Analyzing Spatio-temporal Patterns of Online Public Attentions in Emergency Events: A Case Study of 2009 H1N1 Influenza Outbreak in China

Kainan Cui[1,2], Xiaolong Zheng[2], Zhu Zhang[2], and Daniel Zeng[2]

[1] The School of Electronic and Information Engineering,
Xi'an Jiaotong University, Xi'an, China
[2] State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
kainan.cui@live.cn, xiaolong.zheng@ia.ac.cn

Abstract. Understanding the public attention and perception towards epidemics is critical for public health response. However, the research question concerning the spatio-temporal patterns of public attention and the interactions with media attention and severity of epidemic is still not well studied. Aim to fill this research gap, we chose the H1N1 influenza outbreak in the mainland of China in 2009 as case to study the spatio-temporal patterns of public attention, and their correlations with media attention and severity of epidemic. The results of this paper indicate that public attention and media attention had high correlation from both temporal and spatial perspectives, which can provide us significant insights to understand the collective behavior of massive online users during emergency events.

Keywords: Information diffusion, public attenion, emergency response, influenza, epidemic outbreaks.

1 Introduction

The development of crisis situations was significantly affected by the rapid diffusion of information and opinions through Internet[1][2][3][4][5]. As recent studies suggested that the general public's attention and perception regarding epidemics also plays important role in the evolution process of epidemics[6], studying temporal and geospatial patterns of online public attention can uncover useful mechanism of information diffusion during emergency event to facilitate public health response and possibly prevention measures.

Many researches show that open source information has great application potential for improving the situational awareness during emergency situation. There are already lots of attempts and applications using different online data, such as search engine logs[7][8][9], news[10][11], blogs [12], micro-blog [13] and wiki [14], to predict the number of infections, track the transmission rate and study the response behavior for influenza outbreak. However, the question concerning the temporal and spatial

X. Zheng et al. (Eds.): ICSH 2014, LNCS 8549, pp. 44–50, 2014.

patterns of public attention and the interaction with media attention and severity of epidemic remain largely unexplored.

In order to fill this important research gap and provide implications for decision making during emergency, we chose the H1N1 influenza outbreak in the mainland of China in 2009 as case in this paper. In particular, we performed time series analysis on the online news and corresponding comments at first. Then we performed spatial analysis to study the spatial pattern and evolution process. Our results indicate that public attention towards emergency event had high correlation and similar temporal, geospatial patterns with media attention. The correlation between public attention and severity of epidemic was not stable which needs further study.

The rest of this paper is organized as follows. In Section 2, we begin with a brief introduction to the dataset we used in this study. We report our empirical analysis result and statistical results in Section 3. We conclude our paper by summarizing the findings and discussing several key issues in Section 4.

2 Data and Methods

2.1 Dataset

We chose the H1N1 influenza outbreak in the mainland of China in 2009 as the case to study the public attention during emergency events. Our dataset is composed by three parts, which are news data, comment data and epidemic data. We obtained news data (referred as SINA-ND) and comment data (referred as SINA-CD) from Sina news (news.sina.com), which is a famous online media offering a full array of Chinese-language reports and information.

Sina news provided a special repost regarding H1N1 outbreak, which is available at http://news.sina.com.cn/z/zhuliugan/index.shtml. We developed a customized crawler for collecting data. After removing duplicated data, there are 3303 news covering the period from April 26, 2009 to August 10 2010 and 75878 comments covering the period from May 27, 2009 to August 11, 2010.

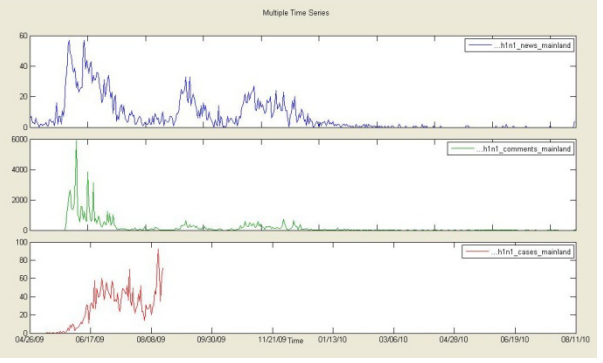


Fig. 1. The comment number distribution of messages by different genes

The epidemic data (referred as CH-ED) used in our study is obtained through an authority website, which serves 31 province-level regions in the mainland of China and covers the period from May 11, 2009 to July 9, 2009. Fig. 1 shows the time series plot of SINA-ND, SINA-CD and CH-ED.

