Information diffusion on communication networks based on Big Data analysis

Zhou, Nan; Zhan, Xiuxiu; Lin, Song ; Yang, Shang-Hui; Liu , Chuang ; Sun, Gui-Quan; Zhang , Zi-Ke

DOI
10.1108/EL-09-2016-0194

Publication date
2017

Document Version
Final published version

Published in
The Electronic Library: the international journal for the application of technology in information environments

Citation (APA)
Zhou, N., Zhan, X., Lin, S., Yang, S.-H., Liu , C., Sun, G.-Q., & Zhang , Z.-K. (2017). Information diffusion on communication networks based on Big Data analysis. *The Electronic Library: the international journal for the application of technology in information environments*, *35*(4), 745-757. https://doi.org/10.1108/EL-09-2016-0194

Information diffusion on communication networks based on Big Data analysis

Received 27 September 2016
Revised 15 April 2017
Accepted 15 April 2017

Nan Zhou
Hangzhou Normal University, Hangzhou, China
Xiu-Xiu Zhan
Delft University of Technology, Delft, The Netherlands
Song Lin, Shang-Hui Yang and Chuang Liu
Hangzhou Normal University, Hangzhou, China
Gui-Quan Sun
Shanxi University, Taiyuan, China, and
Zi-Ke Zhang
Hangzhou Normal University, Hangzhou, China

Abstract

Purpose – Information carriers (including mass media and We-Media) play important roles in information diffusion on social networks. The purpose of this paper is to investigate changes in the dissemination of information combing with data analysis.

Design/methodology/approach – This work analyzed nearly 200 years of coverage of different information carriers during different periods of human society, from the period of only mouth-to-mouth communication to the period of modern society. Information diffusion models are built to illustrate how the information dynamic changes with time and combined box office data of several movies to predict the process of information diffusion. In addition, a metric is defined to identify which information would become news in the future.

Findings – Results show that with the development of information carriers, information spreads faster and wider nowadays. The correctness of the metric proposed has been validated.

Research limitations/implications – The structure of social networks influences the dissemination of information. There are an enormous number of factors that influence the formation of hotspots.

Practical implications – The results and conclusion of this work will benefit by predicting the evolution of information carriers. The metric proposed will aid in searching hot news in the future.

Originality/value – This work may shed some light on a better understanding of information diffusion, spreading not only on social networks but also on the carriers used for the information spreading.

Keywords Big data, Communication networks, Information carriers, Information diffusion

Paper type Research paper

This work was partially supported by Natural Science Foundation of China (Grant Nos 61673151 and 11671241), Zhejiang Provincial Natural Science Foundation of China (Grant No. LY14A05001), Zhejiang Qianjiang Talents Project (QJC1302001), Zhejiang Xinmiao Talents Program (2016R423069) and the EU FP7 Grant 611272 (project GROWTHCOM).

The Electronic Library
Vol. 35 No. 4, 2017
pp. 745-757

0264-0473
DOI 10.1108/EL-09-2016-0194

Introduction

Network science has received increasing attention during the past decade due to a considerable number of network structures emerging from real-world paradigms, which can be of great help to facilitate both research and applications of transportation optimization (Mandl, 1980; Veluscek *et al.*, 2015; Zhao *et al.*, 2015), community detection (Fortunato, 2010; Fortunato and Barthelemy, 2007; Hastings, 2006; Lancichinetti *et al.*, 2008; Leskovec *et al.*, 2010; Papadopoulos *et al.*, 2012) and data mining (Chen *et al.*, 1996; Fayyad *et al.*, 1996; Mulvenna *et al.*, 2000; Ordonez *et al.*, 2006). In the field of information transmission, it becomes much easier than ever with the rapid development of tech-connected communication networks (Huang, 1997; Iribarren and Moro, 2009; Zhang *et al.*, 2016). Information networks can be established with different information carriers (Barabási *et al.*, 2002; DiMaggio *et al.*, 2001; Guille *et al.*, 2013; Kossinets and Watts, 2006; Yoo *et al.*, 2016), including newspaper, radio, television, internet and smartphone. Models can be built to understand the mixing effects of those carriers for the evolution of the methodology, purpose and functionality of society's networks. Subsequently, some simple questions are raised:

Q1. What are the patterns of information spread and how will communication networks be in the future?

To answer these questions, there are network-based evaluation tools that use individual communication models and event diffusion data to determine the impact of different carriers. The researchers aim to design effective models to analyse the evolution of communication networks, as well as future capacity. Specifically, the researchers will:

- construct communication networks, considering both individuals and five kinds of information carriers;
- propose models to explore the flow of information and find news; and
- validate the models using collected box office data.

To understand the evolution of the methodology, purpose and functionality of societal communication networks, an information diffusion model was designed using different adoption rates of five respective information carriers (newspaper, radio, television, internet and smartphone), as well as the spreading dynamics on social networks in different periods. Second, the model reliability was validated by using the proposed model to predict the fraction of individuals who would probably watch movies. Finally, the newly developed benchmark method was used to find out which information could become news in the future, other than the traditional way which only considers the information derived from information carriers as news.

