2019 IEEE 5th International Conference on Knowledge-Based Engineering and Innovation (KBEI)
(Iran University of Science and Technology) – Tehran, Iran

Influence Blocking Maximization in Social Network Using Centrality Measures

Niloofar Arazkhani, Mohammad Reza Meybodi
Department of Computer Engineering & Information Technology
Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran
{n.arazkhani, mmeybodi}@aut.ac.ir

Alireza Rezvanian
School of Computer Science
Institute for Research in Fundamental Sciences (IPM)
Tehran, Iran
a.rezvanian@aut.ac.ir

***Abstract*— Online social networks play an important role as a suitable platform for information diffusion. While positive news diffusion on social network has a great impact in people's life, the negative news can also spread as fast as positive ones. To make the social network a reliable place, it is necessary to block inappropriate, unwanted and contamination diffusion. In this paper, we study the notion of competing negative and positive campaigns in a social network and address the influence blocking maximization (IBM) problem to minimize the bad effect of misinformation. IBM problem can be summarized as identifying a subset of nodes to adopt the positive influence under Multi-campaign Independent Cascade Model (MCICM) as diffusion model to minimize the number of nodes that adopt the negative influence at the end of both propagation processes. We proposed Centrality_IBM algorithm based on centrality measures for finding an appropriate candidate subset of nodes for spreading positive diffusion in order to minimizing the IBM problem. Then, we experimentally compare the performance of the proposed algorithm using some centrality measures to choose the appropriate subset of positive influential nodes. The experiments on different real datasets reveal that the closeness centrality measure outperforms the alternative centrality measures in most of the cases.**

Keywords—centrality measure; degree centrality; closeness centrality; betweenness centrality; influence blocking maximization; social network

I. Introduction

With the growing of online social networks (OSN) and the impact of these networks on everybody's lives, the impact of information diffusion and viral marketing has been affected many life style of people and this issue motivated by influence maximization problem in many recent studies [1]–[5]. Based on the power of viral marketing in OSN, it made company managers to do various activities on these networks to increase their company's customers and advertise their new products. Beside, diffusion of misinformation, rumors and other unfavorable information can spread faster with more terrible effects. Statistics [6] show that on some social networks, such as *Twitter* and *Facebook*, the rumor spreads so fast that makes users confused in their decisions. Bad effects of rumor spreading can be dangerous in some cases, for example in August, 2012, thousands of people in *Ghazni* left their houses in the middle of the night in horror after the rumor of an earthquake, which said that a powerful earthquake would hit the town at 5:00 am spreads [7]. Another example is the spread of misinformation about swine flu in *Twitter* [8]. In this case the spread of misinformation caused a huge panic in the population. According to these examples although social networks are the main source of good activities and news for many people, they are not reliable for all time due to such issues.

Therefore, in order to make social networks as a reliable platform for spreading information, it is important to have tools to limit the effect of misinformation. Finding special nodes which have strong connections to the other nodes in the networks can be a good approach. Researches show that finding centrality measures is a good approach to find important nodes in a network [9]. In this paper, our contribution is to find a good candidate subset of positive nodes for spreading positive news in order to minimize the bad effects of negative news. In this case, we propose to use different aspects of central nodes in a network based on degree, closeness and betweenness centralities to find a subset of positive initial spreaders. We experimentally will show that centrality nodes can have a good effect on spreading positive information against negative one by showing number of negative nodes as results.

In information diffusion problem on social networks, different models have been introduce [10]–[14], among them two models namely independent cascading model (ICM) and linear threshold model (LTM) [15] are known as popular and efficient models. In this work, we assumed that both positive and negative propagations are in Multi-campaign Independent Cascade Model (MCICM) [16].

One of the greatest issues in information diffusion in social networks is the influence blocking maximization which is to select a small subset of nodes in a network as the initial nodes to start diffusing positive information against negative ones and block the spread of negative information as much as possible. This set of selected nodes that can spread the positive information against the negative ones. Some studies have been conducted on limiting rumor diffusion.

In this paper, we compute centrality measures of nodes such as degree, closeness and betweenness centralities to find a subset of positive nodes to spread positive information to block the negative one. We compare these methods on five different real datasets. During the execution of the algorithm, a subset of positive influence spreaders is taken according to centrality measures of each nodes and the positive diffusion start to spread on MCICM diffusion model to block negative diffusion which is spreading in the network on the same diffusion model with high priority. The experimental results demonstrated that choosing a subset of nodes based on high centrality measures have a good effect on blocking contamination for spreading positive information due to the strong connecting that a node with these features can have.

