The Study on Information Characteristics Analysis using Information Diffusion Method

Huang li-dong,Shen xiao-chun
Huaiyin Normal University
Huai an, China, 223300

Journal of Digital Information Management

ABSTRACT: *A new kind of comprehensive evaluation method is proposed through analyzing the characteristics of the source of information. We use of three kinds of information transmission network performance report of the son from a source for information. Exploring the effective information diffusion characteristics, we analyzed the characteristics. The results in these characteristics have different characteristics, often preferred different types of information sources, such as a news conference or columns.*

Categories and Subject Descriptors
H.3.4 [Systems and Software]: Information Networks; **H.3.5 [Online Information Services]** Data Sharing

General Terms:
Information Networks, Information Transfer, Blogs

Keywords: Blog, Information Diffusion, Network Analysis, Ranking

Received: 20 March 2013, Revised 11 May 2013, Accepted 18 May 2013

1. Introduction

With the rapid increase of the blog, cost reduction, information disclosure is a huge number and variety of information has become public open to the public on the web site. However, this kind of information is can't use proper, because so many repeated existed and has many useless or useless information in the blog space. Recommend useful information resources attract blog according to user's different search requests, it is necessary to evaluate their characteristic point of view, and the user information navigation.

This paper presents a method to estimate the analysis of the characteristics of the information source of their information transmission network in the blog space son. Information diffusion network is extraction blog climb up a topic or a real event, including influential source of information. This network, a directed acyclic graph, can be divided into the son the influence of network information resources. Because of the shape and structure of each child, the content of the network depends on information source; we can estimate the characteristics of an information source. The information we experience comes to us continuously over time, assembled from many small pieces, and conveyed through social networks as well as other means. The merging of information, network structure, and flow over time opens interesting questions about the large-scale behavior information networks.

Even though the diffusion of information has been an active research area recently [7], [12], [26], [28], modeling the diffusion in social networks has proven to be a challenging task. It is difficult to obtain large scale diffusion data and to identify and track on a large scale the elements, such as recommendations [25], links [27], [28], tags [8], [7], topics [3], phrases or "*memes*" [26], that spread and propagate through networks. Even if one does obtain large scale real-world diffusion data, however, the issue of modeling the underlying process still remains. Traditionally, models of diffusion and cascading behavior have formalized the spread of ideas, information and influence as processes taking place on social and information networks [13], [15], [31], where each individual node is either active (infected, influenced) or inactive, and active nodes can then spread the contagion (information, influence, disease) along the edges of the underlying network. Parameter estimation of such models is challenging due to the heterogeneity of the nodes and data sparsely. Only recently has the availability of large

Type size (pts.)	Appearance		
	Regular	Bold	Italic
6	Table captions,[a] table superscripts		
8	Section titles, tables, table names, first letters in table captions, figure captions, footnotes, text subscripts, and superscripts		
9	References, authors' biographies	Abstract	
10	Authors' affiliations, main text, equations, first letters in section titles		Subheading
11	Authors' names		
24	Paper title		

Table 1. Type Sizes for Camera-Ready Papers

social network and corresponding diffusion data made it possible to estimate such models in practice [14], [30]. When using such models and fitting them to real-world data one makes several assumptions: (a) complete network data is available, (b) contagion can only spread over the edges of the underlying network, (c) the structure of the network itself is sufficient to explain the observed behavior. However, in many scenarios, the network over which diffusion takes place is in fact implicit or even unknown. Commonly, we only observe when nodes got "*infected*" but not who infected them. In case of information propagation, people usually discover new information without explicitly acknowledging the source. In word of mouth and viral marketing settings, we only observe people purchasing products or adopting new behaviors without explicitly. Similarly, in virus propagation, we observe people getting infected without knowing who infected them. Moreover, many times an activation of a node is not just a function of the social network but also depends on many other factors like imitation and regency. For example, people prefer the most recent information, and they discover new information or make decisions by using many different means, like the search engines, media sites, online forums and blogs or employing their social networks. Thus, even though flows of information and influence have traditionally been thought of as diffusion processes over underlying social networks [13], [15], [29], [31] existing models and formulations may be too constrained to capture the complexity of the underlying phenomena.