Literature review

A complex network is an effective theoretical tool for understanding and analysing the structure and dynamic problems of complex systems in general. Numerous studies of information diffusion on social networks have paid attention to theoretical analysis and empirical research (Pei and Makse, 2013; Ramos *et al.*, 2015), where most systems illustrate networks by nodes indicating the individuals and edges denoting the relation of individuals (Castellano *et al.*, 2009). Many constructive works on information spreading have been developed, such as the modelling of the various information diffusion options. In particular, a series of seminal studies were conducted to understand the impact of network topological structure on the diffusion process; for example, Hu and Zhu (2015) demonstrated that the high-impact nodes significantly lift the efficiency of information diffusion. Communities are significant structures of networks, and some studies have noted that appropriate community can improve the rate of information spreading, but that information contagion will be inhibited following the increase of community strength (Luo *et al.*, 2014; Nematzadeh *et al.*, 2014). In addition, a number of studies have focused on spreading dynamics of coupled networks (Li *et al.*, 2015; Wang *et al.*, 2014).

With improved methods of data collection and analysis, a larger number of works were done based on social networks, such as Twitter (Pei *et al.*, 2014), Facebook (Kee *et al.*, 2016), Sina-microblog (Liu *et al.*, 2015) and smartphone networks (Jiang *et al.*, 2013). Many works show that the real spreading process is different from simulations. For example, Goel *et al.* (2012) analysed the information diffusion data from seven systems and found that most information only reaches a minority of people, which is very different from the model result based on network spreading. He *et al.* (2016) uncovered that the meme popularity is achieved by a wide scale of users on its diffusion trace, while just a small number of highly influential users is insufficient.

The majority of the aforementioned works focused on the process of information spreading between individuals, ignoring mass media, which also has significant impact on information outbreaks. Obviously, media plays an important role on the process of information dissemination, and there are a number of studies focused on the impact of media on information diffusion. Peress (2014) investigated the impact of media on trading and price formation by examining national newspaper strikes in several countries, and pointed out that media play a key role in financial markets. In the same vein, Onnela *et al.* (2007) also researched diffusion patterns of millions of anonymized mobile phone users based on a number of call records, and determined that the people who are exposed are significantly more likely to spread information, and do so sooner than those who are not exposed. Bakshy *et al.* (2012) examined the role of social networks in online information diffusion. In addition, other studies on social networks analysed the effect of network structure on information diffusion (Doerr *et al.*, 2012; Huang *et al.*, 2014a; Onnela and Reed-Tsochas, 2010; Yang *et al.*, 2014).

One of the important purposes of researching information diffusion dynamics is to predict the spreading process and control diffusion scale on real systems (Huang *et al.*, 2014b; Iwata *et al.*, 2013). The current research includes two main problems: predicting the dynamics of information propagation based on the social network structure (Anderson *et al.*, 2015; Iribarren and Moro, 2009; Rodriguez *et al.*, 2014) and speculating on the social network structure based on information diffusion (West *et al.*, 2014; Zhou *et al.* (2013)). Bandari *et al.* (2012) exploited the support vector machine (SVM) regression model to predict the news popularity degree and the precision reach up to 84 per cent. Myers and Leskovec (2014) examined the complete dynamics of the Twitter information network and found that the dynamics of network structure can be characterized by steady rates of change, interrupted by sudden bursts.

Based on the analysis above, this work will study the effect of different media on social networks. The model developed will help to predict the size of information diffusion in order to identify hot news.

Model description

To reflect different information spreading patterns when different information carriers emerge, the researchers propose several network-based dynamic models to investigate the evolution process of information diffusion. In this paper, each node in the network will be at one of three states during information diffusion; that is susceptible (S) individuals who will never be affected by information; informed (I) individuals who will accept the information; and recovered (R) individuals who are immunized to the information. The information can only be transmitted through links between susceptible and informed individuals in the network, where the S-state nodes can be informed by its neighbouring I-state nodes with designed probabilities. After the informed individuals (I-state) have transmitted the information to their neighbours, they will soon change to R-state, suggesting that they would not repeatedly broadcast the information any more. Table I shows all the definitions of symbols used in this paper.

Communication network construction

During different periods of social development, the velocity and coverage of information dissemination are significantly different due to the emergence of new information carriers. To model how information spreads during different eras, five periods (i.e. 1870s, 1920s, 1970s, 2000s and 2010s) are considered according to the time when the five information carriers were popular in the population. In these five periods, five different social networks are constructed according to the existence of information carriers. The nodes in the social network are classified into two types: the first is individuals and the second is information carriers. The degrees of the information carriers in the network are determined by their coverage rates in different periods (Table II). The social networks are named for the newest information carrier which emerges during the corresponding period.

