



## A New PageRank-Based Method for Influence Maximization in Signed Social Networks

Maryam Hosseini-Pozveh<sup>a,\*</sup>, Maryam Lotfi Shahreza<sup>a</sup>

<sup>a</sup>*Department of Computer Engineering, Shahreza Campus, University of Isfahan, Iran.*

### ARTICLE INFO.

*Article history:*

**Received:** 22 January 2023

**Revised:** 29 April 2023

**Accepted:** 9 May 2023

**Published Online:** 5 July 2023

*Keywords:*

Influence, Centrality Measure, Distrust, Signed Networks, Node Ranking.

### ABSTRACT

Influence maximization is one of the most important topics in the social network analysis field. As all the social networks can be considered signed, explicitly or implicitly, assessing influence maximization in these networks is inevitable. Due to the NP-hard nature of this problem, the category of node-ranking-based solutions is of concern, where, the PageRank algorithm is outstanding. Original PageRank is merely defined based on the trust relationships and it is not applicable in signed social networks. Upon an agreement on the scheme of trust propagation, where trust propagates step by step in the social network, the two main schemes of distrust propagation are: a) distrust propagates step by step throughout the social network, and b) distrust propagates up to one step of the neighborhood. Despite the claims made by related researches that scheme (b) is the dominant behavior compared to (a); available PageRank algorithms are updated to incorporate scheme (a). In this study, a new PageRank-based method, which adopts scheme (b) to model the distrust-based influence propagation in signed social networks, is proposed. Accordingly, the importance of each node is computed considering that every node propagates the received influence from its trusted neighbors to other nodes, while it blocks the received influence from its untrusted neighbors. Assessments run on the three real datasets reveal the superiority of this proposed method over other existing PageRank algorithms in maximizing influence in signed social networks. The outperformance is between 22% to 46% considering all experimental settings in comparison with the most effective benchmark method.

## 1 Introduction

Social networks constitute a medium, where information, news, innovations, or influence spreads among its members [1]. As to influence spread, influence max-

imization, first introduced by Kemp et al. [1], is an NP-hard optimization problem defined as identifying a seed set of size  $k$  from a social network's nodes to trigger information diffusion based on an information diffusion model to maximize influence spread in the social network. Node ranking-based methods based on centrality measures are well-applied methods in selecting the top- $k$  nodes as the solution to this NP-hard problem. Among these methods, PageRank is a well-known and highly applicable algorithm [2–7].

\* Corresponding author.

Email addresses: [m.hosseini@shr.ui.ac.ir](mailto:m.hosseini@shr.ui.ac.ir) (M. Hosseini-Pozveh), [m.lotfi@shr.ui.ac.ir](mailto:m.lotfi@shr.ui.ac.ir) (M. Lotfi Shahreza)  
<https://dx.doi.org/10.22108/JCS.2023.136555.1119>  
ISSN: 2322-4460



Signed social networks model both positive relationships (friendship/trust) and negative relationships (hostility/distrust) between nodes. In this context, Epinions and Slashdot could be mentioned as real-world signed social networks [8]. The importance of the signed social network analysis is that even ordinary social networks can be considered signed implicitly, as both kinds of positive and negative relationships exist in all social networks. Accordingly, there exists important research filed in the social network analysis domain, with a focus on mining implicit negative relationships from metadata of the social networks, named sign mining or the negative link prediction field [9, 10]. Influence spreading and information diffusion models in these networks are more complex than in ordinary social networks due to the negative relationships' dynamics. There exists an agreed-upon scheme on how the influence spreads through trust relationships, which is a step-by-step propagation through the social network. To be more specific, every individual receiving the influence from their trusted neighbors tries to propagate it to their neighbors who trust or distrust it; therefore, trust-based influence propagates in the whole social network.

Unlike trust relationships, there exists no agreed-upon scheme for the influence spread through distrust relationships. According to the basic studies [11, 12], there exist two available schemes where diffusion through distrust relationships is of concern. According to the first scheme, a node  $v$  may prefer to be influenced in the opposite manner of node  $u$  when  $v$  is influenced by his/her untrusted neighbor  $u$ . Being influenced oppositely is objectified by adopting an opposite opinion: negative in the opposite of positive, and vice versa. The distrust propagation, the focus of this study, is different from the negative opinion propagation issue in social networks. While the negative opinion can be propagated in basic social networks as well, the distrust propagation issue is limited to signed social networks. However, positive and negative opinions propagation are applied here to model the opposite action based on the existence of the distrust relationships next to the trust relationships in the signed social networks. Adopting the opposite opinion leads to distrust propagation in more than one step of the neighborhood. It means that when a node  $v$  adopts the opposite opinion of his/her untrusted neighbor, the neighbors of  $v$  that distrust or trust  $v$ , adopt the opposite or the same opinion of  $v$ 's respectively. This process proceeds for further steps in a similar manner [11, 12]. This scheme is consistent with the balance theory, where, the statements "my enemy's enemy is my friend" and "my enemy's friend is my enemy" hold [8]. However, not all the theories about the signed social networks describe their behavior perfectly; thus,

other theories including weakly balanced theory and status theory can describe some behavioral aspects of the signed social networks more accurately [8].

According to the second scheme, node  $v$  would not tend to propagate the influence received from his/her distrusted neighbor. Accordingly, node  $v$  receives the distrust-based influence from  $u$  but does not propagate this influence to his/her neighbors who trust or distrust  $v$ . It means node  $u$  propagates distrust up to one neighborhood step, or in other words, distrust-based influence is blocked in node  $v$  [11, 12]. Based on this scheme, when a node  $v$  distrusts  $u$ , it discounts all the judgments made by  $u$  [11]. For instance, when  $u$  declares that he/she has seen a movie and may express his/her opinion about it,  $v$  does not express his/her opposite opinion about that movie as a reaction to  $u$ , but ignores that judgment without expressing his/her own opinion. Most online social networks like Twitter, Facebook, and Instagram even provide the ability to "block" the untrusted individuals explicitly or to "mute" them as a milder reaction than blocking. At the same time, many individuals prefer to monitor their untrusted peers for different reasons but do not propagate the received influence from them.

According to the related studies in this context [11, 12], the second scheme is the dominant behavior adopted by individuals in social networks. To date, the proposed PageRank algorithms for the signed network are based on the first scheme [13–15]. In this study, as a contribution, a new PageRank algorithm is proposed and discussed based on the second scheme. The experimental assessments are run on three real datasets to evaluate the effectiveness of this proposed PageRank for influence maximization in signed social networks.

The rest of this article is organized as follows: The related works are reviewed in Section 2. In Section 3, the proposed PageRank algorithm in signed social networks is presented. The details of the experiments on three real-world datasets are presented in Section 4. Finally, the study is concluded in Section 5.

