The Rise and the Fall of a Citizen Reporter

Panagiotis Metaxas
Wellesley College
Computer Science Department
pmetaxas@wellesley.edu

Eni Mustafaraj
Wellesley College
Computer Science Department
emustafa@wellesley.edu

ABSTRACT
Recently, research interest has been growing in the development of online communities sharing news and information curated by "citizen reporters". Using "Big Data" techniques researchers try to discover influence groups and major events in the lives of such communities. However, the big picture may sometimes miss important stories that are essential to the development and evolution of online communities. In particular, how does one identify and verify events when the important actors are operating anonymously and without sufficient news coverage, as in drug war-torn Mexico? In this paper, we present some techniques that allow us to make sense of the data collected, identify important dates of significant events therein, and direct our limited resources to discover hidden stories that, in our case, affect the lives and safety of prominent citizen reporters. In particular, we describe how focused analysis enabled us to discover an important story in the life of this community involving the reputation of an anonymous leader, and how trust was built in order to verify the validity of that story.

Author Keywords
webscience; social computing; social media; citizen reporters; civic media; crisis informatics; crowdsourcing; news; drug war; microblogging; narcotweets; Twitter; Mexico.

ACM Classification Keywords
K.4.3 [Computers and Society]: Organizational Impacts– Computer-supported cooperative work. H.5.3 Groups & Organization Interfaces—collaborative computing, computer-supported cooperative work; K.4.2 Social Issues

INTRODUCTION

The rise of the Social Web has created a new information source: citizen reporters. Social media and networking platforms, including Twitter and Facebook, allow everyone in the world to report what is happening in their neighborhood or the city they live in, in real-time. 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 [14].

Anyone can be a reporter, but 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. Even 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 [3], we generally trust the news organizations and expect that their reporting is credible.

But what about the credibility of citizen reporting? Unlike established news organizations, it lacks the inherent structures that help us evaluate credibility. However, 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, especially when our safety may depend on the information we receive from citizen reporters?

In the last few years, Mexico has seen a significant amount of violence related to the struggle for control by powerful and ruthless drug cartels [16]. The cities in the paths between drug-producing Central and South America and drug-consuming USA have seen an increase in cartel-orchestrated violence and gruesome scenes aiming to intimidate rival groups. Reports bring the number of drug-related deaths to over 60,000 people [23], [25].

The widely held opinion that different branches of the police force were infiltrated and corrupted by cartels forced the government of former President Calderon to bring Army and Marine forces into cities and towns [26]. While the armed forces are trusted by the citizens and have had considerable success in their efforts, the cartels have increased their terrorizing activities against random bystanders, tourists and even whole villages. In particular, they have attacked journalists who have tried to report on the situation. As a result, in many cities there is a self-imposed silence of the press due to the terror [10].

Citizens, however, have found ways to inform themselves of the dangers they encounter daily by using social media. Twitter, in particular, has emerged as the tool of choice for spreading information that can help citizens plan their day. Twitter users report, confirm, comment on, and disseminate information and alerts about the violence, typically as it

WebSci'13, May 2–4, 2013, Paris, France.

unfolds. Many Mexicans reportedly consult Twitter every morning the same way that they consult the weather information in order to plan their day.

In [21] we describe our initial efforts to collect and analyze data produced by the community around the #MTYfollow tag in Monterrey, Mexico (MTY is its airport code). Complementary work is presented in [20] with an overview analysis of a large data set in several Mexican cities including Monterrey. Both of these papers maintain the anonymity of citizen reporters by introducing pseudonyms for the major actors in the data to avoid any risk that could result by revealing their true identities.

In general, it is not trivial to validate claims made by anonymous actors. For reasons that will become apparent, however, we are dropping some of the anonymity in this paper when we were able to verify, to our satisfaction, both the validity of several claims and the identity of the important actors.

The so-called "Big Data" made possible by ease of collecting digital data, including a massive collection of Social Web data, holds a lot of promise for studying the patterns through which Society is operating. However, it also raises concerns about the verification of these observed patterns [7]. In this paper, we demonstrate how superficial analysis at the "Big Data" level can miss important stories hiding under the numbers.

Anonymous Citizen Reporters

This paper is about the emergence of a network of anonymous users in Monterrey, a large city in Northern Mexico, using (mostly) pseudonymous accounts to inform and be informed about events in their city that could expose them to danger. While they also share reports about traffic accidents, the most serious danger they encounter comes from the fight between ruthless drug cartels that compete for control over land routes inside Mexico. Described sometimes as "war", this conflict for control between the cartels and with the Mexican authorities (local, state, and federal police, the Army, the Navy, etc.) has resulted in thousands of casualties as well as widespread fear and uncertainty in the Mexican population [12], [16].

We usually rely on our journalists and news organizations to keep us informed on the dynamics 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 [10].

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. Anyone can be a reporter. However, this poses a new problem: how do we assess the credibility of citizen reporting? [17] 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 these institutions and we expect that their reporting is credible. Citizen reporting lacks the inherent structures that help us evaluate credibility as we do with traditional media reporting, but sometimes citizen reporters might be the only source of information we have. How can we use technology to help us verify the credibility of such reports?

