Product Information Diffusion in a Social Network and Marketing Implications: A Case Study of Huawei Mobile Phone

Ling Zhang
School of Management, Wuhan University of Science and Technology, China; Center for Service Science and Engineering, Wuhan University of Science and Technology, China

Manman Luo
School of Management, Wuhan University of Science and Technology, China

Lijun Zhu
Center for Service Science and Engineering, Wuhan University of Science and Technology, China; Institute of Scientific and Technical Information of China， Beijing， China

Product Information Diffusion in a Social Network and Marketing Implications: A Case Study of Huawei Mobile Phone

Zhang Ling[12]*, *Luo Manman*[1], *Zhu Lijun*[23]
[1]School of Management, Wuhan University of Science and Technology, China
[2]Center for Service Science and Engineering, Wuhan University of Science and Technology, China
[3]Institute of Scientific and Technical Information of China，Beijing，China

Abstract: There is a need to consider how to spread marketing information into the largest area in a social network. In this paper, Tweets of Huawei Mate 9 were collected to analyze users' information behavior such as tweeting, forwarding, and commenting on tweets. First, the network topology is described as topology structure diagram; second, the Independent Cascade Model (ICM) is used for simulating information propagation; and finally, the article discusses how to identify the influential nodes to maximize the spread of business marketing information. The findings show us how to choose the influential nodes in an enterprise's marketing campaign conducted in a social network. The result shows that enterprises must pay attention to official nodes but also accident nodes. We suggest that the enterprise should pay more attention to the individuals who are characterized by their occupation, interests and are influential in a friend circle or interests-oriented circles.

Keywords: Social Network；Information Propagation；Independent Cascade Model

1. INTRODUCTION

Human social relationships are bound by time and space. However, the evolution of information and communication technologies tools have allowed people to inexpensively and reliably share information anytime and anywhere through social media. Companies such as Philips, HP, and Microsoft have adopted seeding strategies that target influential nodes in social networks to launch new products[1]. For example, Twitter, one of the most popular social media techniques, has evolved into a practical means for sharing opinions on almost all aspects of everyday life[2, 3].

Twitter is a popular microblogging service through which users send and receive text-based posts, known as "tweets", consisting of up to 140 characters. In the process of using Twitter, the users' behavior such as forwarding, or commenting can promote the spread of information in a social network. Many researchers conducted their studies on Twitter because the retweeted times is an clear indicator to the diffusion process[4-6]. Besides, today enterprises regard social networks as an important platform for launching new products and receiving the market feedback on a product. Therefore, it is significant to study the information diffusion in social networks to make the viral-marketing strategy successful. Specifically, how to identify the influential in an online marketing campaign, and how these influential are connected in a social network.

In this paper, the characteristics of the diffusion of an enterprise's product information in a social network were analyzed by using the Twitter data. First, the diffusion network topology was visualized and analyzed. Second, the independent cascade model was used to simulate the diffusion of enterprise mobile phone product information. Third, the influential nodes were identified, and the characteristics of influential nodes were investigated, marketing implications were discussed and concluded.

[1] Corresponding author.Zhangling, Email: beauty_ling@hotmail.com

2. LITERATURE REVIEW

2.1 Diffusion models

In recent years, the social network information propagation has become a research hotspot. The research involves network topology analysis[7] , text content analysis, large-scale data processing and so forth. At the same time, in order to express and predict the process of information diffusion in social networks, researchers have proposed information diffusion models[8]. Some frequently used models are independent cascade model (ICM) [9], linear threshold model (LTM) [10] and epidemics model [11]. Guille [12]noted that the diffusion models could be divided into graph based models and non-graph based models according to the propagation rule that considers the interaction between nodes or not. This study aims to find out the influential nodes and explain why they are influential. Hence, we choose the graph-based models to simulate the diffusion process.

Meanwhile, some scholars have carried on the algorithm optimization research on those models. Because of the complexity of social networks, a more scientific approach is to apply the models in some typical network cases. Algerian author Samir Akrouf[13] et al. analyzed the information propagation process and the influence of a set of nodes in two different networks: an egocentric contact network created by explicit relationships from Flickr social service, and an implicit video-commenting network created by commenting relationship from YouTube service. As a result, the research noticed that ICM performed better on implicit networks with stronger ties since it is based on the interactions between nodes. The article gives us new insights that it is useful to estimate the network attribute before choosing the information diffusion model.