II. RELATED WORK

Many studies have been conducted on influence maximization (IM), but little has been done on influence blocking maximization. In this section, we will briefly introduce the prior works regarding rumor controlling in online social networks.

In earlier works [17], the authors introduced blocking a certain number of links in a network to reduce the bad effects of rumor in a network. In [18], *Chen et al.* presented an efficient algorithm to maximize the positive influence when negative propagations appear. The works of [16], [19], [20] consider to block the influence of rumor by initiating a set of protective individuals. The identification of influential users in a social network is a key problem that has received a huge attention in recent years. Early works relied on heuristics techniques to select the set of influential users. *C. Budak et al.* [16] are among the first groups who studied about misinformation contamination problem. In particular, they used multi-campaign independent cascade model and solved the problem of finding a subset of individuals to start propagating “good” campaign to minimize the rumor diffusion in the social network. Ref. [19] and [7] studied this problem under the competitive linear threshold model and the Opportunistic One-Activate-One (OPOAO) model, respectively [21].

Later works on this problem showed a *(1-1/e)* -approximation. The existing approaches in this field are time consuming so cannot control the large social networks. Recent studies on this field focus on two different categories, unknown opponent strategies and comparative influence maximization. In unknown opponent strategies, it is not practical to assume the opponents’ strategies are known, therefore in this category each player does not know about other players’ strategies. In [22], the authors models the competitive influence model problem as a multi-round multi-party game. *Li et al* consider another model of competitive influence model [23]. In this model the strategy space consists of all IM algorithms that can be adopted by players. The goal is to find a Nash equilibrium strategy for each player to maximize each player’s influence.

In comparative influence maximization the diffusion model for comparative influence maximization considers two different kinds of relationships between two diffusions N and P which are (1) competition and (2) complementary. In competition if a node adopts the influence of P, it has a lower probability to adopt N. [24] and in complementary if a node adopts the influence of P, it has a higher probability to adopt N. [25] and [26]. Rumor source detection is another important problem for rumor controlling. In our work, we assumed that the rumor spreader nodes have been detected before and we try to control the rumor propagation. The prior works in this field was primarily focused on the classic susceptible-infected-recovered (SIR) model where the node can be infected by rumor or recovered. [27] studied about rumor source detection and designed a rumor source estimator by concept of rumor centrality. In Ref. [28] provided a detection performance for degree-regular tree networks.

III. PROBLEM DEFFINITION

Influence blocking maximization problem is to find a subset of positive nodes to block negative diffusion. The main target of this problem is to start a positive diffusion to spread positive information to the nodes of the networks which are not still polluted by negative diffusions.

The main challenge is to find an optimal subset of nodes to start spreading positive information for blocking negative diffusion as much as possible. We proposed *Centrality_IBM* algorithm Fig. 1, for finding positive initial nodes. In diffusion process of this work a node can have three different states such as *N-active*, *P-active* and *no-active*. Before start of diffusing, all of the nodes except negative and positive initial subsets of nodes are in *no-active* state. When positive diffusion reaches to a node its state changes to *P-active* and its state stays *P-active* till the end of diffusion process. When negative diffusion reaches a node its state change to *N-active* and it doesn’t change till the end of the both diffusions. If neither positive diffusion nor negative diffusion reaches a node, its state stays no-active. If both diffusions reach a node, negative one has priority.

IV. PROPOSED ALGORITHM

Many of research studies released that the social behavior of users plays an important vital role in effecting other users in the social network. Also, most of research studies have investigated on the social position of users for information diffusion in order to influence maximization; however the role of the social position of users for influence blocking is investigated using the proposed algorithm. To this end, in this paper, we compute centrality measures of nodes such as degree, closeness and betweenness centralities to find a subset of positive nodes to spread positive information to block the negative one. We compare these methods on five different real datasets. During the execution of the algorithm, a subset of positive influence spreaders is taken according to centrality measures of each nodes and the positive diffusion start to spread on MCICM diffusion model to block negative diffusion which is spreading in the network on the same diffusion model with high priority

In this work, we assumed that negative nodes have been already detected and both diffusions have the same diffusion model. Both diffusions start to spread together but negative diffusion has priority against positive one. For example if both negative and positive diffusions reach to a same node, negative diffusion has priority and its state change to *N-active*.