Networks are built from different time periods when new information carriers become commonly used in the population, assuming the size of the population ($N = 5{,}000$) and generating simulation networks with the BA algorithm (Barabási and Albert, 1999). It is noted that one node represents an information carrier in the network, the links issued from this node denote the number of coverage in the population, and the rates of coverage are obtained from actual data

Table I.
Symbol definition

Symbol	Definition
$S^i, S^N, S^R, S^T, S^I, S^P$	Susceptible (individual, newspaper, radio, television, internet, smartphone)
$I^i, I^N, I^R, I^T, I^I, I^P$	Informed (individual, newspaper, radio, television, internet, smartphone)
$R^i, R^N, R^R, R^T, R^I, R^P$	Recovered (individual, newspaper, radio, television, internet, smartphone)
α_{00}	Probability that S^N, S^R, S^T is informed by I^i
α_{01}	Probability that S^i, S^I, S^P is informed by I^i
α_1	Probability that S^i, S^R, S^T, S^I, S^P is informed by I^N
α_2	Probability that S^i, S^N, S^T, S^I, S^P is informed by I^R
α_3	Probability that S^i, S^N, S^R, S^I, S^P is informed by I^T
α_{40}	Probability that S^N, S^R, S^T is informed by I^I
α_{41}	Probability that S^i, S^P is informed by I^I
α_{50}	Probability that S^N, S^R, S^T is informed by I^P
α_{51}	Probability that S^i, S^I is informed by I^P
β_1	Newspaper adoption rates
β_2	Radio adoption rates
β_3	Television adoption rates
β_4	Internet adoption rates
β_5	Smartphone adoption rates
N	Network size
$\langle k \rangle$	Average degree of the networks

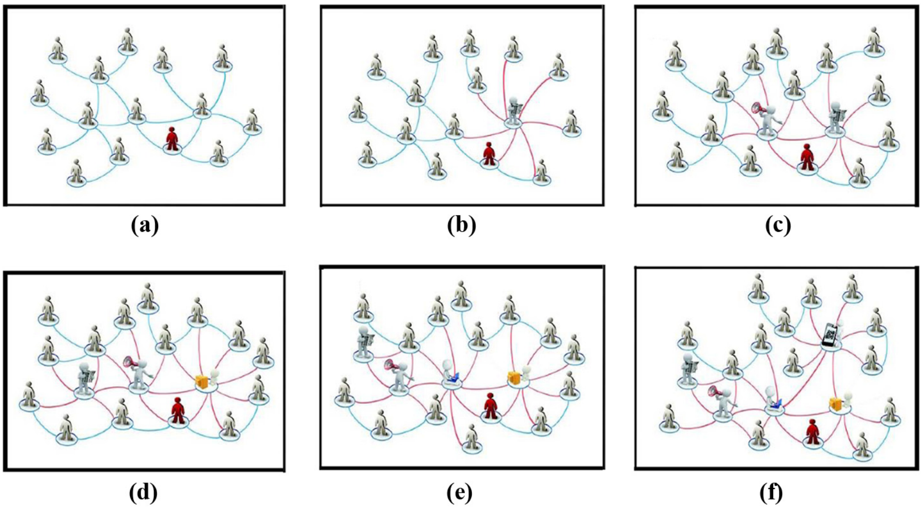
Table II.
Information carriers' adoption rates during different periods

Carrier	1870s (%)	1920s (%)	1970s (%)	2000s (%)	2010s (%)
Newspaper	55.0	48.0	39.0	21.0	14.0
Radio	0.0	6.0	57.0	40.0	34.0
Television	0.0	0.0	75.0	68.0	58.0
Internet	0.0	0.0	0.0	5.8	29.0
Smartphone	0.0	0.0	0.0	0.0	34.0

illustrated in Table II (extracted from historical data at http://media-cmi.com/downloads/Sixty_Years_Daily_Newspaper_Circulation_Trends_050611.pdf and www.internetworldstats.com/emarketing.htm), randomly selected nodes with the rates of coverage in the BA network, finally connected with an information carrier and developed the simulation networks. Figure 1 provides a schematic diagram of the network structures in different periods when different information carriers emerge. However, the individuals' social networks will also change overtime. To describe the network evolution, scale-free networks with tunable clustering coefficients are used (Holme and Kim, 2002; Newman, 2001; Soffer and Vazquez, 2005). The clustering coefficient of node i in the networks is defined as:

$$C_i = \frac{3 \times \textit{number of triangles of i's neighbours}}{\textit{expected number of connected triples of i's neighbours}} \tag{1}$$

The network's clustering coefficient is calculated by averaging the clustering coefficient of all nodes in the network. Table III shows the clustering coefficient of generating simulation networks in each period, and these results are based on equation (1). It is reasonable for the order of the clustering coefficients during different periods because, with the development of



Notes: (a) Individual network; (b) newspaper network: the icon of the individual holding a newspaper represents the newspaper carrier; (c) radio network: the icon of the individual holding a megaphone represents the radio carrier; (d) television network: the icon of the individual watching TV represents the television carrier; (e) internet network: the icon of the individual looking at a computer represents the internet carrier; and (f) smartphone network: the icon of the individual holding a phone represents the smartphone carrier. Figures (a)-(f) show that the network structure is gradually growing more complicated, and the number of different information carriers is more and more. In these pictures, all the red individual icons represent the infected nodes, and the lines indicate the connections between the nodes; in particular, the red lines represent the paths of information diffusion from the informed nodes to the susceptible nodes

Figure 1. Illustration of the communication networks during different periods

tech-connected communications networks, the relationship between nodes become closer than ever before.

The periods considered in the model are as follows:

(1) The period before the emergence of newspapers, when people only transmit information via word-of-mouth (before the 1870s). There is only one way to transmit information during this period: from mouth to mouth. The network size is set to N = 5,000.