2.2 Time Series Analysis Methods

Periodogram [15] (also known as spectral plot) is applied to assess the periodicity of online news and corresponding comments. A peak in the periodogram indicates there is a periodic component near the value corresponding to the peak. To extract the amplitude of a periodic component data, ideal pass filter is applied for the periodogram to filter the raw data, with the frequency interval covering the peak shape in the periodogramls.

To examine the associations between time series of news and comments, news and cases, comments and cases, cross-correlation is used to measure the degree of the linear relationship between two time series with limited time lags (ten days in this study). A specific lag corresponding to a high correlation of two time series may indicate that these two series have certain same influencing factors and a time delay.

2.3 Geospatial Analysis Methods

We chose the Pearson's correlation coefficient to measure the spatial correlation between two geographic distributions. In particular, we computed the correlation value of SINA-ND and SINA-CD, SINA-ND and CH-ED, SINA-CD and CH-ED respectively. We defined $P < 0.01$ as statistically significant. After the correlation analyses, several geospatial distribution plots were generated for visual comparison on geospatial patterns.

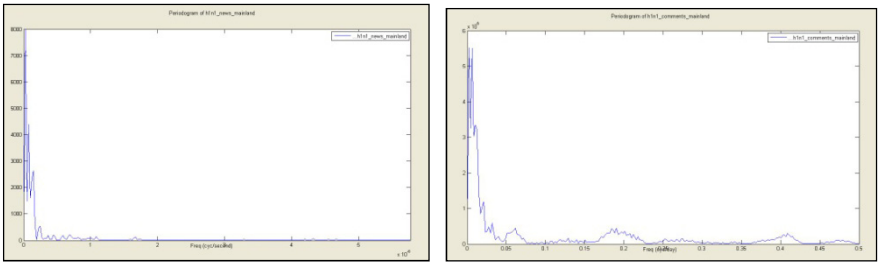
Global Moran's I[16][17] is selected to quantify the geospatial correlation within one geographic distribution. In particular, we computed the association value of SINA-ND, SINA-CD and CH-ED respectively. In order to gain more understanding of the change trends of the geospatial correlation between geographic distributions and the geospatial association within geographic distribution, we select two time snapshot, which are 2009-7-19 and 2009-12-12, for each data.

3 Results

3.1 Time Series Analysis

Fig. 2 shows the periodograms of the SINA-ND and SINA-CD. The main periodic components for SINA-ND and SINA-CD were shown in Table 1., we can observe that the online news and online comments had similar periodic patterns. In particular, except that the periodic component 3 of SINA-CD was obviously longer than the periodic component 3 of SINA-ND, the rest periodic components of SINA-CD and corresponding periodic components of SINA-ND had almost the same period length.

Regarding the amplitude of periodic component, the comments were greater than the news, which means the daily post number of comments is more volatile than daily post number of news,



(a) (b)

Fig. 2. The periodograms of the time series of SINA-ND (a) and SINA-CD (b)

Table 1. A comparison of periodic components for SINA-ND and SINA-CD

	Salient peak	Period(day)	Amplitude
SINA-ND (news)			
periodic component 1	0.002123	471.0094	16
periodic component 2	0.00637	156.9859	12
periodic component 3	0.0127	78.74016	10
periodic component 4	0.0212	47.16981	4
SINA-CD (comments)			
periodic component 1	0.002118	472.1435	440
periodic component 2	0.006356	157.3341	420
periodic component 3	0.0106	94.33962	320
periodic component 4	0.0212	47.16981	200

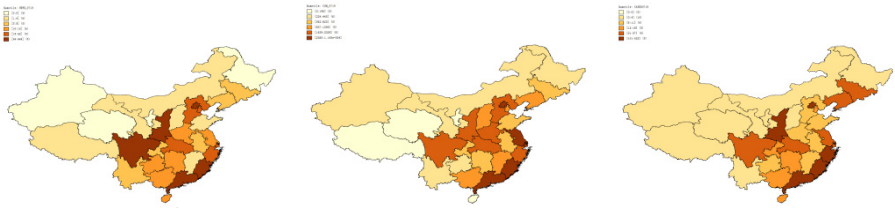
The correlation results and the corresponding time lag were shown in Table 2. These lags are selected to provide the highest correlation for each pair. From Table 2, we can observe significant positive correlation between the news and comments. The correlation degree between news and cases was larger than comments and cases.

