



Analyzing Civil Unrest through Social Media

Ting Hua, Chang-Tien Lu, and Naren Ramakrishnan, Virginia Tech

Feng Chen, Carnegie Mellon University

Jaime Arredondo and David Mares, University of California, San Diego

Kristen Summers, CACI

Mining and analyzing data from social networks such as Twitter can reveal new insights into the causes of civil disturbances, including trigger events and the role of political entrepreneurs and organizations in galvanizing public opinion.

Incidents of civil unrest, such as the recent strike by tens of thousands of Indonesian workers over pay demands (<http://goo.gl/QlgFk8>) and protests by teachers in Mexico against proposed education reforms (<http://goo.gl/FM8nM3>), can undermine a country's stability. A holistic understanding of such incidents, including the events, people, and groups contributing to their occurrence, is thus an important subject in political science research.

Newspapers have traditionally been the primary sources for such analysis. However, daily news accounts typically include only basic factual information about an incident such as the date, location, and number of participants, not details that might reveal the underlying

causes. Figure 1 shows a news article from *Milenio*, a major Mexican newspaper, about a protest in Mexico City calling for the release of wild dogs alleged to have attacked and killed civilians in parts of the city and subsequently captured by the authorities.

Civil unrest can be triggered by any number of factors, either alone or in combination, including actions by the government (controversial legislation, police brutality) or a powerful third party (criminal gang activity), natural disasters that cause widespread human suffering (severe hurricane, major earthquake), and calls to action by "political entrepreneurs"—leaders with a sizable following— or real-world or online political organizations.

Emerging social media offer unprecedented opportunities to track the evolution of civil disorders. In particular, Twitter, as both a social network and information-sharing medium, provides a real-time platform for disseminating news and opinions (H. Kwak et al., "What Is Twitter, a Social Network or a News Media?," *Proc. 19 Int'l Conf. World Wide Web [WWW 10]*, ACM, 2010, pp. 591-600). Analyzing tweets related to specific protests provides insights into the root causes, who the organizers are, and how online expression reflects or contributes to such events.

Tweet analysis involves three steps. The first is to extract the most pertinent tweets associated with an incident from available Twitter data. The next step involves identifying

the key contributing factors: trigger events and relevant political entrepreneurs and organizations. The final step is to correlate these factors to understand how the incident evolved. In the examples that follow in this article, we carried out text processing in the native language (Spanish) but, for presentation purposes, translated the results to English using Google Translate.

EVENT-RELATED TWEET EXTRACTION

To obtain tweets related to events of interest, we developed a ranking algorithm to connect the dots between news reports and tweets. As the top half of Figure 2 shows, we first identify *topic keywords* and *event words* in news accounts.

Topic keywords are those terms most frequently used to describe a topic—for civil unrest, keywords might include “protest,” “march,” and “demand.” Using a database of 9,080 protest events in 10 Latin American countries from January 2011 to September 2013, we

