# Predicting User Engagement on Twitter with Real-World Events

**Yuheng Hu**[1] **Shelly Farnham**[2] **Kartik Talamadupula**[1]
1. IBM Research
2. Third Place Technologies
{yuhenghu,krtalamad}@us.ibm.com, shelly@thirdplacetechnologies.com

**Abstract**

People invest time, attention, and emotion while engaging in various activities in the real-world, for either purposes of awareness or participation. Social media platforms such as Twitter offer tremendous opportunities for people to become engaged in such real-world events through information sharing and communicating about these events. However, little is understood about the factors that affect people's Twitter engagement in such real-world events. In this paper, we address this question by first operationalizing a person's Twitter engagement in real-world events such as posting, retweeting, or replying to tweets about such events. Next, we construct statistical models that examine multiple predictive factors associated with four different perspectives of users' Twitter engagement, and quantify their potential influence on predicting the (i) presence; and (ii) degree – of the user's engagement with 643 real-world events. We also consider the effect of these factors with respect to a finer granularization of the different categories of events. We find that the measures of people's prior Twitter activities, topical interests, geolocation, and social network structures are all variously correlated to their engagement with real-world events.

## 1 Introduction

Social media channels like Twitter and Facebook have emerged as some of the most important platforms for people to report, share, and communicate with others about various types of real-world events. These range from widely-known events (e.g., the U.S Presidential Debate, the Superbowl, and the Oscars) to smaller scale, local events (e.g., a gas leak at 5th and Main, a local parade, or a car accident). Social media has many advantages over the traditional media channels, such as ubiquity, immediacy, and seamless communication in covering real-world events. Given these advantages, social media posts like tweets can typically reflect events as they happen, in real-time. For this reason, recent years have witnessed a growing interest in research that aims to develop tools for real-world event detection and characterization based on social media posts (Sakaki, Okazaki, and Matsuo 2010; Shamma, Kennedy, and Churchill 2009).

Unfortunately, little is understood thus far about the factors that affect people's engagement with real-world events on social media (e.g. posting or exchanging event-related tweets): *Does a person post tweets about an event because they are interested in the topics pertaining to that event? Are they instead engaged because their friends are also posting tweets about it?* Or *is their engagement a reflection of the fact that this is a local event?* Furthermore, *how and to what extent do the different types of events affect the degree of a user's engagement?* Answering these questions holds the key to developing applications as diverse as marketing, political campaigns, and citizen journalism. Consider this: a personalized event recommendation engine can automatically recommend a list of new events (as they are happening) to a user, based on a prediction of that user's Twitter engagement – this can help users learn about and engage with more such events. At the same time, event organizers can take advantage of such a framework to identify potential audiences for their events based on predictions of users' engagement with their event, thereby enabling better and more productive targeted advertising and marketing.

This work aims to answer the questions put forth previously by exploring multiple predictive variables, and quantifying their potential influence on predicting a person's *presence* and *degree* of Twitter engagement with various real-world events. [1] Specifically, we operationalize a person's Twitter engagement with a real-world event as the posting of tweets about that event, including retweets and replies related to the event (e.g. "*OMG massive there is a massive fire right next to Madison Square Garden #pennstation*"). The *presence* of a person's Twitter engagement in response to an event can be defined as the existence of at least one tweet (or RT or mention) that references that event. The *degree* of the person's Twitter engagement is measured by the number of tweets that they post regarding that event; more such tweets indicate that they are more engaged with that event. Inspired by prior theoretical constructs that bridge social science, linguistics, and computer mediated communication, we collect factors that could potentially affect a person's Twitter engagement in real-world events from four broad categories. These are: (i) *Twitter activities* (prior to their engagement with an event), (ii) *tweets' content* (including topical interests), (iii) *geolocation* (the person's geographical proximity to the event), and (iv) *social network structure* (the follow-

[1] Here we are only interested in real-world events that are associated with all of the following: 1) a geolocation (where the event happens), 2) a timestamp (when it happens), and 3) responses on Twitter. Online events and offline events without geolocations and/or timestamps are currently outside the scope of this work.

ers, following, and common neighbors of the person).

We map these dimensions into 17 numeric predictive variables manifested on Twitter that span the volume of tweets produced, burstiness of tweets, frequency of retweets, usage of hashtags, communication mode (direct versus broadcast), topical interest extracted from a user's tweets and those of their following list, geolocation and geographic proximity, and social network structure. We construct two statistical models to assess the relative contributions of these variables towards predicting the presence of a person's Twitter engagement and the degree of that Twitter engagement in 643 real-world events. Additionally, we also perform finer-grained evaluation of our models with respect to different types of real-world events. Using our models, several insights about the aforementioned factors and their influence on predicting the presence and degree of Twitter users' engagement in real-world events are revealed. For example, in terms of the presence of engagement, we find that among all the predictive factors, a user's prior Twitter activity and her social network most significantly impact the presence of the user's engagement with events. Similarly, we also find that measures of topical interest have strong and statistically significant levels of impact on a person's degree of engagement during political, business, sports and sci-tech events (see Sec. 5 for more results and insights).

## 2 Background

**Twitter and Real-World Events** As social media has become prominent in daily life, the evolving ways in which information is generated, viewed, and shared have inevitably transformed people's engagement with real-world events (Kwak et al. 2010). Recent years have witnessed a growing research interest in developing tools for event identification and detection on social media (Sakaki, Okazaki, and Matsuo 2010; Becker, Naaman, and Gravano 2011). In addition, recent research also focuses on making sense of tweets and people's tweeting behavior around various real-world events such as political events (Hu et al. 2012; Diakopoulos, Naaman, and Kivran-Swaine 2010), local events (Hu, Farnham, and Monroy-Hernández 2013), and natural hazard events (Vieweg et al. 2010; Starbird 2013).

Despite the rich literature on Twitter and its role in covering real-world events, to date, we are aware of little research that directly addresses the issue studied in this paper. The most relevant related work is on modeling predictive factors on social media for various other issues such as tie formation (Golder and Yardi 2010), tie break-up (Kivran-Swaine, Govindan, and Naaman 2011), tie strength (Gilbert and Karahalios 2009) and retweeting (Suh et al. 2010).

Our effort differs from this past work in that we are exploring factors that may affect people's Twitter *engagement* in response to *real-world events*. Below, we discuss some background showing how a person's prior Twitter activities (e.g., communicating with others), her tweets' content (e.g., topical interests, linguistic styles), her geographical location, and her social networks relate to her Twitter engagement with real-world events.

