A Study on Influence Maximizing Based on Two Rounds of Filtration Metric in Social Networks

Yang Li[1] and Zhiqiang Wang[2]
(Corresponding author: Zhiqiang Wang)

State Information Center, Beijing 100045, China[1]
Email: liyang_cas@163.com
Beijing Electronic Science and Technology Institute, Beijing 100070, China[2]
Email: wangzq@besti.edu.cn

(Received May 9, 2023; Revised and Accepted Sept. 23, 2023; First Online Apr. 25, 2024)

Abstract

The influence maximization problem is discovering a seed set of nodes in a social network and making the spread as large as possible based on influence propagation. The current related algorithm based on the greedy strategy maintains a better influence propagation but has high time complexity and is not very scalable. This paper proposes a new method to solve the influence maximization problem by reducing the time complexity, called the Two Rounds of Filtration Metric (TRFM) algorithm. The main work is as follows:(1) A regional node metric is proposed based on the local topology to measure the nodes, which reduces the evaluation time. (2) The submodular characteristic is applied to discover the TOP-K seed node set from the candidate node set; meanwhile, the evaluation measurement in the whole network maintains a better influence propagation. The experimental results on the actual data set verify the effectiveness of the TRFM algorithm.

Keywords: Community Division; Greedy Strategy; Independent Cascade Model; Influence Maximization; Social Network

1 Introduction

1.1 Overview of Influence Maximization

Social networks are increasingly integrated into every aspect of our working life by the new generation of information technology. Users can follow the star, make friends, release information, and promote products through social networks such as Weibo, WeChat, Twitter, and Facebook. For example, a company develops a new cell phone and hopes to promote it through some stars to influence more people; As well as a company develops a new APP or a new cell phone and hopes to promote it through some famous bloggers to attract more users to participate by word of mouth, etc. Ultimately, we hope to maximize the influence of other users on social networks.

1.2 Problem of Influence Maximization

The applications mentioned above can be summarized as the influence maximization problem, i.e., we can take a social network graph, for example, where the graph nodes represent users in the social network, the edges of the graph represent user relationships in the social network, which can be described as a problem how to discover the set of k initial nodes in the graph that maximizes the spread of the final influence by given a specified propagation model. Several researchers have carried out extensive research work based on this problem. Kempe first represented the influence maximization study through a discrete optimization problem and proved it to be an NP-Hard problem with the simple greedy algorithm to achieve the optimal solution of (1-1/e). In subsequent research, some researchers keep optimizing the greedy algorithm to improve the performance further. Others propose some heuristic algorithms from scalability and keep advancing to deepen the research.

1.3 The Main Work of The Method in This Paper

In this paper, the search space is reduced by two rounds of node filtration, significantly reducing the running time. The experimental results on the public dataset verify the effectiveness of this paper, and the main work of this paper is as follows:

1) Propose a two-round node filtration method: Through the two-round filtration from the community evaluation and node evaluation method, the node search space is reduced, and the propagation coverage is narrowed.

2) Propose the regional metric of nodes: The local metric of nodes is formed by integrating and evaluating

nodes' neighborhood, radiation, and connectivity attributes.

3) Proposed the greedy algorithm based on submodularity property: Based on the submodularity property, the set of candidate nodes in two rounds is evaluated by the whole network metric, the set of Top-k nodes is found, and which can substantially reduce the time complexity.

The remaining sections in this paper are as follows: Section 2 addresses the review of related research works. Section 3 focuses on the greedy algorithm based on the two-round filtration metric in this paper. Section 4 shows the experimental comparison and result analysis. The final section presents the related conclusions and prospects.

2 Related Work

In the early research process, node degree became the preferred influence node criterion in terms of network structure topology, and it was believed that nodes that might be in the central position in the network or have specific linking properties tend to bring better influence, such as node degree, node centrality, and so on. However, the generative characteristics of scale-free networks determine that such nodes tend to be linked together preferentially, leading to more extensive duplicate coverage of influence propagation.

With the deepening of further research, based on earlier sociological analysis and marketing-related studies, influence propagation models (independent cascade model and linear threshold model) for interactions between users are constructed to evaluate influence, which can get a whole network quantitative perspective by portraying the activation states between nodes. The current research is mainly divided into greedy algorithms and heuristic algorithms.