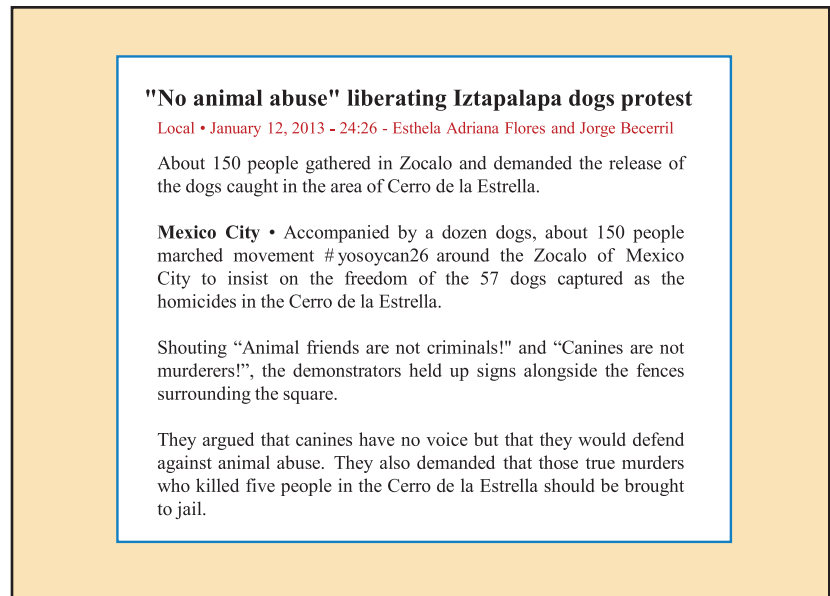


Figure 1. News article from *Milenio*, a major Mexican newspaper, about a protest in Mexico City calling for the release of captured wild dogs alleged to have attacked and killed citizens. Like most such articles, it includes basic facts such as the date of the protest (12 January 2013), its location (El Zócalo, the city’s main public square), and the number of participants (150) but provides little insight into the incident’s underlying causes. The article, originally in Spanish, has been translated into English using Google Translate.

extracted 200 top-ranked topic keywords for each country using information retrieval measures such as TF-IDF (term frequency times inverse

document frequency). The database, provided by MITRE, summarizes news reports from various global news outlets such as the BBC and

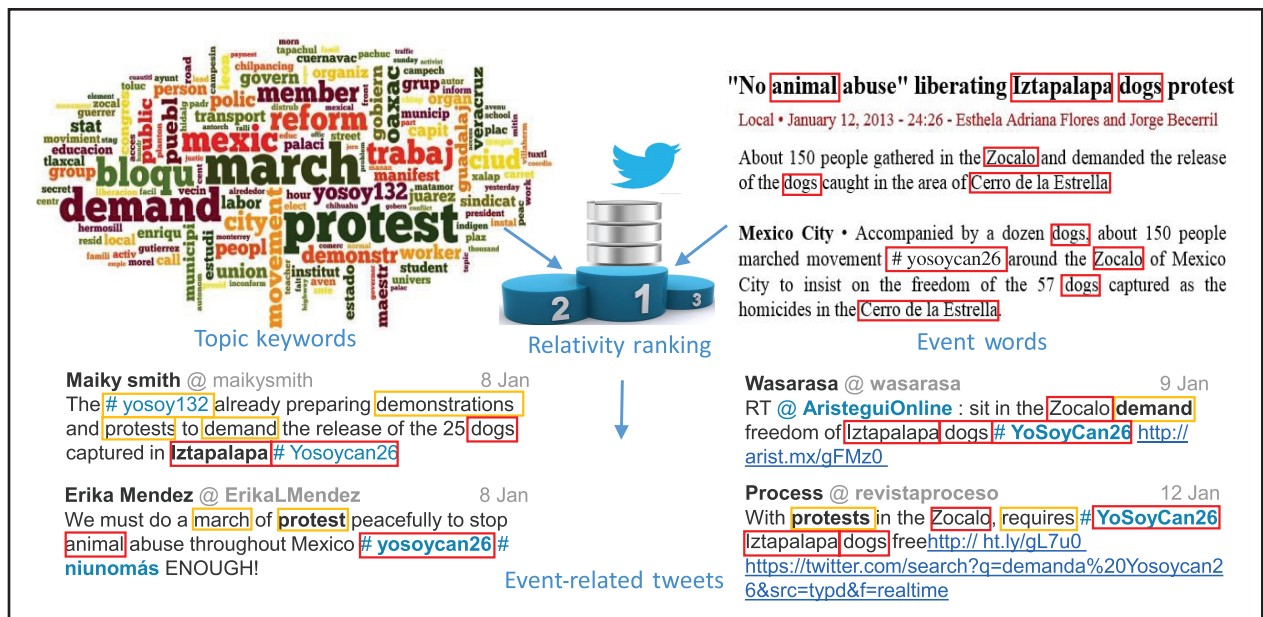


Figure 2. Event-related tweet extraction pipeline. Red boxes indicate event words for the street dog liberation protest in Mexico City and yellow boxes denote topic keywords. The word cloud in the top left shows top-ranked topic keywords for Mexico, generated from a database of 2,141 protest events in that country from January 2011 to September 2013. The tweets, originally in Spanish, have been translated into English using Google Translate.

