Uncovering Hidden Behavioral Patterns in the Era of "We Media": Modeling Spatio-Temporal Dynamics for Twitter News

Hong Huang[1], Han Yu[2], James E. Andrews[1], JungWon Yoon[1], Kelsey L. Burgess[1]
[1]School of Information, University of South Florida, Tampa, FL, USA.
[2]Farmer School of Business, Miami University, Oxford, OH, USA.

Abstract
This research presents a Bayesian statistical model to examine spatio-temporal effects for Twitter use when reporting important events or news. The proposed model tests the Twitter News data surrounding the United States Supreme Court's *Myriad Genetics, Inc.* June 13, 2013 decision and its impact on direct-to-consumer genetic testing and gene patenting. The model demonstrates the sensitivity in distinguishing the behaviours of Twitter users' followers with and without adjusting spatio-temporal effects. It was also found that media professionals' tweets were coming thick and quick, and producing "waves" of engagement of followers. However, grassroots actively participate in tweeting and constantly engage more followers. The model maps tweets across the spatial heterogeneity and temporal evolution in the early and post recognition and discussion of events. These findings demonstrate the importance of spatio-temporal effects to influence professionals or non-professionals for tweeting. The model also guided researchers to detect sub-events with low latency.
Keywords: Twitter news; genetic test; gene patent; Twitter geolocations; Twitter timestamp
Citation: Huang, H., Yu, H., Andrews, J. E., Yoon, Y., & Burgess, K. L. (2017). Uncovering Hidden Behavioral Patterns in the Era of "We Media": Modeling Spatio-Temporal Dynamics for Twitter News. In *iConference 2017 Proceedings* (pp. 704-707). https://doi.org/10.9776/17323

Contact: honghuang@usf.edu

1 Introduction

Over the past several years, Twitter has become one of the most influential social media platforms. Users post "tweets" for a variety of reasons including but not limited to news reporting, calls to action, conversations, information seeking and sharing, along with personal notes like anecdotes and opinions (Honeycutt & Herring, 2009).

Twitter provides constant sentence-length updates that accumulate streams of information in real-time (Hermida 2012). Professional media writers, reporters, and non-professionals can make timely Twitter posts that contain well documented data about timestamps and geographical annotations. The number of followers indicated Twitter user influences (Bakshy et al., 2011). Twitter users such as media professionals and/or grassroots might gain or loss followers differently for tweeting with time and location changes.

Researchers therefore can analyze Twitter streams with new statistical models, which take the unavoidable realistic spatio-temporal effects into consideration for understanding user behaviors, as well as establishing different typologies for Twitter users. The analysis is based on the new model and could add values to data mining on longitudinal and/or spatio-temporal data via a Bayesian nonparametric approach. This approach aims to more substantially map use across the spatial heterogeneity and temporal evolution in a developing event.

In this study, we proposed a non-parametric hierarchical model formulated as a Bayesian inference problem for nonparametric estimation and model selection. The proposed model recognized the effects of spatio-temporal correlations and interactions for Twitter use, then had its theoretical and practical advantages. We used a case of the Twitter data for reporting an important event to demonstrate the model

and related analysis from June 12, 2013, following the Supreme Court's decision to block *Myriad Genetics* from patenting the BRCA gene. Twitter users voiced their thoughts and opinions on this historic moment. On the day of the decision, Tweets mentioning "BRCA" surged as users widely discussed the matter. Various related topics were actively discussed subsequently on Twitter as well.

2 Methodology

In this study, we distinguish observations by space and time. Hierarchical structured additive models provide a very broad and rich framework for complex regression analysis (Fahrmeir et al., 2004; Brezger and Lang, 2006). The models can deal simultaneously with nonlinear time trends, spatial heterogeneity, and unit- or cluster-specific heterogeneity as well as nonlinear covariate effects and the complex interactions between covariates of a different type that inherent our Twitter usage spatio-temporal data. Conceptually, the observations of the number of followers for Twitter i on time t (in second) at location s (in longitude and latitude),yi (t, s), $i = 1, \ldots, n$, are assumed to be a conditionally independent random sample, from a given normal distribution given covariates. The conditional mean μi(t, s) = Eyi(t, s) is linked to a semiparametric additive predictor.

The linear predictor is expressed additively as the sum of the effects of some components that can be interpreted as individual and independent contributions to the number of followers in that region and period:

$$u_i(t,s) = \beta_o + \beta_{media} f(t) + f_{spat}(s_i) + \alpha_i$$

$f(t)$ is a smooth nonlinear time trend, β_{media} is the varying effect of media type, $f_{spat}(s_i)$ is the spatial effects for spatial heterogeneity and α_i is the random effects of Twitter i for individual specific heterogeneity. The intercept term β_o gives a starting number of followers that is shared by all locations, periods, and all types of media. The nonparametric estimation is based on full Bayesian inference via Markov Chain Monte Carlo (MCMC) simulation techniques.