2.1 Introduction to The Progress of Greedy-Based Algorithms

Kempe [17] represented the influence maximization problem as a discrete optimization problem for the first time and obtained the maximized influence propagation by a greedy algorithm. On this basis, Leskovec [21] proposed the celfGreedy algorithm to reduce the number of Monte Carlo simulations by submodular characteristics, reducing the time complexity to a more significant extent. Still, because its search space is the nodes of the entire network topology, the computational performance is affected by the data set. Its worst-case time complexity is approximately equal to that of the original greedy algorithm. In response, Goyal [13] proposed the celfPlusGreedy algorithm to evaluate the influence gain of nodes by further reducing the number of Monte Carlo simulations. Still, the reduced time complexity is more limited.

Subsequently, researchers continued to optimize the algorisms from the topology; Chen [6] proposed the new greedy algorithm to improve the efficiency by pre-deleting edges, which was compared with the celfGreedy algorithm and found to be advantageous only during the first round of computation. Wang *et al.* [28] proposed the CGA and OASNET methods using a greedy algorithm and dynamic programming approach to find the seed nodes. However, the simulation scope is limited to within the community, which reduces the network-wide influence metric.

Later, researchers proposed optimization schemes with different perspectives. Borgs *et al.* [4] proposed a hypergraph-based influence propagation estimation method, which still needs the validation of scene data. Cohen *et al.* [8] proposed to reduce the time complexity by selecting the node with the highest information gain for every round. Laya *et al.* [2]proposed a fuzzy propagation model to deal with the influence maximization (IM) problem. Yang *et al.* [29]proposed an exchange improvement algorithm to improve further the quality of the solution to the non-submodular influence maximization problem. Jie *et al.* [24]proposed a novel influence maximization algorithm of node avoidance based on user interest. Tang *et al.* [23] performed influence evaluation by measuring the lowest boundary of the propagation scope, and the running time was better than the CELF++ algorithm. The running time is better than the CELF++ algorithm. Wang *et al.* [27] proposed the IV-Greedy algorithm based on the multi-path asynchronous threshold model MAT, which can achieve better experimental results on the dataset. Zhou *et al.* [30] reduced the number of Monte Carlo simulations for influence calculation by constructing an upper bound function for the greedy strategy. The experimental results showed that when the size of the seed node set is small, the time complexity is better than the CELF algorithm.

2.2 Introduction to The Progress of Heuristic-Based Algorithms

To further improve the scalability of influence maximization algorithms and better apply them to large-scale social networks, researchers have also proposed some heuristic algorithms, such as Median centrality [3] and k-core [20], etc. Regarding topology, Chen *et al.* [6] proposed the DegreeDiscount method based on the first-order neighborhood influence of nodes, which works better experimentally when the propagation probability is small. Subsequently, Chen *et al.* proposed the LDAG [7] method to select seed nodes by updating the local topology to improve scalability. Still, the experimental results are easily affected by the network topology [15]. Cordasco *et al.* [9] proposed an efficient heuristic for the network structure of the tree, annular graphs, and complete graphs algorithm. They extended it to conduct influence calculations in directed graph network structures [10].

Then propagation paths became the focus of research; for example, Kimura *et al.* [19] proposed SPM/SP1M method based on the shortest path, and Narayanam *et al.* [22] proposed Shapley value-based method, but both algorithms are weak in scalability. Goyal *et al.* [12] offered a way to find the shortest path from the node adjacency region. Galhotra *et al.* [11] proposed a heuristic algorithm based on adjacency paths that reduce the memory overhead compared to the CELF++ algorithm.

Around the relational perspective among nodes, Agha *et al.* [18] studied variable propagation probabilities based on node heterogeneity. They proposed an optimization model that simultaneously constrains the seed set and propagation scope. Wang *et al.* [26] argued for enhancing the consideration of group influence on nodes in multi-relational social networks. Chen *et al.* [5] used reinforcement learning based on the Markov decision process to model the influence problem. Later, the research perspective was gradually expanded, Jiang *et al.* [14] proposed a simulated annealing algorithm to optimize the influence problem; Jung *et al.* [16] performed incremental influence measurement on seed nodes, which can effectively reduce memory overhead and running time.

3 The Proposed Method of This Paper

3.1 Main Ideas

Some improved versions of the greedy algorithm by researchers have reduced the time complexity to some extent. However, the running time is relatively high and needs further improvement in real large-scale social networks. In this paper, we hope to provide a filtering mechanism to evaluate nodes from the community and topological perspectives, which can reduce the more extensive repeated coverage of the propagation scope of the preferentially linked nodes.

