Journal of Computing and Security

Research Article, 2024, Volume 11, Number 1 (pp. 29–37)

http://www.jcomsec.org

Applying Heuristic-Based Greedy Approaches for Influence Maximization-Cost Minimization in Social Networks

Mozhdeh Jalali [a]
Maryam Hosseini-Pozveh [b]
Afsaneh Fatemi [a,*]

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

ARTICLE INFO.

Article history:
Received: 13 November 2023
Revised: 13 March 2024
Accepted: 21 May 2024
Published Online: 23 August 2024

Keywords:
Social network, Influence maximization, Cost minimization, Top-K nodes, Viral marketing.

ABSTRACT

Influence maximization in social networks has been an important research issue in the recent decade. This issue is identifying the most influential individuals in a social network who can convey influence to the largest number of the network's members. However, in influence maximization, the costs of all of the nodes to be selected as seeds are considered the same for the companies that do not hold in the real world. Accordingly, influence maximization-cost minimization has gained attention recently. Available studies have applied multi-objective optimization methods which are time-consuming. Applying the existing approaches of influence maximization for other variants of this problem has been considered in some studies about other multi-objective versions of the influence maximization problem. In this study, extending and well-applying run time-efficient methods of influence maximization are considered to influence maximization-cost minimization. Accordingly, two methods are proposed. The first, Local Lowest Degree Rank (LLDR) is a heuristic-based one which by considering the degree of nodes aims to find the cost-affordable influential nodes with minimum influence overlap among them. The second proposed method, Ratio-aware-CELF-based (RCELF) method, is a Cost Effective Lazy Forward (CELF)-based algorithm which extends CELF as a run-time efficient greedy approach for influence maximization by incorporating the cost function of the nodes into consideration. The proposed methods are evaluated by applying two real-world datasets, Facebook and Last.fm. The results establish the outperformance which in comparison with the most effective benchmark method is between 4% to 32% for LLDR and between 34% to 96% for RCELF.

1 Introduction

In recent years, influence maximization (IM) in social networks has received so much attention. The goal is to find the most influential nodes from the social network graph to make the maximum final spread

* Corresponding author.
Email address: a_fatemi@eng.ui.ac.ir (A. Fatemi)
https://dx.doi.org/10.22108/JCS.2024.139787.1136
ISSN: 2322-4460

[1]. This issue has many applications including recommendation systems [2], disease spreading [3], and rumor control [4]. Viral marketing is one of the most important applications of IM [5]. Today, many businesses use online social networks to conduct their business. Influence maximization helps companies select people to promote their products. Although the advertising of products by influencers on social media is very successful, commercial companies have to pay each person individually for doing this task [6]. In the original influence maximization problem, the cost of all the nodes is considered equal when selecting influential ones [1]. However, the cost of paying different people is not the same as the ones with more friends would be more cost-consuming. Accordingly, an important variant of influence maximization with the goal of minimizing the cost as well is introduced by Yang, & Liu, (2017) [6]. It is called the influence maximization-cost minimization problem (IMCM).

Although there are many studies in the literature considering influence maximization, few works exist with respect to the IMCM problem [6–10] despite its importance. As to the best knowledge of the authors here, all of these methods apply multi-objective optimization methods to solve this problem. These methods are time-consuming in presenting a near-optimal solution and have no guarantee to deliver the best near-optimal one [7]. Extending greedy-based and heuristics-based solutions presented in the literature for the IM problem is considerable to solve other variants of this problem including multiobjective ones, as they have been applied in some related studies [11, 12]. This issue is considered in this study to solve the IMCM problem.

