Proc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET)
9-10 December 2021, Cape Town-South Africa

Influence Maximization: Incorporating Homophily and Topological Overlap

David Oriedi
Department of Computer Science Dedan Kimathi University of Technology Nyeri, Kenya
david.opondo@dkut.ac.ke

Henry Nyongesa
Department of Computer Science Dedan Kimathi University of Technology Nyeri, Kenya
henry.nyongesa@dkut.ac.ke

George Musumba
Department of Computer Science Dedan Kimathi University of Technology Nyeri, Kenya
george.musumba@dkut.ac.ke

2021 International Conference on Electrical, Computer and Energy Technologies (ICECET) | 978-1-6654-4231-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICECET52533.2021.9698454

***Abstract*—Research on Influence Maximization has gained a lot of attention in the recent past. Part of the reason for this is that influence maximization has applications in commercially attractive areas such as word of mouth marketing. A majority of works in influence maximization have relied on information diffusion models that largely ignore the structural properties of the social network. The problem with this is that important attributes of user relationships which are necessary for approximating influence maximization are ignored. Parameters such as Homophily and Topological Overlap are crucial determinants of the level of influence that a user enjoys in the network. This work approximates global influence power of a user by first, considering user interactions, homophily and topological overlap as determinants of node to node relationship strength on a dynamic social graph. Then secondly, computes a global score of influence for each user. We then apply a novel algorithm that approximates influence spread for each influential user. The seed set is built by identifying the most influential users at specific time instances as the social graph evolves.**

***Index Terms*—Influence Maximization, Social influence, Homophily, Topological Overlap, Influential Nodes**

I. Introduction

There has been an exponential growth in social media usage by people across the globe. The net effect of this is enormous social network data that is generated during online interactions among social media users. Among other interests, there has been increased attention given to users who tend to have a substantial amount of influence over other users on social network platforms. This is because influential users get approached from time to time, by companies or political movements to endorse ideas or products in the hope that the online following that such influencers enjoy could translate into increased sales for a product, or widespread acceptance of a political ideology, through word of mouth marketing. From a social network research perspective, this scenario can be modeled using a social graph in which an initial set of influential users is chosen and the rest of the users are then activated or influenced through that initial set of users in a manner that maximizes the benefits. This is called influence maximization.

Influence maximization tries to solve the problem of identifying influential users on the social network such that through them, information can be spread furthest. For example, a company that is introducing a new product in the market may choose to give free samples of such a product to an initial number of influential individuals hoping that these individuals will in turn recommend the product to their friends and so on. The objective of influence maximization is to find a minimum set of users that will achieve the maximum spread of information about, or adoption of, a product or concept.

Given a directed weighted social graph $G = (V, E)$, a diffusion model M and an integer $k < |V|$ being the marketing budget, the influence maximization problem identifies k nodes through which the number of activated nodes throughout the graph is maximized [1]. According to [2], this problem is formally defined as:

$$IM_M(G, k) = argmax_e \subseteq V, |e| = k^{\sigma} M(e, G) \qquad (1)$$

where σ is a function that maximizes the influence spread achieved when the set of nodes in e are activated. In a nutshell, the problem of influence maximization aims to identify a set of k nodes, that is *seed set*, in an online social network with the maximum influence spread. Even though influence maximization has been applied in a number of areas such as viral marketing [3], network monitoring [4], and rumor control [5], there are still major challenges including modelling the process of information diffusion in a social network and dealing with the theoretical complexity of influence maximization problems in general [6].

Kempe $et\ al.$ [1] formalized the problem of influence maximization as a discrete optimization problem and designed approximation algorithms to that effect. However, the problem of influence maximization still remains NP-hard [7], and as a result, there have been numerous algorithmic efforts with a view to devising scalable and efficient influence maximization solutions [8].