3.2 Variable Representation

A social network is modeled using an undirected graph G= (V, E), where node v represents the users in the network and edge e represents the association between users. Table **??** lists the important variables used in this paper; In this paper, S is used as the set of nodes selected to maximize its influence propagation, which also becomes the seed set. $simCas(S)$ represents a stochastic process based on the spread of the node set S's influence; therefore, the result of its influence is also a random set of nodes. The algorithm in this paper uses the graph G and the number k as input to generate a seed set S. The aim is to maximize the influence of other nodes based on the selected seed set.

3.3 Node Evaluation

Community division will help us to filtrate meaningful and dispersed communities, which can avoid repeated coverage of the propagation scope due to the preferential linking of nodes. Hence, the research in this paper involves the work related to community division, and to improve the performance further, this paper adopts Raghavan's [25] label propagation method, which can be achieved in linear time, as the method of community calculation in this paper.

There is some difficulty in evaluating the global attribute metric values of nodes in the whole network, which will consume a lot of running time, so we want to provide a fine-grained method to measure the attributes of nodes. In this paper, we consider candidate node benchmark metrics (BM) by the following three factors: node's adjacency attribute Lv, node's radiation attribute Rv, and node's connectivity attribute Cv.

$$BM(v) = \frac{Lv + Rv}{2} * Cv \tag{1}$$

The variable is described as follows:

- Dv: The node's degree, reflecting the node's number of neighbors.
- Lv: The neighboring metrics of a node, i.e., the ratio of the node's degree to the sum of its neighbor's degree, reflects the strength of the node's influence on neighboring nodes;
- Rv: The radiation metrics of a node, i.e., the ratio of the sum of the node's neighboring degrees to the sum of the community nodes, reflects the radiation strength of the node in the region;
- Cv: The connectivity metrics of a node, i.e., the ratio of the node's betweenness centrality and the sum of the node's betweenness centrality in the community, reflects the node's connectivity strength in the region.

$$Lv = \frac{Dv}{\sum_{W \in N(v)} D(w)} \tag{2}$$

$$Rv = \frac{\sum_{W \in N(v)} D(w)}{\sum_{W \in Com(v)} D(w)} \tag{3}$$

$$Cv = \frac{Bet(v)}{\sum_{W \in Com(v)} Bet(w)} \tag{4}$$

where b_{vw} represents whether node v is connected to node w, 1 if connected, and 0 otherwise. Dv represents the degree of node w, $N(v)$ represents the set of neighboring nodes of node v, $Com(v)$ represents the set of community nodes of node v, and $Bet(v)$ represents the betweenness centrality of node v.

Algorithm 1 The Two Rounds of Filtration Metric (TRFM) Algorithm

Input: Graph G, the amount of seeds k
Output: Top-k vertices

1: initialize $S = \emptyset, S_G = \emptyset$
2: community partition and get z candidate communities: $C_1, C_2, \ldots, C_z$
3: **for** $i = 1$ to z **do**
4:   in community C_i, compute the local value based on Equation (2)
5:   in community C_i, compute the radiation value based on Equation (3)
6:   in community C_i, compute the connection value based on Equation (4)
7:   compute benchmark value bv based on Equation (1), sort the candidate vertex and continually add to the set S_G
8: **end for**
9: **for** each vertex $v \in S_G \backslash S$ **do**
10:   $MG_v = 0$
11:   **for** $i = 1$ to R **do**
12:     $MG_v += |SimCas(S \cup \{v\})|$
13:   **end for**
14:   $MG_v = MG_v/R$
15:   store the vertex v with MG_v into the $Queue\ Q$
16: **end for**
17: sort the $Queue\ Q$ in the descending order
18: $S = S \cup \{$first vertex in $Q\}$ and remove first vertex from Q
19: **for** $i = 2$ to k **do**
20:   **while** true **do**
21:     vf =first vertex, vs =second vertex in Q
22:     **if** vs has not be evaluated in current round **then**
23:       **if** MG_{vf} in current round $>= MG_{vs}$ in **previous / current** round **then**
24:         $S = S \cup \{vf\}$ and remove vf from Q
25:         **break**
26:       **else**
27:         insert the vf into the Q based on its marginal gain MG_{vf}
28:       **end if**
29:     **end if**
30:   **end while**
31: **end for**
32: **return** S

3.4 Metrics and Algorithm Execution

3.4.1 Regional Metrics for Two Rounds

In this paper, we propose Two Rounds of Region Metric (TRFM); we hope to reduce the time complexity by searching the range at the whole network level, and how to locate the nodes' candidate solutions becomes the key. In the first round, we select some scaled subcommunities as candidate communities through community partitioning, which reduces the influence of repeated propagation due to the nodes are often linked together preferentially in the scale-free network; in the second round, we calculate the local attribute metrics through candidate communities to select some nodes with a higher ranking of benchmark metric to join the candidate node set continuously, and then always reduce the search scope to reduce time complexity.