2.2 Influential nodes identification

In the relative study about identifying influencers in a large-scale spreading, it has been accepted that the ability of influencers to initiate a large-scale spreading is attributed to their privileged locations in the underlying social networks[14] [15-18]. The most straightforward measurement of influence is using centrality-based heuristics. In recent years, an increasing number of predictors have been adopted to ranking node's influence in a social network, among which the most universally used ones include degree centrality[19],betweenness centrality[20],k-core [21]and PageRank[22]. Specifically, in the study of information diffusion, some scholars choose influential nodes as initial active nodes and simulate the information propagation process. Li[23] proposed a descriptive diffusion model to take dependencies among the topics into account, to identify the most influential nodes for specific contagion, and they applied the proposed model on an ISIS Twitter dataset, aiming to predicting the diffusion volume. Lu, et al[24]proposed a Score Cumulate model to evaluate the initial influence by using PageRank, and they applied the model in two real-world networks, Facebook and e-print arXiv (a scientific co-author network). Kwak[18]identified influentials on Twitter, and ranked users by the numbers of followers, PageRank and betweenness centrality. Thus, it is proper to identify influential nodes in a social network by the centrality measurements, through a simulation experiment.

3. METHODOLOGY

The diffusion network was constructed through data import and processing procedure, then influentials were identified from performing the simulation experiment. Finally, influential node's background were investigated, as well as tweet content, etc. The methodology could be shown as below:

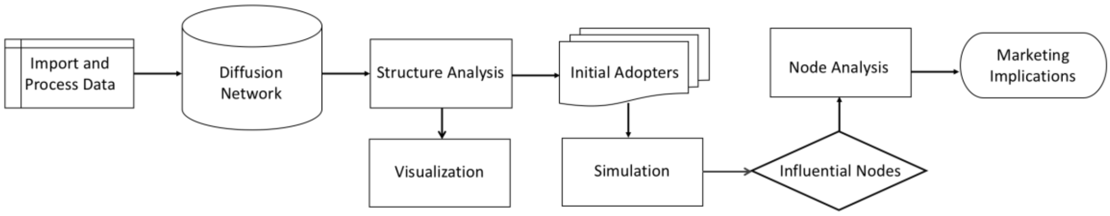

Figure 1. The overview of the methodology

3.1 Data import and processing

There are many social network analysis tools for extracting, analyzing and visualizing social network data. We use NodeXL (a free Excel plug-in) [25] developed by the Social Network Foundation. NodeXL is user-friendly for people who are not able to program. Our tweet data were sourced from Twitter. “Huawei Mate 9” was used as the keyword, searched daily through the node XL from January 3, 2017 to January 13, 2017. Then tweets data that is inaccessible to tweet relationship type were removed. Finally, the extracted diffusion network contains 5791 unique node (users) and 8386 links(relationship) between them. The relationship between vertices (users) include the original tweets (Tweet), comments (Replies to), and mentions (Mention).

3.2 Network structure analysis

After the Twitter diffusion network was constructed, the structure was further analyzed both visually and quantitatively. On one hand, network structure was visualized using a cluster-layout algorithm provided by NodeXL. On other hand, each node’s network metric was calculated, including the betweenness centrality and the PageRank. The nodes were ranked, and the top-ranked nodes in each measurement were selected as the corresponding heuristic initial adopters. The selected initial adopters are applied to the regarding heuristic to simulate information propagation according to independent cascade model.