**Social Activity, Social Capital and Event Engagement** Social capital has been identified as a collection of resources that either an individual or an organization gains through a set of communal norms, networks, and sanctions (Wellman and Wortley 1990). The relationship between social capital and event/civic engagement has long been a research topic (Shah 1998). In particular, many researchers have found that social capital is created when the engagement is "excited" by events and directed toward a particular end or purpose (Hyman 2002). At the same time, prior research has also identified several kinds of social activities and behaviors that can affect social capital on social media. These include directed communications with targeted individuals (e.g., Facebook private messages; Twitter replies, mentions, and favorites), broadcast communications which are not targeted at anyone in particular (e.g., Facebook wall updates or tweets with no "@" in them), and passive consumption of content (Burke, Kraut, and Marlow 2011). Moreover, the volume of social media posts (e.g., total number of tweets in a period) and the posting rate have also been shown to influence social capital (Hutto, Yardi, and Gilbert 2013).

Given the connections between social activity, social capital and event engagement, we empirically test whether a person's prior Twitter activities help in predicting their engagement with an event.

**Topical Interests and Event Engagement** The "endurability" theory (Read, MacFarlane, and Casey 2002) shows that people are likely to remember a good experience and are willing to repeat it. Application of this theory here indicates that a person may be more likely to engage with an event if the topics related to that event are the same as – or at least similar to – the topic that the person is interested in on Twitter.

There are many ways to infer a person's topical interests on Twitter. Perhaps the most straightforward way is based on the content of the person's previous tweets (Abel et al. 2011). Of course, a person's topical interests can be inferred from other resources as well, such as the person's following list (Burgess et al. 2013). This is because, according to the principle of *homophily*, the similarity between individuals leads to a greater potential for interpersonal connections; when establishing connections, people tend to build relationships with others who are like them (McPherson, Smith-Lovin, and Cook 2001). Sharing interests with another person is one form of similarity (Feld 1981) that can be used to build relationships; this can lead to the follow relationship being established.

Here, we empirically study how the topical interests of a person (inferred both from their tweets as well as the users they follow) affect their engagement with events on Twitter.

**Twitter User Types and Event Engagement** Naaman et al. found that there are two basic categorizations of Twitter users: *informers*, who share informational content; and *meformers*, who share tweets about themselves (Naaman, Boase, and Lai 2010). One effective way to distinguish informers and meformers is based on the linguistic styles of their tweets. Those who share information or describe things tend to use more third person pronouns (*She, He, It, They, etc*), while meformers, who post mostly about themselves tend to use first person pronouns (*I, We, Us, etc*) more often.

Here, we *posit* that informers are more likely to engage in events through the posting or sharing of information than meformers. We explore linguistic styles of tweets and examine whether different types can predict a person's event engagement on Twitter.

**Geolocation and Event Engagement** It is known that a person's geographical location (geolocation) significantly affects their social connections and activities in the offline world. Recent research has also found evidence to show that offline geography has a significant impact on user interactions, tie formation, and information diffusion on online social media like Twitter (Kulshrestha et al. 2012). In particular, researchers have discovered that users preferentially connect and exchange information with other users from their own country, and lesser information is exchanged across national boundaries. However, even such transnational links and interactions occur between users in geographically and linguistically proximal countries within their network. Similarly, researchers also identified that geographical proximity plays a key role in trend/innovation adoption (Toole, Cha, and González 2012). Based on these results, we *posit* that a person's geolocation may affect their engagement with real-world events on Twitter if that person's location is geographically proximate to the event's location (e.g., a user may only care about events that happen in their neighborhood).

**Social Networks and Event Engagement** The correlation between social network influence (e.g., network size and social ties) and user engagement has been studied extensively. For example, (de Zúñiga and Valenzuela 2011) showed that the relationship between online and offline network size and people's engagement with civic events is positive . They further found that network structure and social ties (especially weak ties) are determined to be strong predictors of the engagement. There are many different ways to form ties on Twitter, and ties can be formed either directly or indirectly. For example, following a person on Twitter can be seen as a direct tie. In such cases, dyadic properties such as reciprocity play key roles in the process of tie formation. On the other hand, ties can be formed indirectly such as through common network neighbors (known as transitive ties). For example, consider the case where three people form an undirected network: $A$ and $C$ are both friends of $B$, but $A$ and $C$ are not friends. However, as the number of common neighbors (occurrences of $B$) between $A$ and $C$ increases, the likelihood of an $A$-$C$ tie being formed and the corresponding tie strength also increase (Cartwright and Harary 1956). In this paper we *explore* the extent to which these network sizes and tie formations impact a person's engagement in real-world events as compared to the person's Twitter activities, topical interests, user types and geolocation information.

# 3 Data Collection

In this section, we describe our data collection strategy. Note that in order to show how people's Twitter activities, their topical interests, their Twitter user types, their geolocations, and their social networks affect their Twitter engagement with real-world events, we needed to collect: 1) a list of real-world events and their associated tweets, and 2) profiles of Twitter users (who post event-related tweets). Moreover, since we want to evaluate the influence of people's geolocations on their event engagement, we needed to infer the geolocations of both Twitter users and events.

## 3.1 Obtaining Real-world Events and Events' Geolocations

To identify real-world events from tweets, one possible solution is to first obtain an event list directly from newspapers (since reporters often tend to mention the location of the event in their news articles about that event) and then fetch the corresponding tweets. However, such an approach is not applicable for several reasons. First, not every event reported by newspapers is popular/trending on Twitter. As (Hong, Dan, and Davison 2011) pointed out, the popularity of tweets is affected by multiple reasons aside from newsworthiness . Second and more importantly, such an approach will be significantly biased towards larger, more broadly newsworthy events due to the nature of newspapers, which could potentially misguide our analysis. To avoid this, we followed a different approach by first detecting real-world events from Twitter streams, and then inferring their geolocations later.