Table 2. Cross-correlation results among SINA-ND, SINA-CD and CH-ED

	Time lag	Highest correlation
SINA-ND and CH-ED	-10	0.21
SINA-ND and SINA-CD	-1	0.715
SINA-CD and CH-ED	-10	0.125

3.2 Geospatial Analysis

Fig. 3 shows that the three datasets have a similar geospatial distribution. From a qualitative perspective, SINA-ND and SIND-CD had a high correlation of 0.9488, SINA-ND and CH-ED had a high correlation of 0. 872672; SINC-CD and CH-ED had a high correlation of 0.805023.

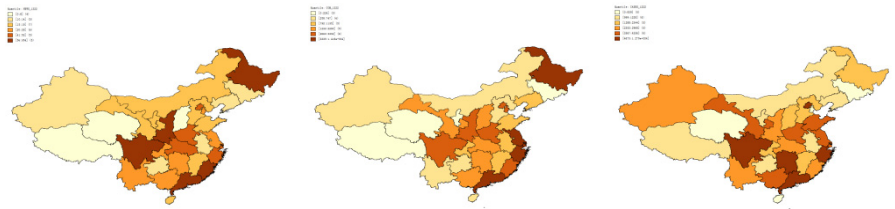


(a) (b) (c)

Fig. 3. Geospatial distribution before July 192, 2009 of SINA-ND (a), SIND-CD (b) and CH-ED (c)

Fig. 4 shows that the SINA-ND and SINA-CD had a similar geospatial distribution. From a qualitative perspective, SINA-ND and SIND-CD had a high correlation of 0.962354 while SINA-ND and CH-ED had a low correlation of 0.1870, SINC-CD and CH-ED had a low correlation of 0.3363.

The statistics of global Moran's I are summarized in Table 3. From this table, we obtain the following observations 1) the offline data (CH-ED-0719 and CH-ED-1222) had higher geospatial association than online data (SINA-ND and SINA-CD) 2) the geospatial association degree for both offline data and online increased over time. 3) The geospatial association degree of online data grew faster than offline data.



(a) (b) (c)

Fig. 4. Geospatial distribution before December 12, 2009 of SINA-ND (a), SIND-CD (b) and CHINA-ED (c)

Table 3. A comparison for the geospatial association of SINA-ND, SINA-CD and CH-ED in different time snapshot

	Moran's I	999 MC simulation		
		E(i)	S_d	Z-score
SINA-ND-0719	0.279	-0.0323	0.826	0.3236
SINA-CD-0719	0.125	-0.0323	0.0863	1.0613
CH-ED-0719	0.439	-0.0323	0.0933	0.0436
SINA-ND-1222	0.313	-0.0323	0.0767	-0.527
SINA-CD-1222	0.421	-0.0323	0.0927	-0.2845
CH-ED-1222	0.444	-0.0323	0.1059	0.0327

4 Conclusions

In this paper we attempt to show the patterns of public attention during epidemic outbreaks from both temporal and spatial perspectives. In particular, we chose H1N1 outbreaks in China mainland as case to examine the periodicity pattern and geospatial association of the media attention (online news) and the public attention (online comments). We also provided temporal and spatial correlation analysis results among media attention, public attention and severity of epidemic.

The results of this study show that the public attention and the media attention had high correlation from both temporal and spatial perspectives, which was agreed with the agenda setting theory. Furthermore the public attention has similar periodicity pattern with media attention, and swung more drastic. This phenomenon could be explained by the herd behavior.

We did not observe stable geospatial correlation between the severity of epidemic and the public attention. The geospatial association of public attention and media attention had increased over time. Further investigation is needed to uncover the mechanism under these empirical results.

One of the limitations in this study is that H1N1 is a brand-new disease affecting the globe which arise great attention from both online media and Web users. However for other diseases or events which Web users concerned not very much, the patterns and features of public attention may vary significantly. However, this work still shed light on the temporal and spatial patterns of public attention during emergency events. Further empirical work is needed to verify our findings.