In a majority of existing works, the social graph construction mostly relies on singular node attributes to create node relationships. The node to node relationship is represented through edge weights in a social graph. The edge weight values have mostly been determined through singular attributes such as stochastic values [1], information propagation actions such as replies, forwards or likes [9], [10] or topological positioning of the node within the network [11]. The problem with singular attributes is that they exclude equally important aspects of the social graph build and therefore may give an incomplete picture of the influence dynamics within the social graph. We therefore propose to define the edge weights through a combination of node attributes such as information propagation actions, homophily and topological overlap. Through this attribute combination, we will be able to establish node relationships that represent both structural and behavioral dimensions of the network.

The most popular information diffusion models are the Independent Cascade (IC) model and the Linear Threshold (LT) model. The two models were initially proposed by Kempe $et\ al.$ [1] for use in solving the influence maximization problem even though the solution turned out to be NP hard. As a result, several improvements of both models have been proposed including Cost Effective Lazy Forward (CELF) [4] which is 700 times faster than the standard greedy algorithm, CELF++ [12], Local Directed Acyclic Graph (LDAG) [13] designed to scale up for large networks, NewGreedy and MixGreedy [13]. However, both IC and LT models rely on stochastic values to represent node to node relationships [14] as opposed to using real values derived from user behaviours as happens in real world networks. In addition, the two information diffusion models are not designed for dynamic social networks [2] and have no clear criteria for choosing the initial seed set k.

The contributions of this paper are as follows: We propose to build a social graph based on a combination of three attributes namely user interactions, homophily and topological overlap. Thereafter, we define a diffusion model in which information spread from one node to the other is effected through information propagation actions such as forwarding, replying or liking the posts of users. Finally, instead of providing an initial seed set as input, our algorithm provides the seed set as one of the outputs. In this way, we provide an answer to the difficult question of selecting seed sets for initial information diffusion. The rest of this paper is organized as follows. Section I introduces the paper, Section II is a summary of related works. In section III, we define and formulate our model. Section IV discusses the results of our experiments while section V concludes the paper.

II. Related Works

Literature on influence maximization is rich with numerous algorithms each of which seek to improve previous ones. In particular, there have been specific efforts in improving the initial greedy algorithm proposed by [1]. The greedy algorithm selects a node with marginal gain increment maximization which gets added to the seed set. However, the greedy algorithm is time consuming when executed on a large social network [15]. In order to improve the speed of the greedy algorithm, several faster algorithms have been proposed including Cost Effective Lazy Forward (CELF) [4], CELF++ [12], NewGreedy [13], and MixGreedy [13]. However, even these improved algorithms cannot deal with large networks because of their time complexity [16]. Cheng $et\ al.$ [17] proposed an algorithm that works by first ranking the nodes then estimates the influence spread, even though the algorithm has large space complexity.

There are several works in which heuristics are applied in determining the influence spreads for influential nodes. For example, experimental results in [1] showed that choosing seed nodes with maximum degrees results in higher node influence spread for the selected nodes although not as high as in greedy approaches. In [13], the degree discount heuristics algorithm was proposed in which for an edge uv, if node u is already in the seed set, then the edge uv will not be considered for degree count. Kimura and Saito proposed the shortest-path influence cascade model being a special case of IC model. In this case, node activation only happens along shortest paths [18].

There are also works that have adopted community detection approaches in detecting influential nodes. The main motivation in this category of approaches is to reduce computational time and improve performance [19]. The idea is to first identify communities within the social network after which influential nodes within the communities are identified. Influence maximization on the whole network is then determined based on the identified set of influential nodes from the communities. In Li $et\ al$, [20], a model known as *Community-diversified influence maximization* model was proposed in which seed nodes are selected from as many different communities as possible based on the communities' previous contribution of seed nodes. Wang $et\ al.$, [21] proposed a community based greedy algorithm for mining top-K influential nodes in which top-k influential nodes were chosen from the communities through dynamic programming. Bozorgi $et\ al.$, [15] applied a community based approach to influence maximization involving competing influencers. In each community, they applied a simple greedy algorithm which uses the *Decidable Competitive Model* (DCM) as the propagation model to find the most influential node in a community at a time.

