

Hiding in Plain Sight: A Tale of Trust and Mistrust inside a Community of Citizen Reporters

Eni Mustafaraj

Computer Science
Wellesley College
emustafaraj@wellesley.edu

Panagiotis Metaxas

Computer Science
Wellesley College
pmetaxas@wellesley.edu

Samantha Finn

Computer Science
Wellesley College
sfinn@wellesley.edu

Andrés Monroy-Hernández

Microsoft Research
Cambridge, MA
andmon@microsoft.com

Abstract

The large volume of user-generated content on the Social Web puts a high burden on the participants to evaluate the accuracy and quality of such content. Researchers have been trying to address this problem by focusing on discovering influential entities inside social platforms to guide their ranking algorithms. Often these entities correspond to people or organizations already influential in the real-world. However, not everywhere in the world is there a right to and protection of free speech. In countries where the traditional media cannot report the truth, anonymity becomes a necessity for citizens who want to exercise their right of free-speech in the service of their community.

But is it possible for anonymous individuals to become influential and gain the trust of a community? In this paper, we discuss the case of a community of Twitter citizen reporters, located in a Mexican city plagued by the drug cartels fighting for control of territory. Our analysis shows that the most influential individuals inside the community were anonymous accounts. Neither the Mexican authorities, nor the drug cartels were happy about the real-time citizen reporting of crime or anti-crime operations in an open social network such as Twitter, and we discovered external pressures to this community and its influential players to change their reporting behavior.

Motivation

We live in an interconnected world. Extended draught or rain in one place of the world affects prices of food everywhere else (AAAS et al. 1989); increase of demand for smart mobile devices in the developed countries is linked to rape, kidnappings and other human rights abuses by militias in the Democratic Republic of Congo (Essick 2001) who control the minerals trade; and the crackdown of maritime transport of illegal drugs in the ports of Florida by US authorities has forced the powerful and ruthless drug cartels to look for land routes inside Mexico, resulting in thousands of dead, fear and uncertainty in the Mexican population, as the cartels fight for control (Greyson 2010).

Two important lessons can be drawn from these examples of interconnectedness: first, economic and political choices

made by citizens in developed countries can substantially affect the lives of people all around the world. Second, no one can afford to ignore what is happening elsewhere with the pretext that it is too far from home. In this interconnected world, nothing is far from home anymore.

We usually rely on our journalists and news organizations to keep us informed on the dynamic of events around the globe, but not every country has a free press or is willing or able to allow the international press to move freely. In some countries, like Mexico, journalists have been killed by organized crime or put under pressure by the authorities to stop reporting on certain events (Committee to Protect Journalists 2010).

The rise of the Social Web has created a new outlet for staying informed: citizen reporting. The different social media and networking platforms, like YouTube, Flickr, Twitter, and Facebook allow everyone in the world to report in real-time what is happening in the place they live. Social media platforms specializing in organizing humanitarian response to disasters, such as Ushahidi, rely on people on the ground to report on situations that need immediate attention (Heinzelman and Waters 2010). Anyone can be a reporter. However, this poses a new problem: *how do we assess the credibility of citizen reporting?* When we read news, we usually choose our information sources based on the reputation of the media organization: BBC, New York Times, Der Spiegel, etc. We trust the news organizations, therefore, we expect that their reporting is credible, though in the past there have been breaches of such trust, and all media organizations have an embedded bias that affects what they choose to report (Baron 2005).

Citizen reporting lacks the inherent structures that help us evaluate credibility as we do with traditional media reporting. But sometimes, citizen reporting might be the only source of information we might have. How can we use technology to help us verify the credibility of such reports?

Challenges and Opportunities

Our starting premise in tackling the problem of technological support for credibility verification is that events do not happen in a vacuum. They are located in a physical place and affect the lives of people living there. Furthermore, we believe that when people are concerned with the quality of their everyday lives and have access to a communication net-

work such as the Web, they will use it to keep alive a conversation about their community and its concerns. Thus, the search for credible sources of information is connected to the discovery of online communities that may be generating reliable information and at the same time be able to verify its credibility.

We do not intend to address every possible online community. We are interested in people trying to live normal lives, while battling adversities caused by a variety of adversaries: oppressive governments, corrupt officials, natural or man-made disasters, organized crime, sectarian wars, etc. The public needs to be informed of such struggles in order to apply coordinated pressure toward their resolution. However, this is easier said than done. Even though access to the Internet has increased dramatically, its use as an activism tool can be interrupted (e.g., the Egyptian government switched the Internet off during the days of demonstrations) or sabotaged (e.g., pro-government Russian activists spammed the Twitter hashtag reporting on the rigged parliamentary election (BBC 2011)). Internet activists can be arrested, like in China (Yang and MacKinnon 2009), or even killed, like in Mexico (Lisosky et al. 2011). In many countries, activists have to balance their instinctive fear of exposure to danger with the need for public attention to their cause in order to gain more support. The Social Web provides a platform to reconcile these two aspects. Making one's own voice heard in a networked public space such as Facebook, Twitter, or YouTube amplifies the message, while preserving various degrees of anonymity for the speaker.