Acknowledgments. The authors would like to thank Zhidong Cao, Saike He, Chuan Luo for their helpful suggestions. This work was supported in part by the following grants: The National Natural Science Foundation of China, Grant No. 71103180, 91124001, 71025001,and 91024030, by the Early Career Development Award of SKLMCCS and by the Ministry of Health under Grant No. 2012ZX10004801.

References

[1] Zeng, D., Chen, H., Lusch, R., Li, S.-H.: Social Media Analytics and Intelligence. Intell.Syst. IEEE 25, 13–16 (2010)

[2] Zheng, X., Zhong, Y., Zeng, D., Wang, F.-Y.: Social influence and spread dynamics in social networks. Front. Comput. Sci. 6(5), 611–620 (2012)

[3] Wang, Y., Zeng, D., Zheng, X., Wang, F.: Propagation of online news: Dynamic patterns. In: IEEE International Conference on Intelligence and Security Informatics, ISI 2009, pp. 257–259 (2009)

[4] Wang, Y., Zeng, D., Zhu, B., Zheng, X., Wang, F.: Patterns of news dissemination through online news media: A case study in China. Inf. Syst. Front., 1–14 (July 2012)

[5] Zheng, X., Zeng, D., Wang, F.-Y.: Social balance in signed networks. Inf. Syst. Front., 1–19 (January 2014)

[6] SteelFisher, G.K., Blendon, R.J., Bekheit, M.M., Lubell, K.: The Public's Response to the 2009 H1N1 Influenza Pandemic. N. Engl. J. Med. 362(22), e65 (2010)

[7] Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature 457(7232), 1012–1014 (2009)

[8] Wilson, K., Brownstein, J.S.: Early detection of disease outbreaks using the Internet. Can. Med. Assoc. J. 180(8), 829–831 (2009)

[9] Luo, Y., Zeng, D., Cao, Z., Zheng, X., Wang, Y., Wang, Q., Zhao, H.: Using multi-source web data for epidemic surveillance: A case study of the 2009 Influenza A (H1N1) pandemic in Beijing. In: 2010 IEEE International Conference on Service Operations and Logistics and Informatics (SOLI), pp. 76–81 (2010)

[10] Collier, N., Doan, S., Kawazoe, A., Goodwin, R.M., Conway, M., Tateno, Y., Ngo, Q.-H., Dien, D., Kawtrakul, A., Takeuchi, K., Shigematsu, M., Taniguchi, K.: BioCaster: detecting public health rumors with a Web-based text mining system. Bioinformatics 24(24), 2940–2941 (2008)

[11] Cui, K., Cao, Z., Zheng, X., Zeng, D., Zeng, K., Zheng, M.: A Geospatial Analysis on the Potential Value of News Comments in Infectious Disease Surveillance. In: Chau, M., Wang, G.A., Zheng, X., Chen, H., Zeng, D., Mao, W. (eds.) PAISI 2011. LNCS, vol. 6749, pp. 85–93. Springer, Heidelberg (2011)

[12] Corley, C.D., Cook, D.J., Mikler, A.R., Singh, K.P.: Text and Structural Data Mining of Influenza Mentions in Web and Social Media. Int. J. Environ. Res. Public. Health 7(2), 596–615 (2010)

[13] Culotta, A.: Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. In: Proceedings of the First Workshop on Social Media Analytics, New York, NY, USA, pp. 115–122 (2010)

[14] Laurent, M., Vickers, T.J.: Seeking Health Information Online: Does Wikipedia Matter? J. Am. Med. Inform. Assoc. 16(4), 471–479 (2009)

[15] Zhang, Z., Zheng, X., Zeng, D.D., Cui, K., Luo, C., He, S., Leischow, S.: Discovering seasonal patterns of smoking behavior using online search information. In: 2013 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 371–373 (2013)

[16] Moran, P.: The Interpretation of Statistical Maps. J. R. Stat. Soc. Ser. B-Stat. Methodol. 10(2), 243–251 (1948)

[17] Cao, Z., Zeng, D., Zheng, X., Wang, Q., Wang, F., Wang, J., Wang, X.: Spatio-temporal evolution of Beijing 2003 SARS epidemic. Sci. China Earth Sci. 53(7), 1017–1028 (2010)