In this study, two new algorithms are proposed to solve the influence maximization-cost minimization problem. The first algorithm, called Local Lowest Degree Rank (LLDR), applies the degree information of nodes to approximate both their influence and cost to IMCM while decreasing the influence overlap among the seed nodes. The second algorithm called the Cost-Effective Lazy Foreward (CELF)-based algorithm to IMCM extends the CELF [13] algorithm as a well-known time-efficient method for IM. Accordingly, the contributions of this study are presented as follows:

- The Local Lowest Degree Rank (LLDR) algorithm is presented to solve the IMCM problem heuristically. In the first phase of this algorithm, a percent of high-degree nodes of the network is selected as the candidate seed set. In the second phase, the nodes with the lowest degree among their neighbors are selected from the candidate set greedily.
- A ratio function is introduced for the IMCM problem as the ratio of influence function and the defined cost function.
- The CELF-based algorithm is proposed by extending the CELF algorithm [13] which selects the seed nodes in a greedy manner optimizing the ratio function. This method is called Ratio-aware-CELF-based (RCELF).
- The proposed algorithms are evaluated on real-world Facebook and Last.fm datasets using degree centrality and the CELF algorithm [13] as the benchmarks.

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

2 Related Works

2.1 Influence Maximization-Cost Minimization

Kempe et al. [14], for the first time, introduced the influence maximization problem as an NP-hard discrete optimization problem. The final number of the influenced nodes by the seed set is named the influence function of IM, or its fitness function. This influence function is proven to have two properties of monotonicity and submodularity. Accordingly, a greedy hill-climbing approach can provide the best near-optimal solution [14]. Such a greedy-based approach is presented by Kempe et al. [14] and is known as one of the ground solutions in the literature to solve the IM problem. However, the time complexity of this method is very high, which is $O(knmR)$, 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 size of the seed set. The presented methods in afterward studies in the IM context can be divided into two categories. The first category covers the methods which are greedy-based to be more run-time efficient in comparison with their ancestors [15, 16] including CELF [13], CELF++ [17], SimPath [18], PMIA [19], UBLF [20], ASIM [21], greedy-based random reverse reachable set method [22], TIM and TIM+ [23], IMM [24], Stop-and-Stare [25], and Tip-Top [26]. The second category covers heuristics-based solutions. The purpose of this category is run-time efficiency. It includes methods such as basic centrality measures [27], degree centrality, closeness centrality, betweenness centrality, and PageRank centrality, which have been and are being applied in the IM problem to determine the top-k nodes with the highest centrality values as well as SingleDiscount and

DegreeDiscountIC [28], diffusion degree [29], Group-PageRank measure [30], local index rank (LIR) [31], and weighted k-shell degree neighborhood method [32].

The influence maximization-cost minimization (IM-CM) problem is proposed for the first time by Yang, & Liu (2017) [6] based on the fact that in the real world, companies not only demand the highest influence spread by the seed set in social networks, but also want it with minimum cost. They considered this problem as a multi-objective optimization problem and accordingly, proposed a multi-objective discrete particle swarm optimization (MODPSO-IM-CM) algorithm to solve it. Afterward, methods to solve this problem have applied the multi-objective optimization methods as well. Accordingly, a multi-objective biography-based optimization strategy by De, & Dehuri (2020) [7], a multi-objective crow search algorithm (MOCSA) by Wang & Zhang (2023) [8], a non-dominated sorting moth flame optimization by Wang, et al., (2023) [10], and NSGA-II as a classical multi-objective optimization algorithm by Qian et al., (2020, August) [9], are extended and applied to solve the IM-CM problem.

Although any multi-objective optimization algorithm may be suitable to find out the non-dominated near-optimal solution for IM-CM [7], other studies about other multi-objective variants of IM show that the available methods presented in a decade to solve the IM problem can be extended and applied to solve other variants of this problem including its multi-objective variants [11, 12]. Moreover, while none of the available methods for IMCM are compared with each other experimentally, only two of them have applied methods for IM as the benchmarks [6, 9].

Based on the presented content above, in this research, the new line of extending methods available for IM to solve the IMCM problem is considered. Accordingly, two methods, a heuristics-based, and a greedy-based are introduced.