The proposed *Centrality_IBM* algorithm is based on finding centrality measures of all the nodes which are not adopted the negative information. Therefore, first we need to compute the centrality measures of nodes and then we need to rank these nodes based on their respective centrality measures. The subset of positive influence nodes are selected among the highly central nodes that are not adopted the negative information.

In following parts, we will provide preliminaries which are useful for the rest of the paper.

A. Graph model

A social network can be modeled as a directed graph $G = \langle N, E\rangle$ consists of N nodes and E edges. Node a is neighbor of node b if and only if there is an edge between them as $e_{ab}\in E$.

B. Diffusion model

Independent cascade model is one of the most well-known diffusion models that have been used in different contexts. In the IC model, a process starts with an initial set of active nodes and unfolds in discrete steps. When a node v activates in step t, it has a single chance to activate each of its inactive neighbor w with probability P_{vw}. The process runs until no more activation is possible.

We used Multi-Campaign Independent Cascade Model (MCICM) which introduces process of two cascades simultaneously in a network [16]. We have two cascades P and N as positive and negative cascades and two initial sets $P(k)$ and $N(k)$. Each node can have three different states such as *P-activate*, *N-activate* or *no-active*. When v actives in step t, by cascade of P (or N), it has only a single chance to *P-activate* (or *N-activate*) its neighbor w. If *N-cascade* and *P-cascade* tries to activate a node at a same time, *N-cascade* has priority. Once a node changed its state to a *P-activate* (or *N-activate*) it will stay unchanged for the rest of the cascade. The process runs until no more activation is possible. The earlier studies using similar diffusion models support the validity of MCICM and its variants. Also, the behavior on this model in real social networks is investigated by *C. Budak et al.* [16].

In following sections three used centrality measures [29], [30] in this paper are introduced.

C. Degree centrality

In a graph $G = \langle N, E\rangle$, degree of a node n is number of edges incident to the node. For a directed graph, there are two different kinds of degrees, out-degree and in-degree. Out-degree of a node is the number of edges from that node to the other nodes and in-degree of a node is the number of edges from other node to that node. For example in Fig. 2 out-degree of node 1 is 3 and in-degree of node 1 is 0. In an undirected graph there is only one kind of degree, for example in Fig. 2 degree of the node 5 is 2. In this work, we have assumed that all of the networks are undirected graphs. In this work, maximum degree is one of the features that is used to find a subset of positive influence nodes.

D. Closeness centrality

Closeness is a measure of centrality in a network. It is calculated as reverse of sum of the length of the shortest paths between the node and all the other nodes in a graph. By this definition for the closeness we can figure it out that the more central a node is, the closer it is to all the other nodes. Closeness was defined by Bavelas [31] that is:

$$C(\nu)=1/\Sigma_{\gamma}\, d(\gamma,\nu) \tag{1}$$

where $d(\gamma,\nu)$ is the shortest path between γ and ν .

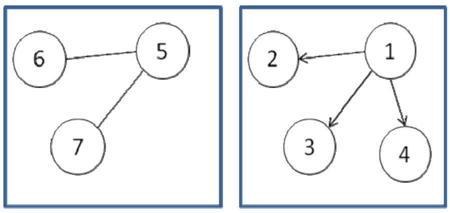

Fig. 1. An exmape for degree centrality

E. Betweenness centrality

Betweenness centrality is a measure of centrality in a graph based on shortest paths. The betweenness centrality for each node is the number of these shortest paths that pass through the node. Freeman (1977) [32] first defined betweenness centrality as follow:

$$g(v)=\Sigma_{s\neq v\neq t}\ \sigma_{st}(v)/\ \sigma_{st} \quad (2)$$

Where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

Algorithm 1. Centrality_IBM

Initialization
Centrality type {Degree, Closeness, Betweenness}
Input: Graph G(V, E)
Output: Subset S
Begin Algorithm
Switch 'Centrality type'
Case 'Degree'
Compute Degree in D
Case 'Closeness'
Compute Closeness in C
Case 'Betweenness'
Compute Betweenness in B
End Switch
Rank nodes with respect their centrality values
S ← Select K nodes with highest values
End Algorithm

Fig. 2. Pseudo-code for the proposed algorithm for finding a subset of positive initial nodes

V. EXPERIMENT

In this section, we evaluate the performance of proposed algorithm using three different centrality measures in comparison with random node selection as base-line algorithm based on five different real datasets with different sizes including Dolphin [33], Jazz [34], Zachary [35], Football and Facebook [36] as well-known real world networks. The descriptions of these networks are given in table 1.