In this paper, we present some techniques that allow us to make sense of the data collected, identify important dates of significant events in them, and direct our limited resources to discover hidden stories that affect the lives and safety of prominent citizen reporters. In particular, we describe how focused analysis enabled us to discover an important story in the life of this community involving the reputation of an anonymous leader in the community. We start by describing the data collection process and the results of the initial automatic analysis, along with its limitations. Next we describe the methods we employed to identify and verify the events we discovered along with interviews of some of the important citizen reporters in this community. In the last section, we outline the characteristics of a semi-automatic system that can enhance the ability of its users to evaluate credibility in real-time sources.

DATA COLLECTION

The initial data for this study, provided by [21], was based on a set of keywords related to Mexico events submitted to the archival service Archivist [1]. The initial dataset analyzed in this paper was based on the selection of the hashtag #MTYfollow (tag capitalization is not important) and consists of 258,734 tweets written by 29,671 unique Twitter accounts, covering 286 days in the time interval November 2010 - August 2011 (see Figure 1).

Due to the opportunistic character of the initial data collection, we knew that the hashtag defining this community was created earlier than the dates of the collected data. We supplemented the initial dataset in several occasions, based on the discoveries we made during the analysis.

First, we performed a series of additional data collection in September 2011 to identify the origin of the hashtag and the follow relationships of the members of the community. In particular, we collected all social relations for the users in the current dataset that were still active in Sept. 2011, 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 #MTYfollow that defines the community we are studying. We also made use of the dataset described in [24] to locate tweets archived in 2009.

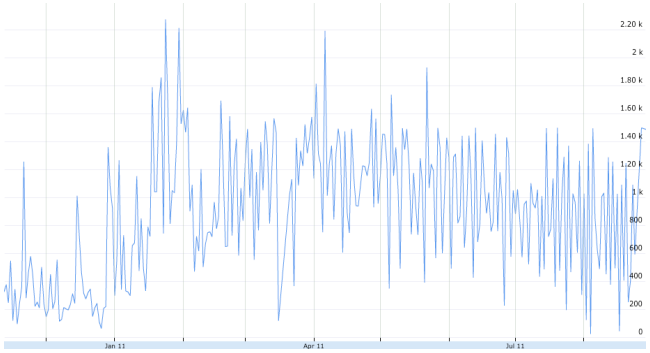

Figure 1. Histogram of the number of tweets collected automatically with hashtag #MTYfollow between Nov. 2010 and Aug. 2011.

For reasons that will be explained in a later section, in June 2012 we performed a third sequence of targeted data collections, seeking accounts that were created in late March, 2011 and had names containing the strings "aguila" and "halcon" (Spanish for "eagle" and "hawk") as well as tweets containing hashtag #aguilasMTY. We collected 654 such accounts and to a further qualified subset of 359 accounts we collected all their social information and their tweets.

INITIAL DATA ANALYSIS

In the initial analysis, we tried to get an overall picture of the data [21]. As is often the case with Social Web data, we noticed that some accounts were tweeting much more frequently than others. We also noticed that the rate of tweets increased when violence was reported in the city (as indicated by the use of keywords for "shooting", "explosion", etc). In some, but not all, instances we were able to match dates with high rates of tweets to published dates of violence in Monterrey, such as the accidental death of two TEC students who were caught in the crossfire between the authorities and the drug cartels on March 22, 2011 [27].

Next, we collected the followers of the accounts that displayed high activity in our data set and we tried to see how they were connected through mutual follow relationship. This analysis resulted in the Gephi graph [2] shown in Figure 2, which is remarkable, and not only for its (accidental) visual resemblance to a human brain. We selected for visualization those accounts that had at least 75 mutual followers in the data set, a semi-arbitrary number indicating higher involvement with the group and interest in each other's postings. The cutoff of 75 was selected for computational reasons and because it resulted in a single strongly connected component of the graph.

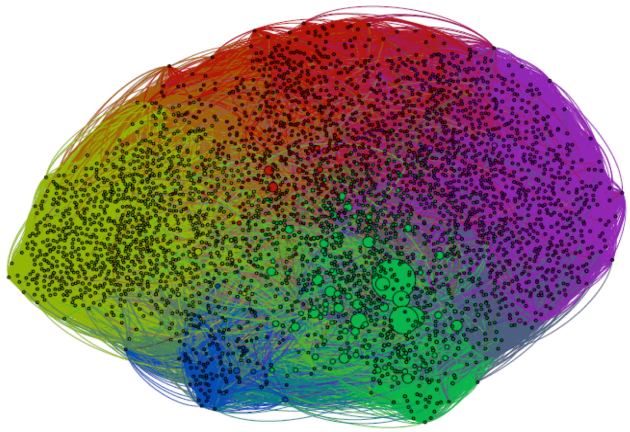

Figure 2. Visualization (in Gephi) of accounts with at least 75 mutual followers of accounts using #MTYfollow in our initial data collection. Three parameters are shown: node size (a circle) indicates size of followers group; node proximity (Euclidian distance between circles) indicates cardinality of common followers; and color (for nodes and edges) indicates subgroups with relatively more followers than others. Of particular interest in this paper are the green and blue groups.

Using Gephi's functions, we drew the graph as follows: Using the Force Atlas 2 force-directed algorithm [11], we visualized node proximity, and gave it its overall shape. Using the Louvain community detection module [5], we detected subgroups that share more connections within themselves than with other groups, and gave it its colors. Five large subgroups emerged. Of particular interest in this paper are the groups colored green and blue, in the lower part of the graph. Finally, using Gephi's logarithmic scale node size drawing method, we drew larger the nodes with more followers. This reveals that the nodes with most followers belong mainly to the green group with few ones in the blue and red group.

