Carnegie Mellon University

Spatial Dynamics of Twitter Data

How the 2011 Tōhoku earthquake affected Japanese the Social Network's patterns

Ramon Villa-Cox
rvillaco@andrew.cmu.edu

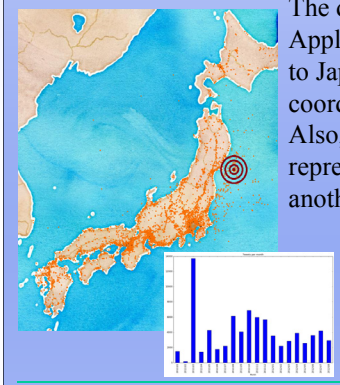
Kathleen Carley
kathleen.carley@cs.cmu.edu

INTRODUCTION:

Numerous studies have confirmed that the spatial dimension plays an important part not only in the topics discussed through social media, but also in the way that these discussions take place. However, little attention has been given to how these patterns change over time or in response to large-scale events. The present work seeks to address this issue, by identifying how user behavior patterns change during and after an extreme environmental event.

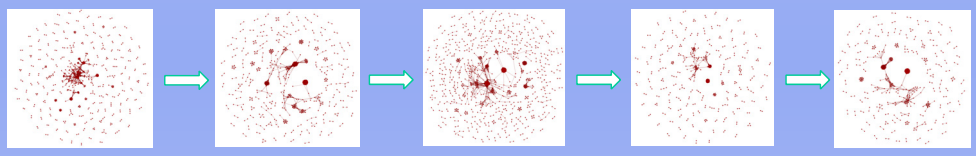
As a case study, we explore the Twitter usage pattern in Japan on 2011, the year where the Tōhoku earthquake struck the country. Though the analysis of 2 years of data collected for the country (2011-2012), we explore how different communication patterns, like the average geographical reach of users and their average movement between tweets, changed over time. Moreover, we explore changes in the relationship between the communication network of users and their spatial disposition in two time periods, the month of the event and the same month one year later.

THE DATA:



The data used in this study was collected from a 10% global feed accessed using the Twitter Streaming Application Programming Interface (API). From this global sample, georeferenced tweets corresponding to Japan were filtered. A origin coordinate was assigned to each user (based on average tweet coordinates), and a distance network was constructed.

Also, we constructed a communication network between users, where a link between two nodes represents whether one mentioned, replied, retweeted or shared a conversation topic (hashtag) with another.

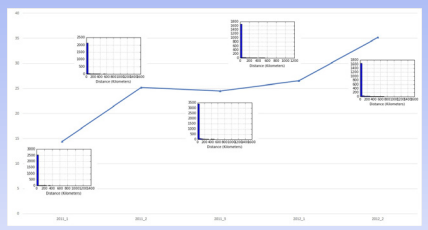


To assess the changes in the network structures, the data was binned into tertiles (3 for 2011 and 2 for 2012).

RESULTS:

To understand how the spatial communication patterns changed through the period studied, we analyzed 3 different dimensions of it.

Average user geographical reach:



Correlation between communication and distance network:

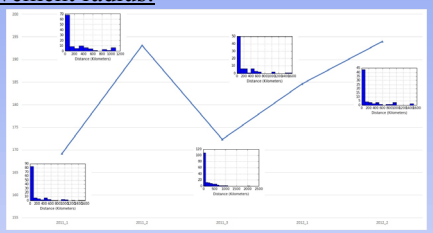
We compared the correlation of the communication network and the user's distance network through Quadratic Assignment Procedure (QAP), both for the different tertiles and to assess changes in user behavior in the month of the earthquake and the subsequent year.

	CORRELATION	P-VALUE
2011_1	-0.003	0.959
2011_2	-0.013	0.98
2011_3	-0.005	0.96
2012_1	-0.01	0.941
2012_2	0.005	0.957

	CORRELATION	P-VALUE
2011	0.063	9.6%
2012	-0.0006	50.1%

Results for the month of the event (11 March to 12 April) in the two sampled years.

User movement radius:



CONCLUSSIONS:

We found evidence that suggest that there is a distinct spatial behavior among users in the presence of extreme events. For example, in the period of the event, users tended to move less between tweets and the geographic reach of an average user was reduced. Finally, there is no significant correlation between the communication network and users' distance network in any of the periods studied, which suggest that, in average, closer users don't tend to interact more through this medium. This seems reasonable considering that these types of social media tend to increase the access of users to others who are farther away.

This work was supported in part by the Office of Naval Research Grant N000141612543 and the Center for Computational Analysis of Social and Organization Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, or the U.S. government.

institute for SOFTWARE RESEARCH