Following section presents experiment results for blocking maximization. All the experiments were launched on a system with hardware configuration of Intel® Core™ i7-4702MQ CPU @ 2.20GHz and 6 GB RAM.

The simulation number of running methods on datasets is 1000 times and average of negative nodes and their variances were taken as results. *Matlab* was used for developing the experiments and MCICM model is the diffusion model of simulating.

A. Experimental result

To verify validity of this algorithm, 5 percent of nodes in each dataset are selected randomly as negative initial nodes, and the positive initial set with different sizes 5, 10 and 20 percent of each dataset are chosen by all methods respectively. The experiment result of blocking effect on Facebook dataset is shown in Fig. 3. As shown in Fig. 3 by increasing the number of positive initial nodes, number of N-active nodes at the end of the diffusion process decreases which shows better results. By increasing the number of initial positive spreaders, closeness method does significantly better than two other methods while betweenness and degree method work almost the same. Choosing random nodes show that centrality methods do better than random.

By running these methods on Jazz dataset, its average and variance Fig.4 shows that closeness and degree methods work significantly same but the average result number of negative nodes for jazz dataset and its variance shows that the betweenness result is not as good as two others approaches. Random method does the worst and it showed the good result of choosing centrality measures in Jazz dataset. Three methods of centrality did same on Football dataset Fig. 5 work significantly same while the bad result of random method shows the good result of centrality methods. In Dolphin Fig. 6 closeness did remarkably better than betweenness and closeness approach in 2 out of 3 results. In Karate Fig. 7 degree method notably did better than closeness and betweenness in 1 out of 3 times of experiments. Fig. 8 shows that closeness takes more time to run.

TABLE 1

Network	Nodes	Edges	Description
Karate	34	78	Zachary's Karate Club Network
Dolphins	62	159	Networks of Dolphins
Football	115	613	Network of Amarican College Football Team
Jazz	195	5484	Network of Amarican musicians
Facebook	4039	88234	Social network