All historical Twitter data was collected from June 12, 2013 to July 2, 2013. Twitter search API was used to test data with the searched key word: "BRCA". The social media monitoring company Simplymeasured (http://simplymeasured.com) was used to harvest the Twitter streams.

3 Preliminary Results

There were a total of 5,649 Twitter posts related to "BRCA" genetic testing, reaching out to nearly 20 million users. These posts were harvested from June 12, 2013 to July 2, 2013. Within one day following the Supreme Court's decision on June 13, 2013, a spike was observed with a data deluge of 2,386 Twitter posts reaching more than 11 million users (Fig 1).

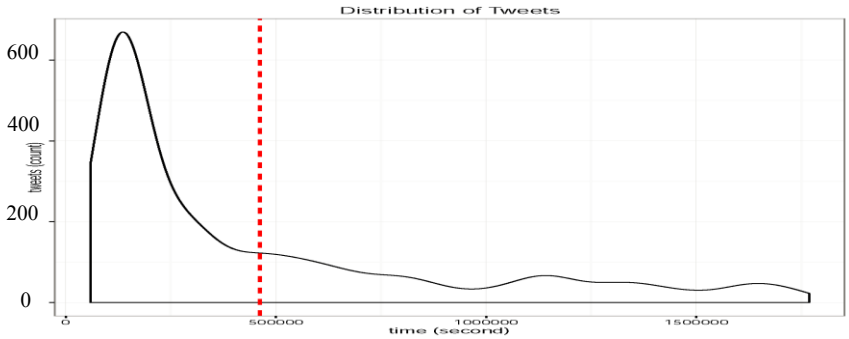

Fig 1. Twitter volumes for "BRCA" genetic testing throughout the time

The model found that without the adjustment of the spatio-temporal effects, the trend of the average number of followers for each twitter user declined throughout the time. However, with the adjustment of spatial and temporal effects, the overall trend of the number of followers increased with fluctuations (Fig 2). Such an observation is similar with the phenomena of the "Simpson paradox" (Wagner, 1985), in which a trend appears in data but disappears or reverses when considering new combinational effects.

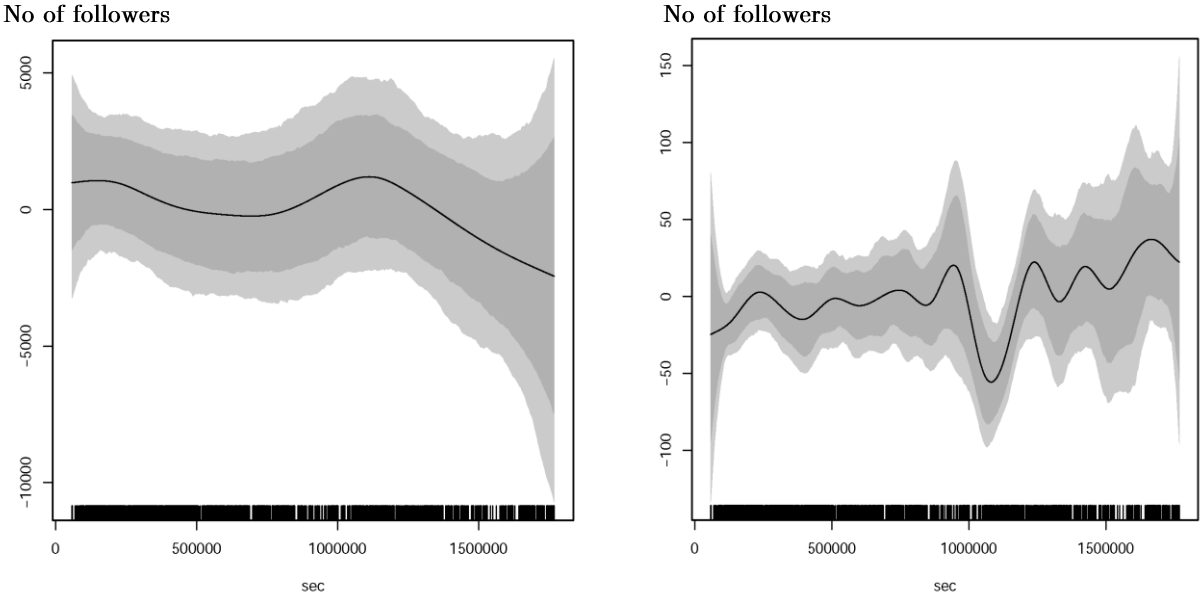

Fig 2. The number of followers for a Twitter user throughout the time, with the unadjusted spatio-temporal effects (left), and adjusted ones (right). Note: The shaped band represents the range of possible values, and the line represents the average value at a time period.