There are a few accounts with many followers in this collection; the largest of all with 57,127 followers is @trackmty (which, in [21] is mentioned with the pseudonym @GodFather). Within the community defined by #MTYfollow that we study and visualize above, this account has 9,079 followers, or 36% of all active members at the time of the data collection.

Then, for @trackmty and other prominent accounts in our data set (that is, accounts with many followers) we decided to look at the histograms of their activity. In particular, looking at the volume of tweets that were sent by @trackmty and that mentioned it (see Figure 3) we see a surprising pattern unlike the pattern of any other prominent account: The volume of its tweets seems to be falling to about a third of its average volume around April 17, 2011, and then drops to zero at the end of July, 2011. Moreover, the volume of tweets containing the string "@trackmty",

that is, tweets that either mention the account or retweeted its contents verbatim, similarly drops by about half around the same time period, though these tweets persist after @trackmty has stopped tweeting. And the two time periods are separated with the highest spike in tweet activity on April 17. This is rather remarkable, given the fact that the histogram of all tweets in the community does not show any such pattern in the tweet rate over time (see Figure 1).

At this point we realized that we needed to examine the events prior to April 17 more closely to make sense of the activity changes for @trackmty.

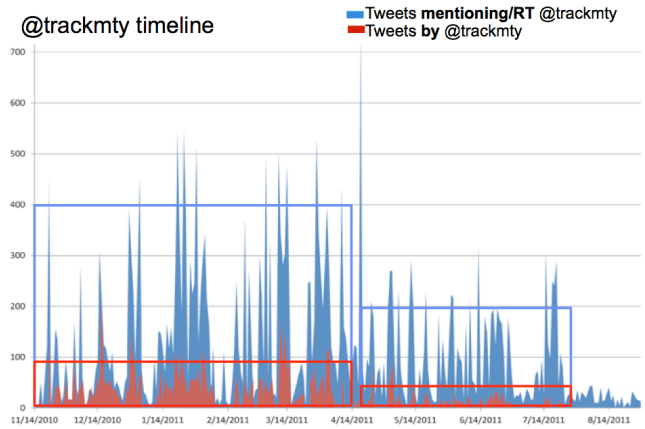

Figure 3. Timeline of tweets related to @trackmty. We observe that this user is tweeting with higher average volume before April 17, 2011 (red rectangles). The community's tweets mentioning it or tweeting its messages (blue rectangles) follows a similar pattern. Notice the spike of activity in mid-April, 2011.

THE LIMITS OF "BIG DATA" ANALYSIS

Up until this point, we had examined the collected data using basic statistical and visualization methods for analyzing large volumes of data without having read any of the contents of the tweets. To employ more sophisticated data mining methods did not seem a promising path for analysis. The answer to the puzzle of activity pattern change we were seeking was *in* the content of the tweets. While our ability to understand the developments without reading specific tweets was diminishing, the questions persisted: What happened shortly before April 17 that resulted in the spike of mentions and reduced activity of @trackmty? What caused the other spikes in January, February, and March 2011? Were the highest peaks the most important in the community's development? And why did @trackmty stop tweeting in late July?

These unexpected patterns in activity provide guidance on where to look for explanations, but for non-Spanish speakers, they cannot help further. However, the initial analysis was already very useful in guiding us to select what to examine. We did not need to read hundred of thousands of tweets serially or at random, since we knew where to look. We decided to engage Spanish-speaking students (see Acknowledgement section) to help us translate a few dozen random tweets posted at around the times of the spikes. We also used the automatic Google translate feature of Chrome [13] to translate a few blog postings.

The actors

Before providing more information on what we discovered through our research, and to make it easier for the reader to follow the description, we will use this subsection to introduce the different actors of the #MTYfollow community.

The News Organizations and Authorities

Reporters in Mexico have often been violently attacked for reporting on the war among drug cartels for control of trafficking passages and cartel conflict with the Army, Marines, and local police. While there are several news organizations reporting on the situation, their reporting is not without gaps. Investigative reporting has diminished due to brutal executions of journalists.

Distrust of the police further complicates the situation on the ground in Monterrey. It is common belief there that local, state and federal police have been infiltrated by the cartels, especially the powerful Golf cartel and the ruthless Zetas. Most people trust the Army and the Marines, but even those institutions are immune to mistrust.

The citizen reporters of #mtyfollow

The majority of the Twitter accounts represented in Figure 2 are citizen reporters. These reporters mainly use pseudonyms out of fear of exposure. In a few occasions, bodies have been found with notes stating that they were executed because they were "talking on the internet" [8]. The term "curators" coined by [20] indicates the more prominent citizen reporters on Twitter. [20] has more information about them, together with some interviews. Curators reportedly spend long hours in altruistic efforts to inform the public of potential dangers in the city, yet they do not often collaborate, as they see other curators as competitors who may "steal their tweets".

The citizen reporter @trackmty

An important actor in the events we are describing is the citizen reporter with account name @trackmty, who has been tweeting since early 2010 with the pseudonym "Melissa Lotzer". She claims to be the originator of the #mtyfollow community, something disputed by some members of the community. She is definitely the most active member of this community, and her 3 twitter accounts (@MelissaLotzer, @monitoreoMNR being the other two) have often come under attack by trolls. She is also the creator of a series of social media accounts named New Revolution of Mexico ("Mexico Nueva Revolucion", or MNR [22]), present on social media platforms that include Facebook, Wordpress, and Twitter. In early 2010,

she organized the MNR twitter group, and in late March 2011 the eagles ("aguilas") group.