For the first step, we adopt the framework mentioned in (Becker, Naaman, and Gravano 2011) to automatically detect real-world events from Twitter. The framework has two stages: first, a clustering algorithm is used to cluster tweets into multiple clusters. Next, for each cluster, a classifier is used to distinguish between real-world events and non-event clusters. More specifically, we use a hierarchical clustering algorithm to cluster tweets, where the distance function between two data points (i.e., tweets) is defined as their topical similarity. We chose this algorithm because it is scalable and does not require *a priori* knowledge of the number of clusters. To calculate the topical similarity, we use topic model LDA (Blei, Ng, and Jordan 2003), a popular machine learning tool for getting topic distributions from tweets. We then use the Jensen-Shannon (JS) divergence on their topic distributions to measure the topical similarity between two tweets. After that, we train a binary classifier to classify the clusters (obtained from the previous step) into two classes: real-worlds events and non-events. Our classifier uses a set of features similar to (Becker, Naaman, and Gravano 2011), including temporal features, social features, topical features and ego-centric features.

To infer the geolocations of the real-word event clusters, we asked annotators to individually read a sample tweet from each real-world event's cluster to gain an understanding of what the event is really about. The annotators were then asked to find the geolocation of the event cluster via search engines by coming up with their own search keywords (e.g., event-related hashtags, timestamps) based on their event understanding. Our assumption here is that many real-world events will be covered by news, blogs, and other media, and their geolocations will often be mentioned. Surprisingly, this simple approach yields results whose quality is good enough for use in this work.

## 3.2 Obtaining Twitter Users' Geolocations

Inferring Twitter users' geolocation based on their tweets has been an emerging research topic in recent years (Cheng, Caverlee, and Lee 2010). In this work, we follow the methods mentioned in (Mahmud, Nichols, and Drews 2014) to infer the geolocations of Twitter users[2]. Specifically, the lo-

[2]One may consider inferring a Twitter user's geolocation based on the information from her Twitter profile, i.e., she may mention

cation inference algorithm uses tweet content, tweeting behavior and other auxiliary information such as time zone to predict the home location of Twitter users. We then verify the extracted location information with the diurnal patterns of the user's tweets (Naaman et al. 2012). For example, most people in New York City will tweet about having dinner and the nightlife between 5:00PM EST to 1:00AM EST. So if a person regularly posts tweets about lunch around 12:00AM EST, they probably are not from the New York City area. Based on our preliminary testing, we found this algorithm together with the diurnal pattern verification yielded stable performance (78.4% for cities).

### 3.3 Constructing the dataset

In practice, we first obtained nearly 2.7 billion English tweets from the Twitter firehose during August of 2014. We then applied the automated event detection algorithm mentioned in (Becker, Naaman, and Gravano 2011) on these tweets to find real-world events. As a result, we obtained 7,468 real-world event clusters.

Next, we needed to infer the geolocations of these event clusters. We hired 20 annotators to read 10 sample tweets from each of their assigned event clusters (each annotator was assigned roughly 373 event clusters) and infer the geolocation (via search engines) based on their understanding of the events. As a result of this step, our annotators were able to infer the geolocations (on city level) for 643 event clusters. Among these 643 event clusters with geolocations identified, 425 events happened in the U.S (e.g., New York City, NY, Beverly Hills, CA, Ferguson, MO), and the rest were in Europe, the Middle East, and Asia.[3]

Finally, based on those 643 events, we obtained a total number of 22,957 Twitter users who posted at least three tweets in response to one of these events. We applied the location inference algorithm (see above) to predict the location (on city level) of each user. Besides, in order to calculate the measures for the predictor variables, we collected all the tweets posted by each user in the most recent six months preceding their first ever event engagement with any of the 643 events used.

## 4 Methods

With 643 events obtained, in this section, we provide more details on the statistical models that are used to predict the presence and magnitude of a person's engagement on Twitter with a given event. We first present the dependent variables used in our predictive models, followed by a description of the predictor variables.

### 4.1 Dependent Variables

*Presence of a person's Twitter engagement in a real-world event:* This is a binary measure that indicates whether or not a person posts, replies to, or retweets tweets in relation to a particular event on Twitter (1: engaged; 0: not engaged).

*Degree of a person's Twitter engagement in a real-world event:* This is a continuous measure that indicates the number of tweets that the person generates (via post, reply to, or retweet) relating to the event.

### 4.2 Predictor Variables

The literature reviewed in the previous section pointed us to five major kinds of predictor variables: Twitter activities, tweets' topics, Twitter user types, geolocation, and social network structure. Using these categories as a guide, we collected 17 variables that are manifested on Twitter as potential predictors of a person's Twitter engagement with a real-world event.

**Variables related to Twitter activities**

*Total number of tweets.* The total number of tweets a person has posted. These tweets include new posts, retweets, and replies.

*Maximum tweets per hour.* The maximum rate of tweets per hour, which captures the "burstiness" of a person's activities.

*Average tweets per hour.* The average rate of tweets per hour, which gives a general idea of a person's level of activity.

*Directed communications.* The number of tweets with "@" (including both @mentions and @replies) plus the number of favorite tweets divided by the total number of tweets. This measure indicates interpersonal activities between the person and other Twitter users.

*Broadcast communications.* The ratio of tweets with no "@" at all in the tweet to total number of tweets in a period.[4]

*Ratio of retweets.* The total number of times a person reposts other Twitter users' tweets, relative to the total number of tweets produced by the person in a period. This measure complements the direct and broadcast communication measures by indicating how often the person interacts with other Twitter users and broadcasts those users' tweets to their own social circle (i.e., their followers).

*Hashtag usage.* This is defined as the ratio of tweets that contain at least one hashtag to the total number of tweets from a person in a period.

**Variables related to tweets' content**

*Topical interests from tweets' content.* This measure is calculated as the topical similarity between two topic distributions: the first is computed based on a person's tweets in a period, while the second is computed based on all the event-related tweets (from other users) posted prior to the person's engagement with that event. In practice, assume a person $u$ has posted $T_u$ tweets in the past three months. Now, assume an event starts at 8:00PM and $u$ engages with this event on Twitter (i.e., user $u$ posts their first event-related tweet) at 8:30PM. Additionally, between 8:00PM and 8:30PM, there are $T_Q$ event-related tweets posted by a set of other users $Q$. We then apply topic model LDA (we set the number of topics $K = 20$ in practice) (Blei, Ng, and Jordan 2003)

her location in her profile. However, this method may result in biased samples since those Twitter users who have relatively more complete profiles, i.e., location, might also be more active in general and tend to have more friends.

[3]Note that most detected event with geolocations were newsworthy and/or major breaking events (e.g., sports game, protests) where the geolocations could be easily inferred from traditional news media by our annotators.

