
Microblogging in Mass Emergency

Sarah Vieweg

University of Colorado
connectivIT Lab
ATLAS Building, Room 320
320 UCB
Boulder, CO 80309
vieweg@colorado.edu

Kate Starbird

University of Colorado
connectivIT Lab
ATLAS Building, Room 320
320 UCB
Boulder, CO 80309
catharine.starbird@colorado.edu

Abstract

Research on the popular microblogging site Twitter is quickly burgeoning. In this paper, we describe data collection, data analysis and briefly discuss findings that have resulted from recent analyses of Twitter communications. Our experience using Twitter as a research site during mass emergency gives rise to several questions regarding our rights as researchers and the expectations of Twitter users.

Keywords

Twitter, microblogging, crisis informatics, computer-mediated communication

ACM Classification Keywords

K.4.2 Social Issues, K.4.3 Organizational Impacts—computer-supported collaborative work.

Introduction

During times of mass emergency, affected populations are suddenly and unexpectedly placed in a position to make decisions and then act on those decisions. However, information about the situation (be it a flood, wildfire, earthquake, etc.) is often challenging and/or cumbersome to obtain.

Recent research [1,2,4,6] shows that during times of mass emergency, many turn to social networking and microblogging sites to gather and distribute timely,

Copyright is held by the author/owner(s).

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

ACM 978-1-60558-930-5/10/04.

relevant information. Twitter in particular has received attention from both people affected by or concerned about mass emergencies, as well as researchers who study computer-mediated communication (CMC) that takes place during such situations.

In this paper, we discuss our research regarding the use of Twitter during mass emergencies, with particular attention paid to our data collection and analysis techniques. The study of Twitter data can be unwieldy and cumbersome. We offer our insight and experience with these data as a contribution to those in the CHI and CSCW communities who are studying this fascinating and thriving field. Additionally, we pose questions that we often encounter during our research.

Data Collection

When a mass emergency occurs, we must quickly determine if, and how, we will study communications within the “Twitterverse.”¹ Once we decide to collect data regarding a specific event, our first step is to use publicly available Twitter search tools, both through their website and 3rd-party applications, to pinpoint relevant search terms based on how Twitterers are communicating about the event. In a second step, we use an in-house software tool that calls on the Twitter Search API to gather tweets containing the search terms. These tweets are then sent to our lab server, and are converted and stored in a MySQL database.

Once the tweets are in database format, we return to the Twitter Search API to collect the entire tweet

¹ Our lab has a relationship with Twitter that allows us to make more than the typically allotted number of “requests” to their servers for tweet data.

streams of each unique user for a specified time period. This allows us to view all of the crisis-related tweets, not just those that contain keywords. It also allows us to see how crisis-relevant tweets fit into broader patterns of Twitter behavior.

Data Analysis

Once we have these data sets (which thus far have included millions of tweets), we narrow them to a manageable size. One method we use is to focus on users who are individuals (i.e. not affiliated with an organization or media outlet) and local to the event, and who have more than three tweets containing a search term in their tweet streams. Thus far, this process of data reduction has yielded data sets that are feasible for manual qualitative analysis.

We then begin the process of detailed analysis using the E-Data Viewer (EDV) [3]. The EDV is visualization and qualitative analysis tool designed specifically for computer-mediated communication data that allows us to parse, analyze, code, and visualize tweets. It is set up so researchers can view tweets temporally, and in the context of user streams.

The EDV allows for both the display and qualitative coding of individual tweets, user streams, and author attributes. As we read tweets, we note themes that emerge in the information being communicated. These themes lead to the collaborative development of categories, which act as the foundation for qualitative coding schemes. Coding schemes can be dynamically evolved and updated by multiple, remote users within the EDV system.

Once we have a good understanding of the types of information Twitterers are sharing, we use the EDV to qualitatively code each tweet using the coding scheme. As we code tweets, we can visualize them to see how certain types of information are being communicated, by whom, and at what time, and compare that to the timing of the emergency event as well as other aspects of the data set.

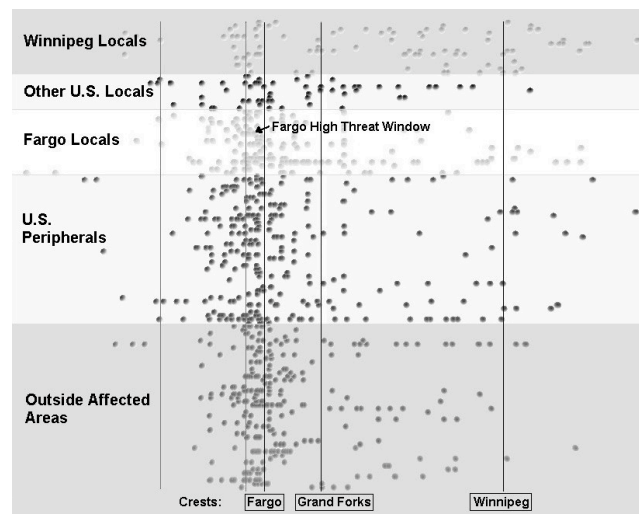


Figure 1. Distribution of search-term containing tweets during the Red River Floods of 2009.