Thus, our basic assumption is that in order to capture the attention of a broader audience, Internet activists have to be active in the social communication platforms that the majority of people use. Their motivation can be described as trying to maximize their message's reach, while minimizing the risk and costs. However, the same platforms that will presumably make it easy for some activists to spread their message, will make it easy for their antagonists to spread rumors, lies, and fear into the public. Anonymity works both ways.

Fortunately, not all messages are created equal in the public opinion. Assuming that serious concern of one's safety can become a message that resonates with the broader public, we should expect the creation of a community around such a message. Inside a community, even anonymous individuals can establish recognizable identities that they can sustain over time. Such *anonymous individuals can become trustworthy if their efforts to serve the interests of the community remain constant over time*.

The contributions of this paper can be described as follows:

1. To the best of our knowledge, this paper presents the first analysis of the practices of a community of Twitter citizen reporters in a life-threatening environment over an extended period of time (10 months).
2. We discover that in this community, anonymity and trustworthiness are coexisting. Because these citizens live in a city troubled by the narcowars that have plagued Mexico since 2006, it is a great example of a community where

anonymity of active participants is crucial, while lack of anonymity may be fatal.

3. We describe a series of network and content based features that allow us to understand the nature of this community, as well as discover conflicts or changes in behavior.

Prior Research

Research on the nature of virtual communities (Jones 1997) has identified as a prerequisite for their creation the need for a platform that can provide the following features: a) interactivity, b) more than two communicators, c) a common-public-place where members can meet and interact, d) and a sustained membership over time. As discussed in (Gruzd, Wellman, and Takhteyev 2011), Twitter not only provides all these features, but it also allows to quantify the nature of each of them. Unfortunately, as the authors point out, one cannot derive the existence of communities from the existence of a suitable platform. In fact, many of the Twitter features, such as lack of search beyond one week, lack of conversational threads, diffused communication, and the large number of participants and messages may work against the notion of a traditional online community. However, Twitter has a unique feature that facilitates on-the-fly creation of communities: the hyperlinked hashtags. While previous research has shown that the majority of Twitter hashtags have a very short half-life span (Romero, Meeder, and Kleinberg 2011), in this paper we analyze the practices of a community of citizens that have been using the same hashtag since March 2010 to report events of danger happening in their city.

Citizen reporting has been studied before as, e.g., in the case of "citizen scientists" providing critical data on species response to climate change (Doughton 2009). The idea of using citizens as sensors in local emergencies, outlining possibilities for capitalizing on more advanced implementations of 911 using social media was discussed in (Shneiderman and Preece 2007). In particular, using Twitter for disaster reporting was studied by (Hughes and Palen 2009; Starbird et al. 2010).

Last year there were numerous news reports on the role that Social Media played in the so-called "Arab Spring", including a recent working paper by (Howard et al. 2011) suggesting "that social media carried a cascade of messages about freedom and democracy across North Africa and the Middle East, and helped raise expectations for the success of political uprising".

Data Collections

The initial data for this study was collected by one of the co-authors, who on November 2010 provided a set of keywords related to Mexico events to the archival service (Archivist 2010). The collection was later divided in separate datasets according to the presence of certain hashtags. The dataset discussed in this paper consists of 258,734 tweets written by 29,671 unique Twitter users, covering 286 days in the time interval November 2010 - August 2011.

The need for data obfuscation

While we would prefer to give further details on the collected data and use them freely in this paper, on ethical grounds, we will protect this community under anonymity, due to potential risk that our research can pose now or in the future. To exemplify the seriousness of the situation, we provide one example out of the many documented in the press of what the lack of anonymity can lead to. On September 27, 2011, the Mexican authorities found the decapitated body of a woman in the town of Nuevo Laredo (near the Texas border) with a message apparently left by her executioners, which starts this way:

Nuevo Laredo en Vivo and social networking sites, I'm The Laredo Girl, and I'm here because of my reports, and yours, ...

Laredo Girl was the pseudonym used by the woman to participate in a local social network that enabled citizens to report criminal activities (Castaneda 2011).

In light of such crimes, our analysis of the citizen reporting activities on Twitter might have the undesired side effect of revealing information that leads in the future to identifying citizens participating in the community. Therefore, we have decided to obfuscate the data in order to not allow the identification of the city and its community. Throughout this paper we will refer to the community with the obfuscated hashtag #ABC_city, which is a substitute for the hashtag present in the tweets of our corpus. We will also substitute the exact text of important tweets with a translation from Spanish to English, so that searching online or with the Twitter API will not lead to unique results.

