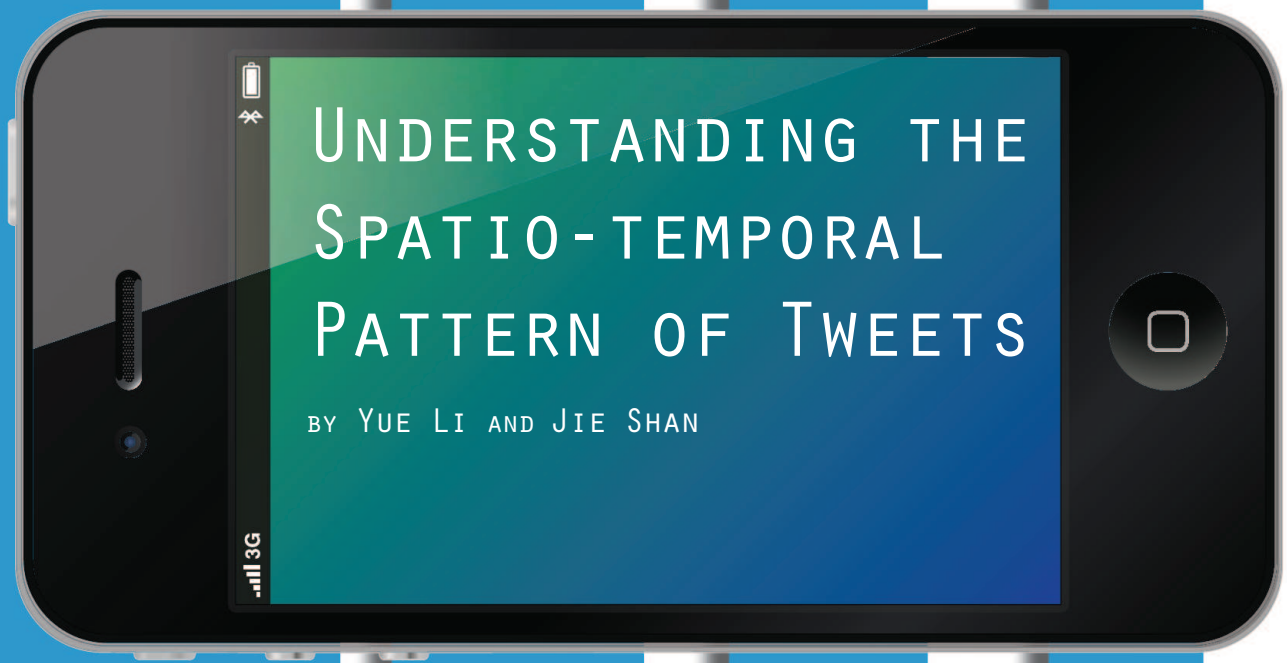

Understanding the Spatio-temporal Pattern of Tweets

by Yue Li and Jie Shan

Introduction

Recent IT technologies and online services are tremendously altering the way people create, use and share geographic information (Elwood, 2008). Enabled by the widespread use of hand-handled GPS, geo-tags, high-resolution graphics and access to the Internet and Web 2.0, spatial data can be collected and produced voluntarily by the untrained general public. This has been termed Geoweb, Volunteered Geographic Information, and neogeography (Goodchild, 2007; Crampton, 2008). Different from the traditional way exclusive to professionals in collecting spatial data, an individual can be seen as an intelligent sensor contributing to local geographic information (Goodchild, 2007). This innovation profoundly impacts the discipline of geography, as well as sociology and politics, by providing innovative solutions other than traditional methods such as focus groups, interviews and surveys (Elwood, 2008; Tsou and Leitner, 2013). The Geoweb, according to Crampton et al. (2013), is not only a simple collection of latitude-longitude coordinates attached with information, but a "socially produced space that blurs the oft-reproduced binary of virtual and material spaces".

continued on page 770

continued from page 769

Twitter, the most popular micro-blogging site, having over 500 million registered users as of 2012 and creating over 340 million tweets per day, has caught the attention of socio-geographic researchers (Lunden, 2012). One important feature about Twitter is its availability from mobile phones, which may have embedded location sensors such as GPS and thus allow users to send messages with their geographic coordinates (Fujisaka *et al*., 2010). With respect to the huge quantity and large diversity of crowds publishing tweets, massive valuable knowledge such as social geographic phenomena can be extracted (Fujisaka *et al*., 2010). Therefore, study of the spatio-temporal pattern of geo-tagged tweets provides important implications for various applications, such as human geography, urban science, location-based services, targeted advertising, content delivery networks, and social media research (Kamath *et al.,* 2013).