(2) The period when newspaper becomes popular in society (around the 1870s). There are two ways to transmit information during this period: word-of-mouth and newspaper. In this period, one media node is added, the newspaper node, into the network, and it is connected randomly to the existing nodes. The degree of the newspaper node is N * 55 per cent, where 55 per cent is the adoption rate of newspapers around the 1870s, and N is the population size. This network is called the newspaper network.

(3) The period when radio becomes popular in society (around the 1920s). There are now three ways to transmit information: word-of-mouth, newspaper and radio. One radio node is added in the newspaper network, randomly connecting the existing nodes with a given degree calculated analogously to Period 3. The adoption rate of newspapers is updated to the up-to-date value. This is now the radio network.

(4) The period when television becomes popular in society (around the 1970s). There are now four ways to transmit information: word-of-mouth, newspaper, radio and television. One television node is added to the radio network, randomly connecting the existing nodes with a given degree calculated analogously to Period 2. The adoption rates of all media nodes are updated to the up-to-date values. This network is the television network.

(5) The period when internet becomes popular in the population (around the 2000s). There are five ways to transmit information: word-of-mouth, newspaper, radio, television and the internet. One internet node is added in the television network, randomly connecting the existing nodes with a given degree calculated analogously to Period 2. The adoption rates of all media nodes are updated to the up-to-date values. This is the internet network.

(6) The period when smartphones becomes popular in the population (around the 2010s). There are six ways to transmit the information: word-of-mouth, newspaper, radio, television the internet and smartphone. One smartphone node is added in the internet network, randomly connecting the existing nodes with a degree calculated analogously to Period 2. The adoption rates of all media nodes are updated to the up-to-date values. This is the smartphone network.

Information diffusion model

As the smartphone network contains both individuals and all type of observed carriers, it is the most complicated case of all models. The investigators thus take the information diffusion process on the smartphone network as an example to describe the detailed diffusion process as follows:

Table III. Clustering coefficients of observed networks during different periods

Network	C
Individual network	0.0450
Newspaper network	0.1093
Radio network	0.1632
Television network	0.2187
Internet network	0.2924

Note: C denotes the clustering coefficient of the networks excluding the information carrier nodes

- Initially, one individual node is randomly chosen as the diffusion seed, indicating it is at the informed state, denoted as $I^i(I^N, I^R, I^T, I^I, I^P)$. All the other nodes are at the susceptible state, denoted as $S^i(S^N, S^R, S^T, S^I, S^P)$.
- Informed process: At each time step, each informed node will inform its susceptible neighbours about the information with the corresponding probabilities described in Table I.
- Recovered process: At each time step, the informed nodes will change to the recovered state $R^i(R^N, R^R, R^T, R^I, R^P)$ after the informed process.
- The steps are repeated until the information spreads to all accessible nodes in the network.

Mathematically, this diffusion process can be illustrated by the differential equations as follows:

$$
\begin{cases}
\frac{dS^i}{dt} = -S^i(\alpha_{01}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{41} n_4 I^I + \alpha_{51} n_5 I^P) \\
\frac{dI^i}{dt} = S^i(\alpha_{01}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{41} n_4 I^I + \alpha_{51} n_5 I^P) - I^i \\
\frac{dR^i}{dt} = I^i \\
\frac{dS^N}{dt} = -S^N(\alpha_{00}\langle k\rangle I^i + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{40} n_4 I^I + \alpha_{50} n_5 I^P) \\
\frac{dI^N}{dt} = S^N(\alpha_{00}\langle k\rangle I^i + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{40} n_4 I^I + \alpha_{50} n_5 I^P) - I^N \\
\frac{dR^N}{dt} = I^N \\
\frac{dS^R}{dt} = -S^R(\alpha_{00}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_3 n_3 I^T + \alpha_{40} n_4 I^I + \alpha_{50} n_5 I^P) \\
\frac{dI^R}{dt} = S^R(\alpha_{00}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_3 n_3 I^T + \alpha_{40} n_4 I^I + \alpha_{50} n_5 I^P) - I^R \\
\frac{dR^R}{dt} = I^R \\
\frac{dS^T}{dt} = -S^T(\alpha_{00}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_{40} n_4 I^I + \alpha_{50} n_5 I^P) \\
\frac{dI^T}{dt} = S^T(\alpha_{00}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_{40} n_4 I^I + \alpha_{50} n_5 I^P) - I^T \\
\frac{dR^T}{dt} = I^T \\
\frac{dS^I}{dt} = -S^I(\alpha_{01}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{51} n_5 I^P) \\
\frac{dI^I}{dt} = S^I(\alpha_{01}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{51} n_5 I^P) - I^I \\
\frac{dR^T}{dt} = I^I \\
\frac{dS^P}{dt} = -S^P(\alpha_{01}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{41} n_5 I^I) \\
\frac{dI^P}{dt} = S^P(\alpha_{01}\langle k\rangle I^i + \alpha_1 n_1 I^N + \alpha_2 n_2 I^R + \alpha_3 n_3 I^T + \alpha_{41} n_5 I^P) - I^P \\
\frac{dR^T}{dt} = I^P
\end{cases}
\tag{2}
$$