Limitations

The archival service used for the data collection retrieves data through the Twitter Search API. This has the downside that Twitter stops giving results if the daily limit of 1500 tweets is reached. In our dataset, we discovered that there are 64 days with 6 or more consecutive hours of missing data (outside the sleeping hours). At the time of this writing, we are working on getting permission to recover the missing data from companies that possess archives of “fire-hose” data. The strict terms of service imposed by Twitter to such companies, make it difficult for researchers not affiliated with these companies to be granted such access.

Another limitation of the Twitter Search API is the lack of information about the sender of the message beyond the screen name and user id. We used the Twitter REST API to retrieve the complete information of tweets based on the tweet IDs collected by the Search API. However, the user information embedded in the received results corresponds to the current moment in time (in terms of number of statuses, followers, followees, description, etc.) and not to the state of the account at the time the tweet was sent. Additionally, if users have in the meantime deleted their tweets or the account, no records of tweets will be retrieved. This turns out to be a real limitation because influential anonymous accounts frequently delete and restore their accounts, concealing the account's history.

Supplemental Data

To supplement our limited original dataset, we performed a series of additional data collection in September, 2011. In particular, we collected all social relations for the users in the current dataset, as well as their account information. We collected all tweets for accounts created since 2009 with less than 3200 tweets, in order to discover the history of the hashtag #ABC_city that defines the community we are studying. We also made use of the dataset described in (O'Connor et al. 2010) to locate tweets archived in 2009.

A Community of Citizen Reporters

On Twitter, everything is in flux. Most hashtags live only a few days (Romero, Meeder, and Kleinberg 2011) and trending topics change every 40 minutes (Asur et al. 2011). The attention of the Twitter audience is constantly shifting from one event to the other, from one meme to the next. There are examples of communities created around topical hashtags, for example, #occupy (the Occupy Wall Street movement) or #tcot (top conservatives on Twitter). The study of their network structure has revealed interesting patterns of information sharing and diffusion. The interesting question is always how such communities are created and how do they manage to capture the attention of other Twitter users.

Searching for the Origins of a Community

Communities in Twitter do not have permanent URLs, and searching for a hashtag will only show the latest week of tweets containing the hashtag. In order to learn about the origin of a hashtag and its community, one needs to look at the history of the users who use it and go back in time to find when it was used for the first time. The first dataset we used for such an exploration is the (O'Connor et al. 2010) dataset. It contains several million of tweets collected in the period May 25, 2009 - Jan 25, 2010. This dataset was collected by accessing the Twitter “gardenhose”, which outputs a set of worldwide tweets corresponding to approximately 10% of Twitter's daily volume collected by a random uniform process. Using the user IDs of accounts in our corpus, we searched the O'Connor corpus for tweets written by these users. We found 579,713 tweets by 6,542 different accounts, with a median of 8 tweets per user. Knowing that 10,489 accounts from the corpus were created before February 2010 (when O'Connor's data collection ended), we were able to collect tweets for 62.3% of existing community membership in 2009, which provides us with a good coverage. Then, we searched for all tweets containing hashtags (found 61,668 tweets) and calculated the frequency distribution of the contained hashtags. It turned out that the most frequent hashtags were global hashtags such as #followfriday, #iranelection, #fb (facebook), #fail, or the Spanish #ioconfieso, which indicates that local communities are firmly embedded in the meme-spreading culture of Twitter. We then compared this set of hashtags with hashtags extracted from #ABC_city corpus to discover hashtags that were being using in 2009 and are still being used in 2011.

One of the most interesting discoveries, which shows how much life has changed in Mexico in 2010, is that the

hashtag #balacera (shooting) appeared only 11 times in the O'Connor dataset of half a million random tweets (written in 2009) as compared to 26,740 times in the #ABC_city dataset, which contains a quarter of million tweets about life in one city (2010–2011). We were able to find an older hashtag related to the city, which we will call #old_ABC, ranked as the 11th most frequent tag in the O'Connor dataset. We manually inspected 725 tweets that contained this hashtag and found that only 3% of tweets were reporting dangerous events in the city. This provides evidence that in 2009 the ABC city was safer than in 2010 and beyond, or that Twitter had not yet become a medium for widespread reporting. Ultimately, none of the tweets in O'Connor dataset contained the hashtag #ABC_city, indicating that it hadn't been created yet.

Using the Twitter API, we created a second historical dataset with the following criteria: all users who had an account since 2009, had tweeted more than five times with #ABC_city during the collection period, and had fewer than 3200 statuses. There were 1098 accounts in our dataset that fulfilled these criteria. The Twitter API returned the complete timelines for 761 of these users, for a total of 922,984 tweets. We searched for tweets with #ABC_city, and this time we were lucky: the tweet mentioning for the first time #ABC_city was the inaugural one, on March 19, 2010:

#YXZ_city #ABC I propose #ABC_city to inform about news and important events in our city.