2.2 Independent Cascade Information Diffusion Model

How the information spreads in the social network is modeled through information diffusion models. Computing the influence spread, and accordingly identifying the influentials in the IM problem and its variants including the IMCM problem is conducted by applying these models. Information diffusion models are considered in two cascade-based and threshold-based categories [33]. Independent-cascade model is one of the main models introduced by Kemp et al., [14] which is applied in this study. In the independent cascade (IC) model, each node v may be activated by each of its neighbors independently based on an activation probability $p(u, v)$. Having the seed set S with size k in iteration 0, activating other nodes is done in some discrete iterations. In every iteration t, each newly activated node u in iteration $t-1$ tries to activate each one of its inactive neighbors with the corresponding probability [14]. It should be noted that every newly activated node u has only one chance to activate each one of its neighbors. The activation process ends when no other nodes can be activated. Accordingly, the final influence spread of the set S under the IC model is the final number of activated nodes.

3 Proposed Methods

3.1 Problem Formulation

Having the S as the seed set with a predefined size of k, the number of final activated nodes by S is shown with $\sigma(S)$. Moreover, the cost of seed set S is defined as follows [6]:

$$C(S) = C_1 + C_2 + ... + C_k \tag{1}$$

where C_i is the cost of seed node number i. This cost is defined in different related studies [6, 8, 9] in different forms, but all of them consider the cost related to the degree of nodes, which means the higher the degree of a node, the higher the cost of that node. Accordingly, the problem would be a multi-objective one as follows [6]:

$$\begin{cases} \max\ \sigma(S) \\ \min\ C(S) = C_1 + C_2 + ... + C_k \end{cases} \tag{2}$$

And eventually, it is equivalent to [6]:

$$min \begin{cases} -\sigma(S) \\ C(S) = C_1 + C_2 + ... + C_k \end{cases} \tag{3}$$

The degree of nodes is discussed as a clue of their cost in viral marketing [34, 35]. Accordingly, in this study, the cost function of a node is considered proportional to its normalized degree as follows:

$$C(u) = \frac{d(u)}{\Sigma_{v \in N} d(v)} \tag{4}$$

where $d(u)$ is the degree of node u, and N is the set of all the nodes of the social network.

3.2 The First Proposed Method: Local Lowest Degree Rank (LLDR)

In this section, a heuristic-based method is proposed considering both maximizing the influence spread and minimizing the cost of seed nodes. This method is inspired by local index rank (LIR) proposed by Liu et al., (2017) [31]. In LIR, the nodes whose degree is higher than the degree of their neighbors are determined as local leaders, and local leaders with the highest degree are selected as influential nodes. This strategy decreases the influence overlap happening as high-degree neighbor nodes are not selected all as the seed nodes.

Algorithm 1 Local Lowest Degree Rank (LLDR) algorithm

Input: A social network $G = (V, E)$, number of seed nodes k.

Output: A k-size set S, Final influence spread, Final Cost.

1: $S = \emptyset, Q = \emptyset$
2: $V' = 10\%$ of high degree nodes of V and their neigbors.
3: Sort nodes of the V' ascending based on their degrees in list Q.
4: **while** $|S| < k$ **do**
5:     $v = Q[0]$
6:     **if** ($v.degree \leq$ degree of its neighbors which exist in Q) and (none of v's neighbors belong to S) **then**
7:         $S = S \cup \{v\}$
8:         $Q = Q - \{v\}$
9:     **else**
10:         $Q = Q - \{v\}$
11:     **end if**
12: **end while**
13: **Return** (S, Final_Spread, Final_Cost)

The algorithm of the proposed Local Lowest Degree Rank (LLDR) is presented in Algorithm 1. In the presented LLDR, two main points are considered: 1) selected influential nodes should also be cost-affordable as possible, 2) the influence overlap among seed nodes should be decreased. According to line 2 of Algorithm 1 in Algorithm 1, the candidate nodes as seeds are 10% of high-degree nodes union with their neighbors. From these candidate nodes, nodes are selected ascending based on their degree in a manner that they have the lowest degree comparing their neighbors in the Q. The value of 10% is selected based on conducted experiments on real-world datasets in this study.

