Characterization of Social Media Response to Natural Disasters

Seema Nagar
Department of Computer Science
Indian Institute of Technology
New Delhi, India
nagar.seema@gmail.com

Aaditeshwar Seth
Department of Computer Science
Indian Institute of Technology
New Delhi, India
aseth@cse.iitd.ernet.in

Anupam Joshi
Department of Computer Science
Indian Institute of Technology
New Delhi, India
joshi@cse.iitd.ernet.in

ABSTRACT

Online social networking websites such as Twitter and Facebook often serve a breaking-news role for natural disasters: these websites are among the first ones to mention the news, and because they are visited by millions of users regularly the websites also help communicate the news to a large mass of people. In this paper, we examine how news about these disasters spreads on the social network. In addition to this, we also examine the countries of the Tweeting users. We examine Twitter logs from the 2010 Philippines typhoon, the 2011 Brazil flood and the 2011 Japan earthquake. We find that although news about the disaster may be initiated in multiple places in the social network, it quickly finds a core community that is interested in the disaster, and has little chance to escape the community via social network links alone. We also find evidence that the world at large expresses concern about such largescale disasters, and not just countries geographically proximate to the epicenter of the disaster. Our analysis has implications for the design of fund raising campaigns through social networking websites.

Categories and Subject Descriptors

H.4.m [**Information Systems**]: Miscellaneous

General Terms

Human Factors, Measurement

Keywords

Social Media, Natural Disaster, Response Analysis

1. INTRODUCTION

Online social networking websites such as Twitter and Facebook often serve a breaking-news role for natural disasters: these websites are among the first ones to mention the news, and because they are visited by millions of users regularly the websites also help communicate the news to a large mass of people. We examine Twitter logs from the 2010 Philippines typhoon, the 2011 Brazil flood and and the 2011 Japan earthquake to examine how news about these disasters spreads on the social network. We notice that although the news is seeded in different disjoint parts of the network, it very quickly cumulates into a giant connected component that comprises more than 90% of the users tweeting about the disaster. Further, the connected component closes on itself quickly and does not provide outreach to a larger set of users who may be following the tweeting users. This indicates that although news may be initiated in multiple places it quickly finds a core community that is interested in the disaster, and has little chance to escape the community via social network links alone. We also examine the countries of the Tweeting users and find that they span the entire world, giving contradicting evidence that the global village actually does express humanitarian concern and Twitter may not be helping transcend geographical boundaries to the extent that is often hyped in media. We believe our work therefore has significance in the design of news and fund raising campaigns on social networking websites that information may not naturally diffuse via social links to prospective donors but may have to be artificially seeded among users in different countries.

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

2. SOCIAL SPREAD OF NEWS

We use the ‘Typhoon Philippines’ and the ‘Brazil Flood’ data sets to examine how news about natural disasters spreads on the social network. We divided each data set into n data points, where the i^{th} data point consists of the cumulative number of tweets posted and the number of corresponding users until the i^{th} hour. For each data point we constructed the authors follower network as mentioned in [2].

Figures 1 and 2 show the fraction of nodes at any point of time in the largest connected component for ‘Typhoon Philippines’ and ‘Brazil Flood’ data sets respectively. We observe that the largest connected component quickly gains more than 90% of users in both the disasters.

Figures 4 and 3 show the modularity values with time for the ‘Brazil flood’ and ‘Typhoon Philippines’ datasets respectively. For both the disasters, the modularity first rises and then drops and saturates. This seems to be because people get to know about the disasters from different information sources and starts posting about it, but as more and more followers start posting then gradually small components merge and increase the modularity. Eventually however the modularity settles down to a constant value indicating that new people added to the component may already be sharing a tightly knit community with others in the component.

Finally, we also define a new quantity called *reach*. *Reach* at any point of time is a measure of the spread of infor-

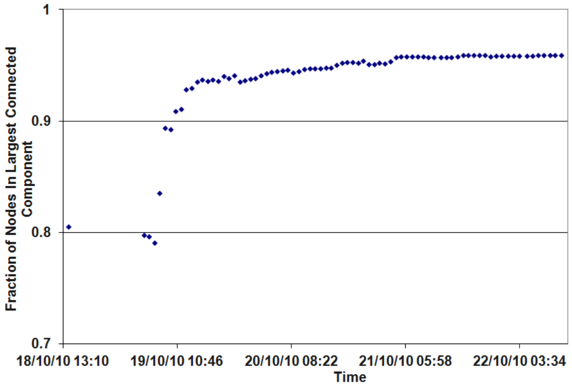

Figure 1: Fraction of nodes in largest connected component in 'Typhoon Philippines' data set