## 2 Related Works

### 2.1 Influence Maximization

Influence-spreading applications including viral marketing in social networks have introduced important issues, one of the most important being influence maximization (IM) [1]. This NP-hard combinatorial optimization problem is to identify a seed set of nodes with a predefined size of  $k$ , where most possible final influence in the social network through an information diffusion model is generated [1]. Independent



cascade (IC) [1], and linear threshold (LT) [1] models are the two primary progressive information diffusion models in IM. Unlike the symmetric models where the state of the nodes can change more than once, in progressive information diffusion models, the state of the nodes can change only once. In this study, the progressive information diffusion models are of concern, because they expose the dynamics of real-world information diffusion more accurately; where the situation of awareness about a piece of news that could not change to not-awareness is a case of progressive information spread [16].

The fitness function of the IM problem, named influence function, is the final number of the nodes influenced by the seed set in the social network [1]. The influence function of the IM problem is increasing monotone and submodular based on the most information diffusion models like IC and LT [1]. A greedy hill climbing-based approach leads to identifying the best near-optimal solution for an NP-hard optimization problem when the fitness function of that problem is increasing monotone and submodular [1]. Accordingly, two main greedy hill-climbing-based frameworks exist in the related literature for solving the IM problem: 1) the Mont-Carlo simulation-based approach presented by Kempe et al. [1], and 2) the random reverse reachable set-based approach presented by Borgs et al. [17]. Improving the run-time efficiency of these two frameworks without losing the effectiveness of result is one of the main research directions in the IM problem context. The extended versions of the Mont-Carlo simulation-based approach [1] in this context are the CELF [18], MixGreedy [19], LDAG [20], CELF++ [21], SimPath [22], PMIA [23], StaticGreedy [24], UBLF [25], and ASIM [26], and extended versions of the random reverse reachable set-based approach [17] are the TIM and TIM+ [27], IMM [28], Stop-and-Stare [29], BCT [30], and TipTop [31]. Because the random reverse reachable set-based approach [17] is more run-time efficient than the Mont-Carlo simulation-based approach [1], the extended versions of the first are more run-time efficient than the second, this makes the TipTop [31] the most run-time efficient among all the mentioned methods.

Although the previous two categories of methods are theoretically supported in presenting the best near-optimal solution for IM, the essence of run-time efficiency and the need to apply scalable methods for IM in large social networks has shifted the researchers' attention to the heuristics-based methods in solving the IM. Heuristics-based methods are more run-time efficient in comparison with previously mentioned methods, while they lack theoretical support in assuring effectiveness considering monotone submodular information diffusion models. Nevertheless, among

these methods some yield results as effective as the previous categories of methods based on experimental assessments [32, 33]. The degree, closeness, betweenness, and PageRank centralities are the widely-applied centrality measures [3] in the IM problem. Accordingly, the  $k$  nodes with the highest centrality values constitute the seed set. Improving the effectiveness of these centrality measures is addressed in some studies including SingleDiscount and DegreeDiscountIC [19], diffusion degree [34], Group-PageRank [35], PRDiscount [36], local index rank (LIR) [37], weighted k-shell degree neighborhood method [38], influence-aware closeness centrality, and influence-aware betweenness centrality measures [39].

Because the greedy-based methods do not assure the effectiveness of results for none-monotone and none-submodular information diffusion models, the heuristics-based methods due to their run-time efficiency become the superior candidates for IM in comparison with greedy-based methods. As most signed information diffusion models lack the two monotone and submodular properties, applying the heuristics-based methods is more practical here.

## 2.2 Information Diffusion Models for Signed Social Networks

In general, information diffusion models are presented by considering the existence of only trust, friendship or, positive relationships in the social networks, while, distrust, enmity, or negative relationships are the essential complementary relationships that must be considered in the social networks' dynamics. As to previously mentioned models, progressive information diffusion models in signed social networks are presented in cascade-based [12, 40–43] or threshold-based [12, 44, 45] categories.

A comprehensive study on signed information diffusion models is presented in [12], where the two cascade-based and two threshold-based information diffusion models are proposed by considering the two schemes about how information may spread through distrust relationships. Based on the first scheme, distrust may spread in more than one step of neighborhood, that is a node tends to adopt the opposite opinion of its untrusted neighbor: negative in opposite of positive, and vice versa. Based on the second scheme, distrust may spread up to one step of neighborhood, that is a node may tend to block the influence received from its distrusted neighbor [12]. According to these proposed schemes, in the cascade-based categories, the sign-aware cascade including blocking nodes (SC-B) model based on the second scheme, Sign-aware cascade (SC) model based on the first scheme, and in the threshold-based category, trust-generated thresh-



old including blocked nodes (TG-T-B) model based on second scheme, and the trust-generated threshold including negative nodes (TG-T-N) model based on the first scheme, are proposed [12]. Comparison of these two schemes through comparing proposed models by considering real-word datasets reveals that second scheme is the one with closer performance to real-world instinct of influence propagation [12].

Among other studies which address the signed information diffusion models, the independent cascade diffusion model in the signed network (named SNIC) [40], influence propagation under the IC model [41], and IC-P model [42]), are cascade-based models according to the first scheme. Notably, these three models are similar to the SC model presented in [12]. The two proposed threshold-based models based on the first scheme, are the linear threshold model extended for signed networks [44], and the polarity-related Linear Influence Diffusion (PLID) model [46]. Considering the second scheme, the PID model [43] is presented as a cascade-based model, which is similar to the SC-B model presented in [12].

Unlike the information diffusion models in unsigned social networks like IC and LT models, the monocity and submodularity of influence function may not hold when distrust relationships are considered next to the trust ones in information diffusion models [12]. Accordingly, the advantage of greedy-based algorithms over heuristics-based algorithms in assuring the best near-optimal solution for influence maximization is not established anymore for signed information diffusion models. This point indicates the importance of applying the heuristics-based methods as being very run-time efficient solutions in this context. Among these, PageRank [14, 46, 47] being one of the most popular and effective methods with many applications for influence maximization in unsigned information diffusion models [35, 36] and the signed ones [13–15], is of concern in this study.

### 2.3 PageRank and its Extensions for Influence Maximization in Signed Social Networks

PageRank algorithm is a popular and effective random walker-based method for ranking nodes in complex and social networks. Through this algorithm, the influence of a node, is measured by considering the influence of its neighbors in the network based on the whole network's link structure [13, 14, 46, 47]. Accordingly, an iterative procedure is applied to calculate a PageRank matrix for a network through equation 1 [14, 46, 47]:

$$PR_{t+1} = d.PR_t.F + [(1-d)/N, \dots, (1-d)/N]^T \quad (1)$$

where,

- $N$  is the number of nodes of the social network.
- $PR_{t+1}$  is the PageRank vector in iteration  $t + 1$ .
- $PR_t = [PR_{1,t}, \dots, PR_{N,t}]^T$ , where  $PR_{i,t}$  is the PageRank value of node  $i$  in iteration  $t$ .
- The matrix  $F$  is obtained based on the following network adjacency matrix:

$$\begin{bmatrix} f_{1,1} & f_{1,2} & \dots & f_{1,N} \\ f_{2,1} & f_{2,2} & \dots & f_{2,N} \\ \dots & \dots & \dots & \dots \\ f_{N,1} & f_{N,2} & \dots & f_{N,N} \end{bmatrix}$$

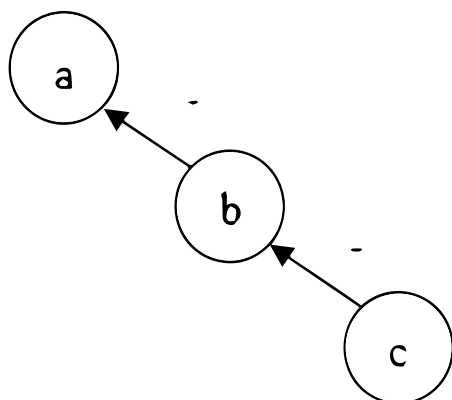
where, if there is a directed edge from node  $j$  to node  $i$ ,  $f_{i,j}$  is  $1/d_j^{out}$ , otherwise it takes the value 0; where  $d_j^{out}$  is the output degree of node  $j$ .