3.3 The Second Proposed Method: Ratio-Aware-CELF-Based (RCELF) Method Algorithm

In the influence maximization-cost minimization, the goal is to maximize the influence spread of the seed set, $\sigma(S)$, and to minimize the cost of the seed set, $C(S)$. Accordingly, the ratio of $(\sigma(S))/(C(S))$ is applied to evaluate the effectiveness of the methods in the studies of this context and other variants of the IM problem which consider the cost value of nodes [6, 11]. This function is both the target optimization function of IMCM and the evaluation function, where the goal is to find a seed set that maximizes the final influence with minimum cost simultaneously. In the proposed method here, this ratio function is applied as the optimization function as well. Accordingly, the ratio function of a set of nodes, S, is presented in equation 5, and the ratio function of a node v is presented in equation 6, as follows:

$$R(S) = \frac{\sigma(S)}{C(S)} \tag{5}$$

$$R(v) = \frac{\sigma(v)}{C(v)} \tag{6}$$

By applying this ratio function, the CELF algorithm [13] as a run-time efficient greedy algorithm to solve IM, is extended here. The proposed algorithm (Algorithm 2) is presented in Algorithm 2, and is called Ratio-aware-CELF-based (RCELF) method. In line 2 of this algorithm, candidate nodes are filtered to have a degree higher than the average degree of the graph in order to prevent preferring nodes with both low influence and low cost to nodes with higher influence but higher cost as well. This is because influence is important here and it should not be sacrificed by cost. This point leads to a decrease in the run time as well.

3.4 Time Complexity of the Proposed Algorithms

Having the degree of nodes, the time complexity of LLDR is computed as follows. The sorting of nodes has the complexity of $O(|V'|log|V'|)$ .The time complexity of lines 4-10 is $O(t)$, where t would be $|V'|$ at maximum. The whole time complexity of LLDR is $O(|V'|log|V'| + |V'|)$.

Having the degree of nodes, the time complexity of RCELF is $O(|V'|mR + t|V'|log|V'|)$, where m is the number of the edges of the social network, R is the number of simulations to compute final influence, and t would be $|V'|$ at maximum.

4 Evaluations

The evaluations are conducted to answer the following question:

- How is the functionality of the two proposed methods in comparison with benchmarks considering the final influence spread, final cost, and the ratio of them?

All the experiments are run on a system with an Intel Core i7 4.20GH processor and 16GB of RAM. The implementations are conducted by applying the Python

Algorithm 2 Ratio-aware-CELF-based (RCELF) algorithm

Input: A social network $G = (V, E)$, number of seed nodes k.

Output: A k-size set S, Final influence spread, Final Cost.

$S = \emptyset, Q = \emptyset$
V' = subset of nodes of the social network which their degree is higher than or equal with $G.average\_degree$.
for each node v in V' **do**
    $v.marg\_gain = R(v)$
    $check_v = 0$
end for
add nodes of V' to Q by $v.marg\_gain$ in descending order
while $|S| < k$ **do**
    $v = Q[0]$
    if ($check_v = |S|$) **then**
        $S = S \cup \{v\}$
        $Q = Q - \{v\}$
    else
        $v.marg\_gain = R(S \cup \{v\}) - R(S)$
        $check_v = |S|$
        Resort Q by $v.marg\_gain$ in descending order
    end if
end while
$Final\_Cost = calculateCost(S)$
$Final\_Spread = \sigma(S)$
Return $(S, Final\_Spread, Final\_Cost)$

programming language and the NetworkX library.