3.3 Simulation

Independent Cascade Model was proposed in the context of marketing by Goldenberg, Libai, and Muller [9].Given a network G = (V, E) where V is the set of vertices, and E is the set of existing edges in the network. A vertex v ∈ V is said to be ***active*** if the information has reached the vertex and was accepted by it. If the information didn’t reach the vertex or the vertex rejected it, then the vertex is said to be ***inactive***. Each inactive vertex tends to become active, and it can switch from inactive to active, but it cannot switch from active to inactive. Given a set of initial active vertices A_0 , vertex v first becomes active in step s and is given a chance to activate each of vertex v’s inactive neighbors with a probability P_{vw} for success. If v succeeds to activate one of its inactive neighbors, say w, in step s+1, then the new active vertex w will be added to A_s to form the new active vertices set A_{s+1},and w will adopt the same activation action to activate its inactive neighbors. Each vertex is only given one chance to activate its neighboring vertices. If v fails to activate w in step s+1, it cannot make further attempts to activate w in subsequent steps. The propagation process ends when there are no more vertices can become active in step s. In our study, the probability of succeeding to activate vertices was set with 0.5.

Selecting an appropriate set of initial active nodes is a key step to simulating information propagation. On one hand, the initial set of active nodes were selected based on high betweenness centrality. On the other hand, high PageRank was considered as supplemental measurement since it is a common measure to gauge the importance of a node. Hence, the metrics of *betweenness centrality* and *PageRank* were adopted to select top-rank node sequence as initial adopters to initiate the diffusion process. If the total number of activated nodes larger, the initial adopters under this measurement is more influential.

4. FINDINGS

4.1 Network structure and visualization

The extracted "Huawei Mate 9" diffusion network for vertices (nodes) contained 5,791 individuals. While the edges which are ties or connections between nodes contained 8,686 connections(edges) Figure 2 is the extracted diffusion network graph which showing the interconnection of users, including retweet, comment relationship. A vertex is created when a user posted an original tweet. An edge is created when user(s) respond to the original tweet: retweet or comment.

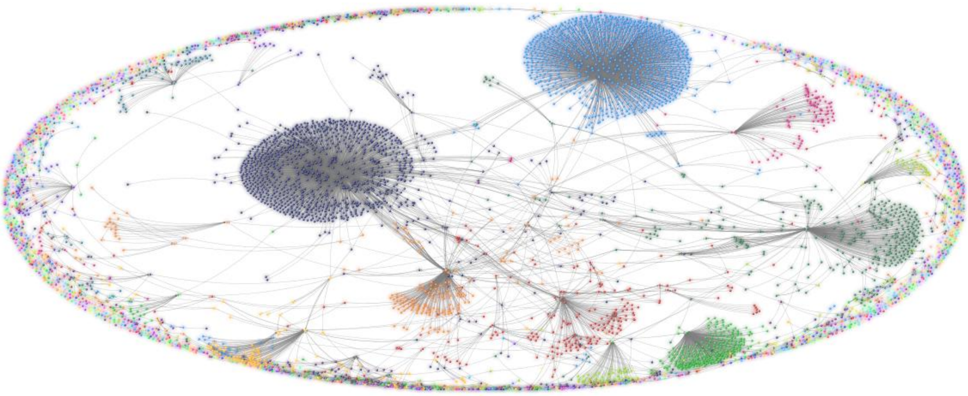

Figure 2. A Network Graph of #Huawei Mate 9

Next, network basic metrics, node's centrality and PageRank have been calculated using NodeXL, as shown in Table 1. These metrics help us characterize the research network as shown in Table 2. At the same time, we ranked the nodes according to the *betweenness centrality* and *PageRank*.

Table 1. Structure Characteristics of the User Networks

Graph Type	Directed Network
Type of relationship	Implicit
Number of vertices	5 791
Number of edges	8 386
Graph density	0.00019
Connected component	1 887
Maximum number of vertices in connected component	3 270
Maximum number of edges connected component	7 496
Diameter	13
Average distance	4. 351 803

Table 2. Metrics statistics

Metric Statistics	Value
Minimum overall degree	1
Maximum overall degree	1 032
Average overall degree	2.889
Minimum out-degree	0
Maximum out-degree	25
Average out-degree	1.448
Minimum in-degree	0
Maximum in-degree	1 031
Average in-degree	1.448
Minimum betweenness	0
Maximum betweenness	5027993.179
Average betweenness	6 197.929

4.2 Simulation results

The diffusion process algorithm for ICM was carried out on this directed network for each heuristic $h \in H$, where $H = \{PageRank, betweenness\ centrality\}$. For each heuristic the diffusion process was run 10 times for a differing number of initial active nodes. There were four sets of initial active nodes consisting of 5, 10, 15, or 20 nodes, respectively $A_0 \in \{5, 10, 15, 20\}$. An average was computed after each run of ten using one of the four heuristics and in of the four sets of initial active nodes. This is the average of active nodes after each run was computed. This average is the influence of the initial set. These results are displayed in Figure 3 which shows the performance of the algorithms used in the ICM on the network.