The Blogs

The vacuum created by the lack of reporting by official news organizations has been filled by several high-visibility blogs. One of the major ones is the Blog del Narcos [6]. Its twitter account @infonarco has over a hundred thousand followers on Twitter, but it does not use the hashtags that would situate it as a large member of the #mtyfollow community. It has been accused of hacking other blogs and has itself been hacked. Recently, this blog was bought by a minor news organization and continues to operate.

There are other relevant blogs belonging to citizens reporting on the drug war and even to trolls, but none as influential as Blog del Narco.

The trolls

These are accounts that appear at several stages in the timeline we examine, targeting the accounts controlled by Melissa Lotzer of @trackmty. They use names intended to confuse the community by spelling similarity, such as @track**rn**ty (with r and n instead of m), and @Meli**s**aLotzer (with one s instead of two). They are promoted consistently by a few accounts in the blue group, suggesting that some of these troll accounts may be aliases.

In particular, a main troll we will call @troll (no need to promote his exact name, which is well known to the more active members of the community) repeatedly attacked and eventually exposed the identity of @trackmty. @troll also has a blog that appears to be connected to one of the authors of Blog del Narco. The blog projects a macho image of its owner, who claims to be a clerk for the local police (not a policeman) and brags that he likes to post pictures of prostitutes, gays, and lesbians.

FOCUSED STUDY OF POSTINGS

We started by translating tweets posted just before specific dates indicated by the six red arrows in Figure 4. These dates were selected by the suggestions of the statistical methods described previously. The order of translation roughly followed the volume of the spikes.

Initially, a few dozen tweets associated with the five highest peaks (i.e., not the last one when @trackmty stops tweeting) were translated. Through this process we developed a theory of what happened to the community and to @trackmty at that time. What follows is a description of some of the major events we discovered by translation. For clarity, they are presented in historical order, not in order of discovery. We should make clear that we only present events that we have verified to our satisfaction, coming from multiple independent resources including interviews, data examination, news reporting, and web archival search.

Events Detected

By the end of 2010 the #MTYfollow community has grown to a relatively stable size. @trackmty has become the largest account in terms of followers among all curators and they all tweet daily about important events, primarily dangerous (not only drug-related) situations, in the metropolitan Monterrey region. The growth of @trackmty in 2010 seems to have been boosted by the fact that the MNR group has gotten credit in the news for publicizing a situation in the small town of Comales, in the neighboring state of Tamaulipas [22]. Close to the border and on the drugs routes to the US, Comales has often suffered during the fights between the drug cartels. In March 2010, Comales seems to have been held captive by a cartel. MNR's open letter to the President of Mexico, publicized in the news and on TV, is associated with the engagement of the Army to free the town.

March 2010 is an important month in the creation of the #mtyfollow community. In fact, several curator accounts and blogs were created in March 2010 including MNR and Blog del Narco.

In early January 2011, a troll account appears with the deliberately confusing name @trackrnty (using lowercase RN due to their visual similarity to lowercase M). This account starts posting copies of previous @trackmty's tweets along with other non-sense in an apparent effort to ridicule the account and confuse the community. We cannot be sure about what prompted the creation of this account, but our guess is that it is more related to jealousy and personal grudges rather than activity by cartel members. While new accounts do not immediately have followers and visibility, the troll's postings are retweeted and promoted by a handful of popular accounts in the blue group (see Figure 2). Soon, @trackmty is complaining about it to the community, and members of the community support her by publicizing that trackRNty is a troll and reporting it to Twitter. Their actions and complaints are effective and in late February the troll account is shut down. These events are reflected in the data by the first and third red arrows in Figure 4.

Between late January and March 2011 several serious crimes shake the city of Monterrey and Mexico in general, and are reflected in the data by the spikes around the second and before the fourth red arrow in Figure 4. Particularly shocking is that two graduate students from the Monterrey Institute of Technology and Higher Education (TEC) are killed in a crossfire between the Army and drug cartels [27]. Additionally, several Mexican mayors are attacked and three are killed within a 2 week period, several car bombs explode allegedly due to the competing Gulf and Zetas cartels, and some US agents are killed or wounded.

An ongoing suspicion among the citizens and the community is that "hawks" ("halcones" in Spanish), informants paid by the cartels, are responsible for the successes of the cartels. A belligerent mayor of a town near Monterrey proposes that citizens become "eagles" ("aguilas"), and counter the hawks [9].

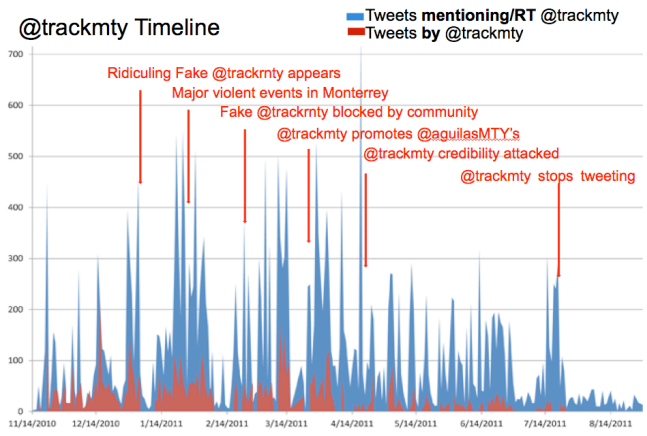

Figure 4. Annotated timeline for tweets related to @trackmty. The red arrows indicate the dates around which we translated tweets, in order based on their volume.