Then, this user reused the new hashtag many times in the following days together with #old_ABC hashtag and others, in order to spread its use:

@userA shootings are being reported in [address]
(good source) #ABC #old_ABC #ABC_city #XYZ_city

In May 11, 2010, the same user who created the hashtag tweeted the following:

@Spammer101 Stop spamming #ABC_city. It's only about important events that might affect our society.

Between May and November 2010 the usage of the hashtag is sparse, with the old hashtags being used more often. An increase in the adoption of #ABC_city starts on November 4th, only a week before the starting period of the #ABC_city dataset. Our assumption is that the hashtag was being kept alive by accounts with a frequent tweeting pattern, for which Twitter restrictions don't allow us to access data going back into the summer of 2010. Now that we know the origins of the community, let's return to the present days and look at its evolving state based on its network properties.

Followees, Followers, and Friends

There are three kinds of explicit network relations in Twitter, despite Twitter's insistence on recognizing only the first two kinds. 1) *Followers* – all the accounts that follow you to read your tweets; 2) *Followees* – the accounts you follow to read their tweets; 3) *Friends* – accounts that mutually follow each-other. Studying each of these relations reveals certain features of the community, which can be used to distinguish it from other Twitter communities.

Followee Relations Out of 29,671 unique users in the corpus, we were able to collect followee information for 24,973 accounts that were active and public in September 2011 (84% of all users in the corpus). There are more than 8,5 million followee links, with an average of 336 followees per user and a median of 162 followees. The total number of unique followees is almost 1,7 million.

Ranking the followees based on the number of relations inside this community serves as an indicator of the attention that this community as a whole pays to other Twitter users. We inspected the top 100 accounts to understand the nature of their popularity. The top account was Mexico's president, Felipe Calderon, followed by the TV news program of the city, and an anonymous citizen reporter to whom we will refer as @GodFather. Four journalists, the city's newspaper, a famous Mexican poet, and a comic's character made up the rest of top ten. Almost half of the accounts in the top 100 are entertainers of Mexican fame, with only a few international superstars such as Shakira or Lady Gaga in the mix. This statistic confirms the widespread perception that a large part of the Twitter appeal derives from its use by celebrities, though it also indicates that each community is interested in its own celebrities. 25 of top 100 most followed accounts belong to local and national journalists and media organizations, compared to 10 for politicians at the state and federal level. In fact, the governor of the state in which ABC city is located (Mexico is a federation of 31 states) ranks at the 45th position in the followees list, one place behind the account of Barack Obama.

This ranking raises some interesting questions. Are Mexican governmental institutions actively using Twitter? Do citizens know about their accounts? If they know about them, but do not choose to follow them, what does that imply about the quality of communication coming from such official sources. Answering these questions requires user studies that are not the focus of this paper. From this paper's perspective, the most important observation is that ten anonymous¹ accounts of citizen reporters made the top 100, with five such accounts ranking ahead of the federal government and governor's Twitter feeds.

Inside-community Followee Relations Ranking all followees is a way to find popular accounts. However, most of these accounts will probably never tweet using the community hashtag. Thus, a better measure of popularity will be a ranking of accounts that are inside the #ABC_city community. That is, instead of ranking 1.7 million accounts based on their frequency of appearance, we rank only the 24,973 accounts of the corpus based on the frequency of appearance in the followee relations of every community member. We find that there are 951,873 followee links to members of the community, which amount to 11% of the overall followee links. In average, every member follows 41 other members of the community (the median is 12), indicating that they have a considerable interest in the community. Only 6% of all members do not follow any members. Further analysis is needed to understand why these "friendless" accounts write

¹Our manual inspection of the accounts couldn't establish a relation to the real-life identities of such accounts.

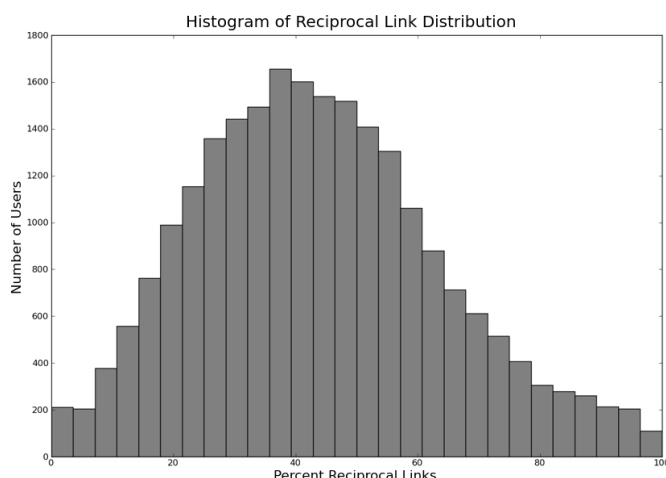


Figure 1: The normal-like distribution of friendship relations (mutual links) in the network of the #ABC_city corpus.

tweets with the hashtag #ABC_city.

