Journal of Computing and Security

Research Article, 2022, Volume 9, Number 2 (pp. 23–38)

http://www.jcomsec.org

New Influence-Aware Centrality Measures for Influence Maximization in Social Networks

Maryam Hosseini-Pozveh [a,*]

[a] *Department of Computer Engineering, Shahreza Campus, University of Isfahan, Iran.*

ARTICLE INFO.

Article history:
Received: 25 June 2022
Revised: 11 September 2022
Accepted: 18 September 2022
Published Online: 10 December 2022

Keywords:
Influence-Based Centrality Measure, Betweenness, Closeness.

ABSTRACT

Influence maximization in social networks is defined as determining a subset of seed nodes where triggering the influence diffusion through the social network leads to the maximum number of final influenced nodes. The tradeoff between the runtime efficiency and effectiveness in the quality of response is the main issue in presenting solutions for this NP-hard optimization problem. Centrality-based methods are applied as a category of efficient heuristic-based solutions to this problem. The two leading causes of losing effectiveness in centrality-based methods are 1) only the link structure and non-awareness of influence diffusion are considered in determining the importance of nodes, and 2) influence overlap exists among selected seed nodes. To address the first cause, an influence-aware betweenness centrality measure is proposed considering both IC and LT models. Moreover, an existing influence-aware closeness centrality measure for LT model is adopted to cover both LT and IC models. To address the second cause, a greedy-based method is proposed by applying influence-aware centrality measures to identify the influential nodes, next to proposing a Jacquard-based measure to overcome the influence overlap problem. The experiments consist of two parts where two real-world datasets are applied: 1) the proposed influence-aware centrality measures are compared with their original versions, and 2) the greedy-based method is compared with benchmark methods. The results indicate the effectiveness of the influence-aware centrality measures and the proposed greedy-based method in maximizing the influence spread in social networks.

1 Introduction

Influence propagates broadly through social networks; it is of essence to consider this behavior. In this context, influence maximization (IM) [1] is one of the primary research challenges. IM is the problem of selecting the K seed nodes which propagate the influence as broadly as possible in the social network graph. In IM, diffusion models are applied to shape influence propagation. Two main diffusion models are the independent cascade (IC) [1] and the linear threshold (LT) [1]. IM is considered a combinatorial optimization problem with NP-hard complexity, where its influence function (the number of final influenced nodes) for both IC and

* Corresponding author.

Email address: m.hosseini@shr.ui.ac.ir (M. Hosseini-Pozveh)

https://dx.doi.org/10.22108/JCS.2022.134191.1100

ISSN: 2322-4460

LT models is increasing monotone and submodular [1]. Due to the existence of these two features, a greedy hill-climbing-based approach, presented by Kempe et al [1], can determine the best near-optimal seed set for this problem. Despite its high effectiveness, the greedy hill-climbing approach has low efficiency regarding its high time complexity. Researchers sought to propose more efficient and scalable methods to maximize the influence of social networks [2, 3]. Accordingly, there exist methods based on the concept of ranking the nodes considering a centrality measure in social networks, where, after ranking nodes, the K top nodes are selected as the seed set. The advantage of these methods is their improving efficiency though their effectiveness is weaker than greedy hill-climbing-based approaches. In this category, degree, closeness, betweenness, and PageRank centralities [4] constitute the well-known classic measures widely applied. Improvement in the effectiveness of mentioned centrality measures is addressed by many researchers. Some of these studies presented the extensions of degree centrality. SingleDiscount and DegreeDiscountIC are heuristics presented by Chen et al. [5] extending the degree centrality. While the SingleDiscount is applicable for all information diffusion models, DegreeDiscountIC is just applicable for the IC model. Local index rank (LIR) [6] is another measure extending the degree of centrality. There exist proposed measures that extend the PageRank measure, including the ones by Xiang et al. [7], Liu et al. [8], and Wang et al. [9], where all are presented subject to a specific information diffusion model. In the research by Hosseini-pozveh et al. [10], an influence-aware closeness centrality is proposed under the IT model.

Regarding the applications of centrality measures in the social network analysis, including IM, the concept of influence-aware centrality measures is extended here by considering the closeness and betweenness centrality.

The contributions of this study are briefed as follows:

- The influence-aware closeness centrality for the LT model, proposed by Hosseini-pozveh et al. [10], is adopted to cover both LT and IC models.
- An influence-aware version of betweenness centrality is proposed for both LT and IC models.
- A greedy-based method is proposed to find the most influential nodes. Accordingly, the proposed influence-aware centrality measures are applied for ranking the nodes, and a Jacquard-based measure is proposed and applied to decrease the influence overlap.
- The experimental assessments are run on two real datasets in two parts: first, the effectiveness of proposed influence-aware centrality measures in comparison with their classic versions is assessed, and next, the effectiveness of the proposed greedy-based method comparing the benchmark methods is assessed.

The rest of this paper is organized as follows. The related works are reviewed in Section 2. The proposed algorithms, including influence-aware closeness and betweenness measures and the greedy-based method, are presented in Section 3. In Section 4, the assessments on influence-aware centrality measures and the greedy-based method are presented. Finally, the study is concluded in Section 5.

2 Related Works

Identifying a K-size seed set of nodes to propagate the influence as broadly as possible in a social network graph based on an information diffusion model is known as the influence maximization (IM) problem [1]. Due to its different applications, such as in viral marketing [1], IM is gaining momentum among researchers. Independent cascade (IC) and linear threshold (LT) models are the two main information diffusion models in the IM domain in social networks. In both models, nodes are in active or inactive states and can change only from inactive to active. An active node means a node that has accepted the influence. At first, all the nodes are inactive except the initial seed set expressed by A_0. The activation process continues in discrete steps until no other activation in the social network is possible [1].

In the IC model, every newly activated node v in step t seeks to activate its inactive neighbors. A neighbor w of node v may be active with the $p_{v,w}$ probability. Without enough social network metadata, this parameter is set to the same value for all existing edges. When multiple newly activated neighbors seek to activate w, an arbitrary random order becomes of concern. If v succeeds in activating w, then w would become active in step $t+1$; otherwise, v would not activate w in future steps [1].

In the LT model, the threshold value for v to be influenced by its all active neighbors is determined by θ_v within the [0, 1] range. This parameter is set to a randomly selected value for every node when there are not enough social network metadata. The influence of a neighbor node w on node v is determined through a $b_{v,w}$, weight where $\sum_{(\text{w neighbours of v})} b_{v,w} \leqslant 1$. At every step t, every node v where its activation formula $\sum_{(\text{w neighbours of v})} b_{(v,w)} \geqslant \theta_v$ is held, will become active in addition to the active nodes of previous steps [1].

The IM optimization problem is of NP-hard time complexity. The fitness function of IM is named the influence function and is equivalent to the final number of influenced nodes by the seed set. The influence functions of IC and LT models have two specific increasing monotone and submodular properties. A greedy hill-climbing approach can provide the best near-optimal solution for an NP-hard optimization problem with a monotone submodular fitness function [1]. In this respect, the greedy hill-climbing approach presented by Kempe et al. [1] is one of the ground solutions in the literature for the IM problem. This algorithm runs in k iterations, each for selecting one of the nodes of the seed set. Throughout the iteration, the node which maximizes influence is selected as the new member of the seed set by evaluating all of the nodes of the social network. The influence function of a seed set is computed through the Monte Carlo simulation method [1]. The disadvantage of this method is its efficiency in the run time. The time complexity of this method, $O(knmR)$, where n is the number of nodes, m is the number of edges, R is the iterations of simulation to compute influence function, and k is the seed set size, is high.