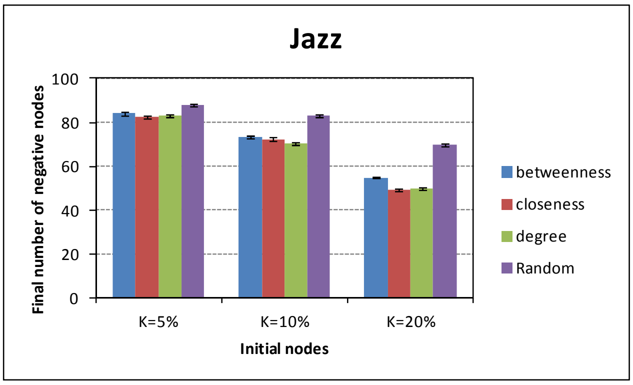

Fig. 3. Jazz

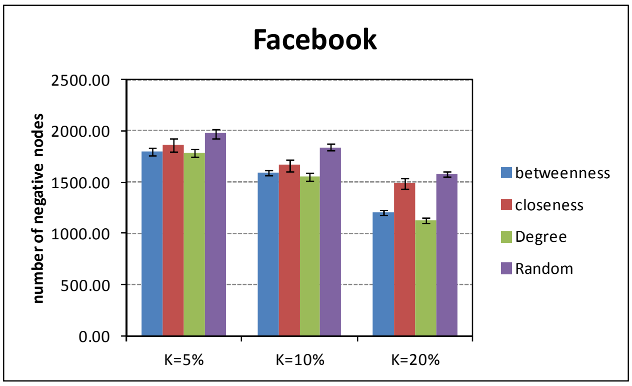

Fig. 4. Facebook

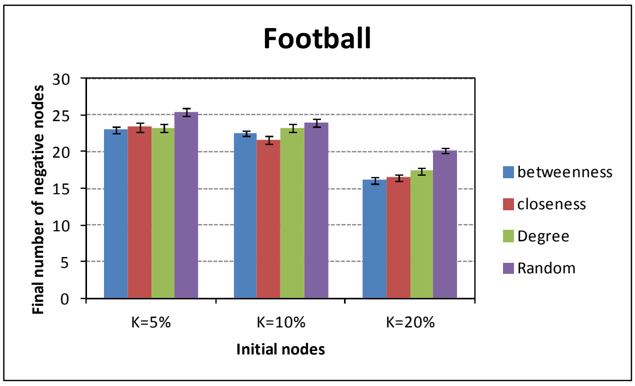

Fig. 5. Football

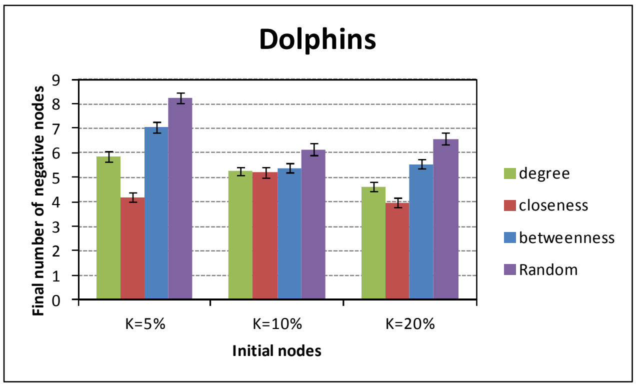

Fig. 6. Dolphin

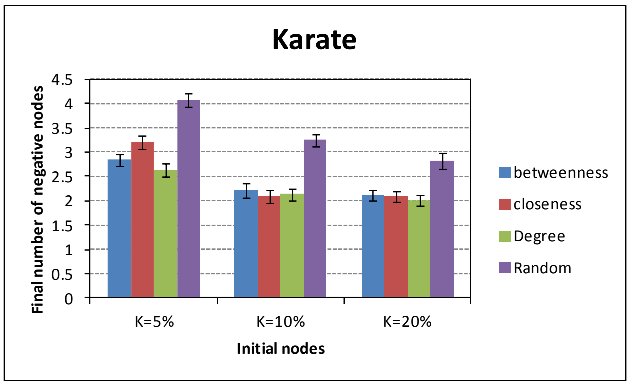

Fig. 7. Zachary

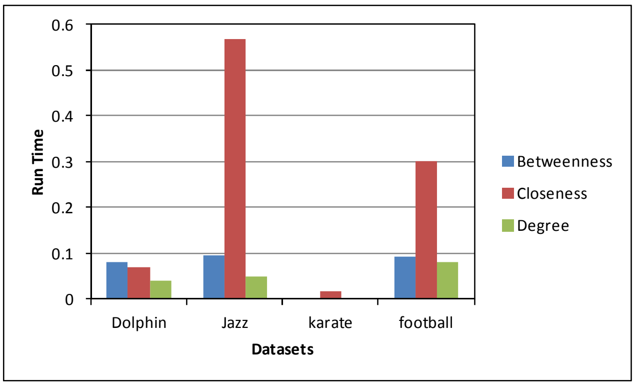

Fig. 8. Runtime

VI. CONCLUSION

In this paper, we proposed *Centrality_IBM* algorithm for solving IBM problem by finding a subset of positive influence initial nodes with high centralities for starting a positive diffusion to block negative one. We used MCICM diffusion model for spreading both negative and positive diffusions. Three different kinds of centralities including closeness, betweenness and degree centralities were used to find a subset of k positive initial nodes in IBM problem. The experiments on five different real datasets with different sizes showed that *Centrality_IBM* does better than random method, so The experiments showed that using centrality is a simple and efficient approach for IBM problem.

REFERENCES

[1] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01*, 2001, pp. 57–66.

[2] T. N. Dinh, Y. Shen, D. T. Nguyen, and M. T. Thai, "On the approximability of positive influence dominating set in social networks," *J. Comb. Optim.*, vol. 27, no. 3, pp. 487–503, 2014.

[3] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '07*, 2007, p. 420.

[4] A. Rezvanian, B. Moradabadi, M. Ghavipour, M. M. Daliri Khomami, and M. R. Meybodi, "Social Influence Maximization," in *Learning Automata Approach for Social Networks*, 2019, pp. 315–329.

[5] M. M. Daliri Khomami, A. Rezvanian, N. Bagherpour, and M. R. Meybodi, "Minimum positive influence dominating set and its application in influence maximization: a learning automata approach," *Appl. Intell.*, vol. 48, no. 3, pp. 570–593, Mar. 2018.

[6] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, 2017.

[7] L. Fan, Z. Lu, W. Wu, B. Thuraisingham, H. Ma, and Y. Bi, "Least Cost Rumor Blocking in Social Networks," in *2013 IEEE 33rd International Conference on Distributed Computing Systems*, 2013, pp. 540–549.