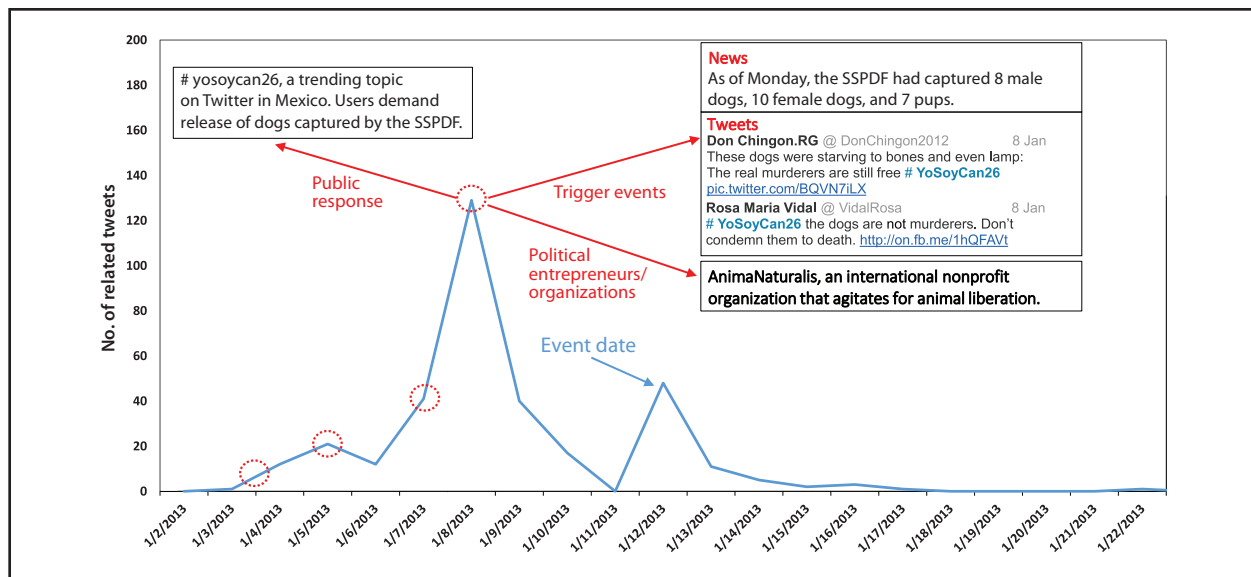


Figure 3. Distribution of tweets related to the street dog liberation protest in Mexico City. Tweets spiked several days before the rally, which is typical of incidents of civil unrest. In contrast, tweets related to breaking news stories and major events such as natural disasters usually spike during the day of the event. The tweets, originally in Spanish, have been translated into English using Google Translate.

CNN as well as major local media such as *Milenio* in Mexico and *ABC Color* in Paraguay.

Event words are context-relevant terms that can help characterize a specific event. In the case of the street dog liberation protest in Mexico City, event words include “dogs” and “Cerro de la Estrella” (the area where authorities captured the animals); these words often appear in accounts of this event but not in those of other civil disturbances.

As the bottom half of Figure 2 shows, we extract tweets containing topic keywords and event words from tweets published around the event date—for example, from 10 days before the event through 10 days after. We primarily consider tweets that include URLs linked to published news reports. Such tweets combine news with user opinions and best reflect Twitter’s hybrid nature as a social network and information-sharing medium.

Our algorithm quantitatively evaluates every tweet’s relevance to the event by considering textual, spatial, and temporal distances

between tweets and event-related news reports. It further clusters tweets based on their content similarities and social ties, and returns the cluster with the largest average relevance score as the set of event-related tweets.

IDENTIFYING CONTRIBUTING FACTORS

Prior research shows that for breaking news stories (M. Hu et al., “Breaking News on Twitter,” *Proc. SIGCHI Conf. Human Factors in Computing Systems* [CHI 12], ACM, 2012, pp. 2751-2754) and major events such as natural disasters (T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors,” *Proc. 19 Int’l Conf. World Wide Web* [WWW 10], ACM, 2010, pp. 851-860), the number of associated tweets typically spikes during the day of the event and drops rapidly soon afterward.

For incidents of civil unrest, however, the number of tweets often begins to rise for several days leading up to the event as awareness of the issue grows, climaxing in

a large burst of tweets before the actual event date. As Figure 3 shows, tweets relating to the street dog liberation protest spiked on 8 January, four days before the rally.

Analyzing such early bursts reveals key contributing factors including trigger events and political entrepreneurs and organizations pushing for protest activity.