- The value of  $d$  is named the damping factor, and 0.85 is a common choice for it in the literature.
- First, at iteration 0,  $PR_{i,0}$  for all the nodes are set to  $1/N$ , next, through some iterations, the final value of  $PR_t$  is computed. Symbol  $t$  is the final iteration if the condition  $\max\|PR_t - PR_{t-1}\| < \theta$  holds.

In the basic PageRank algorithm, all the relationships between nodes are considered trust, friendship, or positive relationships. Because distrust, animosity, or negative relationships exists implicitly in all social networks, there exist social networks that model these relationships explicitly. These social networks are named signed networks. Due to existence of both positive and negative relationships in signed networks, the basic PageRank algorithm is not applicable to compute influence of the nodes. Accordingly, improving the PageRank method to incorporate negative relations into the calculations is addressed in some studies. These studies are reviewed with their limitations here, as follows:

Integrated PageRank is proposed by Chen and Hee [13] where the PageRank is extended to rank the nodes in the influence maximization problem in the signed social networks. In this method, the principal network graph is divided into two positive and negative subgraphs. The positive subgraph contains all the nodes but only positive edges of the social network, while the negative subgraph contains all the nodes but only negative edges; for every node, according to equation 1, the positive PageRank ( $PR_i^+$ ) and the negative PageRank ( $PR_i^-$ ) values are calculated separately in the subgraphs, respectively. The final PageRank value, named the Integrated PageRank, is calculated by applying  $PR_i^+ - PR_i^-$  for all nodes  $i$ ,  $1 \leq i \leq N$ , where  $N$  is the number of the nodes of the social network. According to this idea, the influence of a node is obtained by calculating the positive and negative influ-





**Figure 1.** Part of the Negative Subgraph of a Signed Network.

ence values separately, and subtracting the negative influence value from the positive influence value. Integrated PageRank is an attempt to extend PageRank based on the first scheme about influence diffusion through negative edges. However, a notable drawback exists in the computation of the  $PR_i^-$ . Here the idea is that, how much is the negative influence of a node, is measured considering how much is the negative influence of its negative neighbors based on the whole negative network's link structure. The problem with the corresponding calculations can be illustrated in Figure 1, where node  $a$ , its negative relationship with node  $b$ , and the negative relationship of node  $b$  with node  $c$  are evident. According to Figure 1, these calculations operate against the balance theory where "the enemy of my enemy is my friend" [8] and considers the enemy of the enemy as the enemy, therefore, the assumptions in the balance theory and its corresponding scheme, the first scheme, does not hold.

Mohammadinejad et al.[15] developed a method named POPRank for ranking users in signed social networks. Individuals' personalities, comprised of optimism and pessimism based on positive and negative out degrees of every individual are of concern in POPRank. Accordingly, a modified version of the PageRank is presented incorporating the optimism and pessimism scores of every node in calculations. To compute the POPRank, the social network graph is divided into two subgraphs, one with positive, and one with negative links, where,  $POPRank^+$  and  $POPRank^-$  are computed respectively. The final POPRank for every node is computed as  $POPRank^+ - POPRank^-$ . Even though applying the idea of the personality of the nodes, the exact problem with the Integrated PageRank [13], discussed before, prevails.

Signed PageRank Algorithm (SPA) proposed by Yin.

et al, [14] is another extension of the PageRank to rank nodes to influence maximization. In this algorithm, the belief of every node is updated in every iteration, based on the receiving influence from trusted and untrusted neighbors by considering the positive and negative signs of the edges. The first scheme about distrust propagation in the whole social network is trust propagation, in which when an action of a node  $u$ , which is untrusted for  $v$ , occurs,  $v$  tends to take the action in opposition, which is adopted in this method. Being applied based on the first scheme is the limitation of this method, while second scheme is the dominant behavior in influence spreading through distrust relationships.

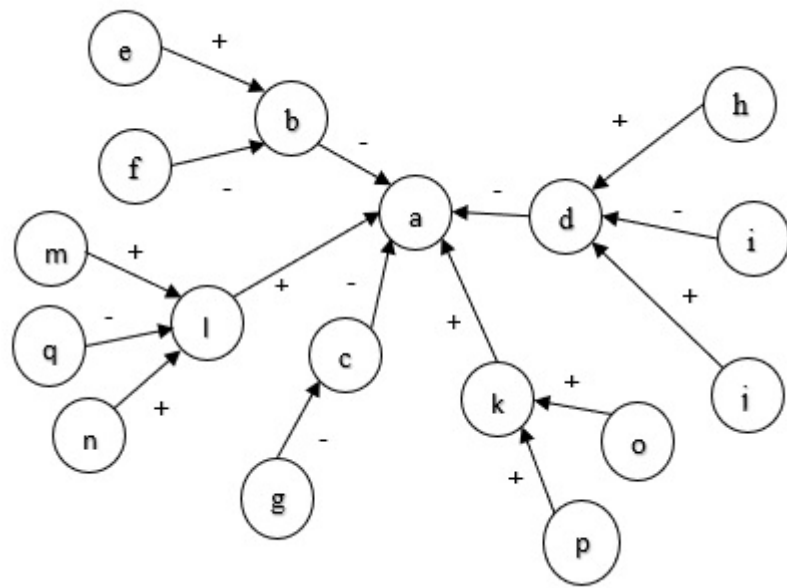
All the extended PageRank algorithms for signed social networks are presented based on the scheme about the influence propagation through negative edges that distrust may spread in more than one step of neighborhood. The dominant behavior in distrust spreading is presented in the second scheme: the tendency of nodes to block the receiving negative influence from their distrusted neighbors. Accordingly, distrust may propagate up to one step of the neighborhood in the social network [11, 12]. In this context, presenting an improved version of PageRank considering a second scheme for influence maximization in signed social networks is addressed in this study.