where $\langle k \rangle$ is the average degree of the network without considering the carriers, $n_1 = \beta_1 N$, $n_2 = \beta_2 N$, $n_3 = \beta_3 N$, $n_4 = \beta_4 N$, $n_5 = \beta_5 N$, definitions of β_1, β_2, β_3, β_4, β_5 are in Table I. By adjusting the adoption rates of different carriers in each period (Table II), the diffusion processes and differential equations of each period can be obtained. For example, in the first period when the only way to transmit information is through word-of-mouth, the adoption rates of the carriers (newspaper, radio, television, internet and smartphone) are all set to zero; hence, in this period, equation (2) reduces to:

$$\begin{cases} \dfrac{dS^i}{dt} = -\alpha_{01}\langle k \rangle S^i I^i \\ \dfrac{dI^i}{dt} = \alpha_{01}\langle k \rangle S^i I^i - I^i \\ \dfrac{dR^i}{dt} = I^i \end{cases} \tag{3}$$

Results and discussion

Spreading patterns in different periods

The proposed models on the corresponding networks described above were used to obtain the evolution process of the fraction of recovered state nodes in each period, shown in Figure 2, where the fraction of recovered nodes can be used to measure the information popularity. It is obvious that information will spread quickly and widely in the smartphone network, followed by the internet network, television network, radio network, newspaper network and individual network, suggesting that, indeed, the advancement of technology improves the speed and breadth of communication. This is in agreement with the fact that people can receive information easier than ever before.

To further validate the effectiveness of the model, a data set of the box office of several famous movies – namely, *Minions*, *Jurassic World* and *X-Men: Days of Future Past* – is collected. These movies were released after 2014, and numerous popular information carriers have already appeared (e.g. WeChat). When a movie is forthcoming or has already been released, the information about this movie would diffuse through mouth to mouth as well as via all the information carriers considered in this paper. Therefore, this is suitable data for the validation of the model. In Figure 3, the model is applied to the smartphone network to fit this movie data, which shows good agreements and, to some extent, demonstrates the reliability of the model.

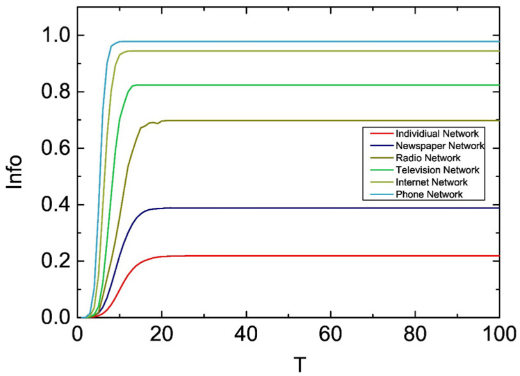
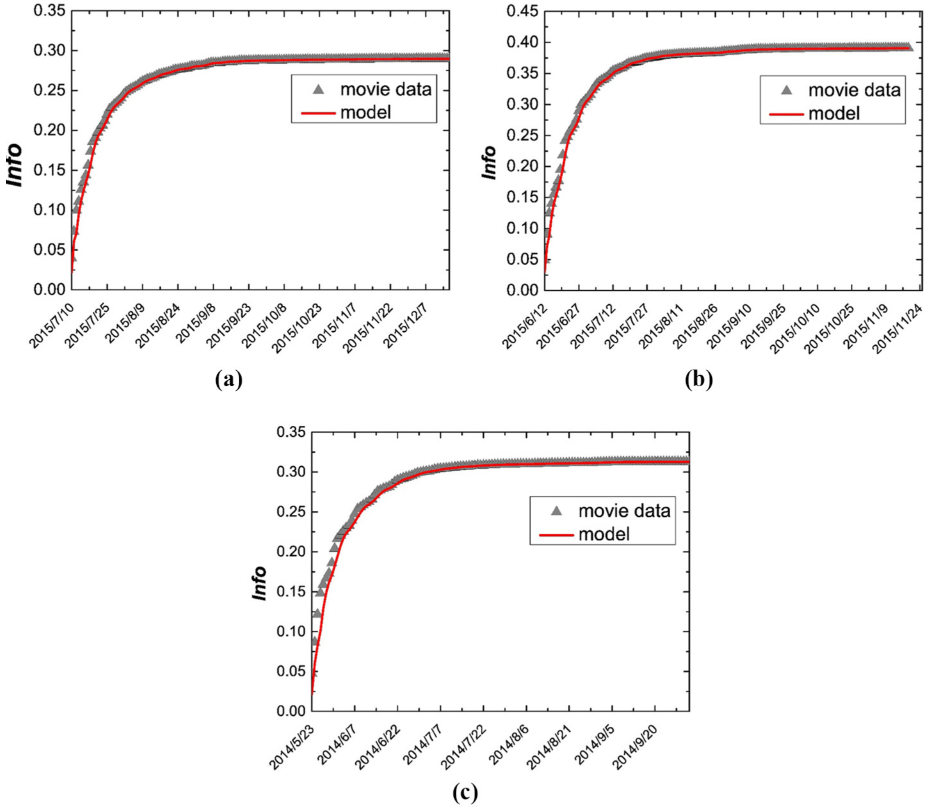