In the case of the street dog liberation rally, the trigger event was the capture of 25 dogs on 7 January by the responsible law enforcement agency, the Secretaría de Seguridad Pública del Distrito Federal (SSPDF). The following day, Twitter users angrily demanded the release of the animals, presumably to be euthanized. They claimed that the dogs are “not murderers” and that the government was ignoring the “real murderers” who “are free.”

Many tweets protesting the SSPDF’s actions included the hashtag #yosoycan26 (“I am dog 26”), which became a trending topic in Mexico on 8 January. This was clearly inspired by #YoSoy132 (“I am [student] 132”), a hashtag associated with a youth-based democratization movement

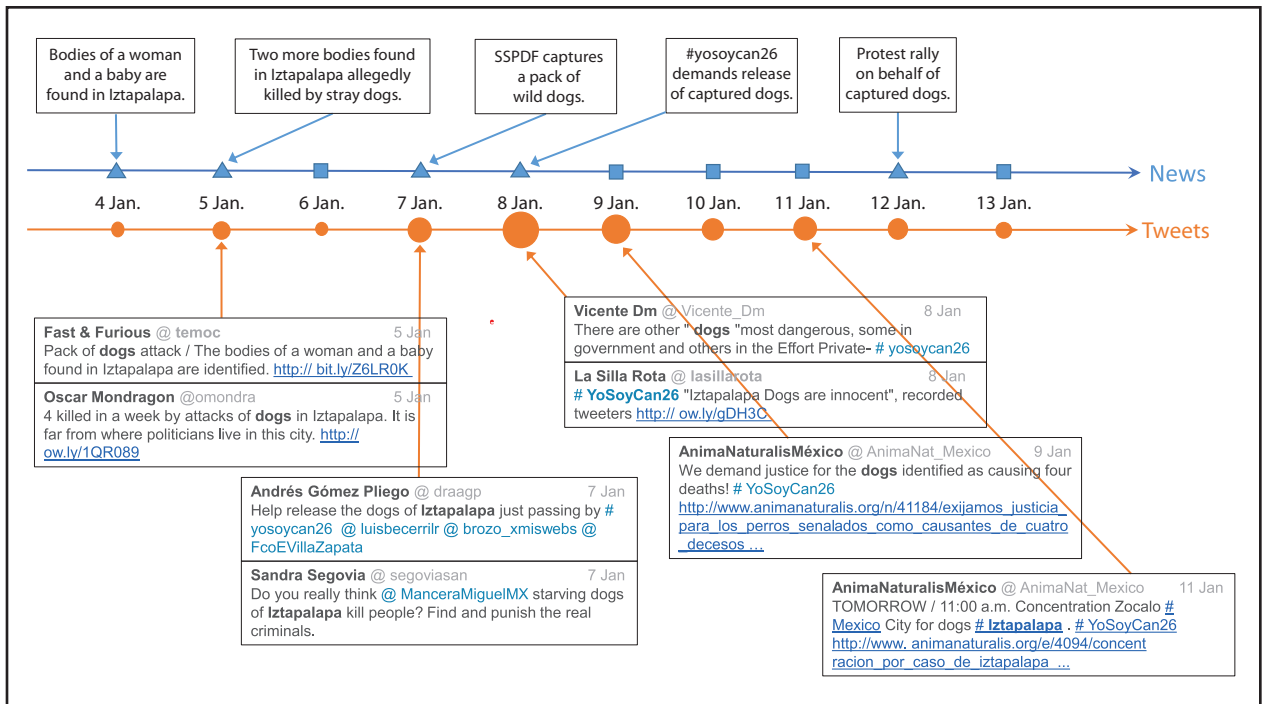


Figure 4. Events leading up to the street dog liberation protest. The blue timeline indicates news reports, while the orange timeline denotes event-related tweets. On the blue line, triangles represent dates with emerging news, and squares are regular dates without emerging news. On the orange line, the size of the circle indicates the relative number of related tweets on the corresponding date. The original tweets, in Spanish, have been translated into English using Google Translate.

that arose in response to alleged media bias and fraud and corruption in the 2012 Mexican presidential election (<http://goo.gl/vth5sN>). In response, animal rights groups such as the international nonprofit AnimaNaturalis helped organize and promote the demonstration.

EVENT EVOLUTION ANALYSIS

Preceding the largest burst of Twitter activity on 8 January were three smaller bursts, indicated by the lower red circles in Figure 3. To explore the reasons for these surges, we created a timeline for the days leading up to the event showing both protest-related tweets (including those without links) and news reports extracted from tweet links, as Figure 4 shows.