### 3 Proposed Method

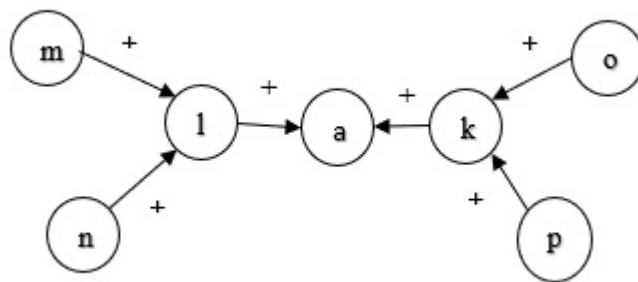
The dominant behavior about influence spreading through negative relationships is that when an untrusted node of  $v$  performs an action, that node  $v$  tends not to do it, or in other words, to block the received influence from propagation to other nodes. In this scheme, the distrust-based influence is spread across the social network in one step at the most [11, 12]. This scheme happens in situations such as when  $u$  declares that he/she has seen a movie and  $v$  that distrusts  $u$ , ignores the judgment made by  $u$  without expressing his/her own judgment on the subject. Based on this notion, a new PageRank algorithm is proposed for signed networks named the Blocked-Based PageRank, BPageRank, with the following details:

- First, the two positive and negative subgraphs are extracted from the main subgraph of the network. The positive subgraph includes all the nodes with only the positive edges of the social network, while the negative subgraph includes all the nodes with only negative edges. These two subgraphs are named  $G^+$  and  $G^-$ .
- As all the PageRank-based measures are computed in iterations, a set of iterations begins where three values are computed for every node  $i$

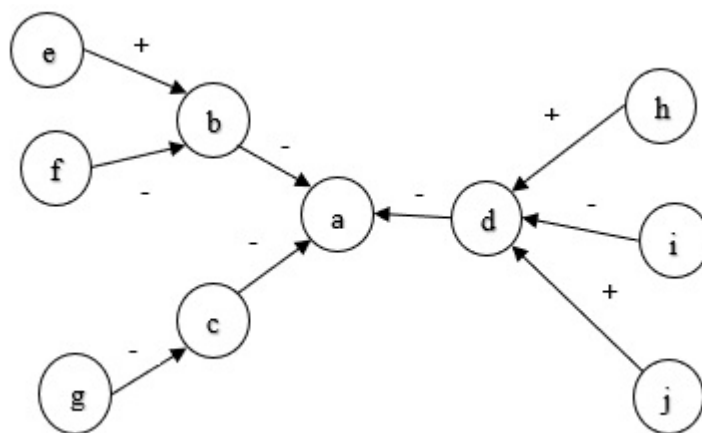




(a)



(b)



(c)

**Figure 2.** a) Example of a Graph  $G$  to illustrate the node  $a$ 's PageRank calculation, b) The graph's view of  $G$ , where the  $PR_a^+$  is calculated c) The graph's view of  $G$ , where the  $PR_a^-$  is calculated.



in every iteration  $t$ : the positive PageRank value  $PR_{i,t}^+$ , the negative PageRank value  $PR_{i,t}^-$ , and the overall PageRank value  $PR_{i,t}$ . The initial values of the  $PR_{i,0}$ ,  $PR_{i,0}^+$ , and  $PR_{i,0}^-$  are set at  $1/N$  for all the nodes.

- For each node  $i$  in  $G^+$ , the positive influence value in iteration  $t$ ,  $PR_{i,t}^+$ , is calculated according to the main PageRank equation (equation 1), but by applying only the positive relationships of the social networks.

Having an example graph  $G$  in Figure 2a which is focused on computing the PageRank value of node  $a$ , its positive subgraph is observed in Figure 2b. The  $PR_{i,t}^+$  would be calculated for the graph in Figure 2a following Figure 2b.

- To calculate the negative PageRank, the following equation is proposed:

$$PR_{i,t+1}^- = d \cdot PR_{i,t} \cdot F^- + [(1-d)/N, \dots, (1-d)/N]^T \quad (2)$$

where, the  $F^-$  is obtained by applying the negative network subgraph expressed in the following matrix:

$$\begin{bmatrix} f_{1,1}^- & f_{1,2}^- & \dots & f_{1,N}^- \\ f_{2,1}^- & f_{2,2}^- & \dots & f_{2,N}^- \\ \dots & \dots & \dots & \dots \\ f_{N,1}^- & f_{N,2}^- & \dots & f_{N,N}^- \end{bmatrix}$$

where,  $f_{i,j}^-$  is set to  $1/d_j^{out,-}$  if there is a directed edge in the negative subgraph from node  $j$  to node  $i$ , otherwise it is set to 0; where  $d_j^{out,-}$  is the negative output degree of node  $j$  (the output degree of node  $j$  in the  $G^-$ ).

According to this equation, the negative PageRank value of node  $a$  is calculated by considering only its negative neighbors. The more the overall PageRank value of a negative neighbor of node  $a$ , the more the negative influence (the negative PageRank value) of node  $a$  in the social network.

An example of a graph's view, where the  $PR_a^-$  is calculated, is shown in Figure 2c., where, while in step one of the neighborhoods, only the negative relationships are considered, in other steps of neighborhoods, all the negative and positive relationships are considered in calculating of this value. This is because in calculating the overall PageRank value of nodes, all the negative and positive income edges are applied, and in  $PR_a^-$  calculation for a node, this overall PageRank value of one-step income negative neighbors of that node is considered.

- After calculating both positive and negative PageRank values in iteration  $t + 1$ , the PageRank value of that iteration for every node  $i$ ,  $PR_{i,t+1}$ , is calculated as  $PR_{i,t+1}^+ - PR_{i,t+1}^-$ . This value

is applied in the  $PR_i^-$  calculations in the next iterations. Moreover, at the end of iterations, this value is adopted as the final PageRank value of the nodes.

- Symbol  $t$  is the final iteration if the condition  $\max_i ||PR_{i,t} - PR_{i,t-1}|| < \theta$ , for all the nodes  $i$  holds.

### 3.1 Time Complexity

In every iteration, the time complexity of computing the  $PR_{i,t}$  for all the  $n$  nodes of the social network, having  $PR_{i,t}^+$  and  $PR_{i,t}^-$ , is  $O(n)$ . In every iteration, computing  $PR_{i,t}^+$  value of all the nodes is  $\sum_{v_i \in N} d_i^{in,+}$ , and the time complexity of computing  $PR_{i,t}^-$  value of all the nodes is  $\sum_{v_i \in N} d_i^{in,-}$ , where  $N$  is the set of the nodes of the social network, and  $d_i^{in,+}$  and  $d_i^{in,-}$  are the positive income degree, and negative income degree of the node  $i$ . Accordingly, the overall time complexity for computing both  $PR_{i,t}^-$  and  $PR_{i,t}^+$  values of all the nodes is  $\sum_{v_i \in N} d_i^{in,+} + \sum_{v_i \in N} d_i^{in,-} = \sum_{v_i \in N} d_i^{in} = m$ . Considering the overall number of iterations as  $t$ , the overall time complexity of this proposed PageRank is  $O((m + n)t)$ . This time complexity is as scalable as the time complexity of the original PageRank; thus, it is proper to be applied in social networks of different sizes.