This article reviews recent socio-geographic research based on micro-blogging services such as Twitter. The spatio-temporal dynamics of geo-tagged tweets in West Lafayette, Indiana, the town of Purdue University, is presented. It is to demonstrate the feasibility of utilizing public location-based social network information, especially geo-tagged tweets, to understand the daily activities and life style of human beings, mostly college students in this case.

Literature Review

Analyzing micro-blog data such as tweets can help discover and understand human behavior patterns and personal lifestyle patterns. Fujisaka *et al.* (2010) analyzed mass movement histories using geo-tagged micro-blog datasets to explore the behavior patterns of individual regions. They proposed two models: an aggregation model to illustrate how many new users enter the region, and a dispersion model to calculate those leaving the region. By using these models, usual and unusual social and natural phenomena can be found, and anonymous crowd mining became feasible (Fujisaka *et al*., 2010). Based on this work, they could detect geo-social events, such as festivals, by comparing micro-blogging data with geographic regularities.

In addition to geo-tags, Hiruta *et al.* (2012) filtered out tweets with content irrelevant to the tagged location to detect events. Similarly, Nakaji and Yanai (2012) took advantage of the visual features of the attached photos to supplement the geo-tag information, and designed a visualization system for real-world events on the online map.

Instead of detecting events from massive geo-tagged tweets (Fujisaka *et al*., 2010; Nakaji and Yanai, 2012), Li *et al*. (2013) related the density of tweets with socioeconomic characteristics of local people to discover the spatial, temporal, and socioeconomic patterns. Similar research has been conducted on the footprints recorded by location sharing services such as Foursquare, Gowalla, and Facebook, where spatial, temporal, social and textural aspects were analyzed to quantitatively assess the human mobility pattern (Cheng et al, 2011).

Different from the above research, where tweets in a given area were collected and analyzed, several researchers focused on the spatial distribution of tweets related with a certain event or topic in a broader area. Crampton *et al*. (2013) focused on the manifestation of one certain event -- the widely reported riots after the University of Kentucky men's basketball team's 2012 championship. They developed a big data analytic engine providing geo-visualization functionality for geotagged tweets, and analyzed the geography of one specific hashtag #LexingtonPoliceScanner, which refers to the online feed of the Lexington Police Department, to assess the ability of using geo-referenced social media data to spatially determine events and the news diffusion over time and space (Crampton *et al*., 2013).

Similar to Crampton *et al.* (2013), Tsou *et al.* (2013) analyzed the spatial distribution of web pages and social media messages with respect to the 2012 Presidential Election. They converted thousands of web pages and millions of tweets related to the keywords "Barak Obama" or "Mitt Romney" into maps. The resulting landscape with different time and keywords were highly correlated to certain major campaign events, leading to the conclusion that this innovative approach is useful for quantitatively studying human activities, social events and human thoughts (Tsou *et al*., 2013).

Ghosh and Guha (2013) mapped the distribution of the tweets related to "obesity". They first extracted the main topics and keywords associated with "obesity" using topic modeling, and explored the spatial patterns of each extracted topic, which were further related to U.S. census data and locations of fast food restaurants. The research offered a prototype for social scientists, especially health geographers, to use large conversational datasets in studying health problems (Ghosh and Guha, 2013).

Motivation and Methodology

To further demonstrate the potential of using geotagged social media data in socio-geographic research, we conducted several analyses into the geo-tagged tweets within West Lafayette, Indiana, United States. It is the most densely populated city in Indiana with a population of 29,596 as of the 2010 census ("2010 Population Finder", 2011). Also, West Lafayette is the home of Purdue University, which has almost 40,000 students as of fall 2012 (Office of Institutional Research, 2013).

This study explores the spatio-temporal pattern of geo-tagged tweets in West Lafayette, and thus infers the human activity and mobility pattern in the area. Particularly, this analysis may shed light on the ac-

tivity patterns of Purdue students, since 93% of young adults (age 18-29) use the Internet, while 72% of them use social networking sites, as reported by a Pew Research Center study (2010).

A total of 4,160 geo-tagged tweets were collected from Twitter using the Twitter Streaming API for one entire week from Thursday, April 11 to Thursday, April 18, 2013. Each tweet is tagged with a latitude and longitude, which are used for the location of the user at the time of posting. Our study evaluates the spatial and temporal distribution of the geo-tagged tweets on weekdays and the weekend using clustering analysis in Esri Maps for Excel.