Again, to understand the appeal to the community of the top 100 ranked accounts, we inspected their Twitter profiles. The top account, @GodFather, has 9,079 followers inside the community, or 36% of all active members. This amounts to 16% of all his audience, he has in total 57,127 followers. @GodFather is an anonymous citizen who has written the largest number of tweets in the corpus (6,675), which make up 25% of all his statuses (26,340). The rest of the top 10 consists of four local news organizations, four other anonymous accounts and one well-known industrialist. In comparison with the make up of the first group of top 100 followees, in this group the number of highly ranked anonymous accounts rises from 10 to 24, while the category of “celebrities” declines from 50 to 22. Bloggers of known identity (mainly reporters of news organizations) make up the third largest group, 22. 10 accounts from this group of 100 were verified by Twitter for their authenticity: among others the state government account, several news organizations and reporters.

Friend Relations A mutual relation in Twitter (the friendship) is significant because it enables the involved accounts to send direct messages to one another. Direct messages offer some privacy to users, though if an account is hacked messages are compromised (unless a user has the habit of deleting them). Communication through direct messages is not visible to researchers or the public and cannot be quantified. However, it is possible to quantify the extent to which such stronger ties exist inside the community by discovering mutual links in the sets of followers and followees. As shown in Figure 1, on average, 40% of user relations are reciprocated.

Figure 2 shows the graph of all members with more than 75 friendship links (we limited the number of nodes for computational reasons) which only reinforces the conclusion that this is a tightly connected community of users. More about this in the next section.

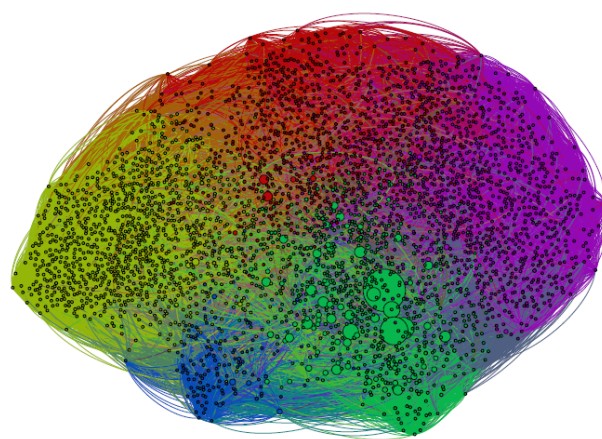


Figure 2: Visualization of the friendship links for nodes with degree 75 or higher. Coloring is produced by the Gephi modularity algorithm that finds communities in a network using the Louvain algorithm (Blondel et al. 2008).

While this examination of the network structure of the community allowed us to draw several insights into the preferences of the community, it tells us little about how the community evolved over time and what keeps them together. For that, we will turn our attention to a series of features related to the volume and content of messages in the #ABC_city corpus.

Community Practices

Each platform of communication facilitates some practices over others. In Twitter, two of the most popular practices are sharing information and retweeting. Other frequent practices are conversations (by starting a tweet with @username), favoring tweets, or creating new memes.

Sharing Information Statistics show that 25% of the total number of daily tweets contain a URL (Rao 2010). This statistic varies based on the interests of Twitter sub-communities. For example, a study of 250,000 tweets related to an important political election, has reported that 43% of tweets contained an URL (Metaxas and Mustafaraj 2010). By contrast, the analysis of the #ABC_city corpus indicated that only 12% of the tweets contain URLs, a low proportion. However, this contrast emphasizes the nature of information being shared by this community: real-time information of dangerous areas to avoid. Understandably, there will be no webpage that will contain such kind of information in real-time. Still, 12% is not a negligible number, thus, we extracted all URLs from tweets, expanded them to full URLs, and created a ranking based on the frequency of the top domain of every tweet. The top three domains are media storage sites such as twitpic.com (4,365 times), yfrog.com (4,232 times), and youtube.com (1,866 times). The next three domains belong to the three local news organizations, which together appear 3,228 times. Twitpic and YFrog are social websites for sharing pictures in Twitter and their dominance in the information sharing prac-

tice suggests that the community prefers citizen reporting to the traditional news reporting. Such a finding is relevant, given that studies have shown that in the Twitter-sphere as a whole, mainstream news organization links are the most shared.