Among the available research directions regarding the IM problem scope [2, 3], the main one is to present efficient and scalable methods for large social network graphs while being effective in the final influence as well. CELF [11], MixGreedy for the IC model [5], LDAG [12], CELF++ [13], SimPath for the LT model [14], PMIA for the IC model [15], StaticGreedy for the IC model [16], UBLF [17], and ASIM for the IC model [18], provide more efficient methods based on the mentioned greedy-based solution.

Borgs et al. [19] presented a new method for IM based on the random reverse reachable set concept, constituting a more efficient solution for this problem than the Mont-Carlo simulation-based solution. In this study, first, a hyper-graph is constructed from the underlying network by generating a few samples. For generating every sample, a random node is selected, and the nodes that influence it are determined and inserted as new edges into the hyper-graph. This process is conducted based on the depth-first search beginning from the randomly selected node and the traverse probability of the edges. Following this, the $top-k$ nodes are selected as the k nodes with the highest degree from this hyper-graph in a greedy order. After selecting each k node, the selected node and its connected edges are removed from the hyper-graph. Some researchers have extended this method to present more efficient methods for IM: TIM and TIM+ [20], IMM [21], Stop-and-Stare [22], BCT [23], and TipTop [24].

The other category of studies in the IM solutions domain is named heuristics-based solutions. The focus is on improving the methods' run time efficiency and scalability in determining the influential nodes. As heuristics-based methods, in theory, they do not assure effectiveness concerning the final influence, but many reveal experimentally competitive results with previously mentioned categories of methods [2, 3]. Basic centrality measures [4], including degree centrality, closeness centrality, betweenness centrality, and PageRank centrality, have been and are being applied to the IM problem to determine the $top-k$ nodes with the highest centrality values. The drawbacks of these basic heuristics methods include 1) only the link structure is considered in computing the value of these measures, not the information diffusion process [2], 2) the selected $top-k$ nodes by these measures may have influenced overlap [2].

SingleDiscount and DegreeDiscountIC, proposed by Chen et al. [5], are the first attempts where the limitations of basic centrality measures are addressed. SingleDiscount is an extended version of the degree centrality measure considering influence overlap issue in a one-step neighborhood. In SingleDiscount, after selecting a node as a seed node, its edges to its neighbors are of no concern when comparing the degree of other nodes to select other seed nodes. DegreeDiscount is based on the degree centrality measure, where, through selecting the seed nodes, a degree discount is assigned to every node considering the number of up-to-now selected seed nodes in its one-step neighborhood and the influence probability of the IC model. Xiang et al. [7] proposed a linear influence model approximating the IC model, followed by applying the PageRank measure to determine the upper bounds of the influence of nodes. Pal et al. [25] proposed a heuristics measure named diffusion degree based on the independent cascade model, where the diffusion degree of every node is the influence of the same node on its neighbors plus the influence of its neighbors on their neighbors considering the influence probability. By adopting the linear influence model [7], Liu et al. [8] proposed a Group-PageRank measure that computes the PageRank of a group of nodes. Accordingly, the Group-PageRank of a set of nodes is the sum of the PageRank of its member nodes with the discount of the influence overlap in one step of the neighborhood between the members. This research proposes a greedy framework, GroupPR, to determine the seed set based on the Group-PageRank measure. Wang et al. [9] proposed the PRDiscount measure, a PageRank-like heuristic scheme based on the weighted cascade (WC) information diffusion model. According to this measure, first, the PageRank of the nodes is computed considering the weight of the edges; next, nodes are selected as the members of the seed set in a greedy manner, where after selecting each seed node, the PageRank of its one-step neighbors are decreased by a discount value. Liu et al. [6] proposed the local index rank (LIR) to rank the importance of nodes. Based on this measure, the nodes with higher degrees than their neighbors are selected as local leaders. LIR decreases the influence overlap though not an influence-aware measure. In Hosseini-pozveh et al. [10], a method is proposed to compute the closeness value of the nodes based on the LT model; then, this measure is combined with a community-detection-based framework for IM. A weighted kshell degree neighborhood method is presented by Namtirtha et al. [26], where both the k-shell and degree are applied but are not influence-aware. Jia et al. [27] proposed a tri-phase-based heuristics method to determine the influential nodes through the IC model, where, in the first phase, the K-shell method is applied to rank the nodes and in the second phase, a weighted degree is applied to distinguish the importance of nodes with the same K-shell values. Finally, the redundant neighbors with low similarity are determined by applying a similarity index between the newly selected seed and its neighbors. The objective of this method, though not influence-aware, is to decrease the overlapping problem.

IM in social networks needs solutions that should be 1) effective considering final influence, and 2) run-time efficient. The hill-climbing-based greedy solution [1] and random reverse reachable set-based greedy solution [19] theoretically assure the best near-optimal solution for IM. Because the second solution constitutes a more efficient ground solution than the first, most new studies with a focus on runtime efficiency, subject to theoretical effectiveness, have proposed methods where the second ground solution is extended. To date, the TipTop [24] is the most runtime-efficient method suggested with effective final influence supported in a theoretical sense. The objective of the third category containing the heuristics-based methods is to determine the seed set efficiently, even if losing quality is anticipated in the final influence. Although these methods lack the theoretical support to assure effectiveness, they are subject to attention in the IM field due to the importance of runtime issues. Presenting methods in this category to gain a proper tradeoff between effectiveness and efficiency is an essential research challenge addressed in this study. In this context, the two closeness and betweenness centralities are assessed as the bases of an effective and efficient method for IM.

3 Proposed Method

Applications of heuristics-based centrality measures in IM make it necessary to present improved versions of these measures. The focus of this study is on the closeness and betweenness centralities. To accomplish this, first, the closeness-aware centrality measure based on the LT model proposed by Hosseini-pozveh et al. [10] is adopted; next, its extension to the IC model is discussed. Moreover, the influence-aware betweenness centrality measure is proposed for both the IC and LT models. These measures are extendable to any information diffusion model. These two measures address the issue that classic centrality measures are presented based on the information of link structure and ignore the information diffusion. The influence-aware closeness and betweenness are global measures because they define the importance of the node globally in the social network. These proposed measures do not address the influence overlap problem. Accordingly, a two-step Jacquard-based measure is proposed. Combing this measure with the influence-aware centrality measure in a greedy-based manner leads to the determination of the effective influential nodes. Consequently, both the influence diffusion and the influence overlap are addressed in this greedy-based method.

3.1 Live-Edge View of IC and LT Information Diffusion Models

Computation of the influence-aware versions of closeness and betweenness is conducted based on the outcomes of the live-edge [1], where every outcome is equivalent to every iteration of the Monte Carlo simulation of the information diffusion process. Constructing a live-edge outcome for the IC and LT models is described as follows:

- Generating the Live-edge outcome for the IC model: First, activation or no activation of all the edges is determined at once according to the influence probability of every edge; next, a sub-graph of the social network is constructed, including all the nodes of the initial social network and the selected edges as the live-edges. In this sub-graph, all the nodes accessible through an available path from node v are considered as being activated by v [1].
- Generating the Live-edge outcome for the LT model: Among the income edges of every node v of the social network, at the most, one is selected at random with $b_{v,w}$ probability, and no edges are selected with $(1 - \sum_{(\text{w neibours of v})} b_{v,w})$ probability, where the nodes w are the incoming neighbors of the node v. These selections are followed by constructing a sub-graph of the social network that includes all the nodes of the initial social network and the selected edges as the live edges. In this sub-graph, all the

Input: Graph $G(V, E)$ of a social network, Information diffusion model (IC or LT), List of the influence probabilities of the edges for IC model or weight probabilities of the edges for LT model.
Output: the obtained values of ICC for all the nodes $v \in V$.