In summary, most of the existing works on influence maximization do not consider all node attributes at the initial level of social graph construction. In addition, even though there have been improvements, most of the algorithms that are derived from the greedy approach still suffer from high run times especially when the social network grows. We propose a discrete approach to selection of seed nodes in which the most influential nodes are identified at discrete time analysis of the social graph. In this way, the efficiency of our proposed algorithm is increased even though the social graph grows. Finally, our algorithm is designed for dynamic social networks with the most influential nodes identified, their respective values of influence spread approximated and other metrics are analyzed and reported at discrete points in time.

III. Model Description

A. Social Graph Description

We model the social network as a social graph $G = (V, E, W)$ in which V is a set of nodes, E is a set of edges and W is a set of edge weights. Initially the social graph is unweighted. To build the edge weight $w(i,j) \in W$, between the nodes i and j, we use three different parameters namely *user interactions, homophily and topological overlap* to dynamically build the edge weight as time evolves.

1) User Interactions: For this component of the edge weight, we consider different types of social action events. Event types include retweets, replies and likes (favorites). Each event e_k of type k, is assigned an importance weight α_k. The weight of importance is a reflection of how significant the event type is as an indicator of the influence of a node. To compute the edge weight using this parameter, we get a weighted sum of all the events at a particular time instance t, that is:

$$w_1(i,j)_t = \sum_{k=1}^{na} \alpha_k \, |\{e_k\}_t| \tag{2}$$

where na is the total number of event types. To show the dynamic nature of these interactions and its effect on node to node relationship strength, the edge weight is updated after every time interval δt.

2) Homophily: Prinstein $et\ al.$ [22] describes the concept of homophily as a case of similar individuals associating with one another more than dissimilar individuals. The similarity is based on attributes such as age group, religion, business or gender [23]. Therefore, to compute the second component of edge weight, we use a node attribute set $a_1, a_2, a_3, ..., a_n$ between the pair of nodes. Although the attributes vary in importance, their similarity is a good measure of how homophily influences associations. For this segment, we consider attributes such as age, gender, religion,and profession. Each of these attributes is associated with an importance factor, a_{ij}, by which the edge weight is increased if the attributes are similar for a pair of nodes. We formulate this increase as shown in equation (3):

$$w_2(i,j)_t = w_2(i,j)_t + a_{ij} \tag{3}$$

3) Topological Overlap: Topological overlap is a measure of the commonality among the neighbors shared by any two nodes [24]. The number of common friends for a pair of nodes can determine how far information travels in the social network. Our social graph is directed and weighted, and influence is inferred towards a node. In other words, we say that a post by node i has influenced node j if node j acts on that post by replying, forwarding or liking it. In this case therefore, our definition of the neighbor of a node is limited to those nodes that have in-links to the node in question. To compute topological overlap, we apply the formular proposed by Ravaz $et\ al$ [24].

$$w_3(i,j)_t = \frac{|n_t(i) \cap n_t(j)| + 1}{min(k_t(i), k_t(j)) + 1} \tag{4}$$

where $n_t(i)$ represents the neighbors of node i at time instance t and $k_t(i)$ is the in-degree of node i at time t. The added value 1 indicates that there is a direct link between node i and node j, otherwise it is omitted.

To compute the edge weight, $w(i,j)$, we combine all the contributing attributes in a linear combination. Each of the edge weight components is weighted with a trade off parameter $\beta \in [\,0,1]$ to balance the three edge weight components. Therefore the final edge weight between node i and node j at time t is:

$$w(i,j)_t = \beta_1 \cdot w_1(i,j)_t + \beta_2 \cdot w_2(i,j)_t + \beta_3 \cdot w_3(i,j)_t \tag{5}$$

where $\beta_1 + \beta_2 + \beta_3 = 1$. Put in other words, $w(i,j)_t$ represents the influence that node i have on node j at time t, also represented as $I(j,i)_t$ from this point hence forth.