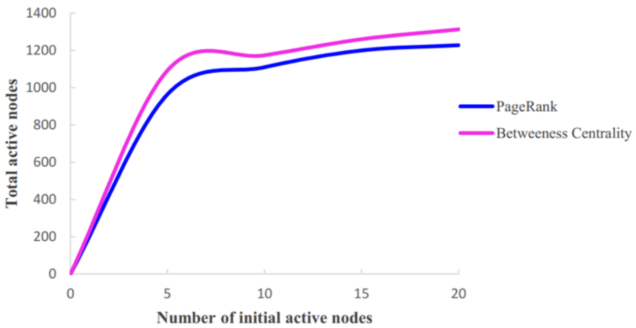

Figure 3. Simulation results for Independent cascade model

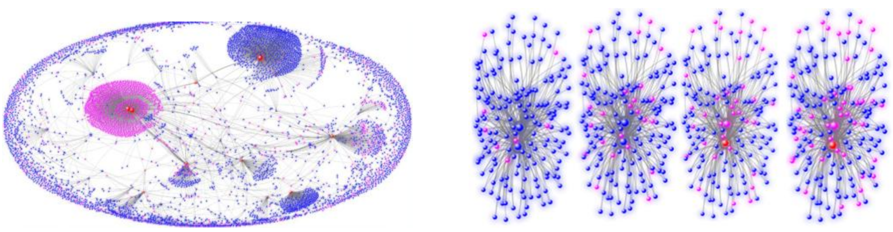

Figure 4. The left is an instance after a diffusion run using the ICM. The right is the activated status of a branch group node “*androidheadline*” change with Different number of initial active nodes. Noted that *red, fuchsia, blue* respectively represents nodes of *initial active, activated, inactive*.

1) It can be seen from Figure 3 the results converge to be proximate and close when the initial active set is chosen based on PageRank and betweenness centrality. The initial active nodes that have high overall degree

activate the largest number nodes: 22.68% of the nodes. The activation effect of high PageRank is not as good as the high betweenness centrality nodes.

2) From Figure 3, the curves become steady when the initial active node sets contain more than seven (7) nodes. The first seven (7) targeted nodes (with high PageRank, betweenness centrality) influence a large fraction of the network. When the initial active nodes are less than seven (7), the activation effect of choosing nodes with high betweenness centrality values seems the same as choosing high PageRank values. When the initial active nodes are larger than seven (7), the activation effect of the node with high betweenness centrality is relatively weakened.

3) Figure 4 visualizes the diffusion process using the ICM (independent cascade model). NodeXL enables us to visualize the nodes based on their metrics. The left part illustrates an instance of diffusion process choosing 10 initial active nodes (in red) based on the high betweenness centrality. By the end of the process, the initial adopters have influenced about 45% of the nodes (infected nodes are in fuchsia). The right part shows the node "*androidheadline*" has not been activated when there are 5 or 10 initials activation nodes. While the number of initial nodes achieved 15 or 20, although node "*androidheadline*" has become the activated node, there is a little diffusion change in the branch group.

4.3 Influential nodes identification

In our study, total number of activated nodes was defined as the influential. Larger the total number of activated nodes, higher the influential based on the regarding initial adopter strategy. Based on the experimental results represented in Figure 3. the heuristic for the diffusion process based on: ***PageRank, betweenness centrality*** converge to be proximate. Top 20 nodes in PageRank, betweenness centrality, and obtained 27 non-repeat nodes were merged. The influential diffusion nodes set S= {*huaweimobile, huaweimobileuk, androidauth, threeuk, youtube, nobunaga_s, huawei, androidheadline, jet, huawei_japan_pr, droid_life, huaweimobileksa, hamadsalleeh, androidcentral, wsj, xataka, techzilla, princepipo, majuzb, khajochi, huaweimobileesp, huaweimobilemy, this_is__e, freeconteston, metrini, mobilenewsmag, rkii2306*}. The nodes into were divided into four categories "Huawei official", "media", "mobile review", and "ordinary user".