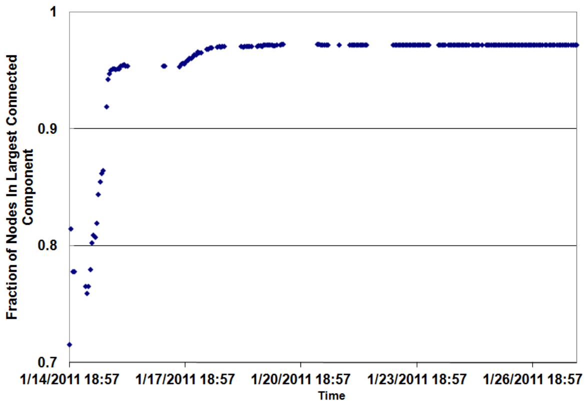

Figure 2: Fraction of nodes in largest connected component in 'Brazil Flood' data set

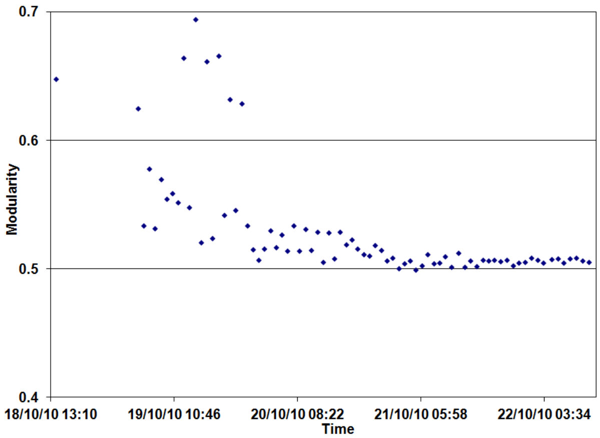

Figure 3: Modularity for 'Typhoon Philippines' data set

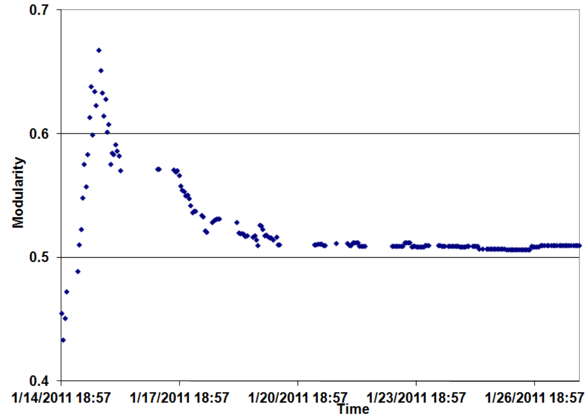

Figure 4: Modularity for 'Brazil flood' data set

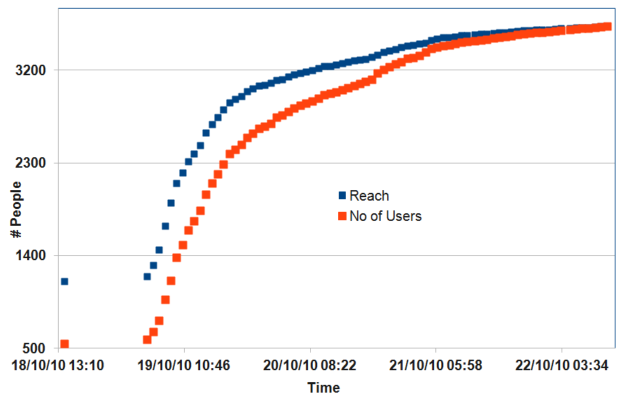

Figure 5: Reach and No of Users posted till that time in 'Typhoon Philippines' data set

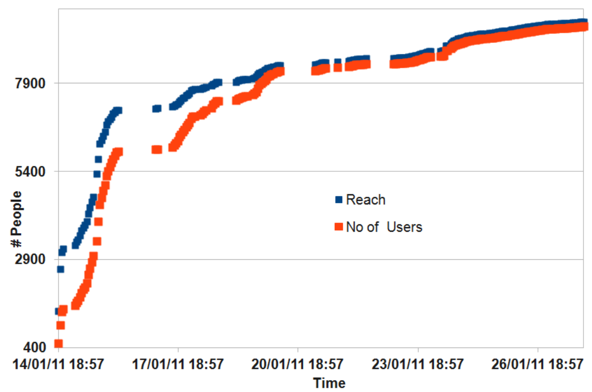

Figure 6: Reach and No of Users posted till that time in 'Brazil flood' data set

Table 1: Number of users with valid locations

Data Set	Total Users	Users With Valid Location
'Brazil Flood'	9488	6378
'Japan Earthquake'	2006655	1302787

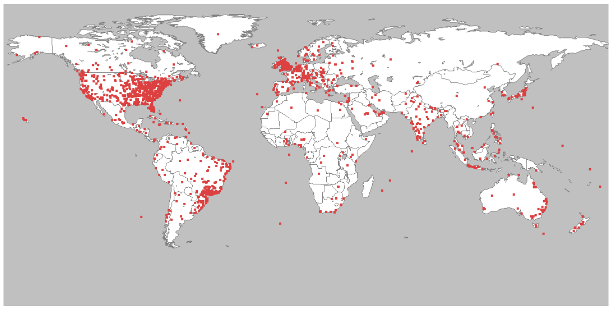

Figure 7: Users Distribution for Brazil Flood

mation about the disaster, defined as the count of people posting about the disaster and their one hop neighborhood. Figures 5 and 6 show the *reach* and the number of users who have posted until that time. We see that the *reach* rises rapidly in the beginning and gradually saturates, indicating that new users are added rapidly but additions slow down eventually. More interestingly, the *reach* finally merges with the number of users, indicating the component closes on itself.