B. Node Influence Power

The node influence power is a measure of the relative influence that the node has over its neighbors. It is a reflection of how the neighbors respond to the postings of a node through their propagation actions such as retweets, likes and replies. We adopt the personalized pagerank algorithm as proposed in [10] in order to calculate the influence power of a node. Although the personalized pagerank algorithm is popular for calculating global popularity of nodes, our model considers propagation actions as a major component of this important formular alongside other parameters i.e homophily and topological overlap. We argue that this consideration will include both the structural and behavioral aspects of node relationships. Therefore, the influence power $I_p(i)$ of a node i is calculated as shown in equation 6.

$$IP(i) = pg_1 + pg_2 \tag{6}$$

Where pg_1= $d \times \left(\sum_{j \in Followers(i)} \frac{I(j,i) \times IP(j)}{Followees(j)} \right)$ and pg_2= $(1-d)\frac{Followers(i)}{N}$;

d is the dumping factor, mostly set at 0.85, although Avrachenkov $et\ al.$ [25] argue that higher values of d might lead to a ranking that is highly sensitive to small pertubations on the structure of the network.

C. Node Influence Zone

An influence zone is an area within the social graph where a node has direct and indirect influence over other nodes. The Node Influence Zone is determined by tracing all the paths that have in-links to a node in the graph. For example if node A has an in-link from node B, and node B has an in-link from node C, which in turn has an in-link from node D, then the nodes B,C and D are said to be influenced by node A. This influence also includes other nodes that may be connected to nodes B, C or D directly or indirectly, so long as such connections are through in-links. Therefore, in the whole graph, any node that has at least one in-link can have an influence zone.

D. *Influence Maximization*

As has been explained in Section I, influence maximization is the process through which the influence of a set of nodes is maximized throughout the social network, usually for commercial purposes. In our model, we identify top most influential node at each time stamp. We therefore have at the end of the model run, a set of topmost nodes from each discrete time period. This set of nodes collectively form the seed node collection that can maximize the influence across the network.

We propose a novel approach in which we begin by identifying candidate nodes within the influence zone based on marginal increases on the average influence within the influence zone. At a given time duration, we loop through the graph to get the node that has the highest marginal change in its average zone influence. This node becomes the candidate node for influence maximization at that time. This process is repeated as the model continues to execute. At a point of saturation, i.e no significant change seen with subsequent results, we get a set of nodes that collectively form the most influential seed set. This set of nodes is what can be used to maximise influence within the network. Algorithm 1 outlines the steps of identifying the seed nodes for influence maximization.

Algorithm 1 Selective BFT for Influence Maximization

while true **do**
$list \leftarrow \emptyset$
for $k \leftarrow 1$ to N **do**
Compute Node Influence, I_n
for $k \leftarrow 1$ to N **do**
Compute Node Local Average Influence, I_{av}
if $I_n > I_{av}$ **then**
Compute Marg. Node Infl. Change, I'_{marg}
if $I'_{marg} > 0$ **then**
Add node to $list$
$CandidateNode \leftarrow Max(list)$
$Q \leftarrow \emptyset$
$Sp(i) \leftarrow \emptyset$
$visited(i)$
$Q.enqueue(i)$
while $(\neg empty(Q))$ **do**
Q.dequeue(i)
foreach $j \in adjacencylist(i)$ **do**
if $(edge.in()) \wedge (\neg visited())$ **then**
$Q.enqueue(j)$
$visited(j)$
$j.parent = i$
$Sp(i) = Sp(i) + 1$
return $Sp(i)$
end

IV. Simulation and Results

The model has been tested on simulated twitter social network data. The simulation generates three different twitter social actions namely retweets, replies and likes (favorites) during the simulated interactions. The users are modelled as nodes while the relationships among the nodes are modelled as edges between the nodes. There are 50 nodes, each having five attributes. In total there are 1000 edges. The model uses the social actions to define relationships among the nodes. In particular, each social action is associated with a weight that shows how important it is in influence definition [10]. A social graph is built that shows the intensity of information exchange among the nodes in the network.

To begin with, we compute the edge weight between each pair of nodes. The edge weight is made up of three different components, i.e user interactions, homophily and topological overlap. The model then proceeds to calculate the influence power for a node in the social network. A node's local influence is contributed to by the neighbors that have in-links to the node. The rationale for this is that we have defined influence in terms of social actions that are directed at the node which is has made a post on the social network.