Method:
1: Initialize R=10,000
2: for each vertex $v \in V$ do
3: $ICC_v = 0$
4: end for
5: for r=1 to R do
6: Generate live edges $Outcome_r$ from G
7: for each vertex $v \in V$ do
8: $dist = 1;$
9: Do
10: $ICC_v = ICC_v + (\sigma(v)_{step=\mathrm{dist}\ \ where\ x}) \times (1/\mathrm{dist})$
11: $dist++$
12: while (There exist any inactivated node which is a candidate for activation in $Outcome_r$)
13: end for
14: end for
15: for each vertex $v \in V$ do
16: $ICC_v = ICC_v / R$
17: end for

Figure 1. ICC Algorithm for LT and IC Models.

nodes accessible through an available path from node v are considered activated by v [1].

3.2 Influence-Aware Closeness Centrality (ICC) Measure

In the closeness centrality measure, the centrality of every node means its closeness to all the other nodes of the social network considering the link structure of the network (Metcalf et al., 2016):

$$CC_v = \sum_{u \in G} \frac{1}{d_{vu}} \tag{1}$$

where d_{vu} is the shortest path distance from node v to node u.

An influence-aware version of this measure according to the LT information diffusion model is proposed in [10], where both the link structure and the influence which propagates through it, are considered in determining the importance of the nodes. According to this measure, after computing a live-edge outcome, the influence-aware closeness of a node v is considered as its closeness to all the nodes accessible from it in the outcome sub-graph where the influence probability of the path from v to the accessible nodes is greater than or equal to the parameter λ. Nodes at the endpoint of the paths with an influence probability less than λ are out of the influence scope of the v. The influence probability of a path is the outcome of multiplying the influence probabilities (or weights) of the edges in that path [10].

This approach can be similarly extended to the IC model by having live-edges of the IC model. The only difference in this algorithm between the two models is how to compute the live-edge outcomes. The algorithm through which the ICC on both LT and IC are computed is presented in Figure 1.

According to the algorithm in Figure 1, first, the number of the live-edge outcomes is initialized in line 1, and the ICC values of the nodes in lines 2 to 4. Next, a loop begins (lines 5 to 14), where R live-edge outcomes are generated (line 6), and for every outcome, the ICC of every node is computed and summed up with its previous value (lines 7 to 13). Eventually, the average of the sum of the ICC values in all R outcomes is computed as the ICC value for every node (lines 15 to 17). Accessible nodes from node v are determined step by step with their distance from node v in the neighborhood based on condition x, which is "the probability of the $path$ $\geq= \lambda$". $\sigma(v)_{(\mathrm{step=dist\ where\ x})}$ (line 10) is the number of nodes in the distance $dist$ of the neighborhood of node v with condition x. It is notable that when condition x is not met for a path, this path is considered blocked and does not extend for further $dist$ values. Based on the equation in line 10, for every $dist$ value, the ICC of a node v, computed up to distance $dist - 1$, is summed up with the $1/dist$ multiplied by the number of reachable nodes in distance $dist$ from node v.

The time complexity of lines 2 to 4, and 15 to 17, is $O(n)$, where n is the number of nodes of the social network. The time complexity of line 6 is $O(m)$, where m is the number of the social network's edges. This line runs in a loop for R iterations, where R is the number of live-edge outcomes, thus, making the time complexity of this part $O(Rm)$. In lines 7 to 13, for every node v, all the edges accessible through it, the m'_v, are counted. The time complexity for all nodes is $m'_1 + ... + m'_n = n\overline{m'}$, where m' is the number of edges originating from nodes in the influence scope of those nodes in the live-edge outcome graph, and $\overline{m'}$ is the average value of m' considering all the nodes. Because of R iterations in a loop, the whole time complexity of this part becomes $O(Rn\overline{m'})$, making the time complexity of the ICC algorithm $O(n+Rm+Rn\overline{m'})$, where n is the number of the social network nodes, m is the number of the social network edges, R is the live-edge outcomes' number, m' is the number of edges originated from a node in the influence scope of that node in the live-edge outcome graph, and $\overline{m'}$ is the average value of m' where all the nodes are of concern. This time complexity is more scalable than the time complexity of the original closeness centrality algorithm, which is $O(nm)$.

3.3 The influence-aware betweenness centrality (IBC) measure

In the betweenness centrality, the centrality of every node is the number of its appearance in the shortest paths between pairs of other nodes of the social network by considering the link structure of the network [28]. This measure is presented as a normalized value expressed through Eq. 2 [28]:

$$BC_v = \sum_{s,t \in G} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{2}$$

where $\sigma_{s,t}$ represents the number of all the shortest paths from node s to node t, and $\sigma_{s,t}(v)$ represents the number of these paths that pass through node v.

The idea of the betweenness measure is to determine the power of a node as a bridge through which the information in a network is crossed [28]. Here, an influence-aware version of this measure considering the LT and IC models is proposed. The IBC is computed based on the live-edge outcomes. Accordingly, the betweenness of a node is approximated by multiplying the number of the outgoing influence-based paths from it, by in the number of the incoming influence-based path to it. In this context, for the directed graphs, next to the live-edge outcomes, their inversed graph is necessary, where the direction of edges is inversed. For undirected graphs, both of these sub-graphs are the same, therefore, for undirected graphs, computations are different.

For a directed graph, for every node v, the number of accessible nodes by v with the condition "the probability of the path $\geqq \lambda$" are determined in both graphs. These two values are named IBC_1_v and IBC_2_v under the outcome sub-graph and its reverse sub-graph, respectively. IBC_1_v is the indicator of the power of v in reaching other nodes in the social network (the number of the outgoing influence-based paths originated from it), and IBC_2_v is the indicator of the power of v to be accessible from other nodes in the social network (the number of the incoming influence-based path to it in the network), considering the influence spread. These two values are combined to determine the betweenness value of the node v as follows:

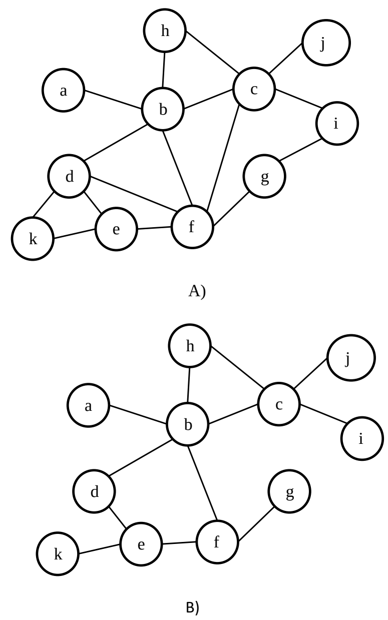

Figure 2. A) a sample graph, B) a sample live-edge outcome generated from the graph (A) based on the IC model with an influence probability of 0.7.

$$IBC = IBC\_1_v \times IBC\_2_v \tag{3}$$

For an undirected graph, the reverse sub-graph is not applied in computations, and the betweenness value of the node v is computed as follows:

$$IBC = \frac{IBC\_1_v}{2} \times \frac{IBC\_1_v}{2} \tag{4}$$

According to Eq. 4, half of the influence-based paths to other nodes are considered incoming and the other is considered outgoing; consequently, the logic of Eqs. (3 and 4) are similar.