[4]It is possible that a tweet could be a broadcast and yet include the "@" symbol, but we disregard this relatively rare occurrence in this work.

on both $T_u$ and $T_Q$ to learn the topic distributions respectively. We then measure the topical interests similarity between the two learned distributions based on JS-divergence. Intuitively, higher similarly indicates that the person's prior exhibited topical interests (reflected from their prior tweets' content) are closer to the event's topics (which are inferred from other people's event-related tweets).

*Topical interests from the person's following list.* This measure is calculated based on the topical similarity between the topics of the tweets written by the people that a person follows, and the event's topics. More specifically, the user's following list's topics are computed using methods mentioned in (Burgess et al. 2013): first, given the following list of a person, we obtain the 200 most recent tweets from each user on that list. Next, we distill topic distributions from these tweets using LDA. On the other hand, as mentioned above, we are also able to get $T_Q$ event-related tweets posted by other users prior to the target person's engagement in the event (i.e., their first event-related tweet). For the tweets $T_Q$, we run LDA (we set the number of topics $K = 20$ in practice) to obtain the same number of topics as the vectors in the previous analysis, and the corresponding topical distribution for each topic. We then compute the similarities between the two topical distributions – one learned from the following list, and the other from the event – using JensenShannon divergence. As with the cosine similarity measure, higher similarity here indicates that the person's topical interests (reflected from the list of people that they follow) are closer to the event's topics.

**Variables related to Twitter user types**

*Meformer.* This is computed as the ratio of meformer tweets to the total number of tweets by a person in a period. As mentioned earlier, we detected meformer tweets based on linguistic styles. More specifically, if a tweet contains any of the 24 self-referencing pronouns (e.g., words like "I", "me", "we", "us") identified in LIWC, then it is classified as a meformer tweet.

*Informer.* This is computed as the ratio of informer tweets to the total number of tweets by a person in a period. We identified informer tweets as those containing any of the 20 third-person pronouns (e.g., words like "He", "She", "it", "them") defined in LIWC. In addition, if a tweet contains either a URL, "RT", "MT", or "via", we deem it an informer tweet as well.

**Variables related to geolocation information**

*Geographical proximity.* The first measure considers the geographical proximity between a person's location and the event's location. As indicated in the previous section, the dataset used in this study only includes Twitter users and events whose geolocations could be identified. Note that it is impractical to model the proximity of two geolocations continuously in terms of their physical distance in miles, because the effect of geographical proximity may not be linearly proportional to physical distance. In fact, such an effect is more likely to be exponential according to recent research (Kulshrestha et al. 2012). Therefore, in practice, we only consider measuring the geographical proximity in terms of two discrete bins: local (distance within $\leq 50$ miles) and non-local (distance $> 50$ miles).

**Variables related to network structure**

*Number of followings.* The number of Twitter users that a person was following.

*Number of followers.* The number of Twitter users who were following the person.

*Followings posted prior.* The number of a person's followings who had already posted event-related tweets before the person posted to that event. As discussed earlier, since following (e.g., A follows B) forms a directed tie, it is possible that the person will be influenced to post tweets when a lot of their followings post about an event prior to their own engagement.

*Average common neighbor prior.* This measure examines the overlaps between the followings of a person $a$ and the followings of user $b$, where $a$ has already engaged in the event on Twitter while $b$ has not. In the context of Twitter, a person's following list often represents their interests. Therefore, the common neighbor factor essentially measures the shared interests between two people. According to triadic closure, such a measure also indicates the tie strength between $A$ and $B$ (Wasserman 1994). In practice, for the user $a$, we compute this feature as $\sum_{b \in B} CommonNeighborhood(a, b)/|B|$, where $B$ is a set of users who have already engaged in the event ($a \notin B$).

*Number of followings about news.* This measure is defined as the total number of a person's followings who are deeply involved in news. To identify such news related accounts, we first obtain Twitter profiles for all of the person's friends. We then look at each profile to check which ones contain news related keywords such as "news", "reporter", "journalist", "TV" and so on. We deem those users news related accounts. One motivation for this measure is that news agencies are often authorities and first-hand resources for reporting events. It is possible that if a person followers a lot of news agency accounts, then they will likely be interested in knowing about and engaging with real-world events.

# 5 Results

In the following section, we first provide descriptive statistics for the variables used in our statistical models. Following this, we present the contribution of these variables in predicting the presence and degree of people's Twitter engagement with real-world events.

## 5.1 Descriptive Statistics

Table 1 shows descriptive statistics (mean, standard deviation) for the number of events that one person engages with, and the event-related tweets that person posts – along with 17 predictor variables – based on the event data we collected in August 2014 (see the 'Data Collection' section). For comparison, we also generate statistics based on an event participant's regular tweets five months prior the the events (i.e., March 2014 to July 2014). We calculate the significance of the difference between these two situations. Note that some of the predictor variables are compared pair-wise, such as topical interests, geographic proximity and so on. Therefore, we only report the pair-wise statistics for the event data.

Also note that we excluded users that were extreme outliers ($z$-score $> 4.0$) with respect to our metrics for activity levels and follower/following counts. As a result, we had a total of 22,170 people (we removed 787 "outlier" users from a total of 22,957 users in our dataset, see Data Collection section) participate, by posting Twitter messages over the course of 643 events. Within these messages, 28% of the messages had hashtags, 48% retweets, 27% direct replies, 33% links, and 68% mentions, indicating that the event participants were highly interactive.

| | Not engaged with events | | Engaged with events | | Difference |
|---|---|---|---|---|---|
| *Measure* | *Mean* | *SD* | *Mean* | *SD* | *p-value* |
| **Event engagement** | | | | | |
| Event count | – | – | 12.1 | 7.78 | – |
| Tweet count | – | – | 3.53 | .322 | – |
| **Twitter activity variables** | | | | | |
| Total tweets | 278.2 | 167.2 | 28.2 | 5.82 | $***$ |
| Max tweets per hr | 6.39 | 5.78 | 6.57 | 5.44 | ns |
| Avg. tweets per hr | 1.14 | .004 | 1.74 | 1.23 | ns |
| Directed communications | 2.81 | 6.4 | 1.83 | 5.33 | $***$ |
| Broadcast communications | .83 | .22 | 1.48 | 1.11 | $***$ |
| Hashtag ratio | .2 | 0.24 | .42 | .006 | $***$ |
| RT ratio | .15 | .41 | .44 | .054 | $***$ |
| **Twitter content variables** | | | | | |
| Topical interests from tweets content | – | – | .25 | .04 | – |
| Topical interests from following | – | – | .11 | .01 | – |
| **Twitter user type variables** | | | | | |
| Meformer tweets | .41 | .14 | .29 | .219 | $***$ |
| Informer tweets | .24 | .23 | .43 | .31 | $***$ |
| **Geolocation variable** | | | | | |
| Geographical proximity | – | – | 318mi | 189.8mi | – |
| **Social network variables** | | | | | |
| Num. followers | 387 | 150.1 | 387 | 150.1 | ns |
| Num. friends | 117 | 109.78 | 117 | 109.78 | ns |
| Followings posted prior | – | – | 4.33 | 5.00 | – |
| Average common neighbor prior | – | – | 10.33 | 10.42 | – |
| News friends | 5.73 | 8.22 | 5.73 | 8.22 | ns |