The local influence value is used in the personalized pagerank algorithm in equation 6 to give a node's influence power. A personalized pagerank algorithm, also used in [10], gives the global representation of a node's influence on the whole social network. With a global influence value, we can tell how influential a node is in comparison with other nodes in the social network. Because our model is dynamic, the influence power of a node is computed at discrete time points, meaning that the influence power of a node at time t_0 may be different from its influence power at time t_1. The idea is to model what happens in real life social networks where one individual or topic can trend or be influential at a time t_0 but perform dismally a few hours or days later in terms of popularity. In other words, our model seeks to predict influential users in the network in a dynamic manner.

The final step in executing the model is to compute influence maximization. Here, we seek to establish how many nodes a given node can influence directly or indirectly across the network. Conceptually, influence maximization is the process of identifying the most influential nodes in the network (also known as seed nodes) for purposes of using them to pass information rapidly and efficiently across the social network. This concept provides companies with an opportunity to quickly market ideas, products or technologies to a huge segment of the population through the influential nodes. Determination of the most influential node is based both on its local and global influence capacities. To begin with, we get the local influence value of a node. Secondly, we calculate the global influence power through equation 6. When selecting nodes for influence maximization, it is not efficient to consider all nodes in the network as this will result in high computation time. Instead, at each discrete time point, we calculate a marginal change in node influence power since the last time stamp. If the change for a given node is positive, we apply a *Breath First Traversal (BFT)* algorithm on the path leading towards this node. The *BFT* algorithm is part of algorithm 1 in this paper. At a given time stamp, the node that is linked to the most nodes, as

TABLE I
NODE INFLUENCE SPREADS AT DIFFERENT TIMES.

Node Influence Spread		
Time Stamp	Node Id	Influence Spread
t_1	45	2
t_2	40	3
t_3	17	2
t_4	21	5
t_5	35	2
t_6	19	17
t_7	27	4
t_8	49	2
t_9	48	38
t_{10}	6	40
t_{11}	16	48
t_{12}	6	48
t_{13}	6	49
t_{14}	6	49
t_{15}	6	49

determined by *BFT*, becomes the most influential node at that time instance.

In summary, to determine the most influential node at time t_i, the social graph is updated with the most recent changes on the three parameters of influence *(user interactions, homophily and topological overlap)*, and the global influence power for each node is calculated. Based on the node influence power values at t_{i-1}, the marginal change in influence power is computed. If the marginal change in influence power is greater than zero, the *BFT* algorithm is applied on the nodes whose marginal change in influence power is greater than zero. This ensures that only nodes that are consistent in posting influential information on the network get to be considered for possible nomination into the seed set. At each time stamp, we only get one influential node, which is the node with the most nodes directly or indirectly influenced.

We will now discuss the results from the simulation experiment. This experiment was run on HP computer laptop EliteBook 840, with Core i7 CPU @2.10 GHz, 16GB RAM running windows 10. The development was done in python with Spyder IDE. Through the output, the experiment seeks to achieve the following: The dynamic identification of the most influential node at a particular time stamp, the comparison of node influence scores of our model and the scores coming from centrality measures.

As shown in Table I, the model dynamically reports the most influential node at different time stamps. As expected, the influence spread values are initially low but progressively grow as more activities are experienced on the social network. The influence spread scores eventually reach a saturation point at which there is no significant change in the values of the influence spread being reported by the nodes. According to these results, node 6 has emerged the most influential overall. However, each time stamp shows largely unique cases of influential nodes. This is what happens in typical real life social networks, as temporal influence varies according to topics, personalities, political subjects etc. Figure 1 shows the evolution of node influence spread values as time progresses.