To quantify the properties of the son network, we focus on three different basic structure means information scattered, information collection, and information transmission. We can estimate the characteristics of information resources, from the three aspects.

Since the network dimension-force theory is put forward in 2004 [1], some scholars some of the main theoretical research in network information according to this theory, and the diffusion according to the theory, Network dimension-force including communication power, Suffer force support information exchange in the entire network, and it is the existing basis of another two forces. The clustering force diffusion and share information the largest amount. The information time-quantity the number of efficiency availability is big and time efficiency is short, from this view in the network space information diffusion can be seen as a process, in this process the information time. The effectiveness and efficiency of the number change over time. In the network dimension-force information diffusion influence of information have facilitation time effect and number efficiency. And the reality of space, the network information transmission in space has much more widely diffusion area and spread much shorter, this means network dimension-force can be convenient. This time-quantity number of effectiveness in the big and time effectiveness is effective short can fully reflect the influence of the network dimension-force information diffusion, so we can describe the network dimension-force influence through research time-quantity effect information diffusion. However, there are two important problems Exist. The first question is, time and efficiency availability is the number two objects; therefore, we need to the two subjects focused on an object to describe network dimension-force influence. The second the problem is, the influence of the information diffusion from the reality of factors space and cyberspace and the influence factors of network space, should go to the skin scribe the influence of network dimension-force respectively. The two problems become key analysis the influence of the network dimension-force information spread. A proper dispersion model is the key to solve the two important problems. In the consultation related in the domestic and foreign literature [11], now have less research information diffusion model in cyberspace, but there are a lot of research diffusion model geographical particles and similar technology information. These studies provide us comparison and references. Now there are three lead factors diffusion model in space. They are complex system model, the maximum entropy model and the transition probability model. Complex system model is the most public model in the spread of information research.

Here we solve these problems will be through the development of a model spread no network in the dominant knowledge is necessary. Not predict which nodes in the network will infect other nodes, we focus on modeling the world influence of diffusion rate through the nodes (implicit) network. The spread of the model often ignored time. In the era of discrete and. Instead, we accurate model not only each node has influence, but also in the spread with the passage of time how to spread in the works. Consider the spread of information network media, in no explicit network dissemination of information. Along with the information transmission, a blog or a web site is "*infected*" when it comes to the information. In this case the personal and websites may in different behavior methods: news agency play amplification, blog for two as early detection and elaborators (or response Chambers), when the mainstream media give a person a kind of dominant. For example, some web sites as "*influential*" or early people. The blog and the mainstream media are driving new content into the system of different way [22], [11], and usually generated content is regarded as a blog the more credible than with the mainstream media [21]. This paper aims to develop understanding rise and decline rate of diffusion mechanism as time goes on. What is causing some information waterfall growth? Big, why others are still small? And what are the part different participants in the spread of the dynamic? We consider the change means diffusion-based framework and establish the view of literature to the social influence [10] [20]. We formulated the linear influence model (LIM) started the number of new infections that node depend on other nodes are infection of the past. We then the number of new infections model node as a function other node of the infected the past time. In these models, each node has a related influence function it. Then the number of node new infections in time t is the influence of the function of infected by the time the node T. Back to our sample information diffusion, us the assumption of the number of web sites (i.e. node) mentioned that specific information depends on other websites mentioned the information in advance.