**Table 1:** Mean and SD values for Twitter users' event engagement, compared to averaged values of these Twitter users' non-event tweeting behavior, and paired sample $t$-tests for the difference. $***\,p < 0.001$, $**\,p < 0.01$, $*\,p < 0.05$.

**Twitter activities** On average, a user engaged in 12.1 events over a month, and they posted 3.53 tweets per event. In terms of burstiness, users posted no more than 6.57 tweets within an hour (average). This seems to indicate that over the course of an event, people tend to post using a stable pace (as avg. tweet per hr is very different between tweets from the event a person engaged in and norma tweets from the person's prior tweets history. The Broadcast Communication shows the average number of tweets that are not directed to any specific person. During events, this rate is significantly higher. Such changes are also reflected in directed communication. The ratios of retweets and hashtag usage to the total number of tweets in a period are moderate for the majority of users retweets comprised about 15% of users' messages, and hashtags were used in about 20% of tweets. Compared to these, we witness significant changes during events – where the ratio of hashtags and retweets increases to 44% and 42% respectively. Combining these discoveries, we conclude that people tend to communicate more with others during an event that they are engaged in, thus showing a deeper involvement and engagement with the topics related to that event.

**Tweet content** In general, users show a fairly diverse range of topics that they post in relation to, which is reflected in and manifested as the relatively low topical similarity to actual event topics. In particular, the topic similarity inferred from a user's tweet content is 0.25, while topic similarity inferred from their followings is 0.11.

**Tweet user types** Besides, nearly half of users' regular tweets are identified as "meformer" (41%), and the "informer" category accounts for 24% of tweets. However, in the context of event engagement, the percentage value of "informer" tweets witnessed a sharp increase to 43%, and "meformer" tweets decreased to 29%. This indicates that people tend to share more information (e.g., through retweets, third person comments about the event) during the course of an event. However, people do also continue posting information about their thoughts and their presence during the event.

**Geolocation** In terms of the geographical proximity between the event participants' location and the event's location, we found that most events were non-local to the event participants – this is reflected in that measure's relatively high value (i.e., 318 miles between the inferred event participants' locations and the events' locations), accompanied with high standard deviation (189.8 miles).

**Social network** The majority of users have an average of 387 followers, and 117 friends. About 4.33 event participants who joined in the event prior to the target user's engagement are the followings of that user. Moreover, for the people who posted prior to the target user but are not part of the following set, it is seen that there are around 10 common friends between those users and the target user. This indicates that one-hop weak ties do exist between event participants. Later we will demonstrate the strength of these predictors.

## 5.2 Prediction of presence

We now turn to the core question examined in this study: *to what extent do the 17 variables used predict the presence, and degree, of a person's Twitter engagement with a real world event?*

In order to examine the relative impact of these variables, we first standardized the measures, and then examined whether they predicted a user's participation/engagement using a repeated measures (643 trials, or events) logistic regression. The question of whether or not the user participated was modeled as a binary dependent variable. Table 2 shows the results of this regression. An immediate insight that can be gleaned is that the total tweets posted by a user prior to her event engagement is a significant predictor of whether the user will take up or engage with an event. Specifically, as far as communication oriented tweets are concerned, both directed and broadcast communication are

| | $\beta$ | ***SE*** | ***p-value*** |
|---|---|---|---|
| **Twitter activity variables** | | | |
| Total tweets | .37 | .045 | $< .001^{***}$ |
| Max tweets per hr | .01 | .039 | 0.21 |
| Avg. tweets per hr | -.08 | .033 | 0.11 |
| Directed communications | -.17 | .071 | $< .001^{***}$ |
| Broadcast communications | .04 | .059 | $< .001^{***}$ |
| Hashtag ratio | .09 | .045 | $< .001^{***}$ |
| RT ratio | .069 | .039 | $< .001^{***}$ |
| **Tweet content variables** | | | |
| Topical interests from tweets content | .12 | .039 | .22 |
| Topical interests from followings | .07 | .017 | .17 |
| **Twitter user types variables** | | | |
| Meformer tweets | .06 | .013 | .46 |
| Informer tweets | .02 | .016 | $< .001^{***}$ |
| **Geolocation variables** | | | |
| Geographical proximity | .01 | .032 | 0.59 |
| **Social network variables** | | | |
| Num. followers | -.04 | .023 | $< .001^{***}$ |
| Num. friends | -.07 | .016 | $< .001^{***}$ |
| Followings posted prior | -.02 | .024 | $< .01^{**}$ |
| Average common neighbor prior | -.22 | .015 | $< .01^{**}$ |
| News friends | .30 | .022 | $< .001^{***}$ |

**Table 2:** Prediction of presence: Logistic regression coefficients for standardized variables in simultaneous repeated measures logistic regression predicting participation in events over 643 "trials". Adjusted $R^2 = 0.67$, $***\, p < 0.001$, $**\, p < 0.01$, $*\, p < 0.05$

| | $\beta$ | ***SE*** | ***p-value*** |
|---|---|---|---|
| **Twitter activity variables** | | | |
| Total tweets | .087 | .055 | $< .001^{***}$ |
| Max tweets per hr | .02 | .043 | .32 |
| Avg. tweets per hr | .11 | .033 | $< .01^{**}$ |
| Directed communications | -.17 | .028 | .24 |
| Broadcast communications | -.01 | .076 | $< .001^{***}$ |
| Hashtag ratio | .11 | .045 | $< .01^{**}$ |
| RT ratio | .49 | .012 | $< .01^{**}$ |
| **Tweet content variables** | | | |
| Topical interests from tweets content | .11 | .029 | 0.12 |
| Topical interests from followings | .06 | .02 | 0.21 |
| **Twitter user type variables** | | | |
| Meformer tweets | -.11 | .014 | .28 |
| Informer tweets | .21 | .026 | $< .001^{***}$ |
| **Geolocation variables** | | | |
| Geographical proximity | .02 | .032 | .52 |
| **Social network variables** | | | |
| Num. followers | -.04 | .033 | $< .001^{**}$ |
| Num. friends | -.11 | .046 | $< .001^{**}$ |
| Followings posted prior | .02 | .024 | $< .001^{**}$ |
| Average common neighbor prior | -.02 | .015 | $< .001^{**}$ |
| News friends | .13 | .022 | $< .14$ |