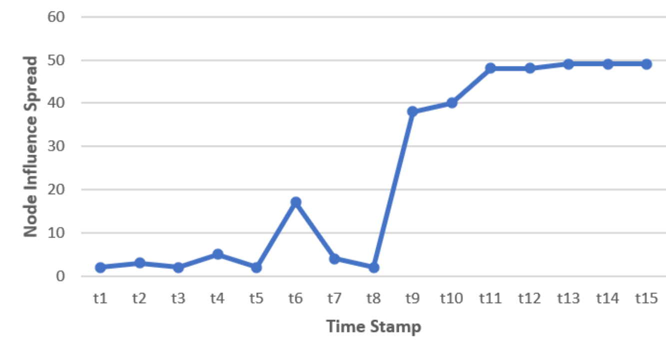

Fig. 1. Dynamic Influence Spread Measures

Notably, a saturation point is reached that shows the influence spread scores not improving beyond a certain point. This shows the modularity attribute of influence growth.

Figure 3 shows how our model compares with centrality measures on the impact of total wij on influence scores for selected nodes. From the figure, it is easy to see that the influence scores coming from centrality measures are not determined by the amount of social actions that take place between the nodes. This is the reason why centrality measures show little variance even though the graph is evolving in structure and behaviour. On the other hand, our model brings to the fore two significant aspects. One, the influence scores posted by the model keep changing thereby reflecting the flactuating nature of wij values - which represents the changing relationship strengths among the nodes in the network - therefore the evolving nature of tie strengths. Secondly, it can be seen that influence scores are not directly determined by values of wij, rather other factors are involved, namely homophily and topological overlap, both of which do not necessarily involve information exchange. In Figure 2, the sharp rise and fall of influence observed with changing number of followers shows that our model relies on the actual amount of information exchanged among the nodes as opposed to just the number of followers. A possible approach to smoothening these sharp rise and fall could be in performing a semantic filtering of the information exchanged among the nodes thereby attributing influence score only to social actions semantically relevant to the post made by a node.

V. CONCLUSION AND FURTHER WORK

In this work, we proposed a new approach for influence maximization on a dynamic social network. The approach employs a variant of the Pagerank algorithm and the Breadth First Traversal to determine the most influential seed set for influence maximization. This approach relies heavily on social actions, homophily and topological overlap - as determinants of node influence through which we have been able to select seed nodes. Simulation results have shown that influence scores computed based on actual social actions on the network reveal more in terms of dynamic changes within the network compared to centrality measures. Future research may consider applying natural language processing in analysing the actual

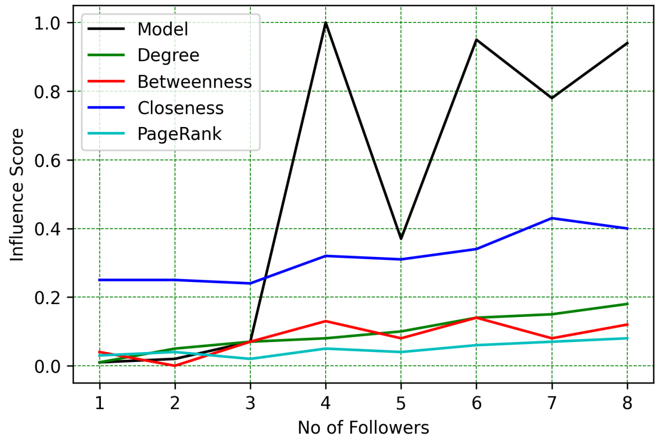

Fig. 2. Comparison of Followers and Influence scores

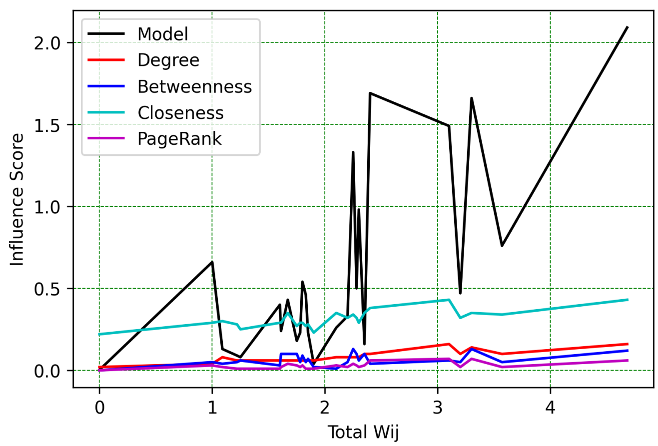

Fig. 3. Wij and Influence Scores

content of the information exchanged between network users in order to define node tie strengths and therefore influence, on the relevance and context of the information exchanged rather than treating all communications between nodes as similar.