## 4 Evaluation

Experiments are run to address the following main question:

- Can the Blocked-Based PageRank algorithm presented based on the second scheme of influence propagation through negative relationships, better determine the importance of nodes in comparison with the benchmark PageRank-based algorithms presented based on the first scheme of the influence propagation through negative relationships?

All the implementations are run in Python, and by applying the NetworkX library. Experiments are run on a 2.5 GHz Intel core i7 processor with 16 GB memory.

Other issues of concern about the experiments are presented as follows:

#### The datasets:

The three real world datasets applied in this study are:

- Wikipedia Requests for Adminship dataset [48]: In this dataset, the information of 10,835 nodes and 159,388 voter/votee relationships are main-



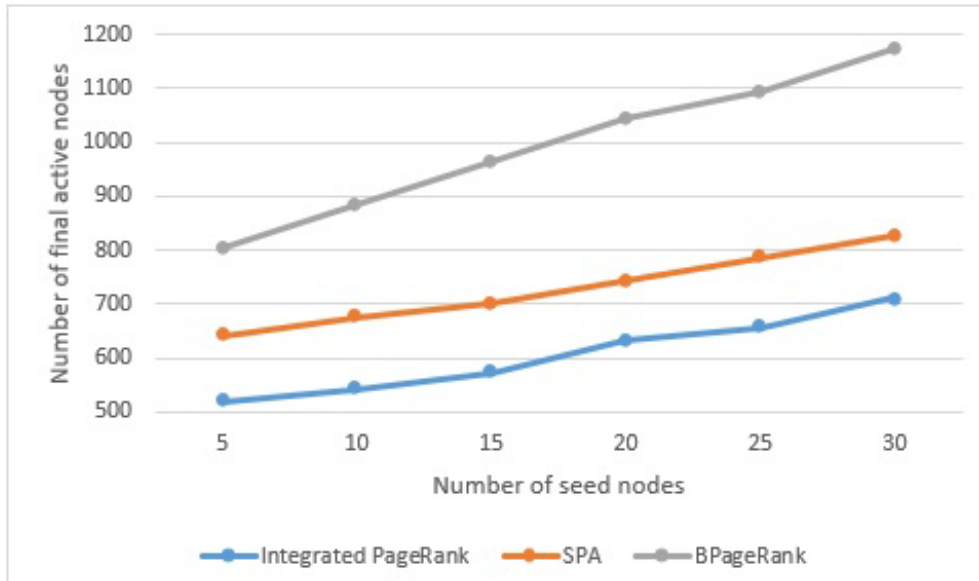


Figure 3. Final Influence Spread on the Wikipedia Requests for Adminship Dataset on the SC-B Model.

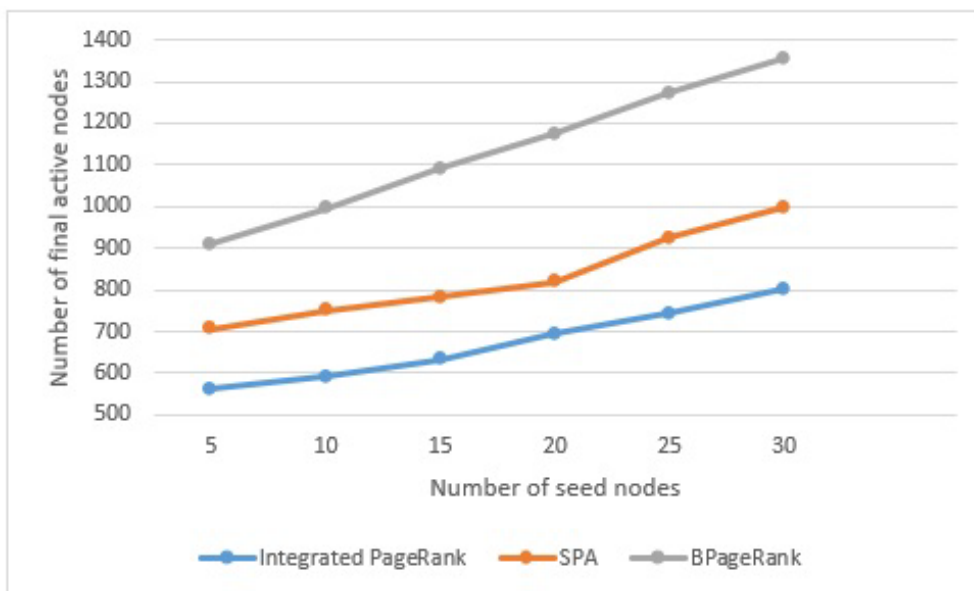


Figure 4. Final Influence Spread on the Wikipedia Requests for Adminship Dataset on the TG-T-B Model.

tained, where, the votes are within the -1 to +1 range.

- Epinions dataset [49]: In this dataset, the information of 132000 nodes and 841372 trust/distrust edges are maintained.
- Slashdot dataset [8]: In this dataset, the information of 82140 nodes and 549202 friend/foe edges are maintained.

#### The benchmark methods:

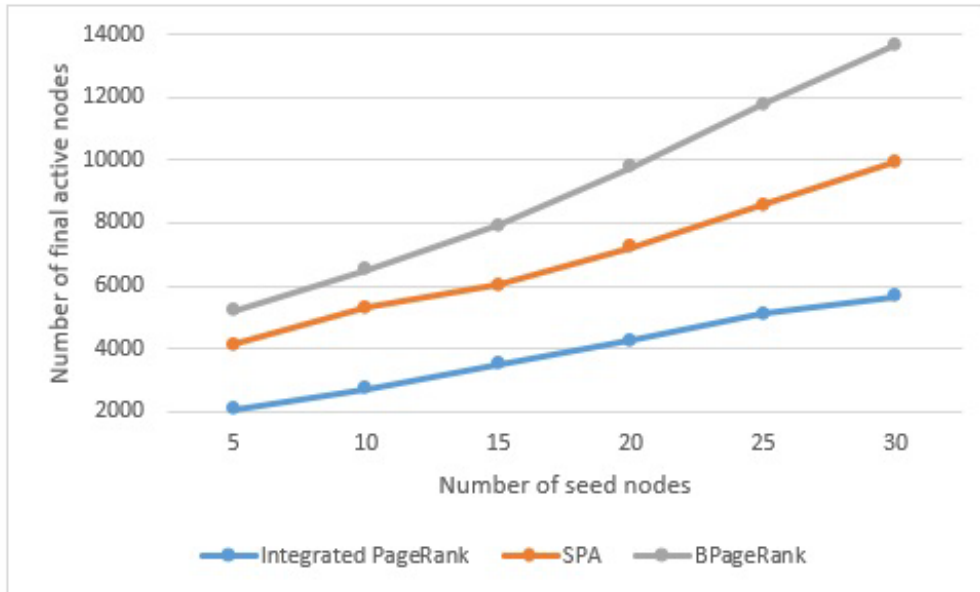
Three algorithms, the Integrated PageRank [13] POPRank [15], and Signed PageRank Algorithm (SPA) [14], which are PageRank-Based algorithms are

presented based on the first scheme of the negative propagation through the distrust relationships. As discussed, both the Integrated PageRank [13], and POPRank [15] methods implement the first scheme improperly. Among them, the Integrated PageRank [13] is selected as one of the benchmark methods. Moreover, the Signed PageRank Algorithm (SPA) [14] is selected as another benchmark method to evaluate the performance of the Blocked-Based PageRank algorithm (BPageRank).