As mentioned, the independent cascade (IC) information diffusion model is applied to calculate the influence of nodes. According to the settings applied in the literature, the influence probability between all pairs of nodes is set to 0.1, and the number of the Monte Carlo iterations is set as 10000 to simulate influence spreading.

4.1 Datasets

Two real-world datasets applied in studies are introduced as follows:

- Facebook [36]: This dataset includes the information of 4039 nodes and 88234 edges of one of the most popular social networking services, Facebook. This dataset is taken from the Stanford University website.
- Last.fm [37]: This dataset includes the information of 7624 nodes and 27806 edges of Last.fm, a music distribution service. This dataset is taken from the Stanford University website. In this dataset, nodes are users from Asian countries that are members of this network. The edges also indicate the existence of a friendly relationship between the nodes.

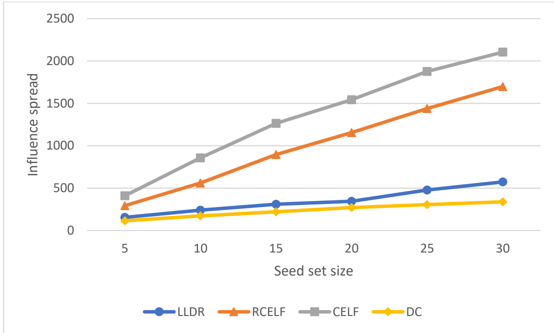

Figure 1. Influence Spread of last.fm Dataset.

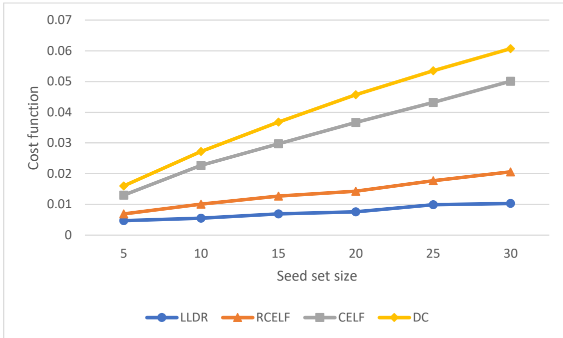

Figure 2. The Cost Function of last.fm Dataset.

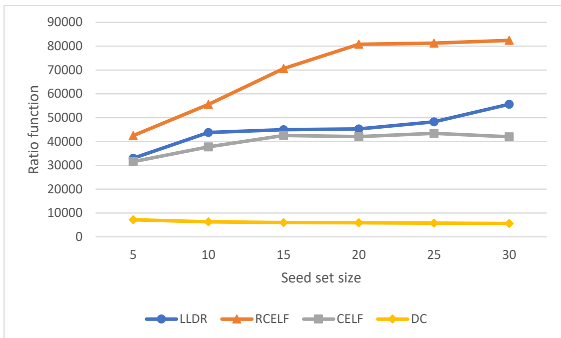

Figure 3. The Ratio Function of last.fm Dataset.

4.2 Evaluation Measures

According to the objective of this study, which is influence maximization-cost minimization in social networks, the proposed methods are evaluated and compared with benchmarks considering three measures: 1) the final influence spread, 2) the total cost of the identified seed set, and 3) the ratio function (defined in equation 5) as the main objective of this study.

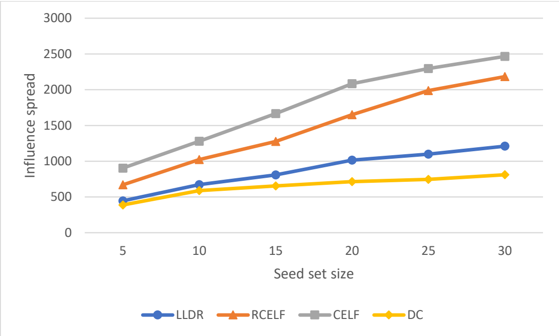

Figure 4. Influence Spread of Facebook Dataset.