Figure 2.
Diffusion patterns on different networks: individual, newspaper, radio, television, internet and smartphone



Notes: (a) *Minions* (release date is 10 July 2015); (b) *Jurassic World* (release date is 12 June 2015); (c) *X-Men: Days of Future Past* (release date is 23 May 2014)

Figure 3.
Evolution comparisons of movies data and model results

Hot news prediction

As is well known, general information does not always reach a large fraction of people, while important news not only arrives to a considerable fraction of people but can also spread faster than other information. Nowadays, mass media (newspaper, radio and television) is not the only way to diffuse news, as much ordinary information can spread to a large fraction of individuals via the internet or personal smartphones (Borgatti *et al.*, 2009; Golosov *et al.*, 2014). Thus, news cannot be just defined as information spread by mass media. Therefore, in this section, an indicator is defined to predict whether information will turn into hot news.

The diffusion process is first simulated 200 times on the smartphone network as 200 different information events; the early stage (set as a maximum step as ten) of the diffusion is used to predict the transmission probability μ by using the benchmark method (Chen *et al.*, 2014), and μ is defined as:

$$\mu = \frac{1}{N_R}\sum_{i \in R}(N_R^i + N_I^i)/k_i \tag{4}$$

where N_R is the total number of recovered nodes in the network, N_R^i, N_I^i, respectively, represent the number of recovered and informed neighbours of node i and k_i is the degree of node i. The distribution of μ is shown in Figure 4. Apparently, most information events should have a low value of μ. The investigators decided that if $\mu > 0.5$, the corresponding information event will have a high probability to become hot news. As a consequence, 41 information events were predicted to become news. In addition, the authors validated the prediction by calculating the final fraction of recovered nodes of the 200 information events, where 87 per cent of the 41 events have a large fraction of recovered nodes.

Conclusion

To better understand the impact of information carriers on the diffusion of information, a susceptible-informed-recovered (SIR) information diffusion model considering the emergence of different information carriers during different periods is built in this work. Based on the simulation results, the authors validated that information diffuses faster and wider with the progress of time. In addition, the model can reproduce the information diffusion process of the box office data of several movies when considering the information carriers in the contemporary era. Furthermore, a metric was defined to identify hot news in the future. The validation from the simulation results shows a positive proof for the metric. In future studies, the investigators will collect more data, such as the data of information dissemination (for example, news, gossip and so on) on social networks to improve the news mining model with extended data.

This work provides a more effective means to reveal the impact of different information carriers on communication networks. In particular, the work provides a detailed method for illustrating the evolution of information diffusion for different information carriers. This work provides a better understanding of the information diffusion process.

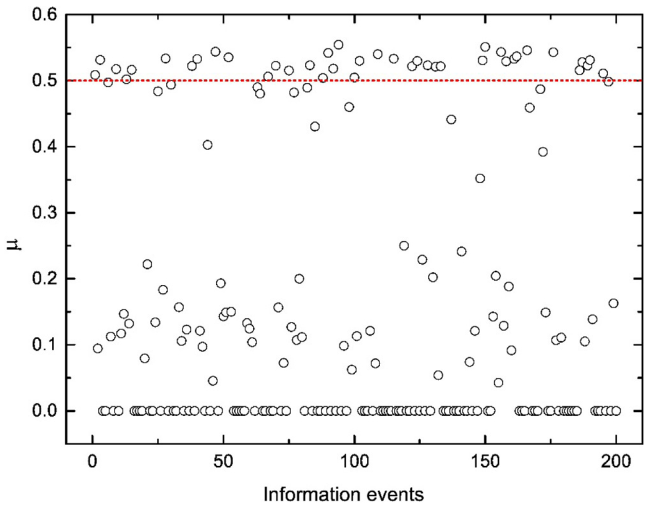


Figure 4. Distribution of transmission probability μ for 200 information events, events with $\mu > 0.5$ would be considered as hot news

References

Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J. and Tiwari, M. (2015), "Global diffusion via cascading invitations: structure, growth, and homophily", *Proceedings of the 24th International Conference on the World Wide Web*, *ACM, New York, NY*, pp. 66-76.

Bakshy, E., Rosenn, I., Marlow, C. and Adamic, L. (2012), "The role of social networks in information diffusion", *Proceedings of the 21st International Conference on the World Wide Web*, *ACM, New York, NY*, pp. 519-528.

Bandari, R., Asur, S. and Huberman, B.A. (2012), "The pulse of news in social media: forecasting popularity", *The Sixth International AAAI Conference on Weblogs and Social Media*, available at: https://arxiv.org/pdf/1202.0332.pdf (accessed 30 May 2017).

Barabási, A.L. and Albert, R. (1999), "Emergence of scaling in random networks", *Science*, Vol. 286 No. 5439, pp. 509-512.

Barabási, A.L., Jeong, H., Nda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002), "Evolution of the social network of scientific collaborations", *Physica A: Statistical Mechanics and its Applications*, Vol. 311 No. 3, pp. 590-614.

Borgatti, S.P., Mehra, A., Brass, D.J. and Labianca, G. (2009), "Network analysis in the social sciences", *Science*, Vol. 323 No. 5916, pp. 892-895.