As observed, a sample of live-edge sub-graph is generated from Figure 2. A, in Figure 2. B. To generate this graph, first, the influence probability of all the edges is set at 0.7 in Figure 2. A, based on the IC model. Next, the edges are selected as live based on 0.7 probability, and unselected edges like the $d-k$ are considered as dead. The IBC values of nodes c, and f are computed based on the live-edge sub-graph as follows:

The λ value is considered as 0.49; thus, all influence-based paths have length two here. According to Fig-

```
Input: Graph G(V,E) of a social network, Information diffusion model (IC or LT), List of the influence
probabilities of the edges for IC model or weight probabilities of the edges for LT model.
Output: the obtained values of IBC for all the nodes v ∈ V.

Method:
1: Initialize R=10,000
2: for each vertex v ∈ V do
3:     IBC_1_v = 0; IBC_2_v = 0 ; IBC_v = 0;
4: end_for
5: for r=1 to R do
6:     Generate live_edges Outcome_r from G
7:     Generate reverse graph of Outcome_reverse_r from Outcome_r
8:     for each vertex v ∈ V under the Outcome_r do
9:         dist = 1;
10:        Do
11:            IBC_1_v = IBC_1_v + (σ(v)_(step=dist where x))
12:            dist ++
13:        while(There exist any inactivated node which is a candidate for activation in Outcome_r)
14:    end_for
15:    for each vertex v ∈ V under the Outcome_reverse_r do
16:        dist = 1;
17:        Do
18:            IBC_2_v = IBC_2_v + (σ(v)_(step=dist where x))
19:            dist ++
20:        while(There exist any inactivated node which is a candidate for activation in Outcome_reverse_r)
21:    end_for
22: end_for
23: for each vertex v ∈ V do
24:    IBC_1_v = IBC_1_v/R
25:    IBC_2_v = IBC_2_v/R
26:    if IBC_1_v = 0
27:        IBC_v = 0
28:    else_if IBC_2_v = 0
29:        IBC_v = IBC_1_v
30:    else
31:        IBC_v = IBC_1_v × IBC_2_v
32: end_for
```

Figure 3. IBC Algorithm for LT and IC Models.

ure 2. B, $IBC\_1_c = 7$, $IBC\_1_f = 8$, therefore, $IBC_c = 49/4$, and $IBC_f = 16$. The ICC values of these nodes according to subsection 3-2 are computed as $ICC_c = 4\times 1 + 3 \times 1/2 = 5.5$, and $ICC_f = 3 \times 1 + 5 \times 1/2 = 5.5$. Here, the IBC determines node f more influential than c, while ICC determines them as influential as the other one.

The algorithm for the directed graph is presented in Figure 3. The only difference in this algorithm between the two LT and IC models is in how to compute the live-edge outcomes. According to this algorithm, first, a number of the live-edges are initialized in line 1, and IBC values of the nodes are initialized in lines 2 to 4. Next, a loop begins (lines 5 to 22), where after generating every live-edge outcome (line 6) and the reverse graph of the outcome (line 7), the IBC_1 and IBC_2 of every node are computed and summed up with their previous value in the outcome (lines 8 to 14) and its reverse graph (lines 15 to 21), respectively. Finally, the average of the sum of IBC_1 and IBC_2 values in all the R outcomes is computed as the IBC_1 and IBC_2 values for every node (lines 24 and 25, respectively). The final IBC value for every node is determined by multiplying the IBC_1 and IBC_2 values of the node (line 31). If a node does not influence any other node, its IBC_1 value would be zero; consequently, its IBC value is set to zero (lines 26 to 27). If the value of IBC_1 is not zero for a node, but its IBC_2 value is zero, the node could be influential proportional to its IBC_1 value; consequently, its IBC volume is set to IBC_1 (lines 26 to 29). Accessible nodes from node v are determined step by step from node v in the neighborhood based on condition x, which is "the probability of the path $\geqq \lambda$". $\sigma(v)_{(\text{step=dist where x})}$ (lines 11 and 18) is the number of nodes in the distance $dist$ of the neighborhood of node v in condition x. It is notable that when condition x is not met for a path, that path is considered blocked with no extension for further $dist$ values.

The time complexity is computed similarly to that of ICC. Both time complexities of lines 2 to 4 and 23 to 31 are $O(n)$, where n is the number of the nodes of the social network. The time complexity of line 6 is $O(m)$, the same as line 7, where m is the number of the edges of the social network. This line runs in a loop for R iterations, where R is the number of live-edge outcomes; therefore, the time complexities of these two parts are $O(2Rm)$. In lines 8 to 14, for every node v,

Input: Graph $G(V, E)$ of a social network, Information diffusion model (IC or LT), List of the influence probabilities of the edges for IC model or weight probabilities of the edges for LT model, number of seed nodes *K*, flag to indicate computing *ICC* or *IBC*.

Output: set *S* as seed set.

Method:

```
1:  Initialize S = ∅
2:  Candidate_ set = V
3:  for each vertex v ∈ V do
4:        Compute IBC_v (or ICC_v )
5:  end_ for
6:  Sort nodes in Candidate _set descending based on IBC_v (orICC_v)
7:  while the seed set size! = k do
8:        remove a node u from the head of the list Candidate_set
9:        check = false
10:       for each vertex v ∈ S do
11:            if OSJ(v,u) ≥ α
12:                 check = true
13:                 break
14:            else if TSJ(v,u) ≥ β
15:                 check = true
16:                 break
17:       end_ for
18:       if (check = false)
19:            insert u into S
20: end _while
```

Figure 4. The proposed greedy-based method (GBM) for IM.

all its accessible edges the m'_v, are counted. The time complexity for all nodes is $m'_1+...+m'_n = n\overline{m'}$, where, m' is the number of edges originating from nodes in the influence scope of those nodes in the live-edge outcome graph, and $\overline{m'}$ is the average value of m'. Because of R iterations in a loop, the whole time complexity of this part is $O(Rn(\overline{m'})$. The time complexity of lines 15 to 21 is the same. The time complexity of the IBC algorithm is $O(n+Rm+Rn\overline{m'})$, where n is the number of the nodes of the social network, m is the number of the edges of the social network, R is the number of live-edge outcomes, m' is the number of edges originated from a node in the influence scope of that node in the live-edge outcome graph (and its reverse graph), and $\overline{m'}$ is the average value of m', considering all the nodes. This time complexity is more scalable than the time complexity of the original betweenness centrality algorithm, which is $O(nm + n^2)$.

3.4 Two-Step Jacquard-Based (TSJ) Measure

The Jacquard is a recognized measure in determining the similarity between two entities. Applying common neighbors in this measure is considered in determining the similarity between the two nods in the social network as the indicator of the influence overlap between them. As influence spreads in the social network, the common neighbors in one step of the neighborhood are not enough to determine the influence overlap; therefore, the two-step jacquard-based (TSJ) measure is presented as follows:

$$TSJ(v,u) = \frac{|N_{v,2} \cap N_{u,2}|}{|N_{v,2} \cup N_{u,2}|} \tag{5}$$

where, $N_{v,2}$ and $N_{u,2}$ are the set of one- and two-step in-neighbors of v and u including them, respectively..

If the $TSJ(v,u) \geq \beta$, the nodes v and u are considered similar in having influence overlap. It is obvious that if there exists influence overlap in the one-step of the neighborhood between two nodes, there would be influence overlap in the two-step of the neighborhood but not vice versa. Accordingly, the one-step jacquard-based (OSJ) measure is tested before the TSJ due to its lower time complexity. This measure is defined as follows:

$$OSJ(v,u) = \frac{|N_v \cap N_u|}{|N_v \cup N_u|} \tag{6}$$

where, N_v and N_u are the set of one-step in-neighbors of v and u, including them, respectively.

The condition here is $OSJ(v,u) \geq \alpha$.

3.5 Greedy-Based Method (GBM) to Determine Influential Nodes

A greedy-based method is proposed which combines the betweenness or closeness centrality measure as the global index for nodes to be influential, accompanied by the TSG measure as the local index to determine the influence overlap between two nodes, the algorithm of which is expressed in Figure 4 According to this algorithm, first, the seed set is set to empty (line 1), and the candidate set of nodes to be assessed as the seed nodes are set to V (line 2). Next, the ICC or IBC values of the social network nodes are computed based on Figure 1 or Figure 3, respectively (lines 3 to 5), followed by sorting the nodes descending based on the computed values (line 6). A loop begins and iterates to identify the k seed nodes (lines 7 to 20), where at each iteration, a node from the top of the sorted list is selected (line 8), and if that node does not have influence overlap with any of the nodes added into seed set in previous iterations, is inserted into the seed set (lines 11 to 19). To determine if there exists an influence overlap between node u and node v, first, the condition $OSJ(v,u) \geq \alpha$ is checked, where $OSJ(v,u)$ is computed based on Eq. 5. If this holds, nodes u and v have influence overlap; otherwise the condition $TSJ(v,u) \geq \beta$ is checked, where $TSJ(v,u)$ is computed based on Eq. 4. If this holds, nodes u and v have influence overlap, otherwise, none.

