E. Cano, N. Ireson, C. Macdonald, I. Ounis, Y. He, and Others. Real-Time Detection, Tracking, and Monitoring of Automatically Discovered Events in Social Media. *Association for Computational Linguistics*, 2014.

[28] A. Pang, G. Cameron, and Y. Jin. Integrated crisis mapping: Toward a publics-based, emotion-driven conceptualization in crisis communication. *Sphera Publica*, 7:81–96, 2007.

[29] B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, jan 2008.

[30] M. J. Paul, M. Dredze, J. P. Michael, and D. Mark. You are what you Tweet: Analyzing Twitter for public health. *Icwsm*, pages 265–272, 2011.

[31] L. Petrecca. After bombings, social media informs (and misinforms), apr 2013.

[32] J. Rogstadius, M. Vukovic, C. A. Teixeira, V. Kostakos, E. Karapanos, and J. A. Laredo. CrisisTracker: Crowdsourced social media curation for disaster awareness. *IBM Journal of Research and Development*, 57(5):4:1–4:13, sep 2013.

[33] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

[34] D. Samuels and C. Zucco. The power of partisanship in Brazil: Evidence from survey experiments. *American Journal of Political Science*, 58(1):212–225, 2014.

[35] R. Sullivan. Live-tweeting terror: a rhetorical analysis of @HSMPress_ Twitter updates during the 2013 Nairobi hostage crisis. *Critical Studies on Terrorism*, 7(3):422–433, 2014.

[36] D. Tripathy and F. J. Daniel. Protests build in New Delhi after child rape, apr 2013.

[37] C. Vaccari, A. Valeriani, P. Barberá, R. Bonneau, J. T. Jost, J. Nagler, and J. Tucker. Social media and political communication: a survey of Twitter users during the 2013 Italian general election. *Rivista italiana di scienza politica*, 43(3):381–410, 2013.

[38] F. Vis. Twitter As a Reporting Tool for Breaking News. *Digital Journalism*, 1(1):27–47, 2013.

[39] C. Wells and K. Thorson. Combining Big Data and Survey Techniques to Model Effects of Political Content Flows in Facebook. *Social Science Computer Review*, pages 1–20, 2015.