Results and Discussion

The geo-tagged tweets of weekdays and weekend days are found to have different geographic patterns. On weekdays, the tweets are concentrated on the Purdue campus and its surroundings as well as a few apartment clusters (Figure 1), whereas on weekends they are relatively more evenly distributed (Figure 2). This phenomenon suggests a flow of Twitter users in West Lafayette move from the Purdue campus to other parts of the city when the weekend comes. Furthermore, this distribution pattern corresponds to the fact that Purdue students attend school during weekdays, and go back home or leave town on weekends, which infers that one big group of Twitter users in West Lafayette are Purdue students, enhancing the feasibility of using Twitter data in student targeted marketing and municipal development.

The temporal pattern of the campus tweets is also significant. On weekdays, the count of geo-tagged tweets varies greatly by time, while there is no considerable variation for the weekends (Figure 3). The number of tweets on weekdays starts to rise at 7:00 am, which corresponds to the fact that the first classes on campus start at 7:30 am. It continues to grow during the day until 4:00 pm. This indicates students tweet a lot on campus between and/or during classes. The peak of the tweet count occurs around noon during the lunch time. It is noticed that the number of tweets starts to decline at 4:00 pm and continues until 7:00 pm, which is likely the time period when students go to the gym or jogging outside. Later the tweet count starts to rise until midnight, when most students are likely finishing the day and getting ready to go to sleep. Fewer and fewer tweets are observed after midnight, with the minimum count reached during 4:00 - 7:00 am. The results in Figure 3 demonstrate that the Twitter users in West Lafayette are most active from 1:00 – 7:00 pm and 10:00 pm – 1:00 am on weekdays. However, the number of tweets is quite stable over time on weekends since only two peaks are observed respectively at noon and 8:00 pm. The above temporal pattern sketches some interesting yet detailed weekly living patterns of college students.

Figures 4 and 5 plot the spatial distribution of tweets at different times during weekdays. The majority of the geo-tagged tweets from 11:00 am – noon on weekdays are clustered on the Purdue campus (Figure 4), whereas tweet clusters from 8:00 – 9:00 pm are away from the campus and rather spread out (Figure 5). Notably, tweets around local restaurants and bars start to emerge (Figure 5). This phenomenon suggests that Twitter users in the evening in West Lafayette move from the Purdue campus to other parts of the city.

continued on page 772

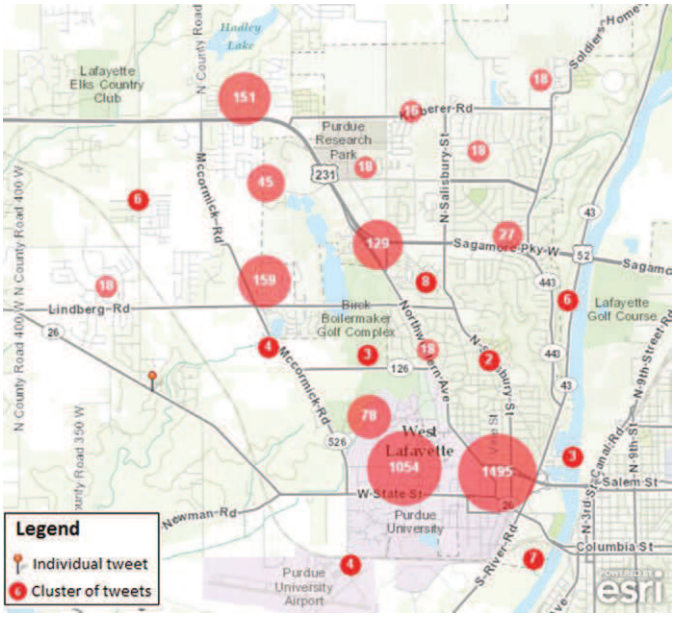

Figure 1. Geo-tagged tweet clusters on weekdays.

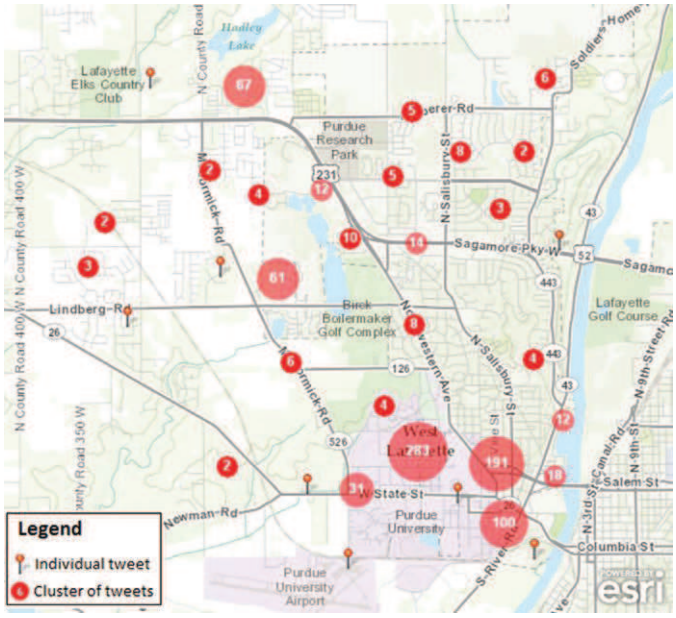

Figure 2. Geo-tagged tweet clusters on weekends.