These events were likely among the main reasons that lead @trackmty to propose the creation of "eagle" twitter users in Monterrey. In late March, she persuades some of her many followers to create new accounts, code-name them "aguilasMTY" and report on the activity of hawks. In the data, this is marked by the fourth red arrow.

Collecting Twitter accounts created shortly before April and containing the strings "aguila," we found that there were dozens such accounts created at that time. Many of these accounts send their first tweet to @trackmty: "reporting for duty". @trackmty also sends the first tweet that uses a new hashtag, #aguilasMTY on March 30, 2011:

Aprendamos a identificar halcones: Motociclistas, carteristas, esquineros, taxis piratas etc. DENUNCIEMOS #mtyfollow #Aguilasmty

(Learn to identify hawks: Motorcyclists, pickpockets, people at corners, pirate taxi drivers, etc. DENOUNCE.)

Part of her new plan is for the eagles to be more proactive and report not only areas of high risk, but also areas of calmness, so that the community gets a clearer idea of what is happening. In a further data collection we found that in the early days of April 2011 there were at least 625 tweets reporting on calm areas ("zonas tranquilas") in the city, half of them sent by @trackmty.

But not everyone in the community is happy with this development. Most of her old followers, including some from the old MNR movement disagree and do not join the new initiative. Some question her motives, accusing her of getting too "bossy". Still, others are pleased by this initiative and start reporting using the new hashtag.

A new account appears in early April named @antihalcones ("anti-hawks" in Spanish) and spends its first day sending fifty-seven tweets accusing @trackmty of working for the Zetas cartel. Their reasoning is that by reporting on calm areas in the city she was effectively crowdsourcing the "eagles" and informing the cartel members about where it was safe to go, away of the presence of the authorities. We remind the reader that the Zetas at that time were wrestling control of the city from the hands of the established Gulf cartel, which is widely believed to be supported by the local police.

The initial accusations of @antihalcones do not propagate far. Only a few of the blue members retweet the message. @trackmty's credibility in the community is currently too strong and does not seem damaged by these tweets. @antihalcones makes another claim, however. It predicts that @trackmty's Twitter account is implicated with a Blackberry smartphone that will supposedly prove her connection to the Zetas. It promises to make more revelations soon.

On April 9, the most serious accusations come from the widely read Blog del Narco [6]. It reports that the authorities found a Blackberry in a safe house, and that this smartphone supposedly connected to a twitter account that had communicated with @trackmty through direct messages. (Recall that direct messaging is enabled automatically by mutual following of two parties, like the thousands depicted in Figure 2.) The blog makes identical accusations as @antihalcones did about reporting of calm areas. The same night, @antihalcones is triumphant: it tweets that its prediction about the existence of the Blackberry working for cartels were "proven." It also promises that soon will give more information about @trackmty and in a threatening tone warns her to prepare for what will follow.

The post by the widely read Blog del Narco proved too strong a blow for @trackmty's reputation to stand. A blog has permanence, compared to the fleeting posts on Twitter, making it possible for the news to spread for days after the blog post appears. After this incident, most of the aguilas accounts stop tweeting. As this news floods the community, there are varying responses: some believe the accusations, but others are skeptical. A new troll, @MelisaLotzer (with one s) re-appears and, aiming to confunse @trackmty's followers, pretends to be @trackmty. It copies and posts old tweets by Melissa, claiming that *it* is the real @trackmty.

The original Melissa @trackmty is slow to react and when she does, she tries to point to her past accomplishments, in particular the creation of MNR and the interviews she has given to several reporters from the US and Spain (REF). But the frequency of her tweeting decreases, along with the community's retweets. Finally, at the end of June, she stops tweeting altogether.

The reaction of the community

The events we discovered from the tweets and news translations made it obvious that we needed to get a better

understanding on the way the community reacted to the accusations. We realized that the overall volume of tweets containing the string @trackmty, as in Figure 4, was not giving us this information, and that we needed separate the verbatim retweets from the simple mentions. We know from past research (e.g., [18]) that verbatim retweets indicate that the conveyance of trust in the original sender through the retweeter's rebroadcast of the message to its followers. Very rarely do people retweet messages verbatim with which they disagree. The mentions, on the other hand, show some interest in the account being mentioned, but do not necessarily imply trust. In Figure 5, we graph the data as we did in Figure 4, but making a distinction between retweets and mentions. We believe that this image is striking in revealing the change of mood by the community towards @trackmty. Notice that before March 2011, the community's messages related to @trackmty follows closely its activity through retweets, indicating a certain level of trust and attention to what she posts. Starting in April, however, the volume of retweets falls, while the volume of mentions increases dramatically. Some recovery of the earlier pattern that occurs in the summer is probably related to the short-term memory that real-time media exhibits, and in the end of the attacks on Twitter by her accusers.

What we did not find during the translation of the tweets, but discovered later during interviews with members of the community is that her tormentors succeeded in uncovering the person behind the pseudonym Melissa Lotzer. In late July, they reveal her real identity, her photograph, friends and home address. (Her home and neighborhood can be viewed via Google Street View.) We describe next how we discovered this, and how we evaluated the validity of our findings through interviews.