References

[1] 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*, pp. 137–146, 2003.

[2] M. Jaouadi and L. B. Romdhane, "Influence maximization problem in social networks: An overview," in *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1–8, IEEE, 2019.

[3] 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*, pp. 57–66, 2001.

[4] 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*, pp. 420–429, 2007.

[5] 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*, pp. 665–674, 2011.

[6] Y. Li, J. Fan, Y. Wang, and K.-L. Tan, "Influence maximization on social graphs: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 10, pp. 1852–1872, 2018.

[7] S. Chen, J. Fan, G. Li, J. Feng, K.-l. Tan, and J. Tang, "Online topic-aware influence maximization," *Proceedings of the VLDB Endowment*, vol. 8, no. 6, pp. 666–677, 2015.

[8] C. Aslay, L. V. Lakshmanan, W. Lu, and X. Xiao, "Influence maximization in online social networks," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 775–776, 2018.

[9] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," in *International conference on knowledge-based and intelligent information and engineering systems*, pp. 67–75, Springer, 2008.

[10] M. Azaouzi and L. B. Romdhane, "An efficient two-phase model for computing influential nodes in social networks using social actions," *Journal of Computer Science and Technology*, vol. 33, no. 2, pp. 286–304, 2018.

[11] R. Geisberger, P. Sanders, and D. Schultes, "Better approximation of betweenness centrality," in *2008 Proceedings of the Tenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pp. 90–100, SIAM, 2008.

[12] A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++ optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web*, pp. 47–48, 2011.

[13] 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*, pp. 199–208, 2009.

[14] D. Oriedi, C. de Runz, Z. Guessoum, A. A. Younes, and H. Nyongesa, "Influence maximization through user interaction modeling," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pp. 1888–1890, 2020.

[15] A. Bozorgi, S. Samet, J. Kwisthout, and T. Wareham, "Community-based influence maximization in social networks under a competitive linear threshold model," *Knowledge-Based Systems*, vol. 134, pp. 149–158, 2017.

[16] L. Qiu, W. Jia, J. Yu, X. Fan, and W. Gao, "Phg: A three-phase algorithm for influence maximization based on community structure," *IEEE Access*, vol. 7, pp. 62511–62522, 2019.

[17] S. Cheng, H. Shen, J. Huang, W. Chen, and X. Cheng, "Imrank: influence maximization via finding self-consistent ranking," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 475–484, 2014.

[18] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," in *European conference on principles of data mining and knowledge discovery*, pp. 259–271, Springer, 2006.

[19] M. Jalayer, M. Azheian, and M. A. M. A. Kermani, "A hybrid algorithm based on community detection and multi attribute decision making for influence maximization," *Computers & Industrial Engineering*, vol. 120, pp. 234–250, 2018.

[20] J. Li, T. Cai, K. Deng, X. Wang, T. Sellis, and F. Xia, "Community-diversified influence maximization in social networks," *Information Systems*, vol. 92, p. 101522, 2020.

[21] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1039–1048, 2010.

[22] M. J. Prinstein and K. A. Dodge, *Understanding peer influence in children and adolescents*. Guilford Press, 2008.

[23] H. Zardi, L. B. Romdhane, and Z. Guessoum, "A multi-agent homophily-based approach for community detection in social networks," in *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, pp. 501–505, IEEE, 2014.

[24] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, "Hierarchical organization of modularity in metabolic networks," *science*, vol. 297, no. 5586, pp. 1551–1555, 2002.

[25] K. Avrachenkov, N. Litvak, and K. S. Pham, "A singular perturbation approach for choosing the pagerank damping factor," *Internet Mathematics*, vol. 5, no. 1-2, pp. 47–69, 2008.