Random the cellular automata model is established is the main representative above type. This model can be reflect the reality of the position is extremely through diffusion simulation, but can't describe a specific spread process. By focusing on information inflow and outflow in the diffusion of space. This model can describe quantitative information diffusion, but it only concerns the spread of the special space and can't describe concrete diffusion process to take place on a node. Transition probability model is usually used to study technology diffusion. There is some representative research such as technology innovation diffusion. The efficiency of the information and quantity in Cartesian coordinate system effectiveness and her they use to describe the intersection of the influence network dimension-force. Her research can reflect the influence of network dimension-force intuitive but still less quantitative. This paper is based on the further investigation in the above research. There are two major kinds of point's innovation as follows.

2. Extraction of Information Diffusion Network

We present a method to extract an information diffusion network from the search results of a blog search engine as follows:

2.1 Searching and Crawling Blog Entries for a Topic

Analysis information diffusion exactly, we should only related information extraction a theme or a real-the affairs of the world. Therefore, we are looking for in blogs and inquire about a topic or an event, in the blog contains in search results to climb. We use blog search engines such as yahoo Japan and Technorati Japan. Because a blog search engine results returned by their creation time, we use a two-step approach to collect in the blog over a period of time. At the start of the search, we are looking for in the blog search range climbed to the blog search engine to collect old in the blog search created before time began. Next, we are looking for new blog climb to keep up with the changes intermittent days, or weeks.

2.2 Extraction of Hyperlinks in Body of Blog Entry

The content of the link only extraction related blog post, we identify the body and extraction blog link. Because blog entry is producing different blog service or projects, often similar in structure, extraction for the following four types of areas of the main body of the blog.

1) A region enclosed with a pair of special comments used for Google AdSense.

2) A region enclosed with a pair of "*div*" elements, whose class attribute name is a specified one for the blog service.

3) A region enclosed with a pair of "*div*" or "*td*" elements with body text similar to text specified in an RDF's "*dc*: *description*" attribute.

4) A selected region enclosed with a pair of "*div*" or "*td*" elements in consideration of text length, the amount of punctuation, and ratio of anchor text.

3. Information Diffusion Analysis

3.1 Identifying Sub network that Information Source Affects

To estimate the influence of each information source upon the blog space, we identify the sub network that is reachable from the information source in an information diffusion network and analyze its information diffusion properties.

An information diffusion network is $G = (V, E)$. V is a set of nodes, and a node of V is v_i ($i = 1, ..., |V|$). E is a set of edges, and a directed edge of E from a node v_i to a node v_j is e_{ij} ($j = 1, ..., |V|$).

The sub network $G_k = (V_k, E_k)$ that is reachable from an information source $v_k (1 \le k \le |V|)$ is defined as

$$V_k = \{v_i \mid v_i = v_k \vee directed\_path\,(v_k, v_i)\} \qquad (1)$$

$$E_k = \{e_{ij} \mid v_i \in V_k \vee v_j \in v_k\} \qquad (2)$$

Where $directed\_path\,(v_k, v_i)$ is true if there is a directed path from v_k to v_i. V_k always contains an end node v_j of e_{ij} if V_k contains the start node. Furthermore, we handle cases where does not contain but contains to consider the influence of information gather structures (described below).

The relation between information sources and subnetworks is shown in Figure 1. A node indicated by a star is an information source, and the region enclosed by dashed lines is its sub network. There may be an overlap region between two sub networks as in the case of information sources 1 and 2. A sub network may also include another sub network as in the case of information sources 2 and 3.

3.2 Basic Structure Units of Information Diffusion Networks

To analyze the properties of sub networks, we consider directed 2-edge connected sub-graphs, which are basic structure units of an information diffusion network and its sub-networks.