Retweeting The most active practice in which the users of the #ABC_city community are engaged is retweeting. In fact, 50% of all tweets are retweets. Retweeting is especially popular with occasional members. For 75% of users that have only one tweet in the corpus, that one tweet is a retweet. A frequent pattern that we noticed in the retweets is shown in the following example²:

RT @userA @GodFather #ABC_city It is quiet in the transit zone of [address] 6:20pm

Thus, Twitter users were reporting through replies to one another and these replies were being retweeted, and sometimes commented, when they contained questions, instead of reports. The influential members of the community were widely retweeted while they themselves became the centers of information provided by other, anonymous and eponymous citizens. In turn, they retweeted the information they received, while in many occasions they first asked for verification about the credibility of such information, but turning to their audiences. By doing so, they were showing a concern for credibility which, we believe, increased the trust that their followers felt about them. Asking for verification of crucial information has been used in other high risk situations. (Meier 2011) discusses the risks of unverified information faced by Ushahidi. Andy Carvin, the NPR journalist who has been following the “Arab Spring” events, has also been using verification as a way to solving the credibility problem by crowd sourcing (Farhi 2011).

Past research has shown that retweeting is indicative of agreement between the original sender and the retweeter (e.g., (Metaxas and Mustafaraj 2010; Conover et al. 2011)). Over time, retweets are effectively providing information about a community of social media users that are in agreement on specific issues. Otherwise, the chance of a community member retweeting a message of an opposing political community is under 5%.

Since retweets involve a relation between two users: the original sender and the retweeting user, we can create a network of such relations for all retweets in the corpus. This retweet graph is shown in Figure 3. The distances among nodes is dictated by a force directed algorithm known as the “ForceAtlas2” method in (Gephi 2010), which is an improved version of the well-known (Fruchterman and Reingold 1991) graph drawing algorithm. The coloring was created by applying the Louvain algorithm (Blondel et al. 2008) that seeks subgraphs with high modularity, that is, subgraphs (referring to communities) that have many more connections among the members of the community than to outside members of the community.

Zooming in inside this graph, as shown in Figure 4, reveals the most influential nodes in the community, which we

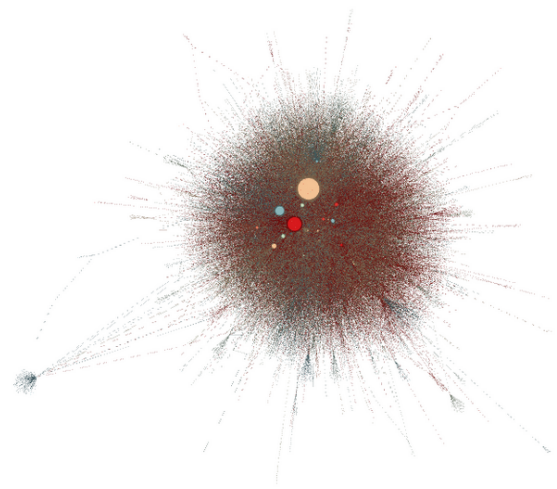


Figure 3: The retweet graph reveals a large component that is actively involved in retweeting, with smaller star-like components at the fringes. Closer examination reveals that the stars at the fringes were occasional retweeters of famous users (e.g., entertainers) and could easily be identified and excluded from our analysis. The nodes have been drawn in size relative to their in-degree, that is to the degree that their messages had been retweeted, revealing a small number of accounts that rose to prominence in the community.

identified as the anonymous citizen reporters. The biggest node belongs to @GodFather.

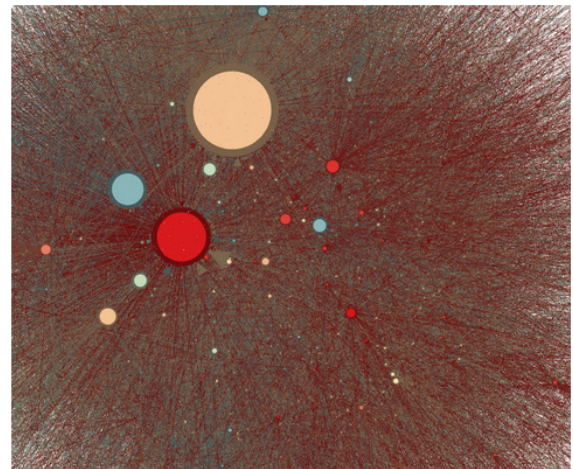


Figure 4: A closer look at the core of the community reveals 13 nodes that have a larger share of their messages retweeted. The spatial proximity of these nodes enforced by the (Fruchterman and Reingold 1991) algorithm indicates that they were also retweeting each other (as opposed to the nodes in the periphery of the retweet graph).

Frequency of Communication As it is well established, many activities in social media platforms conform to the

²This is an example of a tweet which is a sent directly to @GodFather by @userA, and then was retweeted by @GodFather

long-tail distribution: a small number of users produces the larger amount of the content, while the majority of the users contributes rarely. This community is no different, 49% of the users have sent only one tweet (5.6% of the overall volume), while a small group of 786 users (2.52%) have generated 55.4% of the corpus volume by sending 100 tweets or more. Independent of the fact that a user is prolific or mostly silent, it appears that the community becomes more active whenever there is news of shootings taking place in the city. Figure 5 reveals this pattern. The activity of three groups of users (users who have written less than 5 tweets, between 5 and 100 tweets, and more than 100 tweets) is shown together with the frequency of the appearance of the word shooting (balacera in Spanish) overlaid.