The image above shows a dynamically created visualization of tweets authored by individual Twitterers during the Red River Flooding event. Tweets from individual user streams are displayed horizontally, with time moving left to right. The crests at the different locations are marked with vertical lines. User streams are sorted by the geographic location of the Twitterer. Researchers can click on individual spheres, which

represent tweets, to read tweet content and view or change tweet and author codes.

Textual Analysis

In addition to annotating tweets with coding categories, we also perform detailed textual analysis. This involves going beyond an overall understanding of what information Twitterers are relaying. One goal of our research is to inform techniques for the automatic extraction of relevant data during times of mass emergencies in real-time. As such, a further step we take is to look for linguistic indicators of phenomena that can be used to inform the development of information extraction (IE) tools.

Thus far, this deeper analysis has involved uncovering geo-location and geo-referencing information included in tweets, as well as linguistic phenomena such as markedness and implicature [6].

Twitter Churn

In our study of the Red River Flooding event [4] we found that less than 10% of search-term tweets contained original information. Most tweets broadcasted information that was available in other places in the Twittersphere and the surrounding web. These tweets either repeated information, linked to information on external websites, or were retweets. We termed this re-use behavior, “derivative information distribution,” or, more informally, “the churn”. We see the Twitter churn as an important characteristic of this information space; through original and derivative activity, users shape the space, both creating and countering the intense noise stirred by millions of users.

An additional aspect of Twitter behavior is the broadcast of retweets. Our research suggests that retweets, a user-driven adaptation, act as an informal recommendation system [4]. We have found that individuals local to an emerging crisis retweet information that may be valuable to other locals, and that these retweets show a preference for information originating in traditional media and emergency response organizations [5].

Work in Progress

Along with developing tools to automatically analyze tweets in real or near-real time, we are addressing other research questions. This research includes attempting to understand how network connections affect message content, user stream behavior and information spread during crises, and developing strategies for tracking “conversations” between users and using the EDV to analyze them in context.

Emerging Questions

Several questions have arisen during our work on Twitter use during mass emergency, including issues of ethics and privacy. As researchers, we strive to act consciously while respecting Twitter users. Below are some questions we struggle with.

Q1: Are users aware of the public nature of their Twitter broadcasts? The Twitter terms of use explicitly state “What you say on Twitter may be viewed all around the world instantly. You are what you Tweet!”² But do users understand their tweets are being archived and analyzed? Do we as researchers have the right to study Twitter communications without consent?

² www.twitter.com/tos. Retrieved November 17, 2009.

Q2: Do we anonymize Twitter communications or cite them? Users seem to shape communications for audience consumption. We have also seen evidence suggesting users feel a sense of ownership of their tweets. Do they want to be cited? Do we treat tweets as private communications or public broadcasts?

Conclusion

There are several ways to approach the study of Twitter. In this paper, we have provided a glimpse of how we manage the vast amounts of data produced by Twitter, the results of our analyses, and questions we encounter as we continue to research the Twitterverse.

Acknowledgements

This work is supported by the US National Science Foundation IIS-0546315 and IIS-0910586, and an NSF Graduate Fellowship awarded to Starbird, but does not represent the view of the NSF.

References

- [1] Qu, Y., Wu, P.F., & Wang, X. Online Community Response to Major Disaster: A Study of Tianya Forum in the 2008 Sichuan Earthquake. In *Proc HICSS 2009*.
- [2] Shklovski, I., Palen, L., & Sutton, J. Finding Community Through Information and Communication Technology in Disaster Events. In *Proc CSCW 2008*, ACM Press. 127-136.
- [3] Starbird, C. E-Data Visualization for Qualitative Research. Interactive Poster presented at *CSCW 2008*.
- [4] Starbird, K., Palen, L., Hughes, A.L. & Vieweg, S. Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In *Proc. CSCW 2010*, ACM Press. Forthcoming.
- [5] Starbird, K., & Palen, L. Pass It On: Retweeting in Mass Emergency. In submission to *ISCRAM 2010*.
- [6] Vieweg, S., Hughes, A.L., Starbird, K., & Palen, L. Microblogging During Two Natural Hazard Events: What Twitter May Contribute to Situational Awareness. In *Proc. CHI 2010*, ACM Press. Forthcoming.