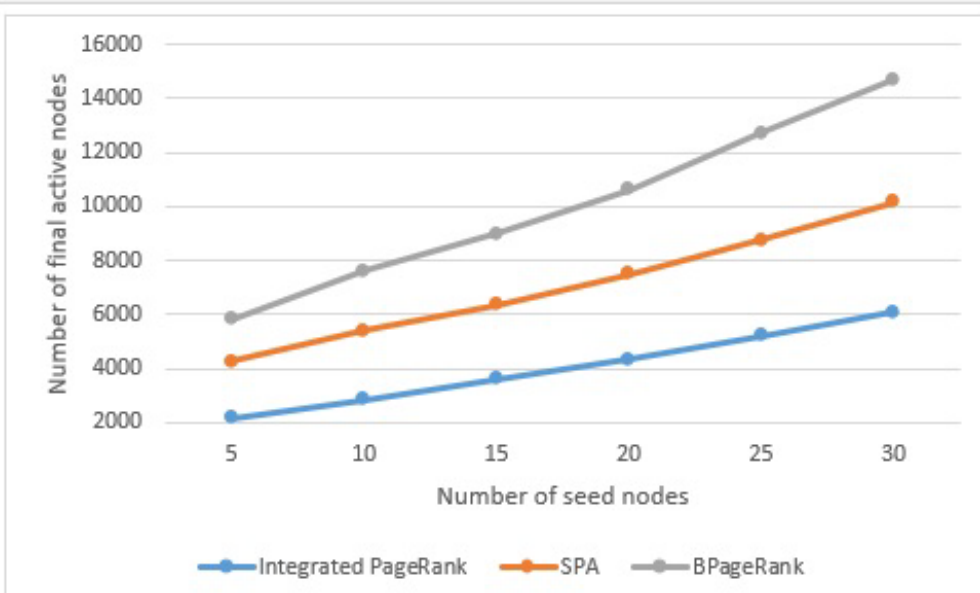
#### The comparison measure:

To compare the subject methods, the number of





**Figure 5.** Final Influence Spread on the Epinions Dataset on the SC-B Model.



**Figure 6.** Final Influence Spread on the Epinions Dataset on the TG-T-B Model.

the final active nodes obtained through simulating the influence spread based on an information diffusion model, activated by the selected seed nodes, is of concern. Accordingly, one cascade-based information diffusion model, SC-B [12], and one threshold-based information diffusion model, TG-T-B [12] are adopted here.

### Results and Discussion:

The evaluations are made through the following steps:

- (1) The top  $k$  nodes are identified through each

- mentioned method on every dataset separately.
- (2) The selected top seed nodes are applied to trigger the information diffusion subject to SC-B and TG-T-B information diffusion models.
- (3) The final activated nodes are counted and are applied in comparing the methods.

The results of final influence are presented in Figures 3 and 4 for Wikipedia Requests for the Adminship dataset, in Figures 5 and 6 for Epinions dataset, and in Figures 7 and 8 for Slashdot dataset by considering SC-B and TG-T-B models respectively.

As observed in these figures, this proposed method,



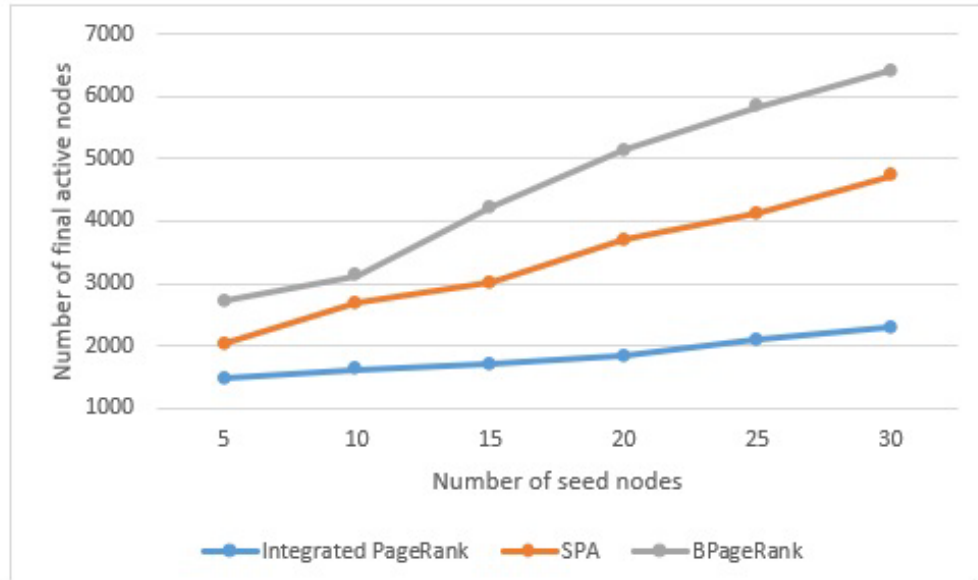


Figure 7. Final Influence Spread on the Slashdot Dataset on the SC-B Model.

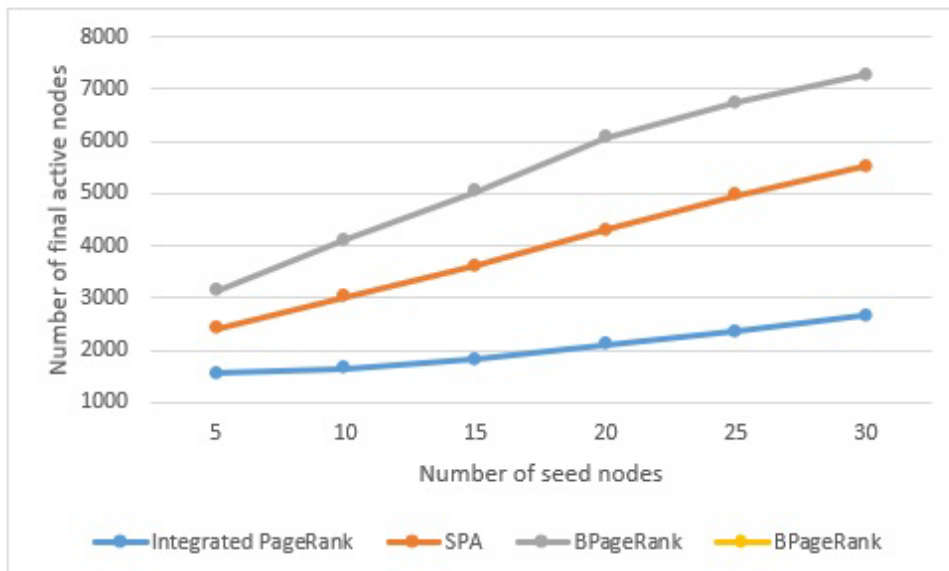


Figure 8. Final Influence Spread on the Slashdot Dataset on the TG-T-B Model.