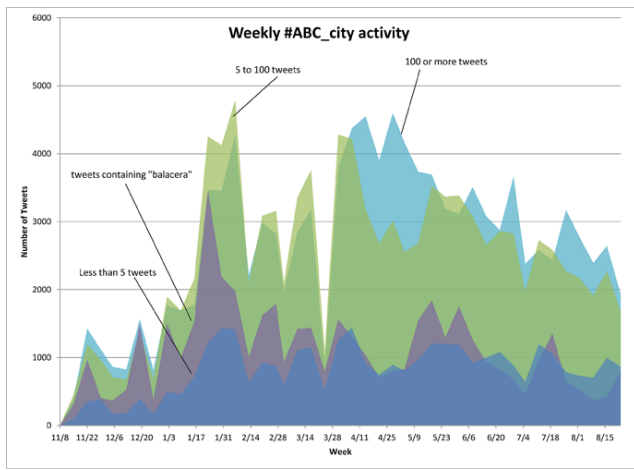


Figure 5: Tweeting activity of three groups of users with different tweeting patterns overlaid with the frequency of appearance for the word “balacera” (shooting). All three groups have an increase in activity, matching the ups of the balacera distribution. There is only one discrepancy, in April–May 2011, related to an event explained in the next section.

While for most of the time the three groups show a conforming pattern of activity (in terms of lows and peaks), there is a period of anomaly in the time interval April–May 2011, where there is more activity from the group of the prolific users. This corresponds to an important event for the community itself and we will discuss that event in the following section.

Who to Trust?

In April 2011, a blog that reported news about organized crime in Mexico, published a story about @GodFather, accusing him of working for the drug cartels. The charges were that he had created a network of citizen reporters that were reporting on the movement of army and police or alternatively on the safety and tranquility of other areas. This later information was allegedly being used by criminals to avoid law enforcement, while at the same time putting in risk the citizens in the area.

We found evidence in the data of new Twitter accounts created in early April with the sole purpose of criticizing @GodFather and other users that were supplying information to him through the #ABC_city hashtag, see example:

#ABC_city Hello. We are back to denounce @GodFather for working for Cartel X so that they can claim ABC.

@GodFather didn’t publicly address the charges, however, as Figure 6 indicates, he reduced his involvement with the community, until stopping it completely. After a silence period that lasted several months, @GodFather returned, closed his account, and opened a new one (under another name). He tweets occasionally with the #ABC_city hashtag, but currently has less than 1/10 of his previous following. He gets frequent inquiries about his previous account and he explains that the account was hacked and he closed it. Because the alleged hacking happened outside the time period of our data collection, we don’t have a way to verify it without access to historical archives of tweets.

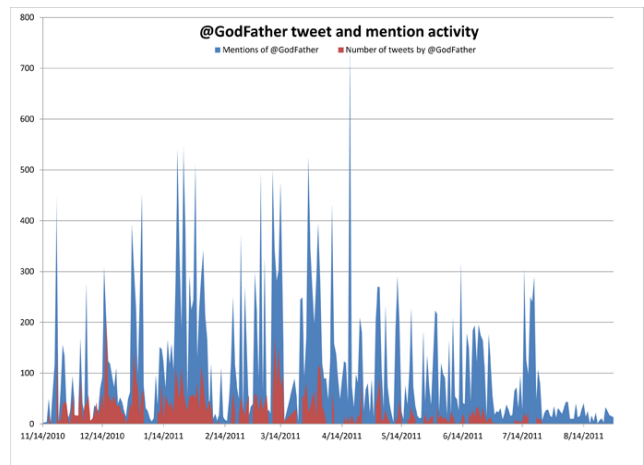


Figure 6: Daily distribution of tweets for the anonymous account @GodFather and its daily mentions in tweets by other members of the community. In April 2010, he was accused by newly created anonymous accounts of working for the criminal organization. After that event, he decreased his involvement in the community and at the end of July stopped tweeting altogether.

Discussion

Though research with Twitter data has captivated the interest of many researchers, especially those who attend ICWSM, there are almost no long term studies about communities of practice in Twitter. The lack of such research makes it difficult to compare the different statistics and features we discussed in this paper with those of other known communities. While certain graph-theoretic metrics might be applied to such communities, especially when trying to discern many of them in a large network such as Twitter, they lack a semantic that can help us understand the dynamics of relations and practices inside the community. We hope that

other researchers will be able to apply some of the features described here to communities of their interest, so that in the future comparative studies about the nature of Twitter communities can be performed.