From the above observations we claim that disjoint communities on the social network quickly merge together into a giant connected component (Figures 1 and 2, Figures 4 and 3), but the component is poorly connected with the rest of the network (Figures 5 and 6).

3. GEOGRAPHICAL SPREAD OF CONCERN

We use the 'Brazil Flood' and the 'Japan Earthquake' data sets for analysis in this section. We extracted user location from their profile information on Twitter. Although a study [1] indicates that only 66% of Twitter users specify a valid location, we feel that the sample distribution is random and representative of actual user participation.

Figures 7 and 8 show the user distribution on a world map. We can see from the two figures that people concerned about the disaster are spread all across the world. Countries from Europe and North America, and Australia are common in both the disasters. There are a few differences among Asian countries though, for example, the users from China did not post about the flood in Brazil but did mention the Japan earthquake, which is quite interesting to observe.

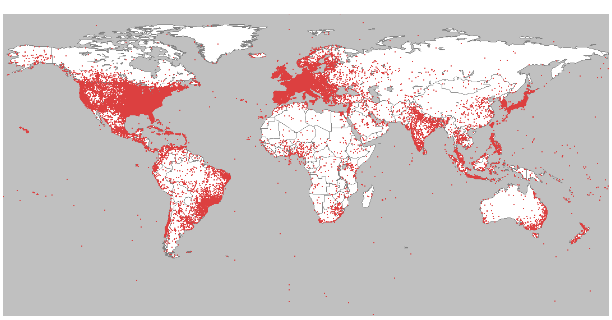

Figure 8: Users Distribution for Japan Earthquake

From the above observations, we can conclude that several countries and those not necessarily geographically proximate to the disaster region, do express concern on Twitter about the disasters.

4. RELATED WORK

In [4] Barabasi et al. study evolution of the collaboration network. In another paper, [7] Duncan et al. study evolution of the email network of a large university. They found that the aggregate network properties appear to approach an equilibrium state, whereas individual properties are unstable. In our work, we study the evolution not of the network itself, but of the component users interested in a particular event.

Significant research has been conducted on content analysis of information discussed on social media sites. Grinev et al. [6] demonstrate TweetSieve, a system that obtains news on any given subject by sifting through the Twitter stream. Along similar lines, Twinner by Abrol et al. [3] identifies news content of a query by taking into account the geographic location and the time of query. Go et al. [5] exploit the fact that people use emoticons to express their sentiment on events happening around them, and show that tweets that have positive or negative emoticons attached as labels are an effective way for distant supervised learning. We have not done detailed content analysis of the tweets, but have only considered tweets matching a specific topic. We do plan to do deeper analysis in the future.

Information diffusion on social media has also been studied in several contexts. Yang et al. [8] and [9] Zhao et al. have proposed information diffusion models for social media based on users influencing other users. We have not done any predictive work on the direction of information propagation and influence, but feel that our analysis can be utilized in several applications. For example, in the design of fund raising campaigns, the closing of the giant component on itself indicates that information may not naturally diffuse via social links to different parts of the social network, but may have to be artificially seeded in different parts of the graph.

5. CONCLUSIONS

In this paper, we investigate the dynamics of diffusion of information on social networks and the spread of concern in the world about natural disasters. We notice that although news is seeded in different disjoint parts of the network, it very quickly cumulates into a giant connected component that comprises more than 90% of the users tweeting about the disaster, and the component closes on itself by remaining disconnected with the rest of the Twitter network. We also examine the countries of the Tweeting users and find that interest spans the entire world.

6. REFERENCES

[1] http://asc-parc.blogspot.com/2011/01/further-details-on-location-field.html.

[2] Twitter rest api, https://dev.twitter.com/docs/api.

[3] S. Abrol and L. Khan. Twinner: Understanding news queries with geo-content using twitter. 2009.

[4] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. Apr. 2001.

[5] A. Go, R. Biyani, and L. Huang. Twitter sentiment classification using distant supervision. 2009.

[6] M. Grineva, A. Boldakov, L. Novak, and A. Syssoev. Tweetsieve: Sifting microblogging stream for events of user interest ? 2009.

[7] G. Kossinets and D. J. Watts. Empirical Analysis of an Evolving Social Network. *Science*, 311(5757):88–90, Jan. 2006.

[8] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *In Proceedings of the International Conference on Data Mining*, 2010.

[9] J. Zhao, J. Wu, and K. Xu. Weak ties: Subtle role of information diffusion in online social networks. *Phys. Rev. E*, 82(1):016105, Jul 2010.