Initially, news media reported that the bodies of a woman and a baby were found in Iztapalapa, a borough in Mexico City's Federal District, on 4 January. The news,

however, attracted little attention among Twitter users. The following day, the government announced that the deceased, along with two others, were killed by wild dogs. Twitter users showed more interest, but many tweets simply reported the news headline.

The turning point occurred on 7 January, when the government announced that it had captured a pack of stray dogs it suspected of committing the attacks. Tweets doubting whether the dogs were responsible began spreading quickly among Twitter users. In addition, tweets with the hashtag #yosoycan26 began to appear, calling on the government to free the dogs and, as one Twitter user suggested, "find and punish the real criminals." Within the same day, this hashtag became one of the most popular trending topics on Twitter in Mexico.

Protest-related tweets peaked on 8 January. Some tweets were limited to expressions of sympathy for the

captured animals, such as "dogs of Iztapalapa are innocent," but others expressed strong dissatisfaction with the government: "There are other 'dogs' more dangerous, some in government and others in the private sector." These antigovernment tweets attracted great public attention, such that the news media reported #yosoycan26 as a trending topic on Twitter in Mexico later that day. These reports helped spread #yosoycan26—and the protestors' cause—to a broader audience.


After 8 January, the volume of relevant tweets began to decrease, as no new information was reported after then. The following day, however, AnimaNaturalis joined the tweeting campaign, demanding "justice for the dogs identified as causing four deaths." On 11 January, the organization called for real-world action: "TOMORROW/ 11:00 a.m. Concentration Zocalo #Mexico for dogs of #Iztapalapa." About 150 people responded to the

call and gathered at the city square, some with placards bearing the hashtag #yosoycan26. By that time, the number of captured dogs had risen to 57.

Three days after the protest, the government initiated adoptions for 23 of the dogs that were puppies and determined to be unrelated to the attacks. It also indicated that adoptions for the other 34 adult dogs would be initiated soon.

Mining and analyzing data from social networks can reveal new insights into the causes of civil disturbances, including trigger events and the role of political entrepreneurs and organizations in galvanizing public opinion. Twitter in particular has become the dominant medium for organizing demonstrations, since organizers can instantly reach out to thousands of people. Traditional media also continue to play an important role as original sources

of information and by helping to spread trending topics from Twitter to the wider population.

Our research results point the way to future work. It would be especially useful to more fully explore the underlying dynamic processes and characteristics of political entrepreneurship. The fact that Twitter is both a social network and an information-sharing medium permits analysis of the different means of communication that political entrepreneurs use, and which appear to be the most or least effective. Given the large number of instances of civil unrest, it might be possible to distinguish different types of political entrepreneurs as well as events. 

Acknowledgments

This work is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract D12PC00337. The views

and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US government.

Ting Hua is a doctoral candidate in computer science and a research assistant in the Spatial Data Management Lab at Virginia Tech. Contact her at tingh88@vt.edu.

Chang-Tien Lu is an associate professor of computer science at Virginia Tech and vice chair of the ACM Special Interest Group on Spatial Information (ACM SIGSPATIAL). Contact him at ctl@vt.edu.

Naren Ramakrishnan, Discovery Analytics column editor, is the Thomas L. Phillips Professor of Engineering at Virginia Tech and director of the university's Discovery Analytics Center. Contact him at naren@cs.vt.edu.

Feng Chen is a postdoctoral research fellow at Carnegie Mellon University's H. John Heinz III College. Contact him at fchen1@cmu.edu.

Jaime Arredondo is a joint doctoral program student in the Division of Global Public Health and a graduate researcher in the Center for Iberian and Latin American Studies (CILAS) at the University of California, San Diego. Contact him at jarredon@ucsd.edu.

David Mares is a professor of political science and holds the Institute of the Americas Chair for Inter-American Affairs at the University of California, San Diego. Contact him at dmares@ucsd.edu.

Kristen Summers is the chief technology officer of the Advanced Knowledge Solutions Division at CACI, an information solutions and services company supporting national security missions and government transformation for intelligence, defense, and federal civilian clients. Contact her at ksummers@caci.com.



Selected CS articles and columns are available for free at <http://ComputingNow.computer.org>.