The time complexity of the greedy-based algorithm includes the time complexity of $(O(n+Rm+Rn(\overline{m'}))$ for computing the ICC or IBC values, $O(nlogn)$ for sorting the nodes based on the computed ICC or IBC values, the time for selecting the k seeds where is computed as follows:

To compute the OSJ and TSJ values, the set operations, union and intersection, in PYTHON are applied. Accordingly, having two sets of p and q sizes, the time complexity of union would be $O(p+q)$, and the time complexity of intersection would be $O(min(p,q))$. In OSJ, every node is considered a set with its neighbors as its members; therefore, its degree is the size of this set. The number of nodes that are assessed to find the k seed nodes is considered as $k^{'}$, with their degrees as $d_{1,k'}, d_{2,k'}, \ldots, d_{k',k'}$. The degree of the k nodes determined as the seed nodes are $d_{1,k}, d_{2,k}, \ldots, d_{k,k}$. It is assumed that the degrees of nodes to be assessed as the seed set in each iteration are less than the degrees of nodes selected for this iteration as the seed set. The time complexity of the compared nodes with $d_{1,k'}$ and $d_{1,k}$ degrees is $O(2d_{1,k'}+d_{1,k})$. Consequently, the time complexity of the compared node with $d_{1,k'}$ degree is $(2d_{1,k'}+d_{1,k})+(2d_{1,k'}+d_{2,k})+..+(2d_{1,k'}+d_{k,k})$=$O(2kd_{1,k'}+k\overline{d}_k)$, where $\overline{d}_k$ is the average of $d_{1,k}, d_{2,k}, \ldots, d_{k,k}$ values. Accordingly, the whole time complexity of all compared nodes is $(2kd_{1,k'}+k\overline{d}_k)+(2kd_{2,k'}+k\overline{d}_k)+...+(2kd_{k',k'}+k\overline{d_k}) = O(2kk^{'}\overline{d}_{k'}+kk^{'}\overline{d}_k)$, where $\overline{d}_{k'}$ is the average of $d_{1,k'}, d_{2,k'}, \ldots, d_{k',k'}$ values. In TSJ, the number of $k^{'}$ nodes' step-one and step-two neighbors are $l_{1,k'}, l_{2,k'}, \ldots, l_{k',k'}$ and the same of nodes' k are $l_{1,k'}, l_{2,k'}, \ldots, l_{k',k'}$. It is assumed here that these numbers for nodes to be assessed as the seed set in each iteration are less than of nodes selected up to this iteration as the seed noes. The time complexity here is computed as $O(2kk^{'}\overline{l}_{k'}+kk^{'}\overline{l}_k)$, similar to the previous one where $\overline{l}_k$ is the average of $l_{1,k}, l_{2,k}, \ldots, l_{k,k}$ values and $\overline{l}_{k'}$ is the average value of $l_{1,k'}, l_{2,k'}, \ldots, l_{k',k'}$ values. The complete time complexity of this part is $O(2kk^{'}\overline{d}_{k'}+kk^{'}\overline{d}_k+2kk^{'}\overline{l}_{k'}+kk^{'}\overline{l}_k)$.

4 Experiments

Experiments are run to address the following main questions:

- Can the influence-aware centrality measures better predict the importance of nodes in comparison with their classic versions?
- Can the proposed greedy-based method improve the influence spread compared to benchmark methods?

All the implementations are run in Python. Experiments are run on a 2.5 GHz Intel core i7 processor with 16 GB memory.

Datasets: The two real-world data sets which are applied here are:

- Astro Physics collaboration network (ca-AstroPh) [29]: an undirected graph containing 18772 nodes and 198110 edges. In this social network, the scientific collaborations between authors who have submitted their papers to the Astro Physics category are modeled.
- Epinions social network (soc-Epinions1) [30]: a directed network containing 75879 nodes and 508837 edges. In this social network, the who-trust-whom relationships between the members of the general consumer review site, Epinions.com, are modeled.

Both datasets are available in the SNAP Stanford datasets.

Benchmark methods: The benchmark methods are selected by considering 1) the assessment of both the effectiveness and efficiency of the proposed methods and 2) their applicability in both the LC and IT models:

- Degree centrality (DC) [4]: In this centrality measure, the top nodes with the highest degree among all social network nodes are selected as the seed nodes.
- Closeness centrality (CC) [4]: In this centrality measure, the top nodes with the highest closeness value among all social network nodes are selected as the seed nodes.
- Betweenness centrality (BC) [28]: In this centrality measure, the top nodes with the highest betweenness value among all social network nodes are selected as the seed nodes.
- SingleDiscount (SD) [5]: The most well-known efficient heuristics method in the IM context, which selects seed nodes based on their degrees in a manner that the degree of every node is decreased by 1 when its neighbor is selected as a seed node. Notably, SingleDiscount is more effective than Local index

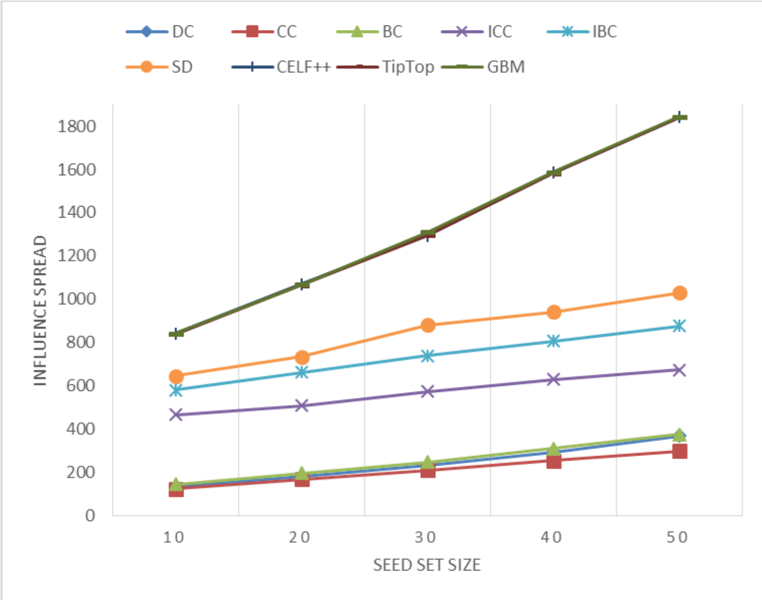

Figure 5. Final influence spread on the ca-AstroPh dataset on the IC model.

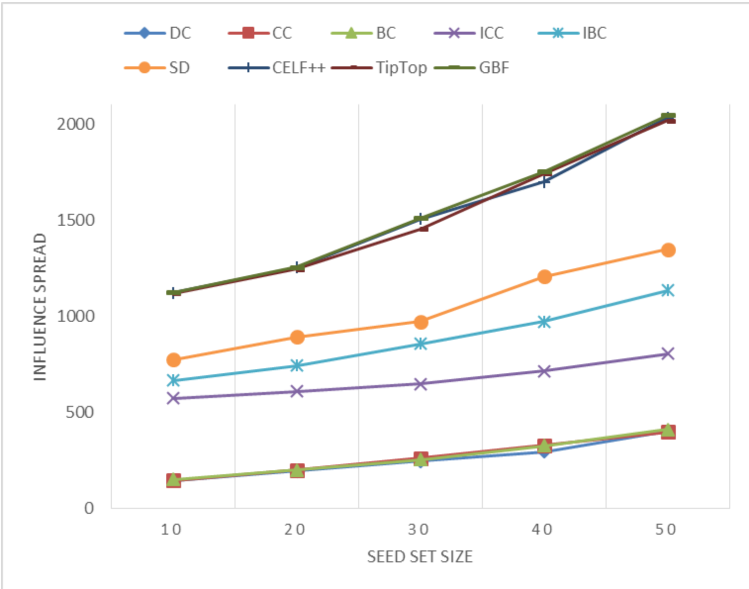

Figure 6. Final influence spread on the ca-AstroPh dataset on the LT model.

rank (LIR) [6] and weighted k-shell degree neighborhood method [26], as the two heuristics-based methods applicable for both IC and LT models.