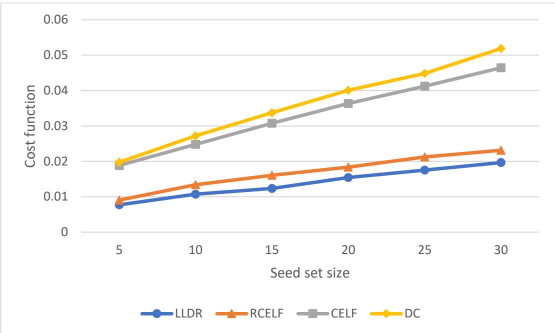

Figure 5. The Cost Function of the Facebook Dataset.

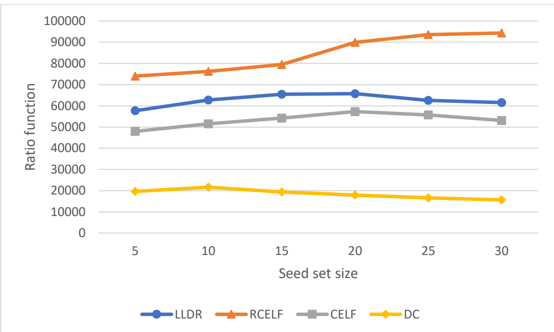

Figure 6. The Ratio Function of the Facebook Dataset.

4.3 Benchmarks

Benchmark methods applied here are: Degree centrality (DC) [27], and CELF algorithm [13]. Degree centrality is a well-known heuristic approach for IM. As this method selects k nodes with the highest degree as seeds, the cost function of it would be the worst in comparison with any other method. Accordingly, it can be a special benchmark considering the influence spread and ratio of it in comparisons. The CELF algorithm is a time-efficient greedy approach to IM that identifies the best near-optimal solution for the IM problem due to the monotonicity and sub-modularity of the influence function of basic information diffusion models [14, 17, 33]. Considering its influence function, it can be another special benchmark algorithm in this context.

4.4 Evaluation Results

The evaluations are conducted with different seed set sizes from 5 to 30 [6, 10, 14, 33]. The evaluation results are presented in the following in Figures 1-6.

The evaluation results of the proposed methods considering influence spread for Last.fm and Facebook are compared in Figures 1 and 4, respectively. As it is evident, CELF has the best and DC has the worst functionality in this respect, and RCELF outperforms LLDR. LLDR outperforms DC as it considers the influence overlap among seed nodes although it selects nodes with lower degrees than DC, but as it selects nodes with lower degrees among 10%, its final influence spread is not as good as RCELF. The RCELF method avoids selecting high-degree individuals as much as CELF by considering the cost of nodes. This issue leads to less final influence spread than the CELF method.

The evaluation results of the proposed methods considering cost function for Last.fm and Facebook are compared in Figures 2 and 5, respectively. DC and CELF have the first and second worst cost functions as it was predictable. The cost function of LLDR is better in comparison with RCELF. It is because LLDR targets low-degree nodes among candidate nodes, but RCELF aims to maximize the ratio function which leads to selecting seed nodes with higher cost and higher influence as well as how the ratio function of them is maximized.

The evaluation results of the proposed methods considering the Ratio function for Last.fm and Facebook are compared in Figures 3 and 6, respectively. Accordingly, RCELF produces a higher ratio than the LLDR method and the other benchmarks. By filtering nodes, it selects nodes that have a degree above the average degree. Using this filter prevents the RCELF algorithm, which seeks to find influential nodes with the least costs, from choosing very low-degree nodes because nodes with a very low degree will not have a good influence on the network. Combining filtering nodes and considering costs in the RCELF algorithm increases the ratio value for the seed set, while the nodes with properly produced final influence are selected. This means that the RCELF method is proper to maximize influence spread and minimize total costs simultaneously.

As is demonstrated, the proposed methods behave similarly in both the Facebook and Last.fm datasets.