Castellano, C., Fortunato, S. and Loreto, V. (2009), "Statistical physics of social dynamics", *Reviews of Modern Physics*, Vol. 81 No. 2, pp. 591-646.

Chen, D.B., Xiao, R. and Zeng, A. (2014), "Predicting the evolution of spreading on complex networks", *Scientific Reports*, Vol. 4, Article No. 6108.

Chen, M.S., Han, J. and Yu, P.S. (1996), "Data mining: an overview from a database perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8 No. 6, pp. 866-883.

DiMaggio, P., Hargittai, E., Neuman, W.R. and Robinson, J.P. (2001), "Social implications of the Internet", *Annual Review of Sociology*, Vol. 27 No. 1, pp. 307-336.

Doerr, B., Fouz, M. and Friedrich, T. (2012), "Why rumors spread so quickly in social networks", *Communications of the ACM*, Vol. 55 No. 6, pp. 70-75.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), "The KDD process for extracting useful knowledge from volumes of data", *Communications of the ACM*, Vol. 39 No. 11, pp. 27-34.

Fortunato, S. (2010), "Community detection in graphs", *Physics Reports*, Vol. 486 No. 3, pp. 75-174.

Fortunato, S. and Barthelemy, M. (2007), "Resolution limit in community detection", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104 No. 1, pp. 36-41.

Goel, S., Watts, D.J. and Goldstein, D.G. (2012), "The structure of online diffusion networks", *Proceedings of the 13th ACM Conference on Electronic Commerce*, *ACM, New York, NY*, pp. 623-638.

Golosov, M., Skreta, V., Tsyvinski, A. and Wilson, A. (2014), "Dynamic strategic information transmission", *Journal of Economic Theory*, Vol. 151, pp. 304-341.

Guille, A., Hacid, H., Favre, C. and Zighed, D.A. (2013), "Information diffusion in online social networks: a survey", *ACM SIGMOD Record*, Vol. 42 No. 2, pp. 17-28.

Hastings, M.B. (2006), "Community detection as an inference problem", *Physical Review E*, Vol. 74 No. 3, p. 035102.

He, S., Zheng, X. and Zeng, D. (2016), "A model-free scheme for meme ranking in social media", *Decision Support Systems*, Vol. 81 No. 1, pp. 1-11.

Holme, P. and Kim, B.J. (2002), "Growing scale-free networks with tunable clustering", *Physical Review E*, Vol. 65 No. 2, p. 026107.

Hu, H. and Zhu, J.J. (2015), "Social networks, mass media and public opinions", *Journal of Economic Interaction and Coordination*, Vol. 10 No. 1, pp. 1-19.

Huang, C.F. (1997), "Principle of information diffusion", *Fuzzy Sets and Systems*, Vol. 91 No. 1, pp. 69-90.

Huang, H., Tang, J., Wu, S. and Liu, L. (2014a), "Mining triadic closure patterns in social networks", *Proceedings of the 23rd International Conference on the World Wide Web*, *ACM, New York, NY*, pp. 499-504.

Huang, J., Li, C., Wang, W.Q., Shen, H.W., Li, G. and Cheng, X.Q. (2014b), "Temporal scaling in information propagation", *Scientific Reports*, Vol. 4, Article No. 5534.

Iribarren, J.L. and Moro, E. (2009), "Impact of human activity patterns on the dynamics of information diffusion", *Physical Review Letters*, Vol. 103 No. 3, p. 038702.

Iwata, T., Shah, A. and Ghahramani, Z. (2013), "Discovering latent influence in online social activities via shared cascade Poisson processes", *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *ACM, New York, NY*, pp. 266-274.

Jiang, Z.Q., Xie, W.J., Li, M.X., Podobnik, B., Zhou, W.X. and Stanley, H.E. (2013), "Calling patterns in human communication dynamics", *Proceedings of the National Academy of Sciences*, Vol. 110 No. 5, pp. 1600-1605.

Kee, K.F., Sparks, L., Struppa, D.C., Mannucci, M.A. and Damiano, A. (2016), "Information diffusion, Facebook clusters, and the simplicial model of social aggregation: a computational simulation of simplicial diffusers for community health interventions", *Health Communication*, Vol. 31 No. 4, pp. 385-399.

Kossinets, G. and Watts, D.J. (2006), "Empirical analysis of an evolving social network", *Science*, Vol. 311 No. 5757, pp. 88-90.

Lancichinetti, A., Fortunato, S. and Radicchi, F. (2008), "Benchmark graphs for testing community detection algorithms", *Physical Review E*, Vol. 78 No. 4, p. 046110.

Leskovec, J., Lang, K.J. and Mahoney, M. (2010), "Empirical comparison of algorithms for network community detection", *Proceedings of the 19th International Conference on the World Wide Web*, *ACM, New York, NY*, pp. 631-640.

Li, W., Tang, S., Fang, W., Guo, Q., Zhang, X. and Zheng, Z. (2015), "How multiple social networks affect user awareness: the information diffusion process in multiplex networks", *Physical Review E*, Vol. 92 No. 4, p. 042810.

Liu, C., Zhan, X.X., Zhang, Z.K., Sun, G.Q. and Hui, P.M. (2015), "How events determine spreading patterns: information transmission via internal and external influences on social networks", *New Journal of Physics*, Vol. 17 No. 11, p. 113045.