the BPageRank, outperforms SPA and Integrated PageRank methods in all datasets and for both cascade-based and threshold-based information diffusion models. This outperformance is between 22% to 46% for different seed size, for different datasets subject to different information diffusion models in comparison with SPA, and between 54% to 187% for different seed sizes, for different datasets, and subject to different information diffusion models in comparison with Integrated PageRank. In this regard, at seed size=30, BPageRank outperforms SPA at 42% and 35% subject to SC-B and TG-T-B models respectively, and outperforms Integrated PageRank

at 65% and 69% subject to SC-B and TG-T-B models respectively, for Wikipedia dataset. At seed size=30, BPageRank outperforms SPA at 37% and 44% subject to SC-B and TG-T-B models respectively, and outperforms Integrated PageRank at 142% and 141% subject to SC-B and TG-T-B models respectively, for Epinions dataset. At seed size=30, BPageRank outperforms SPA at 36% and 32% subject to SC-B and TG-T-B models respectively, and outperforms Integrated PageRank at 179% and 174% subject to SC-B and TG-T-B models respectively, for Slashdot dataset. The difference in percentage increase is due to the different characteristics of the networks, and



the different dynamics of the information diffusion models. These findings are consistent with the basic studies [11, 12] about the dominant behavior of individuals confronting the receiving influence from their distrusted neighbors. Blocking the negative influence in the way which is considered in BPageRank, yields more accurate results in comparison with SPA which considers propagating negative influence in more than one step when identifying influential nodes. Intenerated PageRank yields less effective results because the negative influence propagation in more than one step of neighborhoods is not applied properly.

## 5 Conclusions

Considering the importance of influence maximization in signed social networks, this study presents an approach for ranking nodes based on the PageRank algorithm to propagate influence as broadly as possible in these networks, by triggering the top-k nodes. Diffusion through negative relationships could be interpreted in two ways: 1) the first scheme, which is based on the fact that nodes accept the opposite opinion with an untrusted neighbor and trigger the spreading of this piece of influence all across the network, and 2) the second scheme, which is based on the fact that nodes tend to block the influence received from an untrusted neighbor and stop propagating it in the social network. To the best of the authors' knowledge, the three existing PageRank methods which have been proposed for IM in signed social network, are based on the first scheme, while the second scheme describes the dominant behavior about influence propagation through distrust relationships in signed social networks. Accordingly, a new PageRank method, named the BPageRank, is proposed based on the second scheme. The evaluations on three existing real data sets reveal the outperforms of this method over the existing benchmark methods.

Assessing the other existing theories about signed social networks, including weakly balanced theory and status theory, to improve the effectiveness of this proposed method is a future study of concern. Applying the influence awareness in PageRank-based methods to IM in signed social networks is another important research theme. Assessing the applications of this proposed algorithm in other areas of signed social networks could be another research issue.

## References

- [1] D. Kempe, J. Kleinberg, and E. 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*, pages 137–146, 2003. doi:10.1145/956750.956769.
- [2] LC. Freeman. *Centrality in social networks: Conceptual clarification*. Social network: critical concepts in sociology. Londres: Routledge, 2002.
- [3] A. Landherr, B. Friedl, and J. Heidemann. A critical review of centrality measures in social networks. *Wirtschaftsinformatik*, 52:367–382, 2010. doi:10.1007/s11576-010-0244-0.
- [4] U. Kang, S. Papadimitriou, J. Sun, and H. Tong. Centralities in large networks: Algorithms and observations. In *Proceedings of the 2011 SIAM international conference on data mining*, pages 119–130. SIAM, 2011. doi:10.1137/1.9781611972818.11.
- [5] S. Banerjee, M. Jenamani, and DK. Pratihari. A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62: 3417–3455, 2020. doi:10.1007/s10115-020-01461-4.
- [6] X. Gong, H. Yu, and T. Yu. Literature review on the influence of social networks. In *SHS Web of Conferences*, volume 153, page 01009. EDP Sciences, 2023. doi:10.1051/shsconf/202315301009.
- [7] F. Gammoudi, M. Sendi, and MN. Omri. A Survey on Social Media Influence Environment and Influencers Identification. *Social Network Analysis and Mining*, 12(1):145, 2022. doi:10.1007/s13278-022-00972-y.
- [8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1361–1370, 2010. doi:10.1145/1753326.1753532.
- [9] J. Tang, Y. Chang, Ch. Aggarwal, and H. Liu. A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)*, 49(3):1–37, 2016. doi:10.1145/2956185.
- [10] M. Ozer, MY. Yildirim, and H. Davulcu. Negative link prediction and its applications in online political networks. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pages 125–134, 2017. doi:10.1145/3078714.3078727.
- [11] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412, 2004. doi:10.1145/988672.988727.
- [12] M. Hosseini-Pozveh, K. Zamanifar, and AR. Naghsh-Nilchi. Assessing information diffusion models for influence maximization in signed social networks. *Expert Systems with Applications*, 119: 476–490, 2019. doi:10.1016/j.eswa.2018.07.064.
- [13] Sh. Chen and K. He. Influence maximization on signed social networks with integrated