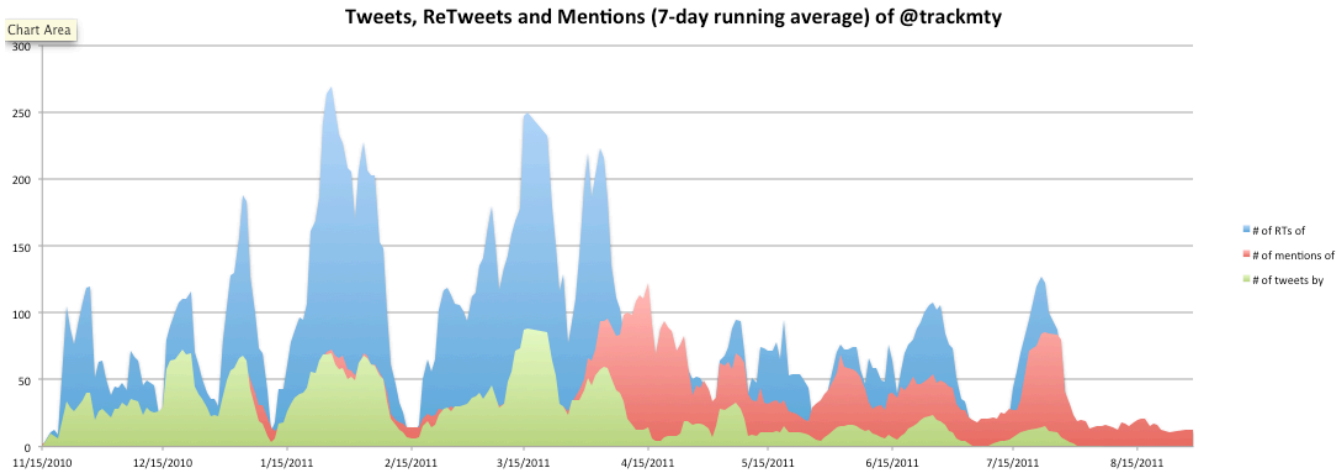

Figure 5: Histogram (7-day averages) of the number of tweets by @trackmty (green/front), the number of verbatim retweets of her tweets (blue/back) and mentions of her account (red/middle).

INTERVIEWS

Data analysis was very useful in identifying specific dates of events to study, and the study of specific tweets on those dates helped us develop a theory of what probably happened to @trackmty. A question in our minds throughout was, how can we be sure that this was really what happened? Was it really the attempt of the creation of the aguilas group that brought her downfall? Can one verify the theory developed by piecing together statistical analysis of a timeline and translation of a few dozen tweets?

We had never spoken to any of the individual citizen reporters that were sending thousand of tweets over several months to check whether our theory was in accordance with their perception of what happened. What if we had missed some major developments in the tweets we did not translate or those that were missing from our spotty collection?

We wanted to interview members of the community. But why would they talk to us, people they do not know, and reveal details that could put them or their friends in danger? The difficulty in interviewing members of this community is described in [20] and reveals the curators' mistrust towards people outside the community, even those claiming to be Mexican nationals and researchers in leading US universities. How could we gain their trust and have them respond to our theory? And how could we further check the validity of their responses?

The break came almost serendipitously. In July 2012, we were asked to present our earlier work [21] at a colloquium at Harvard's Berkman Institute. This earlier work consisted of a general analysis of the tweet volume and friends relationships. It did not include any of the theory we presente in the previous section. However, the talk was broadcasted live and recorded for future viewing, giving its record some permanence. Berkman's colloquia are widely

advertised and attract attention even outside the US. We guessed that some of the attendees might tweet about the talk, as it often happens, using Twitter, the medium that the #MTYfollow community uses constantly. Though it was a long shot, we decided to present not only our published work, but also much of the theory we had developed. Worried that we might accidentally reveal something that might expose @trackmty to danger, we used a pseudonym for her, the hashtag and the city. We figured that if curators from the #MTYfollow community listened to the talk, they would recognize the situation we describe and might react by telling us whether we got the story right or not.

Indeed, by the end of the talk [4] there were several tweets by the #mtyfollow community that had shown active interest in our presentation, some offering to help with any questions we might have. Through a common Twitter acquaintance we were able to establish direct messaging with a couple accounts that were in the green and red groups (Figure 2) and interview them.

Establishing trust in our communication with the interviewees did not turn out to be difficult. From their perspective, we had proven our credentials beyond doubt through the Berkman talk and our web presence. From ours, we knew details of their Twitter activity through several months of data and could verify their responses. Below we describe our interview with one curator from the green group, who knew the community's events well because she had joined #MTYfollow in its early stages. Our discussion began using direct messages –short synchronous messages– a rather inconvenient format for an interview. The discussion quickly moved to email and long asynchronous discussions over an extended period of time.

Interview with a curator

Our first interviewee had recognized the person we described in the Berkman talk and acknowledged that we had gotten the story right. @trackmty was, indeed the person we had described in the talk under the pseudonyms "Alex" and @GodFather, the citizen reporter that was attacked by trolls and blogs. Apparently the reason for the attack, in her opinion, was that she had become too prominent in the community, daring to even claim that she was the creator of the #MTYfollow tag, while others disagreed, claiming that they had that honor[1].

The profile of this curator matched those described in [20] (though she was not one of them). She is a young woman who usually spends long hours in front of communication devices. She was moved to become a curator by her concern for the community, and anxiety about the state of affairs in Mexico, and in Monterrey in particular. She was also idealistic, hoping for an end of the violence and the defeat of the cartels. While she had initially joined the MNR movement, she was disappointed by @trackmty's organization of the aguilas and had distanced herself from Melissa. Yet she is sure that there is no connection between Melissa and Zetas or other drug cartels. She also worried about the aggressiveness of the trolls in uncovering Melissa's identity, photograph and home address; she would not want the same to happen to her.