In a time when social networking platforms such as Facebook and Google+ are pushing to force users to assume their real-life identities in the Web, we think that it is important to provide examples of communities of citizens for which maintaining their anonymity inside such networks is essential. But being anonymous makes one more susceptible to denigration attacks from other anonymous accounts, leaving the other members of community in the dilemma of who to trust.

This paper is only a first in a series of papers that will deal with the nature of communities of citizen reporters, especially with the issue of trust and source credibility. The analysis described here raises more questions than is able to answer, since established means to tackle the problem are yet to be defined. Additionally, more and better data are needed to answer some of such questions, especially in terms of being able to track events over time while they are happening, as opposed to going back after the fact.

Acknowledgments

The work of E. Mustafaraj and P. Metaxas was partially supported by the NSF grant CNS-1117693. We are grateful to Yesenia Trujillo for manually verifying accounts of Mexican users and translating tweets.

References

- AAAS; INSA; IRRI; and ICAR. 1989. Famine and national and international food prices. In *Climate and food security*. Archivist, T. 2010. archivist.visitmix.com.
- Asur, S.; Huberman, B.; Szabo, G.; and Wang, C. 2011. Trends in social media : Persistence and decay. In *ICWSM-11*.
- Baron, D. 2005. Persistent media bias. *Journal of Public Economics* 90.
- BBC. 2011. Russian twitter political protests 'swamped by spam'. <http://www.bbc.co.uk/news/technology-16108876>.
- Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 10. <http://arxiv.org/abs/0803.0476>.
- Castaneda, M. M. 2011. Woman's decapitation linked to web posts about mexican drug cartel. <http://bit.ly/yTobHp>.
- Committee to Protect Journalists, C. 2010. Attacks on the press 2010: Mexico.
- Conover, M.; Ratkiewicz, J.; Francisco, M.; Goncalves, B.; Flammini, A.; and Menczer, F. 2011. Political polarization on twitter. In *In Proc. of ICWSM '11*. AAAI Press.
- Doughton, S. 2009. Climate researchers seek citizen scientists. <http://bit.ly/wQ8sw3>.
- Essick, K. 2001. Guns, money and cell phones. <http://bit.ly/wQ39Pm>.
- Farhi, P. 2011. Nprs andy carvin, tweeting the middle east. <http://wapo.st/y4piQY>.
- Fruchterman, T. M. J., and Reingold, E. M. 1991. Graph drawing by force-directed placement. *Software: Practice and Experience* 21:1129–1164.
- Gephi. 2010. Forceatlas2, the new version of our home-brew layout. <http://bit.ly/yGXRhJ>.
- Greyson, G. W. 2010. Calderon's anti-drug strategy. In *Mexico: narco-violence and a failed state?* Transaction Publishers.
- Gruzd, A.; Wellman, B.; and Takhteyev, Y. 2011. Imagining twitter as an imagined community. *American Behavioral Scientist* 55(10).
- Heinzelman, J., and Waters, C. 2010. Crowdsourcing crisis information in disaster-affected haiti.
- Howard, P. N.; Duffy, A.; Freelon, D.; Hussain, M.; Mari, W.; and Mazaid, M. 2011. Opening closed regimes. what was the role of social media during the arab spring? <http://bit.ly/xb1BPK>.
- Hughes, A., and Palen, L. 2009. Twitter adoption and use in mass convergence and emergency events. In *ISCRAM-09*.
- Jones, Q. 1997. Virtual-communities, virtual settlements and cyber-archaeology: A theoretical outline. *Journal of Computer Mediated Communication* 3(3).
- Lisosky, J. M.; Henrichsen, J. H.; Henrichsen, J. R.; and Cramer, C. 2011. Don't shoot the messenger: Journalists who risk everything to tell stories of conflict. In *War on Words: Who Should Protect Journalists?*
- Meier, P. P. 2011. Verifying crowdsourced social media reports for live crisis mapping: An introduction to information forensics. <http://bit.ly/kstN2t>.
- Metaxas, P. T., and Mustafaraj, E. 2010. From obscurity to prominence in minutes: Political speech and real-time search. In *WebSci10: Extending the Frontiers of Society On-Line*.
- O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; and Smith, N. A. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proc. of 4th ICWSM*, 122–129. AAAI Press.
- Rao, L. 2010. Twitter seeing 90 million tweets per day, 25 percent contain links. <http://tcn.ch/y8QnPY>.
- Romero, D.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of WWW Conference*.
- Shneiderman, B., and Preece, J. 2007. 911.gov. *Science* 315.
- Starbird, K.; Palen, L.; Hughes, A.; and Vieweg, S. 2010. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *CSCW-10*.
- Yang, G., and MacKinnon, R. 2009. The power of the internet in china: Citizen activism online. In *Far Eastern Economic Review*.