- pagerank. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, pages 289–292. IEEE, 2015. doi:10.1109/SmartCity.2015.86.
- [14] X. Yin, X. Hu, Y. Chen, X. Yuan, and B. Li. Signed-PageRank: An efficient influence maximization framework for signed social networks. *IEEE Transactions on Knowledge and Data Engineering*, 33(5):2208–2222, 2019. doi:10.1109/TKDE.2019.2947421.
- [15] A. Mohammadinejad, R. Farahbakhsh, and N. Crespi. Employing personality feature to rank the influential users in signed networks. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 346–353. IEEE, 2016. doi:10.1109/BDCloud-SocialCom-SustainCom.2016.59.
- [16] Y. Ni, L. Xie, and Z. Liu. Minimizing the expected complete influence time of a social network. *Information Sciences*, 180(13):2514–2527, 2010. doi:10.1016/j.ins.2010.03.012.
- [17] Ch. Borgs, M. Brautbar, J. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 946–957. SIAM, 2014. doi:10.1137/1.9781611973402.70.
- [18] J. Leskovec, A. Krause, C. Guestrin, Ch. 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*, pages 420–429, 2007. doi:10.1145/1281192.1281239.
- [19] 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*, pages 199–208, 2009. doi:10.1145/1557019.1557047.
- [20] 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*, pages 88–97. IEEE, 2010. doi:10.1109/ICDM.2010.118.
- [21] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. Celf++ optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48, 2011. doi:10.1145/1963192.1963217.
- [22] A. Goyal, W. Lu, and VS. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *2011 IEEE 11th international conference on data mining*, pages 211–220. IEEE, 2011. doi:10.1109/ICDM.2011.132.
- [23] Ch. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25:545–576, 2012. doi:10.1007/s10618-012-0262-1.
- [24] S. Cheng, H. Shen, J. Huang, G. Zhang, and X. Cheng. Staticgreedy: solving the scalability-accuracy dilemma in influence maximization. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 509–518, 2013. doi:10.1145/2505515.2505541.
- [25] Ch. Zhou, P. Zhang, J. Guo, and L. Guo. An upper bound based greedy algorithm for mining top-k influential nodes in social networks. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 421–422, 2014. doi:10.1145/2567948.2577336.
- [26] S. Galhotra, A. Arora, S. Virinchi, and Sh. Roy. Asim: A scalable algorithm for influence maximization under the independent cascade model. In *Proceedings of the 24th International Conference on World Wide Web*, pages 35–36, 2015. doi:10.1145/2740908.2742725.
- [27] Y. Tang, X. Xiao, and Y. Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 75–86, 2014. doi:10.1145/2588555.2593670.
- [28] Y. Tang, Y. Shi, and X. Xiao. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1539–1554, 2015. doi:10.1145/2723372.2723734.
- [29] Hung T Nguyen, My T Thai, and Thang N Dinh. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In *Proceedings of the 2016 international conference on management of data*, pages 695–710, 2016. doi:10.1145/2882903.2915207.
- [30] H. Nguyen, M. Thai, and TN. Dinh. A billion-scale approximation algorithm for maximizing benefit in viral marketing. *IEEE/ACM Transactions On Networking*, 25(4):2419–2429, 2017. doi:10.1145/2882903.2915207.
- [31] X. Li, J. Smith, T. Dinh, and MT. Thai. Tip-top:(almost) exact solutions for influence maximization in billion-scale networks. *IEEE/ACM Transactions on Networking*, 27(2):649–661, 2019. doi:10.1109/TNET.2019.2898413.
- [32] Y. Li, J. Fan, Y. Wang, and KL. Tan. In-



- fluence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872, 2018. doi:10.1109/TKDE.2018.2807843.
- [33] S. Banerjee, M. Jenamani, and D. Pratihari. A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62:3417–3455, 2020. doi:10.1007/s10115-020-01461-4.
- [34] S. Pal, S. Kundu, and CA. Murthy. Centrality measures, upper bound and influence maximization in large scale directed social networks. *Fundamenta Informaticae*, 130(3):317–342, 2014. doi:10.3233/FI-2014-994.
- [35] Q. Liu, B. Xiang, E. Chen, H. Xiong, F. Tang, and J. Yu. Influence maximization over large-scale social networks: A bounded linear approach. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 171–180, 2014. doi:10.1145/2661829.2662009.
- [36] Y. Wang, B. Zhang, V. Athanasios Vasilakos, and J. Ma. Prdiscount: A heuristic scheme of initial seeds selection for diffusion maximization in social networks. In *Intelligent Computing Theory: 10th International Conference, ICIC 2014, Taiyuan, China, August 3-6, 2014. Proceedings 10*, pages 149–161. Springer, 2014. doi:10.1007/978-3-319-09333-8\_17.
- [37] D. Liu, Y. Jing, J. Zhao, W. Wang, and G. Song. A fast and efficient algorithm for mining top-k nodes in complex networks. *Scientific reports*, 7(1):1–8, 2017. doi:10.1038/srep43330.
- [38] A. Namtirtha, A. Dutta, and B. Dutta. Weighted kshell degree neighborhood: A new method for identifying the influential spreaders from a variety of complex network connectivity structures. *Expert Systems with Applications*, 139:112859, 2020. doi:10.1016/j.eswa.2019.112859.
- [39] M. Hosseini-Pozveh. New influence-aware centrality measures for influence maximization in social networks. *Journal of Computing and Security*, 9(2):23–38, 2022. doi:10.22108/jcs.2022.134191.1100.
- [40] W. Liu, X. Chen, B. Jeon, L. Chen, and B. Chen. Influence maximization on signed networks under independent cascade model. *Applied Intelligence*, 49:912–928, 2019. doi:10.1007/s10489-018-1303-2.
- [41] J. Sheng, L. Chen, Y. Chen, B. Li, and W. Liu. Positive influence maximization in signed social networks under independent cascade model. *Soft Computing*, 24:14287–14303, 2020. doi:10.1007/s00500-020-05195-x.
- [42] D. Li, Zh. Xu, N. Chakraborty, A. Gupta, K. Sycara, and Sh. Li. Polarity related influence maximization in signed social networks. *PloS one*, 9(7):e102199, 2014. doi:10.1371/journal.pone.0102199.
- [43] C. Qu, J. Bi, and G. Wang. Personalized information diffusion in signed social networks. *Journal of Physics: Complexity*, 2(2):025002, 2021. doi:10.1088/2632-072X/abd5a9.
- [44] X. He, H. Du, M. Feldman, and G. Li. Information diffusion in signed networks. *PloS one*, 14(10):e0224177, 2019. doi:10.1371/journal.pone.0224177.
- [45] D. Li and J. Liu. Modeling influence diffusion over signed social networks. *IEEE Transactions on Knowledge and Data Engineering*, 33(2):613–625, 2019. doi:10.1109/TKDE.2019.2930690.
- [46] S. Kim and S. Lee. An improved computation of the pagerank algorithm. In *Advances in Information Retrieval: 24th BCS-IRSG European Colloquium on IR Research Glasgow, UK, March 25–27, 2002 Proceedings*, volume 2291, pages 73–85. Springer, 2002.
- [47] W. Xing and A. Ghorbani. Weighted pagerank algorithm. In *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004.*, pages 305–314. IEEE, 2004. doi:10.1109/DNSR.2004.1344743.
- [48] R. West, H. Paskov, J. Leskovec, and Ch. Potts. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2:297–310, 2014. doi:10.1162/tacl.a\_00184.
- [49] P. Massa, K. Souren, M. Salvetti, and D. Tomasoni. Trustlet, open research on trust metrics. *Scalable Computing: Practice and Experience*, 9(4), 2008.





**Maryam Hosseini-Pozveh** received her B.Sc. degree in software engineering and her M.Sc. degree in software engineering from the University of Isfahan, Isfahan, Iran, in 2005 and 2009, respectively. She obtained her Ph.D. degree in the software engineering field from the University of Isfahan in 2016. Currently, she is an assistant professor in the Department of Computer Engineering, at the Shahreza Campus, University of Isfahan. Her main areas of research interest include Complex networks, Data Science, and Advanced algorithms.



**Maryam Lotfi Shahreza** received her Master's degree in Computer Engineering (software) at Isfahan University of Technology in 2006. She received her Bachelor's degree in Computer Engineering (Algorithm and computational sciences) at the University of Tehran in 2009. She received her Ph.D. in Computer Engineering (software) at Isfahan University of Technology in 2018. Her research interests include issues related to Data Sciences, Text mining, Complex Networks, Bioinformatics, and Computational Biology, especially Systems Biology and Pharmacology Sciences.