5. DISCUSSIONS

The simulation result is consistent with the Akrouf's[13] experiment results, which also applied the independent cascade model to an implicit network. It indicates that the propagation rule of independent cascade model is proper to fit the diffusion network in marketing domain.

In this study, 27 influential nodes were identified, and classified the nodes into four categories according to their relationships with Huawei's new product "Huawei Mate 9", namely, "official", "media", "mobile review", and "ordinary user". Not all influential diffusion nodes are Huawei's official platform, information media, digital products reviewers, some of them are ordinary users and their influence we didn't expect. They also promote the information diffusion in our case. Thus, influential nodes unexpected arise in the event were called as "*accident node*". In this paper, "*accident node*" are {*nobunaga_s, princepipo, majuzub, metrini}*.

The average rank of Node "*nobunaga_s*" is 6. It is account of a Japanese actor, who published one tweets on January 5. The tweet's content is "I heard that Huwawei Mate 9's performance is very well, playing games without crashing, I would like to chose Huawei Mate 9 to be my next mobile phone." This tweet immediately caused more than 200 forwarding and nearly 1900 "like" (thumbs up). The node "m*ajuzub*" has high betweenness centrality rank, it is a Japanese religion scholar, who had released a tweet on January 7. The tweet says, "I am very interested in the new Huawe Mate9 which equipped with a Leica camera", and this tweet was responded by Huawei's Japanese official "*huawei_japan_pr"*. The node is interested in the Huawei's mobile and pay attention to the mobile purchase information constantly. The tweet content of "*nobunaga_s*" and "m*ajuzub*" is shown in Figure 5. From this case, it is worth for Huawei market staff to further consider the marketing opportunities of the user's identity of scholar. Moreover, the official's response behavior may draw user's attention on mobile product and lead to users' purchase behavior eventually.

Figure 5. "*accident node*" tweets content

From the view of average ranking of node "*rkii2306*", it just an ordinary user, but its betweenness centrality ranking is very high. Betweenness centrality essentially reveals how important each node is in providing a "bridge" between different parts of the network[14]. The higher the betweenness centrality, the stronger the intermediary effect of the node, and the stronger dependence of other nodes. The node's identity is Japanese android lovers, who have strong interest in the new mobile phones on the market. The reason of its high betweenness centrality is because that it locates in the two core node branches group. One branch is based on Huawei official node "*huawei_japan_pr*" as core node, another branch is based on accident node "*nobunaga_s*" as core node. The local network structure where node "*Rkii2306*" is shown in Figure 6.

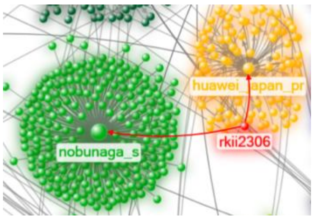

Figure 6. Node "*rkii2306*" in the Diffusion Network

6. CONCLUSIONS

This article used the NodeXL to visualize a product information diffusion network. The network topology represents the users' tweeting, forwarding, commenting relationships. Meanwhile, the article used independent cascade model (ICM) to simulate the information diffusion process. To identify the influential nodes, betweenness centrality and PageRank were considered as the measurements. The simulation results show that in the initial stage of the information, Huawei's official and media have an ability to trigger a large cascade. Then in the middle stage, some hub nodes such as mobile review, ordinary users emerge, leading to small-scale subsequent information dissemination. It is worth to explore the influence of those "*accident node*". In our case, some accident nodes are scholars, game lovers and so on. They can guide target customer groups, promote enterprise marketing effectively; accordingly promote a large scale of information dissemination. Based on the discussions, we suggest that the enterprise should pay more attention to the individuals who are characterized by their occupation, interests and are influential in a friend circle or hobby-oriented circles.