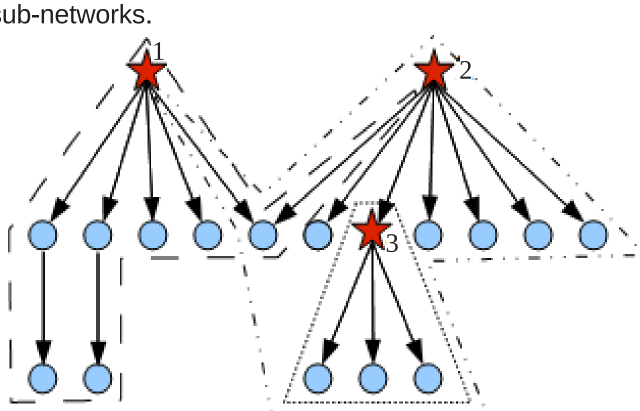

Figure 1. Relation between information sources and their sub-networks

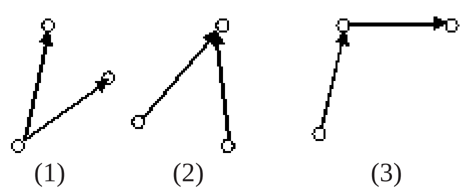

Figure 2. Directed 2-edge connected sub graphs

In a directed acyclic graph such as an information diffusion network, as described in this paper, there are three types of directed 2-edge connected sub graphs as basic structure units: a pair of edges from the same node (Figure 2-(1)), a pair of edges to the same node (Figure 2-(2)), and a pair of edges in which the start node of one edge is the end node of the other edge (Figure 2-(3)). Each basic structure unit is related to the basic phenomena of information diffusion: information scattering, information gathering, and information transmitting. Therefore, we call them information scatter structures (Figure 2-(1)), information gather structures (Figure 2-(2)), and information transmit structures (Figure 2-(3)) in this paper.

3.3 Quantification of Information Diffusion Properties

To quantify the information diffusion properties of each information source on the blog space, we analyze the basic structure units in each sub network.

The number of each basic structure unit that is contained in the sub network $G_k = (V_k, E_k)$ that is reachable from an information source $v_k (1 \le k \le |V|)$ is defined as

$$N_s(G_k) = \sum^{v_i \in V_k} \frac{d_{out}(v_i) \times (d_{out}(v_i) - 1)}{2} \qquad (3)$$

$$N_g(G_k) = \sum^{v_i \in V_k} \frac{d_{in}(v_i) \times (d_{in}(v_i) - 1)}{2} \qquad (4)$$

$$N_t(G_k) = \sum^{v_i \in V_k} d_{in}(v_i) \times d_{out}(v_i) \qquad (5)$$

Here, $d_{in}(v_i)$ is the in-degree of a node v_i and $d_{out}(v_i)$ is the out-degree of a node .

To make the basic structure units comparable with each other, we normalized them by using the number of nodes $|V_k|$ of a sub network G_k. Then, the information scatter degree $P_s(G_k)$, information gather degree $P_g(G_k)$, and information transmit degree $P_t(G_k)$ are defined as

$$P_s(G_k) = \frac{N_s(G_k)}{|V_k|} \qquad (6)$$

$$P_g(G_k) = \frac{N_g(G_k)}{|V_k|} \qquad (7)$$

$$P_t(G_k) = \frac{N_t(G_k)}{|V_k|} \qquad (8)$$

We use these three measures of information diffusion performance level information source.

The three measures have different characteristics. Dispersion degree of quantitative information has many wide attracted from blog attention. This is similar to the in-degree in a hyperlink network, but information is scattered degree of target not only adjacent node, but also contains subnet. The degree of information gathered, often source of information is co-referenced a variety of sources of information. Information transmission is of quantitative information spread by how many steps degree reputation. Note that bloggers tend to bypass the nodes and the source link among middle node information, if not to give you additional information.

4. Evaluation

4.1 Description of Data Sets

We evaluated the identification ability and the characteristics of $P_s(G_k)$, $P_g(G_k)$ and $P_t(G_k)$. Table 1 shows the datasets used for evaluation. Note that we used Japanese queries for collecting blog entries but translated them to English in this table. Each query was selected according to an event in the real world as follows.

1) iPhone 3G came to the Japanese market.