5 Conclusions

Influence maximization-cost minimization is an important variant of influence maximization as it incorporates the cost of nodes for companies as well as their influence when identifying the seed set. Available methods apply multiobjective optimization methods in this regard which are time-consuming. Extending the existing time-efficient methods of IM context, two methods are presented for IMCM. The first proposed method, called Local Lowest Degree Rank (LLDR), selects the nodes with the lowest degree among their neighbors from the candidate nodes which are a percent of high-degree nodes of the social network. The second proposed method, called Ratio-aware-CELF-based, extends the CELF algorithm for IM in the context of IMCM. The algorithm optimizes the ratio function between the influence spread and the cost function of the nodes. In identifying the influential people in the proposed Ratio-aware-CELF-based method, nodes below the average degree are removed. This point leads the algorithm to find nodes with acceptable influence. Incorporating cost into the seed identification phase makes these proposed algorithms to be applied in real-world social networks. The evaluations on two real-world datasets establish the superiority of the proposed methods in IMCM.

Assessing other more time-efficient methods of IM to be applied in the IMCM context is one of the future research directions. Considering other network structure properties including community structure as well as the influence propagation as a special property of social networks in defining cost function and solving IMCM problem is another future work of this study.

References

[1] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872, 2018. doi:10.1109/TKDE.2018.2807843.

[2] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation: a generative model approach. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 671–680, 2012. doi:10.1145/2348283.2348373.

[3] Shunyu Yao, Neng Fan, and Jie Hu. Modeling the spread of infectious diseases through influence maximization. *Optimization letters*, 16(5):1563–1586, 2022. doi:10.1007/s11590-022-01853-1.

[4] Didier A Vega-Oliveros, Luciano da Fontoura Costa, and Francisco Aparecido Rodrigues. Influence maximization by rumor spreading on correlated networks through community identification. *Communications in Nonlinear Science and Numerical Simulation*, 83:105094, 2020. doi:10.1016/j.cnsns.2019.105094.

[5] Huimin Huang, Hong Shen, Zaiqiao Meng, Huajian Chang, and Huaiwen He. Community-based influence maximization for viral marketing. *Applied Intelligence*, 49:2137–2150, 2019. doi:10.1007/s10489-018-1387-8.

[6] Jie Yang and Jing Liu. Influence maximization-cost minimization in social networks based on a multiobjective discrete particle swarm optimization algorithm. *IEEE Access*, 6:2320–2329, 2017. doi:10.1109/ACCESS.2017.2782814.

[7] Sagar S De and Satchidananda Dehuri. Multi-objective biogeography-based optimization for influence maximization-cost minimization in social networks. In *Biologically Inspired Techniques in Many-Criteria Decision Making: International Conference on Biologically Inspired Techniques in Many-Criteria Decision Making (BITMDM-2019)*, pages 11–34. Springer, 2020. doi:10.1007/978-3-030-39033-4_2.

[8] Ping Wang and Ruisheng Zhang. A multi-objective crow search algorithm for influence maximization in social networks. *Electronics*, 12 (8):1790, 2023. doi:10.3390/electronics12081790.

[9] Fulan Qian, Cunliang Zhu, Xi Chen, Shu Zhao, and Yanping Zhang. Prnsga-ii: A novel approach for influence maximization and cost minimization based on nsga-ii. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 396–403. IEEE, 2020. doi:10.1109/ICBK50248.2020.00063.

[10] Chao Wang, Liangliang Ma, Lei Ma, Joel Weijia Lai, Jie Zhao, Lu Wang, and Kang Hao Cheong. Identification of influential users with cost minimization via an improved moth flame optimization. *Journal of Computational Science*, 67: 101955, 2023. doi:10.1016/j.jocs.2023.101955.

[11] Cheng-Te Li, Mei-Yuan Huang, and Rui Yan. Team formation with influence maximization for influential event organization on social networks. *World Wide Web*, 21:939–959, 2018. doi:10.1007/s11280-017-0492-7.