- CELF++ [13]: This method is the best representive of hill-climbing greedy-based methods where the run time efficiency is of concern.
- TipTop [24]: This method is the best representive of random reverse reachable set concept-based methods where the run time efficiency is of concern.

Results and discussion: The results of the evaluation of effectiveness and efficiency of the proposed influence-aware closeness centrality (ICC) measure, influence-aware betweenness centrality (IBC) measure, and the proposed greedy-based method (GBM) in comparison with benchmark methods are presented. The results of effectiveness in final influence spread are presented in Figures 5 and 6 for ca-AstroPh and in Figures 7 and 8 for soc-Epinions1 considering IC and LT models respectively. The run-time efficiency results are presented in Figures 9 and 10 for ca-AstroPh and in Figures 11 and 12 for soc-Epinions1 considering IC and LT models respectively.

In Figures 5 - 12, it is observed that both influence-aware centrality measures outperform their classic ver-

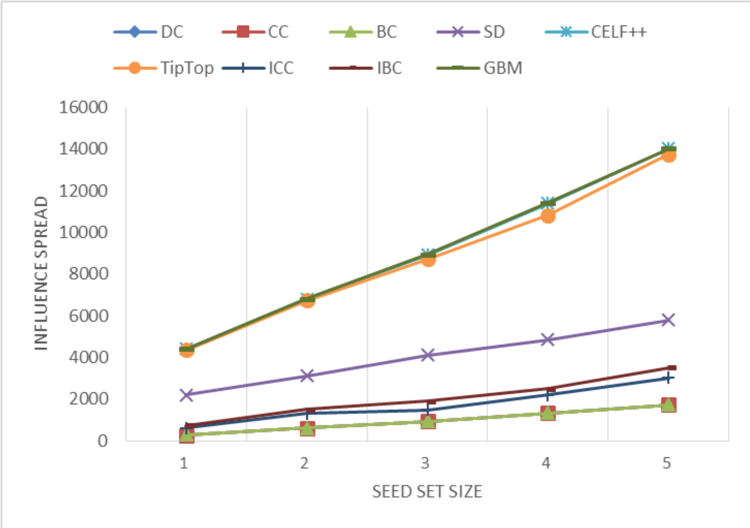

Figure 7. Final influence spread on the soc-Epinions1 dataset on the IC model.

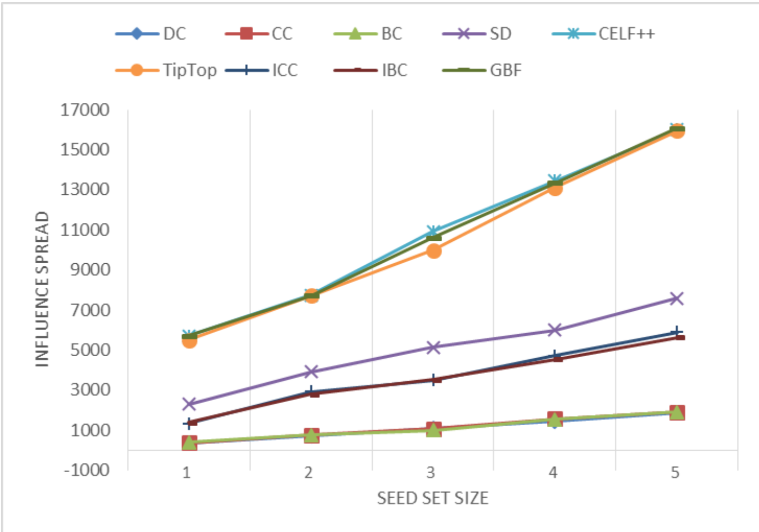

Figure 8. Final influence spread on a soc-Epinions1 dataset on the LT model.

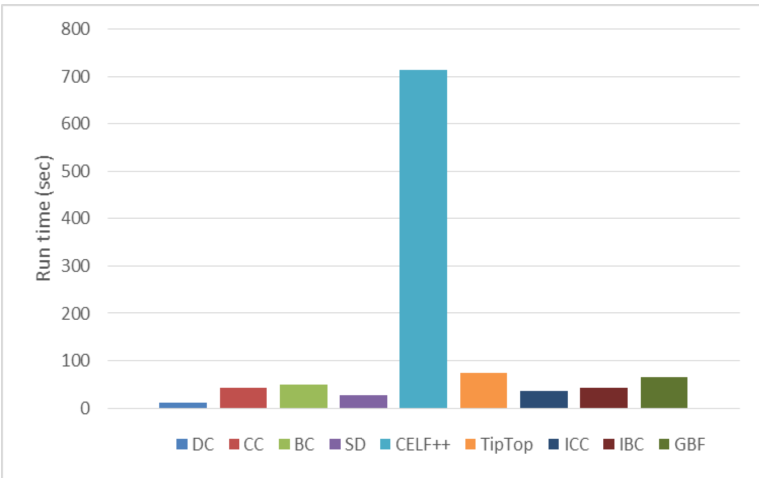

Figure 9. Run-time on a ca-AstroPh dataset on the IC model.

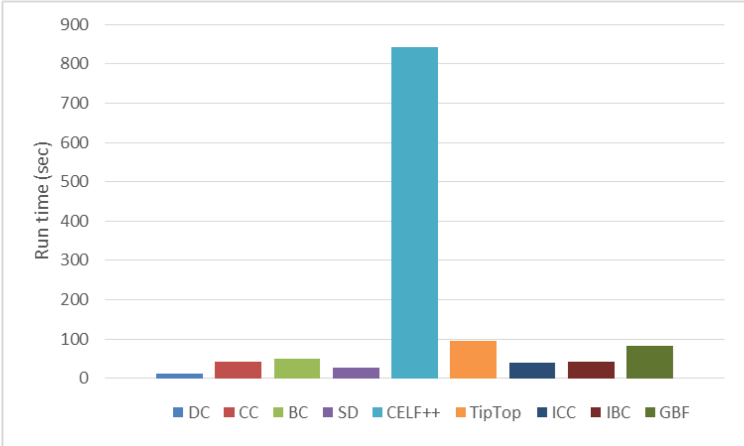

Figure 10. Run-time on a ca-AstroPh dataset on the LT model.

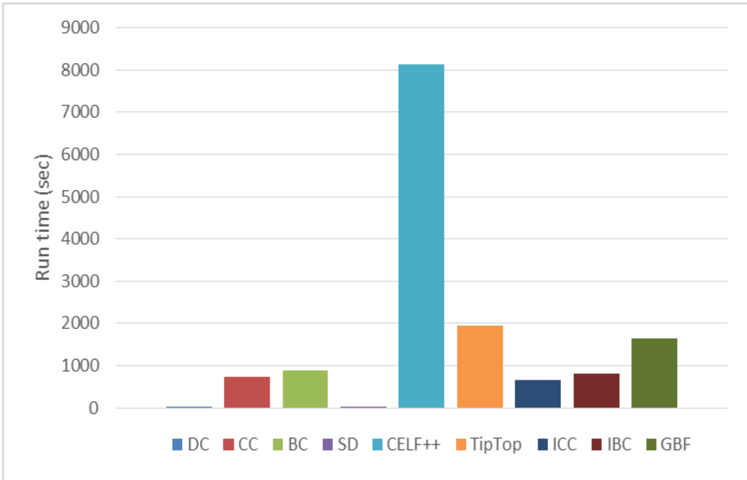

Figure 11. Run-time on a soc-Epinions1 dataset on the IC model.