2) Google Street View was released in Japan.

3) Controversy and cancellation of WaiWai, which was a column of the Mainichi Daily News newspaper.

4) Copyright infringement of Chidejika, which is the official mascot of the National Association of Commercial Broadcasters in Japan.

5) Google Japanese input was released.

6) Apple announced the iPad.

The period was extracted by analyzing collected blog entries.

The number of blog entries was calculated after removing duplicates and invalidly dated blog entries caused by incorrect clock settings. The number of hyperlinks in the bodies of blog entries was calculated after URL filtering as described in Section IV-E2.

Table 2 shows the basic information of information diffusion networks that are extracted from data sets. The number of information sources (ISs) depends on the choice of T. If a smaller T is chosen, the total number of information sources will be increased but incorrect information sources will also be increased. In this work, we chose a T for which the number of information sources, which are extracted using T, is between 30 and 50. The number of extracted nodes of each data set is not high in comparison to the number of blog entries included in the data set. Therefore, we can use the complex algorithm analysis the network because their size is not very big. However, there is a problem, extraction of network scale is too small, and analysis of data set if the topic not becomes popular. The ratio of the number of the nodes in the blog activity depends on the topic of data set. It may be the impact on the blog space topic. A not active topics, a large and complex network not extraction method is useless. However, this is not a serious problem, because there are not many information sources and such a topic a user can investigate them easily.

4.2 Visualization

Understand their characteristics and the relationship between the information transmission network and information resources and blog, our system can use visual information transmission network Jung (Java universal network/graph framework) [12]. Visualization result for each information transmission network as shown in figure 3. The network structure is by Fruchterman Reingold layout algorithm [13], this is a force-directed layout algorithm. Each circle means a node is a blog entry or a web page, link on the blog. Each of the size is of the circle of the decision out-degree node. Every arrow refers to a directed edge direction spread, opposite of a hyperlink.

5. Discussion

Pay attention to the spread of information related to the content and nature of the attributes of the source of information. Each basic structure of meaning is the most important choice of ranking, and three attributes is not absolutely, because of different types of news articles and blog exist. Therefore, we analyzed three different information diffusion characteristics from a different perspective.

First, we analyze the ranking results of dataset 6, which is the latest data, in detail. In $P_s(G_k)$, the top three information sources are the official iPad Japanese site, iPad English site, and Apple site. In total, five official sites exist in the top 10. In $P_g(G_k)$, six news articles about the announcing of the iPad exist in the top 10. In $P_t(G_k)$, the first ranked is an Apple iPad event report. Eight news articles, i.e., four reports about the Apple's event and impressions of the iPad, two discussions on the trademark of the iPad, and two predictions about the next iPhone and iPhone OS, exist in the top 10. Tables 6 and 7

	Quer	Period	# entries	# links
1	iPhone	08/7/10–17	10, 801	24, 638
2	Google Street View	08/8/5–25	1, 376	5, 342
3	Mainichi Daily News	08/8/11–25	3, 863	13, 535
4	Chidejika	09/4/28–5/12	1, 179	5, 915
5	Google Japanese Input	09/12/3–15	2, 166	21, 591
6	iPad	10/1/27–2/7	5, 703	47, 678

Table 1. Data Sets Used For Evaluation

	T	#ISs	#nodes	#edges
1	9	50	732	810
2	5	43	420	517
3	6	34	406	516
4	5	48	503	749
5	5	42	1,047	1,486
6	10	49	1, 155	1, 429

Table 2. Basic Information about Information Diffusion Networks

show that both $P_s(G_k)$ and $P_t(G_k)$ tend to give priority to news articles, but Table 3 shows that the ranking result of $P_g(G_k)$ is different from $P_t(G_k)$. News articles recommended by $P_g(G_k)$ are press releases or quick reports, and news articles recommended by $P_t(G_k)$ are event reports or columns.