**Table 3:** Prediction of degree: Linear regression coefficients for standardized variables in simultaneous repeated measures logistic regression predicting participation in events over 643 "trials". Adjusted $R^2 = 0.56$, $***\, p < 0.001$, $**\, p < 0.01$, $*\, p < 0.05$

good indicators, albeit in opposite senses. The coefficients for those variables seem to indicate respectively that lower directed communication or higher broadcast communication correlate directly with higher engagement. This is fairly intuitive, since directed communication tends to be among a user's friends and about non-event topics, and in most cases can only be seen by the mentioned users; while broadcast communication is intended for a wider audience consisting of all of the user's followers. Finally, both the ratio of hashtags used and the ratio of retweets are positive indicators of event engagement; this is easy to see since RTs and hashtags respectively are two key ways in which a user can signal their active interest and affiliation with an event.

As far as the tweet content variables and Twitter user types variables are concerned, we did not find evidence of the topical interests being good predictors of engagement with events that display those same topics. However, we will show later in our analysis that when the tweets are broken down by topic and not considered as a single monolithic set, these topic-specific correlations become stronger predictors of engagement. As regards meformer versus informer tweets, the meformer tweets are not very good predictors of engagement, which is obvious since such tweets mostly involve the user talking about things that are highly personalized and hyper-local to their own lives. Informer tweets, on the other hand, display a positive correlation to engagement; since such tweets are usually in the third person, this result combined with the broadcast communication considered previously indicate that a user who posts such tweets will usually engage with something that multiple other users are also interested in (hence an event as against a personalized happening).

As concerns geolocation, we did not find any significant evidence – in contrast with prior research (Kulshrestha et al. 2012) – that the geographical proximity has any effect on a user's engagement with an event. This would seem to indicate that users will choose to engage with an event whether or not it is "local" (in their surrounding vicinity) or non-local.

Finally, where the social network variables are concerned, we find that all of the variables are predictors with at least some degree of significance (and some more so than others). Interestingly, the only positive correlation is with the number of new friends. A further manual inspection revealed that most of the news friends' posts actually are occurring before the user starts contributing messages and engaging with the event. This indicates that users are inspired and motivated to engage with events when they see tweets from news agencies relating to those events on their timelines. However, this only goes so far – as the negatively correlated variables show, a large number of friends/followers and neighbors may bring down awareness, engagement, and subsequent participation (i.e., their coefficients are negative). We argue that this can be possibly attributed to a variety of factors. Some of these may include cognitive overload on the part of the target user, higher noise, posts being perceived as less personal, and most importantly, a perception that the topic is already sufficiently covered, e.g., posted by friends (thus reducing an "informer" user's motivation in engaging with it).

## 5.3 Prediction of Degree

To further explore the relative impact of these variables in predicting the *degree* of prediction in new events, we performed a linear regression, using participation levels in past events to predict the level of participation in a final, target event. The results are shown in Table 3.

We find that the most significant predictors of the degree of a user's engagement happen to be the social network variables, followed by the twitter activity variables. Specifically, the only social network variable that shows a significant positive correlation is the number of posts from the user's friends prior to the user's engagement with the event, which can be explained in terms of the activity that a user sees on their timeline with regard to that event. However, as in the previous case, increases in the user's network size seem to dampen the degree of engagement somewhat (which can be attributed to many of the same reasons described previously). The participant's own past tweet content seemed to have no significant effect on the predicted degree of engagement, save for the total and broadcast tweets, which offer a historical window into how active the user was in general.

### 5.4 Prediction of Degree w.r.t Different Topics

Finally, we are interested in understanding how, and to what extent, the decomposition of events into their constituent topics affects the performance of our predictors (for predicting the degree of people's engagement). To this end, the first task is to infer topics from our event clusters. We obtained six popular event categories from a news agency: politics & business, technology & science, entertainment, sport, local, and odd news. We then asked 30 annotators to code the event clusters manually, and resolved conflicts later Note that we only allow one label for a given event. Subsequent to this, we ran the linear regression again – these results are displayed in Table 4.

Interestingly, we observe that some predictors do indeed change with respect to different event topics. For example, we witness that for events related to politics & business, the effect of social activities such as the total number of tweets and the max number of tweets per hour exhibits a higher $\beta$ value when contrasted with the findings from the general events in Table 3. We also found that the effect of a person's topic interest is stronger for politics, business and sports events, but relatively lesser for entertainment events. These results suggest that people who are devoted to politics and sports tend to be more recognizable and explicit (e.g. political junkies, business analysts, and followers; sports fans). However, entertainment and science & tech events may consist of event topics – and subsequently user engagement – that varies thick and fast. As regards news users and tweets, following these becomes imperative for politics & business, tech & science, sports, and entertainment; while friends usually tend to post before users engage with local events and odd events. More generally, these results demonstrate the different pathways of information within a social network structure such as Twitter's. For news events, people first learn about them (and thus engage with them) via information posted from news accounts; if they find the event and its topics interesting, they tend to intensify the level of engagement. However, for local events and odd news, people tend to get engaged more via their friends' tweets, and thus the effect of information from friends is shown as more important. Finally, one of the most interesting contrasts occurs with respect to geolocation and geographic proximity – in Tables 2 and 3, geolocation information rarely affects the presence and degree of event engagement. However, when we look at the effect of geolocation with respect to the various event topics, it is shown to be more important for sports and local events. This makes complete intuitive sense: an overwhelming number of users tend to care, to a very large extent, about their own local sports teams and about local events that they may directly affect them.

## 6 Discussion and Implications

At the beginning of this paper, we posed five important questions relating to the engagement of users on social media with real-world events; and whether such engagement (and its level) could be effectively and practically predicted based on information available from that social media. In this section, we consider possible answers to those questions that are suggested by the data and revisit the related theories to examine our answers.