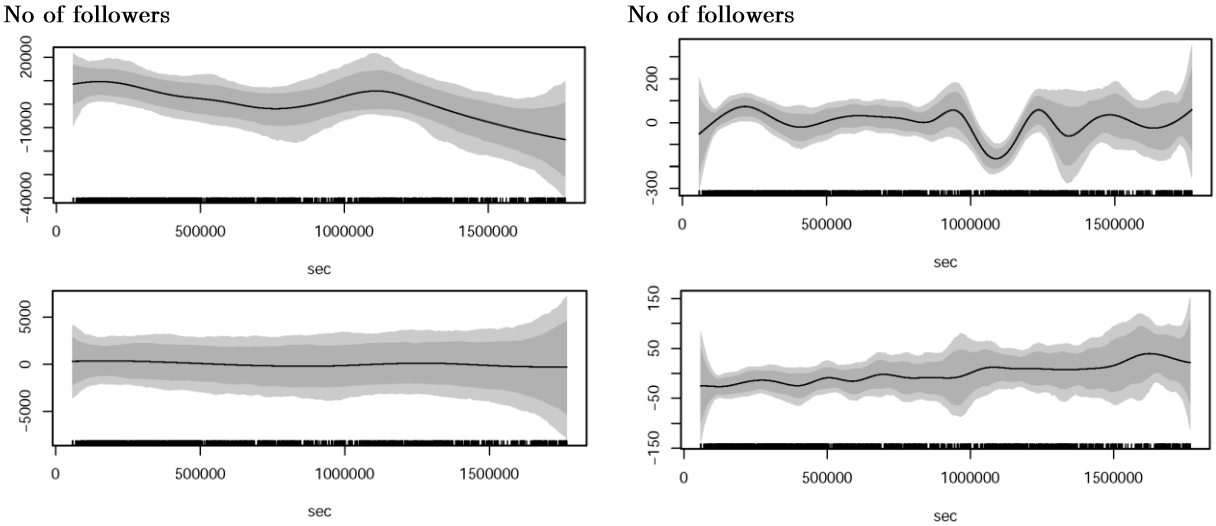

Fig 3. The change of number of followers throughout the time for professional media reporters (top) and nonprofessional (down) with spatio-temporal un-adjustment (left) and adjustment (right).

Media professionals such as journalists, bloggers and media workers, along with non-professionals share and exchange information on the Twitter platform. Without considering spatio-temporal effects, the overall trend of the number of followers for media professionals showed a decline along with the time passed, but for the number of non-professionals' followers remained the same. Interestingly, with the adjustment of spatio-temporal effects, the number of followers for professionals showed "waves" of decreases and increases; the overall pattern for number of followers showed a progressive increase for non-professionals (Fig 3).

4 Conclusion and Future Work

Hermida in 2012 indicated, "Twitter provides a mix of news, information and comments, usually connected to current reality, but without an established order." Spatio-temporal issues play critical roles for people in examining the tweeting behaviors. The proposed Bayesian statistical model can be used for discovering the patterns of tweeting behaviours in the real world, taking full consideration of spatio-temporal effects.

In the era of "We Media" (Gillmor, 2006), everyone is a news distributor. Media professionals' tweets for reporting news are coming thick and fast, creating new "waves" to engage users. Twitter grassroots actively participate in the event reporting and make a continuous progress in engaging more users. Further studies will be conducted to test more Twitter data sets with the proposed model. The model also helps detect sub-events with low latency which indicated by the changes of the number of Twitter followers, and explore fine-grained user behaviors for Twitter use when reporting news. It may help plan campaigns for promoting the events and identify user concerns and interests when using Twitter.

5 References

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74). ACM.

Brezger, A., Lang, S. (2006): Generalized structured additive regression based on Bayesian P-splines. Comput. Stat. Data Anal. 50, 967–991.

Fahrmeir, L., Kneib, T., Lang, S. (2004): Penalized structured additive regression for space-time data: a Bayesian perspective. Stat. Sin. 14, 731–761.

Gillmor, D. (2006). *We the media: Grassroots journalism by the people, for the people*. " O'Reilly Media, Inc.".

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the US Congress. *Journal of the American Society for Information Science and Technology*, *61*(8), 1612-1621.

Hermida, A. (2012). Tweets and truth: Journalism as a discipline of collaborative verification. *Journalism Practice*, *6*(5-6), 659-668.

Honeycutt, C., & Herring, S. C. (2009). Beyond microblogging : conversation and collaboration via Twitter. *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* (pp. 1–10). IEEE.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, *60*(11), 2169-2188.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, *62*(2), 406-418.

Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician*, *36*(1), 46-48.