During the interview, we asked several questions about events we could use to verify through our collection, and that increased our trust in her reporting. The uncovering of Melissa's identity was something that we had missed until that time. Recall that we were translating tweets guided by the height of data peaks, and this event was not among them (Figure 4), though it was definitely a remarkable event in the data. Once we knew where to look, we were able to locate the relevant tweets. Even though the links to Melissa's photographs had been deleted, we were able to locate them through web archival search [15].

Interview with @trackmty

Our interviewee did not want us to reveal her own identity or her pseudonym, but she thought that Melissa would not mind since her identity had already been revealed. We, therefore, decided to establish contact with Melissa. But who was the *real* owner of the original @trackmty? We knew from our data that there were two Melissa's on Twitter, one spelling the name with one s and the other with two. We also knew that the original @trackmty account had been deleted in late 2011 and a new, empty account without any tweets was in its place. Other, almost identically named accounts existed, some of them likely belonging to trolls. We did not want to contact them so as not to alert them that someone was looking for Melissa. We knew that we wanted to locate the one who was also the original owner of the MNR blog.

Again, using the credentials of our talk we established a connection with @MelissaLotzer who proved to us that she could post on the MNR blog [22]. Like the other curators, she is a young woman whose motives were to be engaged and lead the community. These motives came out of the desperation from the security situation in Mexico. An idealist, she was excited by her early recognition in the letter that publicized about the alleged Comales siege and the support of the community in the early attacks by trolls. She thought she could push the cause even further. She acknowledges that the movement of the aguilas turned out to be her major mistake, but she disagrees with the argument that the reporting of calm areas were really helping the cartels. We think she makes some valid points when she says:

"I suppose that in the case of a crime the gunmen would rather be focused on their Nextels listening to their bosses or hawks about where to go. Based on the info released on

[1] In [21] we describe how we discovered that the first proposer of the tag was a minor member of the community, but this detail was lost in the collective memory over time. The original proposer had left the community early on, probably unaware of the effect of his/her actions.

the media, the organized crime modus operandi is to have hawks on every corner and certainly they do not need my reports to see where the police is or which area is calm or empty."

With her new account, @melissalotzer, she has been building back her connections to the community she cares about. At the time of this writing, she has eight thousand followers. She has certainly become wiser through her experience, though she still believes that the aguilas movement was a good idea despite the troubles it caused her.

"I'm sure the aguilas movement was a good idea and I do not regret of making the aguilas – in fact if I had the time, I would restart it again."

She is also not shy describing the change she believes she has made in some citizens' lives:

"I'm completely sure that trackmty was the reason why many people started using twitter. I receive comments daily by followers that are opening a twitter account to a family member just to follow me […] They tell me: please take care of my mom, she will be reading your tweets, she will not be reporting cases because she doesn't know how to use a blackberry or so. Many similar cases like that happen every day."

In her interview, Melissa was very forthcoming, providing more information that we could possibly verify by the data and analysis that is publicly available. Nothing we could not deduce and verify through our data is included in this paper: it is not our purpose just to report a story, no matter how interesting. This is, rather, a case study of the limitations and the intricacies of data analysis. Melissa's objective when she agreed to give us the interview, we believe, was to re-establish some of her lost credibility with the community, but since she is always aware of the personal risks associated with the dangerous situation in Mexico, she does not want us to use her real name.

SEMI-AUTOMATIC TOOLS TO ENHANCE TRUST

An underlining theme of this paper and our overall research [19] is how one can evaluate the credibility of information one receives through real-time media, such as Twitter. This becomes important especially in situations where one is immediately called upon to act based on information one receives. In the case of Mexican narcotweets, for example, what should an individual do upon receiving a tweet from an anonymous informant that a particular intersection through which he is about to pass may be dangerous? Maybe he plays it safe and avoids the area, just in case. But what if his children's school is located in that area? What if the information he receives indicates that the area is, instead, safe? Is it really so, and if yes, for how long? Or, what may have already occurred in the area that makes his anonymous informant interested in reporting that it is now safe?

To address the overall research aim, we have put forth several projects that will help the citizen evaluate the quality of information they receive. In particular, we propose:

- *Establishment of new metrics* that will help users evaluate the trustworthiness of information they receive, especially from real-time sources, and when it demands immediate attention and action. We are looking at real-time algorithms that maintain a *trail of trustworthiness* for every piece of information the user receives.
- *Monitor the evolving ways in which information reaches users*. This monitoring will help us be informed of the changes introduced by search engines and social media companies as they try to improve their services.
- *Establish a personalizable model* that captures the parameters involved in the determination of trustworthiness of information in real-time sources by applying machine learning and data mining algorithms. We design online algorithms that support the estimation of quality via the maintenance of *trails of trustworthiness* that each piece of information carries with it, either explicitly or implicitly. Of particular importance is the fact that these algorithms should help maintain *privacy* for the user's trusted network.
- Design algorithms that can *detect attacks* on real-time sources. For example one can automatically detect bursts of activity related to a subject, source, or non-independent sources. Using an appropriate interface, this information can be useful to the end user as well.