Our contribution is that independent cascade model was applied to a marketing empirical diffusion network, and found it is proper to fit the product information diffusion process. We also found it significant to consider "*accident node*". And the tweet content related to novel technology could attract more participation in ordinary users. However, the limitation exists. First, the diffusion model is basic and simple, there is much room for improvement. For example, the probability of succeeding could be adjusted according to the node's attributes and local network properties. Second, we didn't consider the culture influence on the diffusion process. In this case, the accident nodes are from Japan. It is worthwhile to further comprehend the nationality to improve the diffusion model, or to propose a more comprehensive measurement of the influentials.

ACKNOWLEDGEMENT

Research was sponsored in part by the National Social Science Fund Project "Study on dynamic optimization mechanism of information diffusion in social networks", Agreement Number 15CTQ029.

REFERENCES

[1] Libai, B., Muller, E., & research, R. P. J. O. M. (2013). Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *journals.ama.org*, *50*(2), 161–176.

[2] Richardson, M., & Domingos, P. (2002). Mining knowledge-sharing sites for viral marketing (pp. 61–70). Presented at the the eighth ACM SIGKDD international conference, New York, New York, USA: ACM Press.

[3] Kaplan, A. M., & Haenlein, M. (2012). Social media: back to the roots and back to the future. *Journal of Systems and Information Technology*, *14*(2), 101–104.

[4] Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, *29*(4), 217–248.

[5] Ferrara, E., & Yang, Z. (2015). Measuring Emotional Contagion in Social Media. *PLOS ONE*, *10*(11), e0142390–14.

[6] Tsugawa, S., & Ohsaki, H. (2017). On the relation between message sentiment and its virality on social media, 1–14.

[7] Udanor, C., Aneke, S., & Ogbuokiri, B. O. (2016). Determining social media impact on the politics of developing countries using social network analytics. *Program*, *50*(4), 481–507.

[8] Even-Dar, E., & Shapira, A. (2007). A Note on Maximizing the Spread of Influence in Social Networks. In *Internet and Network Economics* (Vol. 4858, pp. 281–286). Berlin, Heidelberg: Springer Berlin Heidelberg.

[9] 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*, *12*(3), 211–223.

[10] Granovetter, M. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology*, *83*(6), 1420–1443.

[11] Hethcote, H. W. (2000). The Mathematics of Infectious Diseases. *SIAM Review*, *42*(4), 599–653.

[12] Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: a survey. *ACM SIGMOD Record*, *42*(2), 17–28.

[13] Akrouf, S., Meriem, L., Yahia, B., & Eddine, M. N. (2013). Social Network Analysis and Information Propagation: A Case Study Using Flickr and YouTube Networks. *International Journal of Future Computer and Communication*, 246–252.

[14] Watts, D. J., & Dodds, P. S. (2007). Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, *34*(4), 441–458.

[15] Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, *6*(11), 888–893.

[16] Pei, S., & Makse, H. A. (2013). Spreading dynamics in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2013*(12), P12002.

[17] Yoganarasimhan, H. (2011). Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics*, *10*(1), 111–150.

[18] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? (p. 591). Presented at the the 19th international conference, New York, New York, USA: ACM Press.

[19] Albert, R., Jeong, H., & Barabási, A.-L. (1999). Internet: Diameter of the World-Wide Web. *Nature*, *401*(6), 130–131.

[20] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

[21] Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, *5*(3), 269–287.

[22] Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1-7), 107–117.

[23] Li, Q., Kailkhura, B., Thiagarajan, J. J., Zhang, Z., & Varshney, P. K. (2016, November). Influential Node Detection in Implicit Social Networks using Multi-task Gaussian Copula Models. *arXiv.org*.

[24] Lu, F., Zhang, W., Shao, L., Jiang, X., Xu, P., & Jin, H. (2016). Scalable Influence Maximization under Independent Cascade Model. *Journal of Network and Computer Applications*, 1–26.

[25] M, S., A, C., N, M.-F., B, S., E, M. R., J, L., & C, D. (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016. *https://www.smrfoundation.org*. Retrieved March 19, 2018.