	Degree	$P_s(G_k)$	$P_g(G_k)$
$P_s(G_k)$	0.99	-	-
$P_g(G_k)$	0.18	0.16	-
$P_t(G_k)$	0.49	0.43	0.31

Table 3. Information diffusion networks are mainly constructed

Second, we analyze the ranking results of dataset 2, which has many CGMs. In $P_s(G_k)$ and $P_g(G_k)$, there is no CGM in the top 5, and there are two CGMs in the top 10. In $P_t(G_k)$ there are four CGMs in the top 5, and there are six CGMs in the top 10.

Therefore, authoritative information such as an official web site tends to have a higher value in $P_s(G_k)$. Objective and useful material for blogging, such as a news article, tends to have a higher value in $P_g(G_k)$. Subjective and controversial information that grows gradually through word of mouth, such as an impression or a column, tends to have a higher value in $P_t(G_k)$.

6. Conclusions

These characteristics have different characteristics, often preferred different types of information sources, such as a news conference or columns. We don't use the text information are discussed in this paper. A com-the method and the method based on text link-based would be useful to improve the quality of the search results. In addition, the future of the mission is to decide how to design a sorting function in a real information system.

References

[1] Stoilova, L., Holloway, T., Markines, B., Maguitman, A. G., Menczer, F. (2005). GiveALink: Mining a semantic network of bookmarks for web search and recommendation, *In*: Proceedings of the 3rd International Workshop on Link Discovery, p. 66–73

[2] Schenkel, R., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J. X., Weikum, G. (2008). Efficient top-k querying over social-tagging networks, *In*: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 523– 530.

[3] Wasserman, S., Faust, K. (1994). Social Network Analysis — Methods and Applications. Cambridge University Press.

[4] Watts, D. J., Strogatz, S. H. (1998). Collective dynamics of '*small-world*' networks, *Nature*, 393 (4) 440–442, June.

[5] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. (2002). Network motifs: Simple building blocks of complex networks, *Science*, 298 (5594) 824–827, October.

[6] Adamic, L. A., Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog, *In*: LinkKDD '05: Proceedings of the 3rd International Workshop on Link Discovery, p. 36–43.

[7] Adar, E., Adamic, L. A. (2005). Tracking information epidemics in blogspace, *In*: WI '05: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, p. 207–209

[8] Ishikawa, Takashi. (2010). The effect of transitive linking on information diffusion in dynamic acquaintance networks. *In*: Proceedings of the 3rd IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, p. 61-66.

[9] Renato Guseo, Mariangela Guidolin. (2010). Cellular automata with network incubation in information technology diffusion. *International Journal of Physica A: Statistical Mechanics and its Applications*, 389 (6) 2422-2433.

[10] Yao Juan, Helfert Markus. (2010). A framework for information diffusion over social networks research: Outlining options and challenges. *In*: Proceedings of the 5th International Conference on Software and Data Technologies, p. 491-494.

[11] Zhao Jinlou. (2006). Network dimension-force based on original philosophy and management innovations. Harbin Engineering University Dissertation for the Degree of D. Management, p. 46-75.

[12] Yang Jaewon, Leskovec Jure. (2010). Modeling information diffusion in implicit networks. *In*: Proceedings of the 10th IEEE International Conference on Data Mining, p. 599-608.

[13] Nicolás, F., Lori, Carlos Santos. (2010). Use of Shannon information in treatment of high resolution diffusion MRI. *In*: Proceedings of 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, p. 2714-2717.

[14] James Bessen. (2005). Patents and the diffusion of technical information. *International Journal of Economics Letters*, 86 (1) 121-128.

[15] Sinan Sinanovic, Don H Johnson. (2007). Toward a theory of information processing. *Journal of Signal Processing*, 87 (6) 1326-1344.

[16] Supplementary material: http://tinyurl.com/2dxltml.