[8] E. Morozov, "Swine flu: Twitter's power to misinform," *Foreign policy*, 2009.

[9] K. F. Stanley Wasserman, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[10] W. Chen, Y. Yuan, and L. Zhang, "Scalable Influence Maximization in Social Networks under the Linear Threshold Model," in *2010 IEEE International Conference on Data Mining*, 2010, pp. 88–97.

[11] Y. Mashayekhi, M. R. Meybodi, and A. Rezvanian, "Weighted estimation of information diffusion probabilities for independent cascade model," in *2018 4th International Conference on Web Research (ICWR)*, 2018, pp. 63–69.

[12] W. Chen, Y. Wang, and S. Yang, "Efficient Influence Maximization in Social Networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, pp. 199–208.

[13] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Extracting influential nodes on a social network for information diffusion," *Data Min. Knowl. Discov.*, vol. 20, no. 1, pp. 70–97, 2010.

[14] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, 2010, p. 1029.

[15] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*, 2003, p. 137.

[16] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 665–674.

[17] M. Kimura, K. Saito, and H. Motoda, "Minimizing the Spread of Contamination by Blocking Links in a Network.," in *AAAI*, 2008, vol. 8, pp. 1175–1180.

[18] W. Chen *et al.*, "Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate," in *Proceedings of the 2011 SIAM International Conference on Data Mining*, 2011, pp. 379–390.

[19] X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model," in *Proceedings of the 2012 SIAM International Conference on Data Mining*, 2012, pp. 463–474.

[20] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz, "Containment of misinformation spread in online social networks," in *Proceedings of the 4th Annual ACM Web Science Conference*, 2012, pp. 213–222.

[21] S. Li, Y. Zhu, D. Li, D. Kim, and H. Huang, "Rumor restriction in online social networks," in *2013 IEEE 32nd International Performance Computing and Communications Conference (IPCCC)*, 2013, pp. 1–10.

[22] S.-C. Lin, S.-D. Lin, and M.-S. Chen, "A learning-based framework to handle multi-round multi-party influence maximization on social networks," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 695–704.

[23] H. Li, S. S. Bhowmick, J. Cui, Y. Gao, and J. Ma, "Getreal: Towards realistic selection of influence maximization strategies in competitive networks," in *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, 2015, pp. 1525–1537.

[24] W. Lu, W. Chen, and L. V. S. Lakshmanan, "From competition to complementarity: comparative influence diffusion and maximization," *Proc. VLDB Endow.*, vol. 9, no. 2, pp. 60–71, 2015.

[25] H.-C. Ou, C.-K. Chou, and M.-S. Chen, "Influence maximization for complementary goods: Why parties fail to cooperate?," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 1713–1722.

[26] G. Tong, W. Wu, S. Tang, and D.-Z. Du, "Adaptive influence maximization in dynamic social networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 112–125, 2017.

[27] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?," *IEEE Trans. Inf. theory*, vol. 57, no. 8, pp. 5163–5181, 2011.

[28] Z. Wang, W. Dong, W. Zhang, and C. W. Tan, "Rumor source detection with multiple observations: Fundamental limits and algorithms," in *ACM SIGMETRICS Performance Evaluation Review*, 2014, vol. 42, no. 1, pp. 1–13.

[29] A. Rezvanian and M. R. Meybodi, *Stochastic Social Networks: Measures and Algorithms*. LAP LAMBERT Academic Publishing, 2016.

[30] A. Rezvanian, A. M. Saghiri, S. M. Vahidipour, M. Esnaashari, and M. R. Meybodi, "Learning Automata for Complex Social Networks," in *Recent advances in learning automata*, 2018, pp. 279–334.

[31] A. Bavelas, "Communication patterns in task-oriented groups," *J. Acoust. Soc. Am.*, vol. 22, no. 6, pp. 725–730, 1950.

[32] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[33] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations," *Behav. Ecol. Sociobiol.*, vol. 54, no. 4, pp. 396–405, 2003.

[34] J. A. Frieler and M. Pfleiderer, *Inside the Jazzomat - New Perspectives for Jazz Research*. Schott Campus, 2017.

[35] W. W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, 1977.

[36] J. Leskovec and J. J. Mcauley, "Learning to Discover Social Circles in Ego Networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 539–547.