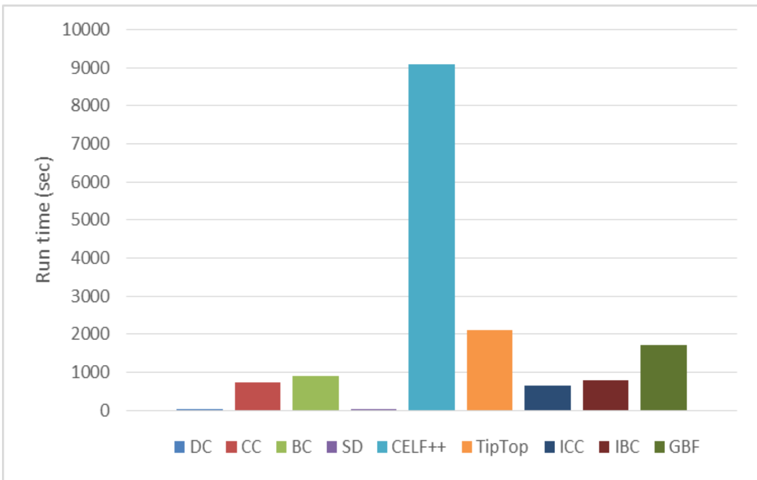

Figure 12. Run-time on a soc-Epinions1 dataset on the LT model.

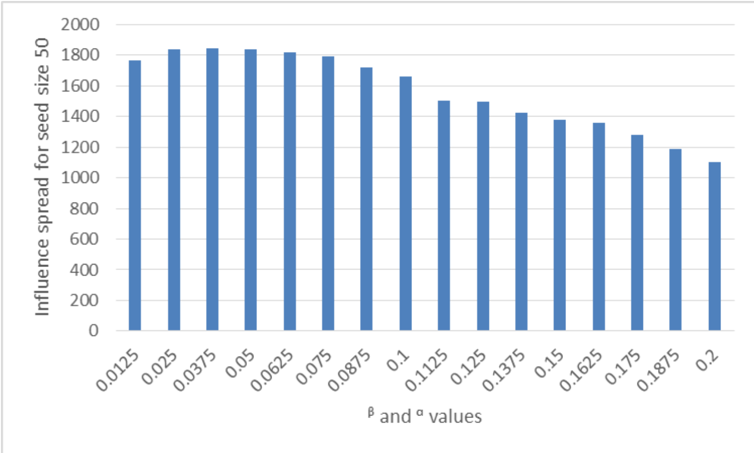

Figure 13. The impact of different values of β and α values on the final influence of the ca-AstroPh dataset on the IC model.

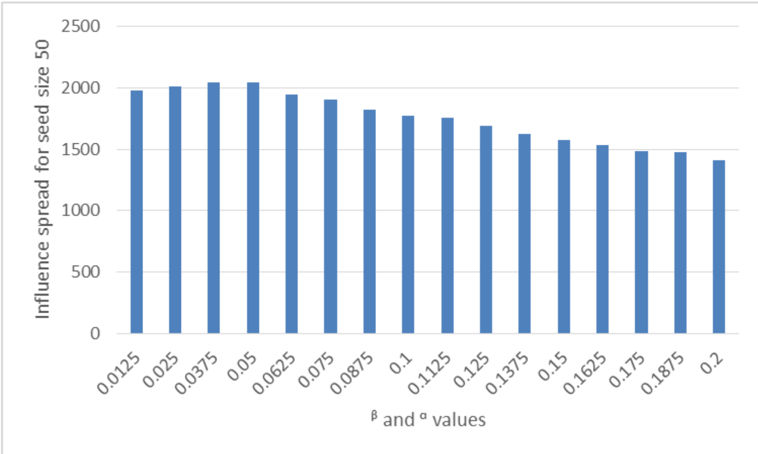

Figure 14. The impact of different values of β and α values on the final influence of the ca-AstroPh dataset on the LT model.

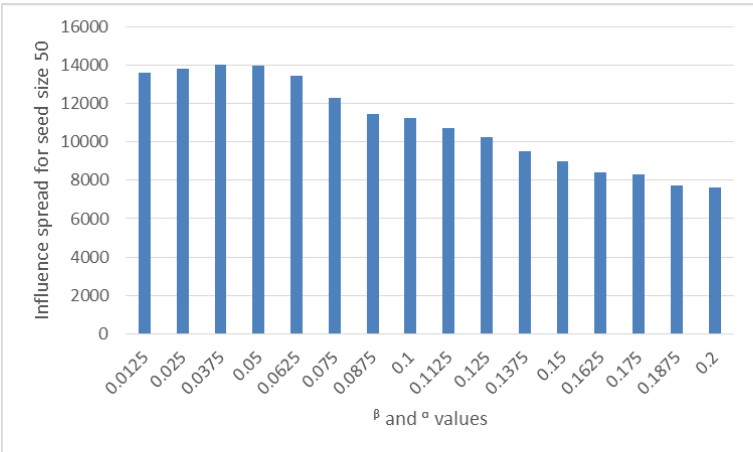

Figure 15. The impact of different values of β and α values on the final influence of the soc-Epinions1 dataset on the IC model.

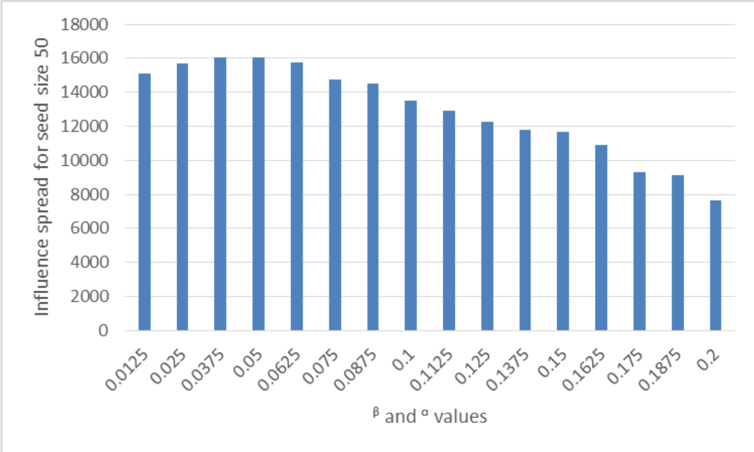

Figure 16. The impact of different values of β and α values on the final influence of the soc-Epinions1 dataset on the LT model.

sions and the degree centrality method regarding influence spread. IBC is more effective than ICC; therefore, IBC is applied in GBM experiments. As to run-time, this proposed greedy-based method (GBM) outperforms CELF++ from 391% to 980% and TipTop from 12% to 23% with effectiveness as well as them considering all the applied models and datasets. The effectiveness of this proposed method outperforms degree discount as well. As to seed size 50, it improves from 78% to 142% considering all the applied models and datasets.

The three classic centrality measures are computed merely based on the link structure. To determine the importance of the nodes, in the computations of betweenness and closeness centralities, all the shortest paths of the network are of concern, while each node may only be influential in a specific scope of the social network, not all of it. Note that not all shortest paths are influence transporter, and considering all of such paths decreases the accuracy in determining the influential nodes. By considering only the influence-based paths in the computations rather than all the paths, the influence-aware versions of closeness and betweenness measures outperform all the classic centrality measures. However, the existence of the influence overlap problem in these two influence-aware centralities leads to its lower effectiveness in comparison with the SD method which addresses the influence overlap problem in-degree centrality. Notably, influence overlap may increase when this influence propagates step by step in several overlapping influence-base paths originating from seed nodes. By addressing the influence overlap problem of influence-aware betweenness and closeness measures, GBM presents the final influence as effective as the CELF++ and TipTop, while its lower runtime makes it more scalable for IM in social networks. Based on the run experiments, GBM outperforms all the counterparts considering both run-time and effectiveness of final influence.