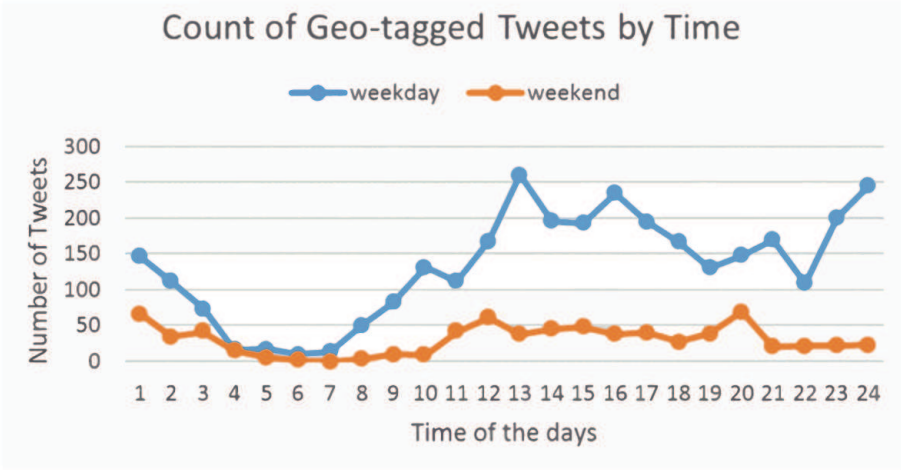

Figure 3. Count of geo-tagged tweets by time.

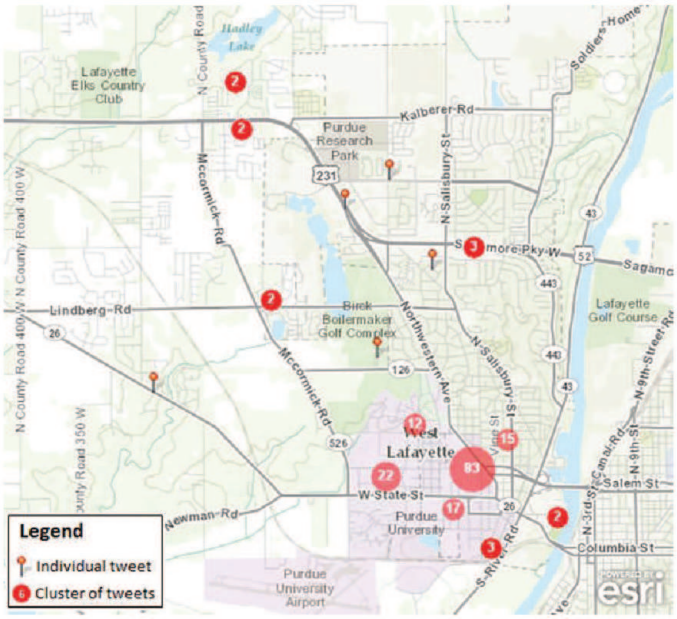

Figure 4. Geo-tagged tweet clusters from 11:00 AM–noon on weekdays.

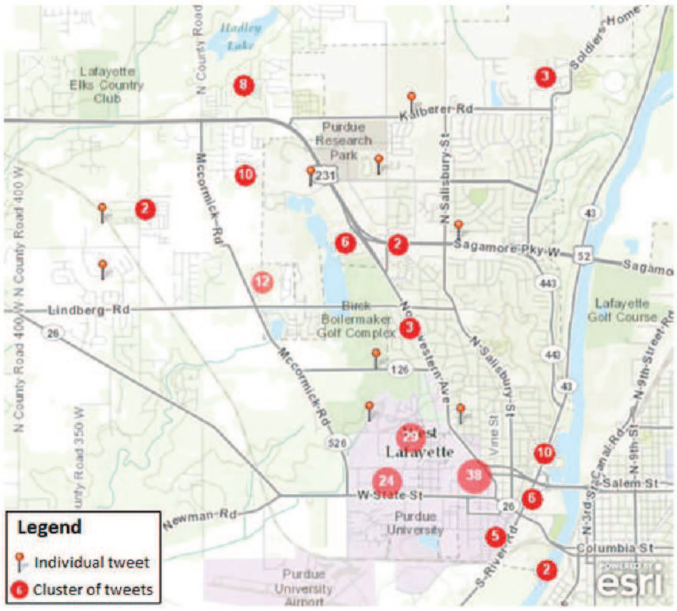

Figure 5. Geo-tagged tweet clusters from 8:00–9:00 PM on weekdays.