[12] Hung T Nguyen, My T Thai, and Thang N Dinh. A billion-scale approximation algorithm for maximizing benefit in viral marketing. *IEEE/ACM Transactions On Networking*, 25(4):2419–2429, 2017. doi:10.1109/TNET.2017.2691544.

[13] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie 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.

[14] David Kempe, Jon Kleinberg, and Éva 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.

[15] Mehdi Azaouzi, Wassim Mnasri, and Lotfi Ben Romdhane. New trends in influence maximization models. *Computer Science Review*, 40: 100393, 2021. doi:10.1016/j.cosrev.2021.100393.

[16] Suman Banerjee, Mamata Jenamani, and Dilip Kumar Pratihar. A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62:3417–3455, 2020. doi:10.1007/s10115-020-01461-4.

[17] 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.

[18] Amit Goyal, Wei Lu, and Laks 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.

[19] Chi Wang, Wei Chen, and Yajun 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.

[20] Chuan Zhou, Peng Zhang, Jing Guo, and Li 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.

[21] Sainyam Galhotra, Akhil Arora, Srinivas Virinchi, and Shourya 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.

[22] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan 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.

[23] Youze Tang, Xiaokui Xiao, and Yanchen 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.

[24] Youze Tang, Yanchen Shi, and Xiaokui 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.

[25] 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.

[26] Xiang Li, J David Smith, Thang N Dinh, and My T Thai. Tiptop:(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.

[27] Andrea Landherr, Bettina Friedl, and Julia Heidemann. A critical review of centrality measures in social networks. *Business & Information Systems Engineering*, 2:371–385, 2010. doi:10.1007/s12599-010-0127-3.

[28] Wei Chen, Yajun Wang, and Siyu 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.

[29] Sankar K Pal, Suman 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.

[30] Qi Liu, Biao Xiang, Enhong Chen, Hui Xiong, Fangshuang Tang, and Jeffrey Xu 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.

[31] Dong Liu, Yun Jing, Jing Zhao, Wenjun Wang, and Guojie Song. A fast and efficient algorithm for mining top-k nodes in complex networks. *Scientific reports*, 7(1):43330, 2017. doi:10.1038/srep43330.

[32] Amrita Namtirtha, Animesh Dutta, and Biswanath 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.

[33] Maryam Hosseini-Pozveh, Kamran Zamanifar, and Ahmad Reza 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.

[34] IC Morărescu, Vineeth S Varma, L Buşoniu, and Samson Lasaulce. Space–time budget allocation policy design for viral marketing. *Nonlinear Analysis: Hybrid Systems*, 37:100899, 2020. doi:10.1016/j.nahs.2020.100899.

[35] Ronak Doshi, Ajay Ramesh, and Shrisha Rao. Modeling influencer marketing campaigns in social networks. *IEEE Transactions on Computational Social Systems*, 10(1):322–334, 2022. doi:10.1109/TCSS.2022.3140779.

[36] Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25, 2012.

[37] Benedek Rozemberczki and Rik Sarkar. Characteristic functions on graphs: Birds of a feather, from statistical descriptors to parametric models. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1325–1334, 2020. doi:10.1145/3340531.3411866.

Mozhdeh Jalali holds a Master's degree in Computer Engineering from the University of Isfahan and currently serves as a software engineer at Irisa company. Her main areas of research interest include data analysis, graph analysis, and social networks.

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.

Afsaneh Fatemi received her B.Sc. from the Isfahan University of Technology, Isfahan, Iran, in 1995, and her M.Sc. and Ph.D from the University of Isfahan, Isfahan, Iran in 2001 and 2012 respectively, all in Computer Engineering. Currently, she is an associate professor in the Faculty of Computer Engineering at the University of Isfahan, and a member of UI Big Data research group. Her main research interests lie in the fields of Complex Networks, Question Answering and Dialogue Systems, and applications of LLMs.