Parameters Adjusting The values assigned to different parameters of the methods adopted in the experiments in Figures 5 - 12 are discussed as follows:

- Most of the existing social network datasets, including those applied in this study, lack the sufficient metadata necessary to learn the influence probabilities of the edges. Accordingly, all the influence probabilities of the edges are set to the same value for the IC model. The proper adopted value of this parameter in the related literature is 0.05 [1–3] which is applied in this study.
- Due to the lack of datasets' metadata, the influence weight of the edges for the LT model is computed as follows:

  When the degree of node v is d_v and the degree of node u is d_u, the influence weight of edge $v-u$ is $1/d_u$, and the influence weight of edge $u-v$ is $1/d_v$. For directed graphs, these degrees are replaced by incoming degrees [1–3].
- The λ parameter for the IC model is set to 0.000125. Accordingly, all the influence paths are with length three. This value is determined by assessing different values for this parameter and its impact on the final influence considering all datasets.
- The λ parameter for the LT model is set to 0.0001. This value is determined based on an assessment of different values for this parameter.
- β and α are the parameters to determine the influence overlap. Different values between 0 and 1 are assessed for both IC, and LT models at different seed set sizes to identify the effect of these parameters on the final influence quality effectiveness. Determining high values for these parameters leads to a decrease in the final influence. The final influence for different values of β and α between 0 and 0.2, at 50 seed set size on the two applied data sets are

shown in Figures 13- 16.

As observed, when both of these parameters are set at around 0.0375 to 0.05, the best results are obtained. Decreasing or increasing more, leads to an increase in influence overlap, therefore, the final influence spread would decrease. These results are true for other seed sizes as well.

5 Conclusions

Centrality-based measures are efficient solutions for IM problem despite losing effectiveness to gain efficiency. Improving the effectiveness of the closeness and betweenness centrality measures in IM is studied in this research. For this purpose, first, the influence-aware versions of these measures, ICC and IBC, are proposed for both IC and LT models; next, a greedy-based method is proposed to address the influence overlap issue in the ICC and IBC by applying a newly proposed two-step jacquard-based measure. Experiments are run first to evaluate these proposed influence-aware centrality measures and, next, to compare this proposed greedy-based method with the well-known benchmarks on the real datasets. The effectiveness of the influence-based centralities becomes evident in comparison with the degree and original centralities. The results indicate the efficiency and effectiveness of these proposed methods in maximizing the spread of influence in social networks.

Improving the run time efficiency is the main challenge for future tasks through which the overlap problem would be solved more efficiently.

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*, page 137–146. ACM, 2003. doi:10.1145/956750.956769.

[2] Y. Li, J. Fan, Y. Wang, and K. Tan. Influence Maximization on Social Graphs: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852 – 1872, 2018. ISSN 1558-2191. doi:10.1109/TKDE.2018.2807843.

[3] S. Banerjee, M. Jenamani, and D. K. Pratihar. A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62(9):3417–3455, 2020. doi:10.1007/s10115-020-01461-4.

[4] A. Landherr, B. Friedl, and J. Heidemann. A Critical Review of Centrality Measures in Social Networks. *Business & Information Systems Engineering*, 2(6):371–385, 2010. doi:10.1007/s12599-010-0127-3.

[5] D. Kempe, J. Kleinberg, and É. Tardos. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 199–208. ACM, 2009. doi:10.1145/1557019.1557047.

[6] 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.

[7] B. Xiang, Q. Liu, E. Chen, H. Xiong, Y. Zheng, and Y. Yang. Pagerank with priors: An influence propagation perspective. In *Twenty-Third International Joint Conference on Artificial Intelligence*, page 2740–2746. Citeseer, 2013.

[8] Q. Liu, B. Xiang, E. Chen, H. Xiong, F. Tang, and J. X. 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*, page 171–180. ACM, 2014. doi:10.1145/2661829.2662009.

[9] Y. Wang, B. Zhang, A. V. Vasilakos, and J. Ma. PRDiscount: A Heuristic Scheme of Initial Seeds Selection for Diffusion Maximization in Social Networks. In *International Conference on Intelligent Computing*, page 149–161. Springer, 2014. doi:10.1145/2661829.2662009.

[10] M. Hosseini-Pozveh, K. Zamanifar, and A. R. Naghsh-Nilchi. A community-based approach to identify the most influential nodes in social networks. *Journal of Information Science*, 43(2):204–220, 2017. ISSN 1741-6485. doi:10.1177/0165551515621005.

[11] 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*, page 420–429. ACM, 2014. doi:10.1145/1281192.1281239.

[12] 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.

[13] 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*, pages 47–48. ACM, 2011. doi:10.1145/1963192.1963217.

[14] A. Goyal, W. Lu, and L. V. Lakshmanan. SIMPATH: An Efficient Algorithm for Influence Max-

imization 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.

[15] C. Wang, W. Chen, and Y. Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012. doi:10.1007/s10618-012-0262-1.

[16] 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*, page 509–518. ACM, 2013. doi:10.1145/2505515.2505541.

[17] C. 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. ACM, 2011. doi:10.1145/2567948.2577336.

[18] S. Galhotra, A. Arora, S. Virinchi, and S. 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. ACM, 2015. doi:10.1145/2740908.2742725.

[19] C. 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. ISBN 978-1-61197-340-2. doi:10.1137/1.9781611973402.70.

[20] 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. ACM, 2014. doi:10.1145/2588555.2593670.

[21] 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*, page 1539–1554. ACM, 2015. doi:10.1145/2723372.2723734.

[22] H. T. Nguyen, M. T. Thai, and T. 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. ACM, 2016. doi:10.1145/2882903.2915207.

[23] H. T. Nguyen, M. T. Thai, and T. N. Dinh. A Billion-Scale Approximation Algorithm for Maximizing Benefit in Viral Marketing. *IEEE/ACM Transactions on Networking*, 25(4):2419 – 2429, 2017. ISSN 1558-2566. doi:10.1109/TNET.2017.2691544.

[24] X. Li, J. D. Smith, T. N. Dinh, and M. T. Thai. TipTop: (Almost) Exact Solutions for Influence Maximization in Billion-Scale Networks. *IEEE/ACM Transactions on Networking*, 27(2):649 – 661, 2019. ISSN 1558-2566. doi:10.1109/TNET.2019.2898413.

[25] S. K. Pal, S. Kundu, and C. Murthy. Centrality Measures, Upper Bound, and Influence Maximization in Large Scale Directed Social Networks. *Fundamenta Informaticae*, 130(3):317–342, 2014. ISSN 1558-2566. doi:10.3233/FI-2014-994.

[26] 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.

[27] W. Jia, L. Yan, Z. Ma, and W. Niu. TPH: A Three-Phase-based Heuristic algorithm for influence maximization in social networks. *Journal of Intelligent & Fuzzy Systems*, 39(3):4393–4403, 2020. doi:10.3233/JIFS-200383.

[28] O. Green, R. McColl, and D. A. Bader. A Fast Algorithm for Streaming Betweenness Centrality. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 11–20. IEEE, 2012. doi:10.1109/SocialCom-PASSAT.2012.37.

[29] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007. doi:10.1145/1217299.1217301.

[30] M. Richardson, R. Agrawal, and P. Domingos. Trust Management for the Semantic Web. In *International Semantic Web Conference*, pages 351–368. Springer, 2003. doi:10.1007/978-3-540-39718-2_23.

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.