*Does a person post tweets about an event because they are interested in the topics pertaining to that event?*

Our analysis confirms that this is indeed the case. To highlight this, we point the reader to the analysis concerning prediction of presence and degree (Table 2 and 3), and the contrast with the similar prediction analysis given a breakdown of the events into different topics (Table 4). In the former case, there is no significant indicator of correlation from the content of a user's tweets to their engagement with an event. However, in the latter, there is a marked increase in the significance of the correlation between the content of tweets related to events in specific topics, and the user's engagement with those events (e.g., politics & business, tech & science, and sports). This is exactly what the "endurability" theory (Read, MacFarlane, and Casey 2002) proves: people are likely to remember a good experience and are willing to repeat it. In other words, people like to repeatedly talk about the topics that they are most familiar with/interested in. So, they will show deeper engagement in those specific topics, in contrast to boarder and more general topics.

*Are they instead engaged because their friends are also posting tweets about it?* The answer to this is positive as well, conditioned on the type of event that the user is engaging with. We have shown in the previous section that certain kinds of events – local events, as well as odd news – users tend to engage more due to their friends (following list) posting content relating to those events prior to the user's own engagement. This verifies the discoveries by Zuniga et al. (de Zúñiga and Valenzuela 2011) network structure and social ties (especially weak ties) are determined to be strong predictors of the civic engagement. We also extend their theory by discovering the social network and time affects on the engagement with real-world events (indeed, some events are about civic issues).

*Perhaps they are just a very active user of Twitter?* The degree to which a user was active on Twitter (the number of tweets posted by them) does indeed show a strong correlation across all cases to their predicted engagement with an event. This correlation seems to be agnostic of the type of event (as against the previous two questions, above), and hence it seems likely that more active users are more likely to be interested and engaged in a new event, across the board. This finding validates our earlier conjecture that these activities will first directly affect people's engagement in events on social media; such engagement will later in-

| | politics and business | | tech and science | | entertain | | sports | | local | | odd events | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ | ***p-value*** | $\beta$ | ***p-value*** | $\beta$ | ***p-value*** | $\beta$ | ***p-value*** | $\beta$ | ***p-value*** | $\beta$ | ***p-value*** |
| **Twitter activity variables** | | | | | | | | | | | | |
| Total tweets | .39 | $***$ | .074 | $**$ | .071 | $**$ | .066 | $**$ | .078 | $**$ | .11 | $**$ |
| Max tweets per hr | .12 | $**$ | .02 | ns | .01 | ns | .08 | ns | .04 | ns | .02 | ns |
| Avg. tweets per hr | .03 | ns | .04 | ns | .11 | ns | .07 | ns | .08 | ns | .12 | ns |
| Directed communications | -.12 | ns | -.12 | ns | .11 | $**$ | -.17 | $**$ | .01 | ns | -.101 | ns |
| Broadcast communications | -.02 | $***$ | -.04 | $***$ | .02 | $***$ | -.11 | $**$ | -.02 | $***$ | -.08 | $***$ |
| Hashtag ratio | .08 | $*$ | .09 | $**$ | .11 | $***$ | .21 | $***$ | .07 | $**$ | .106 | $***$ |
| RT ratio | .09 | $**$ | .06 | $**$ | .071 | $**$ | .087 | $***$ | .08 | $*$ | .19 | $**$ |
| **Tweet content variables** | | | | | | | | | | | | |
| Topical interests from tweets content | .22 | $**$ | .12 | $**$ | .02 | ns | .62 | $***$ | .12 | ns | .12 | ns |
| Topical interests from followings | .08 | $**$ | .07 | $**$ | .07 | ns | .54 | $**$ | .081 | ns | .067 | ns |
| **Tweet content variables** | | | | | | | | | | | | |
| Meformer tweets | -.06 | ns | -.09 | ns | -.05 | ns | -.01 | ns | .01 | ns | -.02 | ns |
| Informer tweets | .21 | $***$ | .102 | $**$ | .12 | $***$ | .08 | $***$ | .09 | $***$ | .11 | $***$ |
| **Geolocation variables** | | | | | | | | | | | | |
| Geographical proximity | .01 | ns | .02 | ns | .01 | ns | .20 | $***$ | .42 | $**$ | .00 | ns |
| **Social network variables** | | | | | | | | | | | | |
| Num. followers | -.18 | $**$ | -.12 | ns | -.21 | ns | -.104 | $**$ | .04 | $*$ | .07 | $***$ |
| Num. friends | -.11 | $**$ | -.21 | $*$ | .08 | $*$ | .107 | $**$ | .11 | $*$ | .02 | $**$ |
| Followings posted prior | -0.2 | ns | -.02 | ns | -.12 | ns | -.11 | ns | .71 | $**$ | .14 | $***$ |
| Average common neighbor prior | -.12 | $*$ | -.02 | $*$ | -.15 | $**$ | -.02 | $**$ | .011 | $***$ | .033 | $**$ |
| News friends | .51 | $***$ | .39 | $**$ | .22 | $***$ | .21 | $**$ | .06 | $**$ | .04 | $**$ |

**Table 4:** Prediction of degree of Twitter engagement given different topics: Linear regression coefficients for standardized variables in simultaneous repeated measures logistic regression predicting participation in events over 643 “trials”, $***\ p < 0.001$, $**\ p < 0.01$, $*\ p < 0.05$

directly affect social capital. Our finding extends existing literatures on the relationship between social media activities and social capital (Burke, Kraut, and Marlow 2011; Hyman 2002) by exploring the role of user engagement.

*Is their engagement a reflection of the fact that this is a local event?* The answer to this question reverts to the pattern of dependence on the kind of event observed in the answers to the first two questions. There are certain *kinds* of events that can be classified as engaging to a user primarily due to their local nature – as described in the previous section, these tend to be sports and local events. The connection to local events is obvious and trivial; a user in New York City is unlikely by and large to care about events that are happening in (say) far-off Tulsa, Oklahoma. For sports, it is likely that users within a given geographical area are more likely to care about teams that call that particular area home (although of course there will always be outliers; however, our analysis is focused on the typical user).

*How and to what extent do the different topics of events affect the degree of a user's engagement?* The answer to this question can be found in the aggregation of the answers to all of the previous questions – it does certainly seem like the different genres of events (even among the typical genres that we considered) affect the degree, and nature, of a user's engagement with an event. While engagement with political & business, science & technology and sports events seems to depend more on the content of past tweets (both of the user as well as the people they follow), engagement with local and odd events tends to correspond more closely with the user's social networks.