[17] Yang, J., Leskovec, J. (2010). Patterns of temporal variation in online media Technical Report, *Stanford Infolab*.

[18] Adar, E., Adamic, L. A. (2005). Tracking information epidemic in blog space. *In*: Web Intelligence, p. 207–214.

[19] Bailey, N. T. J. (1975). The Mathematical Theory of Infectious Diseases and its Applications. Hafner Press, 2nd ed..

[20] Box, G. E. P., Jenkins, G. M. (1994). Time Series Analysis: Forecasting and Control. Prentice Hall.

[21] Brown, J. J., Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14 (3).

[22] Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. *In*: ICWSM '10.

[23] Cha, M., Mislove, A., Gummadi, K. P. (2009). A measurement driven analysis of information propagation in the flickr social network. *In*: WWW'09.

[24] Coleman, T. F., Li, Y. (1996). A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM J. of Optimization,* 6 (4).

[25] Friedkin, N. (1998). A Structural Theory of Social Influence. Cambridge University Press.

[26] Gill, K. E. (2004). How can we measure the influence of the blogosphere? Workshop on the Weblogging Ecosystem.

[27] Goetz, M., Leskovec, J., Mcglohon, M., Faloutsos, C. (2009). Modeling blog dynamics. *In*: ICWSM.

[28] Goldenberg, J., Libai, B., Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word of-mouth. *Marketing Letters*, 3 (12) 211–223.

[29] Goyal, A., Bonchi, F., Lakshmanan, L. (2010). Learning influence probabilities in social networks. *In*: WSDM 10.

[30] Granovetter, M. S. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83 (6) 1420–1443.

[31] Harsin, J. (2006). The rumour bomb: Theorising the convergence of new and old trends in mediated U.S. politics. Southern Review: Communication, *Politics and Culture*, 39 (1).

[32] Hartigan, J. A., Wong, M. A. (1979). Algorithm AS 136: A kmeans clustering algorithm. *Journal of the Royal Statistical Society*. Series C (Applied Statistics), 28 (1) 100–108.

[33] Hill, S., Provost, F., Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 21 (2) 256–276.

[34] Hoerl, A. E., Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 42 (1) 80–86.

[35] Jackson, M. O., Golub, B. (2007). Naive learning in social networks: Convergence, influence and wisdom of crowds. Working paper, June.

[36] Johnson, T. J., Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the internet influence credibility perceptions of weblogs among blog users. *Journalism & Mass Communication Quarterly*, 81 (3) 622–642.

[37] Katz, E., Lazarsfeld, P. (1955). Personal influence: The part played by people in the flow of mass communications. Free Press.

[38] Kovach, B., Rosenstiel, T. (1999). Warp Speed: America in the Age of Mixed Media. Century Foundation Press.

[39] Lawson, C. L., Hanson, R. J. (1995). Solving least squares problems. 3rd edition.

[40] Leskovec, J., Adamic, L. A., Huberman, B. A. (2006). The dynamics of viral marketing. *In*: EC '06.

[41] Leskovec, J., Backstrom, L., Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. *In*: KDD '09.

[42] Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M. (2007). Cascading behavior in large blog graphs. *In*: SDM'07.

[43] Liben-Nowell, D., Kleinberg, J. (2008). Tracing information flow on a global scale using Internet chain-letter data. *PNAS*,105 (12) 4633–4638.

[44] Rogers, E. M. (1995). Diffusion of Innovations. Free Press.

[45] Song, X., Chi, Y., Hino, K., Tseng, B. L. (2007). Information flow modeling based on diffusion rate for prediction and ranking. *In*: WWW '07.

[46] Watts, D. J. (2002). A simple model of global cascades on random networks. *PNAS*, 99 (9) 4766–5771.

[47] Watts, D. J., Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34 (4).

[48] Wu, F., Huberman, B. A. (2007). Novelty and collective attention. *PNAS*, 104 (45) 17599–17601.