continued from page 771

Summary

Our study explored the potential of using geo-tagged micro-blogging service such as Twitter in socio-geographic research. We conducted a study on the geo-tagged tweets in West Lafayette and evaluated the user activity pattern behind the tweets by analyzing the spatio-temporal dynamics. This study suggests the possibility of using geo-tagged tweets in local market research and promotions, human mobility analysis, and even education regulation in a "college town" such as West Lafayette. It helps understand the living patterns and style of college students. Future work will use larger data sets, consider the content of the micro-blogs by semantic analysis, topic modeling, and content analysis, aiming to track the spread of ideas and thoughts in the local area. We will develop a framework of extracting spatio-temporal social patterns from geo-tagged tweets on a city scale to help social researchers, demographic surveyors, market researchers, advertising designers, and policy makers.

References

2010 Population Finder, 2010. United States Census Bureau, retrieved July 31st, 2013 from http://www.census.gov/popfinder/?fl=18.

Cheng, Z., J. Caverlee, K. Lee, & D. Z. Sui, 2011. Exploring millions of footprints in location sharing services, ICWSM, 2011, 81–88.

Crampton, J., 2008. Cartography: Maps 2.0, Progress in Human Geography, 33(1): 91–100.

Crampton, J. W., M. Graham, A. Poorthuis, T. Shelton, M. Stephens, M.W. Wilson, M. Zook, 2013. Beyond the geotag: Situating 'big data' and leveraging the potential of the geoweb, Cartography and Geographic Information Science, 40(2), 130–139.

Elwood, S., 2008. Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS, GeoJournal, 72, 173–183.

Fujisaka, T., R. Lee, K. Sumiya, 2010. Discovery of user behavior patterns from geo-tagged micro-blogs, ICUIMC' 10 Proceedings of the 4th International Conference on Uniquitous Information Management and Communication, January 14-15, 2010, Suwon, Republic of Korea.

Ghosh, D., and R. Guha, 2013. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System, Cartography and Geographic Information Science, 40(2), 90–102.

Goodchild, M. F., 2007. Citizens as sensors: The world of volunteered geography, GeoJournal, 69, 211–221.

Hiruta, S., T. Yonezawa, M. Jurmu, & H. Tokuda, 2012. Detection, classification and visualization of place-triggered geotagged tweets, In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (pp. 956–963), New York, NY, USA: ACM. doi:10.1145/2370216.2370427.

Kamath, K. Y., J. Caverlee, K. Lee, Z. Cheng, 2013. Spatio-Temporal Dynamics of Online Memes: A Study of Geo-Tagged Tweets, The International World Wide Web Conference, Rio de Janeiro, Brazil.

Lee, R., K. Sumiya, 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection, *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, November 02-02, San Jose, California [doi>10.1145/1867699.1867701].

Lenhart, A., K. Purcell, A. Smith, & K. Zickuhr, 2010. Social media & mobile internet use among teens and young adults, *Pew Internet & American Life Project*.

Li, L., M. F. Goodchild, B. Xu, 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr, *Cartography and Geographic Information Science*, 40(2), 61–77.

Lunden, I., 2012. Analyst: twitter passes 500m users in June 2012, 140m of them in US; Jakarta 'biggest tweeting city. Retrieved July 29, 2013 from http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/.

Nakaji, Y., and K. Yanai, 2012. Visualization of real-world events with geotagged tweet photos, In 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) (pp. 272–277), Presented at the 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). doi:10.1109/ICMEW.2012.53.

Purdue University Data Digest, 2012. Retrieved August 1, 2013 from http://www.purdue.edu/datadigest/fastfacts/pg1.html.

Tsou, M., J. Yang, D. Lusher, S. Han, B. Spitzberg, J.M. Gawron, D. Gupta, L. An, 2013. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): A case study in 2012 US Presidential Election, *Cartography and Geographic Information Science*, DOI:10.1080/15230406.2013.79973.

Authors

Yue Li
li1050@purdue.edu

Jie Shan
jshan@purdue.edu

School of Civil Engineering, Purdue University
West Lafayette, IN 47907, USA