**Limitations** Although the data that we use and the results produced from that data seem to imply some rather strong conclusions, certain limitations of the study must also be considered when going forward. The first of these is the categorization of events: although the categories we use in this study are quite general, and capture a large portion of the posts on Twitter, arguments can certainly be made in support of finer-grained categories that will support more nuanced analysis with respect to users' potential engagement with events. Additionally, the event detection and classification process that is currently used by us can be further improved – both to classify events better, and to allot events across different categories (as against just a single category, as is the case currently). We also did not consider people's personality in the study. It is possible that certain personality (e.g., openness and extraversion) may affect people's event engagement. Finally, in this study, we did not consider the fact that there may exist different *kinds* of target users when engagement with events is under consideration. While we did partition a target user's following list coarsely (in terms of friends, news accounts, etc.), the target users themselves may also be distributed across various categories that exhibit some correlation (and hence predictive power) with respect to event engagement.

## 7 Conclusion

In this paper, we developed statistical models of people's Twitter engagement with real-world events. Categories of engagement predictors were conceptually developed, operationalized, and assessed for their relative impact on users' engagement presence, and the degree of that engagement. We explored the relative impact of multiple measures collected from four different user perspectives: prior Twitter activity, tweets' content, geolocation or geographic proximity, and social network structure. In particular, we found several key factors that predict the users' presence in engagement with real- world events, including total number of tweets, communication modes, friends' engagement in events, etc. We also examined the effects of these predictors in predicting the degree of engagement. We also examined the effects of these factors with respect to the different types of events predicated on their topics. We concluded that users' prior activities, as well as their social network structure, can be very good predictors for both the presence and the degree of their engagement with real-world events. Given a finer granularity of events (according to their topics), the content of tweets and the geographic proximity provide additional predictive power with respect to different event categories.

## References


Abel, F.; Gao, Q.; Houben, G.-J.; and Tao, K. 2011. Analyzing user modeling on twitter for personalized news recommendations. In *User Modeling, Adaption and Personalization*. Springer. 1–12.

Becker, H.; Naaman, M.; and Gravano, L. 2011. Beyond trending topics: Real-world event identification on twitter. *ICWSM* 11:438–441.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3:993–1022.

Burgess, M.; Mazzia, A.; Adar, E.; and Cafarella, M. 2013. Leveraging noisy lists for social feed ranking. In *Seventh International AAAI Conference on Weblogs and Social Media*.

Burke, M.; Kraut, R.; and Marlow, C. 2011. Social capital on facebook: Differentiating uses and users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 571–580. ACM.

Cartwright, D., and Harary, F. 1956. Structural balance: a generalization of heider's theory. *Psychological review* 63(5):277.

Cheng, Z.; Caverlee, J.; and Lee, K. 2010. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, 759–768. ACM.

de Zúñiga, H. G., and Valenzuela, S. 2011. The mediating path to a stronger citizenship: Online and offline networks, weak ties, and civic engagement. *Communication Research* 38(3):397–421.

Diakopoulos, N.; Naaman, M.; and Kivran-Swaine, F. 2010. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, 115–122. IEEE.

Feld, S. L. 1981. The focused organization of social ties. *American journal of sociology* 1015–1035.

Gilbert, E., and Karahalios, K. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 211–220. ACM.

Golder, S. A., and Yardi, S. 2010. Structural predictors of tie formation in twitter: Transitivity and mutuality. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, 88–95. IEEE.

Hong, L.; Dan, O.; and Davison, B. D. 2011. Predicting popular messages in twitter. In *Proceedings of the 20th international conference companion on World wide web*, 57–58. ACM.

Hu, Y.; John, A.; Seligmann, D. D.; and Wang, F. 2012. What were the tweets about? topical associations between public events and twitter feeds. In *ICWSM*.

Hu, Y.; Farnham, S. D.; and Monroy-Hernández, A. 2013. Whoo.ly: Facilitating information seeking for hyperlocal communities using social media. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*, 3481–3490. ACM.

Hutto, C.; Yardi, S.; and Gilbert, E. 2013. A longitudinal study of follow predictors on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 821–830. ACM.

Hyman, J. B. 2002. Exploring social capital and civic engagement to create a framework for community building. *Applied Developmental Science* 6(4):196–202.

Kivran-Swaine, F.; Govindan, P.; and Naaman, M. 2011. The impact of network structure on breaking ties in online social networks: unfollowing on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1101–1104. ACM.

Kulshrestha, J.; Kooti, F.; Nikravesh, A.; and Gummadi, P. K. 2012. Geographic dissection of the twitter network. In *ICWSM*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, 591–600. ACM.

Mahmud, J.; Nichols, J.; and Drews, C. 2014. Home location identification of twitter users. *TIST*.

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 415–444.

Naaman, M.; Zhang, A. X.; Brody, S.; and Lotan, G. 2012. On the study of diurnal urban routines on twitter. In *ICWSM*.

Naaman, M.; Boase, J.; and Lai, C.-H. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 189–192. ACM.

Read, J.; MacFarlane, S.; and Casey, C. 2002. Endurability, engagement and expectations: Measuring children's fun. In *Interaction Design and Children*, volume 2, 1–23. Shaker Publishing Eindhoven.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, 851–860. ACM.

Shah, D. V. 1998. Civic engagement, interpersonal trust, and television use: An individual-level assessment of social capital. *Political Psychology* 19(3):469–496.

Shamma, D. A.; Kennedy, L.; and Churchill, E. F. 2009. Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM workshop on Social media*, 3–10. ACM.

Starbird, K. 2013. Delivering patients to sacré coeur: collective intelligence in digital volunteer communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 801–810. ACM.

Suh, B.; Hong, L.; Pirolli, P.; and Chi, E. H. 2010. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Social computing (socialcom), 2010 ieee second international conference on*, 177–184. IEEE.

Toole, J. L.; Cha, M.; and González, M. C. 2012. Modeling the adoption of innovations in the presence of geographic and media influences. *PloS one* 7(1):e29528.

Vieweg, S.; Hughes, A. L.; Starbird, K.; and Palen, L. 2010. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1079–1088. ACM.

Wasserman, S. 1994. *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Wellman, B., and Wortley, S. 1990. Different strokes from different folks: Community ties and social support. *American journal of Sociology* 96(3):558.