Currently, we develop algorithms that calculate the trust in an information trail based on a score that is affected by the influence and trustworthiness of the informants and the independence of their social connections [19]. The knowledge and experience from the #MTYfollow community has provided us with an important test case.

CONCLUSION

Technical papers analyzing collections of data often focus on the statistical and data mining methods that can reveal hidden patters of communication and interaction between social actors. However, it is neither easy to verify the validity of newly discovered patterns, nor certain that the data collected is complete. In analyzing the tweets around a popular hashtag used by users who worry about their personal safely in a Mexican city we found that one must go back and forth between collecting and analyzing many times while formulating the proper research questions to ask. Further, one must have a method of establishing the ground truth, which is particularly tricky in a community of – mostly – anonymous users.

We saw that it is possible for trust to emerge even around an anonymous leader in the community, as expressed by the willingness of community members to act based on anonymous informants and follow advice that may affect their lives. However, trust gained anonymously is flimsy and can be damaged fast through coordinated attacks. And, as it has been observed throughout human history, prominence brings envy.

On the technical side, we were impressed by the remarkable performance of the Louvain algorithm [5] in clustering the groups of trolls from the citizen reporters. In the future, we plan to examine the extent to which semi-automatic algorithms can measure the trustworthiness of the information users receive via real-time sources.

ACKNOWLEDGMENTS

The authors would like to thank Samantha Finn and Elize Huang for the development of many of the figures in the paper, Susan Tang, Yesenia Trujillo, Danaë Metaxa-Kakavouli, and the automatic translation feature of Google's Chrome browser for translating the tweets and blogs from Spanish, and all who helped by providing helpful critical comments during the development of this document. We gratefully acknowledge the CNS-1117693 grant from NSF.

REFERENCES

1. Archivist, T. (2010). http://archivist.visitmix.com
2. Bastian, M. et al. (2009). Gephi: an open source software for exploring and manipulating networks. In ICWSM09. http://bit.ly/VA0Hkx
3. Baron, D. (2005). Persistent media bias. Journal of Public Economics 90. http://bit.ly/XYLynL
4. Berkman Center Talks (2012) Narcotweets: Reporting on the Mexican Drug War using Social Media, July 10. http://bit.ly/LRDVvS
5. Blondel, V. D et. al. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 10. http://arxiv.org/abs/0803.0476
6. Blog del Narco (2011). Delincuencia organizada infiltrada en redes sociales. BlogdelNarco.com. http://bit.ly/UK85qW
7. boyd, danah and Kate Crawford (2012). "Critical Questions for Big Data." Information, Communication, & Society 15:5, p. 662-679.
8. The Guardian (2011). Woman's decapitation linked to web posts about Mexican drug cartel. 25 September http://bit.ly/yTobHp
9. CNN Mexico (2011). "After God, is the Army": Mayor Garcia, Nuevo Leon http://bit.ly/XNKpib
10. Committee to Protect Journalists, C. (2011). Attacks on the press 2011: Mexico. http://bit.ly/YJBAN2
11. Gephi. (2010). ForceAtlas2, the new version of our home-brew layout. http://bit.ly/yGXRhJ
12. Greyson, G. W. (2010). Calderon's anti-drug strategy. In Mexico: narco-violence and a failed state? Transaction Publishers.
13. Google automatic translation. http://translate.google.com
14. Heinzelman, J., and Waters, C. (2010). Crowdsourcing crisis information in disaster-affected Haiti. US Institute of Peace. http://bit.ly/WZsLdd
15. Internet Archive (accessed June 1, 2012) http://archive.org
16. Los Angeles Times (2013) Mexico's Drug War (Last accessed Jan 29, 2013) http://lat.ms/Srnpdg
17. Meier, P. (2011). Verifying crowdsourced social media reports for live crisis mapping: An introduction to information forensics. iRevolution blog http://bit.ly/kstN2t
18. Metaxas, P. and Mustafaraj, E. (2010). From obscurity to prominence in minutes: Political speech and real-time search. In WebSci10. http://bit.ly/qInTmW
19. Metaxas, P. and Mustafaraj, E. (2012) Trails of Trustworthiness in Real-Time Streams (Extended Summary) DIST/CSCW'2012 http://bit.ly/spbYis
20. Monroy-Hernández et. al. (2013) The new war correspondents: The rise of civic media curation in urban warfare. CSCW'13, Feb. 23-27, San Antonio, TX, 2013
21. Mustafaraj et. al. (2012) Hiding in plain sight: A tale of trust and mistrust inside a community of citizen reporters. ICWSM'12, June 4-8, Dublin, Ireland, http://bit.ly/JbaAeW
22. MX Nueva Revolucion blog (2012) Narcotuits: Un estudio de Microsoft, Wellesley College y Harvard http://bit.ly/Wk7i12
23. NYTimes (2012). Mexico Updates Death Toll in Drug War to 47,515, but Critics Dispute the Data. http://nyti.ms/WkmzyS
24. O'Connor, B. et. al. (2010). From tweets to polls: Linking text sentiment to public opinion time series. Proc. of ICWSM'10, 122–129. AAAI Press.
25. Reuters. Special Report: If Monterrey falls, Mexico falls. (Last accessed 12 Jan 2013) http://reut.rs/VSVKRz
26. Wikipedia. Mexican Drug War. (Last accessed 12 Jan 2013) http://bit.ly/YJBaq1
27. Wall Street Journal. (2010) Killings Take Drug War to Mexico Elite. http://on.wsj.com/WGpYEp