Luo, S., Xia, H. and Yin, B. (2014), "Continuous opinion dynamics on an adaptive coupled random network", *Advances in Complex Systems*, Vol. 17 Nos 3/4, p. 1450012.

Mandl, C.E. (1980), "Evaluation and optimization of urban public transportation networks", *European Journal of Operational Research*, Vol. 5 No. 6, pp. 396-404.

Mulvenna, M.D., Anand, S.S. and Büchner, A.G. (2000), "Personalization on the net using web mining: introduction", *Communications of the ACM*, Vol. 43 No. 8, pp. 122-125.

Myers, S.A. and Leskovec, J. (2014), "The burst dynamics of the Twitter information network", *Proceedings of the 23rd International Conference on the World Wide Web*, *ACM, New York, NY*, pp. 913-924.

Nematzadeh, A., Ferrara, E., Flammini, A. and Ahn, Y.Y. (2014), "Optimal network modularity for information diffusion", *Physical Review Letters*, Vol. 113 No. 8, p. 088701.

Newman, M.E. (2001), "Scientific collaboration networks, I: network construction and fundamental results", *Physical Review E*, Vol. 64 No. 1, p. 016131.

Onnela, J.P. and Reed-Tsochas, F. (2010), "Spontaneous emergence of social influence in online systems", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107 No. 43, pp. 18375-18380.

Onnela, J.P., Saramäki, J., Hyvönen, J., Szab, G., Lazer, D., Kaski, K., Kertsz, J. and Barabási, A.L. (2007), "Structure and tie strengths in mobile communication networks", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104 No. 18, pp. 7332-7336.

Ordonez, C., Ezquerra, N. and Santana, C.A. (2006), "Constraining and summarizing association rules in medical data", *Knowledge and Information Systems*, Vol. 9 No. 3, pp. 1-2.

Papadopoulos, S., Kompatsiaris, Y., Vakali, A. and Spyridonos, P. (2012), "Community detection in social media", *Data Mining and Knowledge Discovery*, Vol. 24 No. 3, pp. 515-554.

Pei, S. and Makse, H.A. (2013), "Spreading dynamics in complex networks", *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2013 No. 12, p. 12002.

Pei, S., Muchnik, L., Andrade, J.S. Jr, Zheng, Z. and Makse, H.A. (2014), "Searching for superspreaders of information in real-world social media", *Scientific Reports*, Vol. 4, Article No. 5547.

Peress, J. (2014), "The media and the diffusion of information in financial markets: evidence from newspaper strikes", *The Journal of Finance*, Vol. 69 No. 5, pp. 2007-2043.

Ramos, M., Shao, J., Reis, S.D.S., Anteneodo, C., Andrade, J.S., Havlin, S. and Makse, H.A. (2015), "How does public opinion become extreme?", *Scientific Reports*, Vol. 5, Article No. 10032.

Rodriguez, M.G., Leskovec, J., Balduzzi, D. and Schölkopf, B. (2014), "Uncovering the structure and temporal dynamics of information propagation", *Network Science*, Vol. 2 No. 1, pp. 26-65.

Soffer, S.N. and Vazquez, A. (2005), "Network clustering coefficient without degree-correlation biases", *Physical Review E*, Vol. 71 No. 5, p. 057101.

Veluscek, M., Kalganova, T., Broomhead, P. and Grichnik, A. (2015), "Composite goal methods for transportation network optimization", *Expert Systems with Applications*, Vol. 42 No. 8, pp. 3852-3867.

Wang, W., Tang, M., Yang, H., Do, Y., Lai, Y.C. and Lee, G. (2014), "Asymmetrically interacting spreading dynamics on complex layered networks", *Scientific Reports*, Vol. 4, Article number: 5097.

West, R., Paskov, H.S., Leskovec, J. and Potts, C. (2014), "Exploiting social network structure for person-to-person sentiment analysis", *Transactions of the Association for Computational Linguistics*, Vol. 2 No. 9, pp. 297-310.

Yang, Z., Yuan, N.J., Xie, X., Lian, D., Rui, Y. and Zhou, T. (2014), "Indigenization of urban mobility", *Physica A: Statistical Mechanics and its Applications*, Vol. 469, pp. 232-243.

Yoo, E., Rand, W., Eftekhar, M. and Rabinovich, E. (2016), "Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises", *Journal of Operations Management*, Vol. 45 No. 1, pp. 123-133.

Zhang, Z.K., Liu, C., Zhan, X.X., Lu, X., Zhang, C.X. and Zhang, Y.C. (2016), "Dynamics of information diffusion and its applications on complex networks", *Physics Reports*, Vol. 651, pp. 1-34.

Zhao, H., Bai, X., Zheng, D. and Wang, F. (2015), "Vehicle route assignment optimization for emergency evacuation in a complex network", *Journal of Intelligent & Fuzzy Systems*, Vol. 28 No. 1, pp. 107-116.

Zhou, K., Zha, H. and Song, L. (2013), "Learning social infectivity in sparse low-rank networks using multidimensional Hawkes processes", *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, Vol. 13, pp. 641-649.

Corresponding author
Zi-Ke Zhang can be contacted at: zhangzike@gmail.com