Based on the "diminishing returns" property of the submodular function, the Marginal Gain obtained by adding a node v to the set S cannot be smaller than the marginal gain obtained by adding the same element v to the parent set T of S, denoted as $f(S \cup \{v\}) - f(S) >= f(T \cup \{v\}) - f(T)$. Based on the property of "diminishing returns," the number of evaluations is reduced, and the computational performance is improved by comparing the marginal gain value of the current round with the previous round. The time complexity is $O(N)$ in the optimal case and $O(KNRM)$ in the worst case. Therefore, in this paper, the algorithm also introduces the idea of submodular characteristics [21] to reduce the number of Monte Carlo simulations, which can be in the algorithm in two steps:

Find the first seed node: In the first round of calculation, the influence gain value is calculated for the set of filtered nodes and stored in the queue in reverse order, and the first node of the line is the first seed node we found, and then the node is removed at the head of the queue.

Find the remaining set of $k-1$ seed nodes: Continue to evaluate the marginal gain of nodes in each round and update the queue by comparing the influence gain of the first node in the current round with the influence gain of the second node in the previous round, if the gain value of the first node is more significant, we select the first node as the seed node. Otherwise, we insert the node into the corresponding position in the queue according to its influence gain value. Then we iterate the comparison of influence gain and update the node queue until the remaining $k-1$ seed nodes are found.

3.4.2 The Execution Process of The Algorithm

The algorithm in this paper is shown in Algorithm 1. Line 2 performs community division. Lines 4, 5, and 6 calculate the nodes' neighboring metrics, radiative metrics, and connectivity metrics in the current community. The baseline attributes of the nodes in the current community are calculated in line 7, sorted, and added to the candidate node set. In lines 9-18, the evaluation of influence gain is computed for each node in the candidate node-set, and the 1st seed node is found. In lines 19-29, the evaluation calculation of the influence gain values of the nodes stored in the queue is iterated until the $k-1th$ seed node is found.

Complexity analysis: Line 2 community division consumes $O(M)$. Lines 3-8 consume $O(MCNC)$. Lines 9-18 consume $O(zRNC)$, and lines 18-32 compute statistically for the remaining $k-1$ node sets with optimal time complexity of $O(NC)$ and worst time complexity of

$O(kzRNC)$. Overall, the optimal time complexity is thus $= O(M) + O(MCNC) + O(zRNC) + O(k) = O(zRNC)$ and the worst time complexity is $= O(M)+O(MCNC)+O(zRNC) + O(kzRNC) = O(kzRNC)$.

4 Experimental Analysis

We conduct our experiments on publicly available datasets and compare them with current heuristics and greedy algorithms to verify the effectiveness of the proposed method in two aspects: the range of influence and the running time.

4.1 Experimental Setup

Operating system: Windows 10, processor: Intel (R) i5 1.8GHz, memory: 32G.

4.1.1 Experimental Data Set

The experimental data were obtained from the dataset of Arxiv, a paper collaboration network [1], where a node represents that the user published a paper and an edge represents those two users co-authored the paper. The dataset is as follows:

- Dblp: DBLP academic paper collaboration dataset, where the number of nodes is about 14,485 and the number of edges is 37,026.
- GrQc: A collaborative dataset of papers in general relativity and quantum cosmology, where the number of nodes is about 5,242 and the number of edges is 14,485.
- Hep: A combined dataset of articles in high energy physics, where the number of nodes is about 15,233 and the number of advantages is 31,380.
- Phy: A collaborative dataset of papers in the field of physics, where the number of nodes is about 14,997 and the number of edges is 57,866.

Table 2: Statistics of four real-world networks

DataSet	#Vertice	#Edge
Dblp	14,485	37,026
GrQc	5,242	14,485
Hep	15,233	31,380
Phy	14,997	57,866

We extracted the structure of four types of paper collaboration networks from the arXiv paper literature, each node in the network represents an author, and each edge represents the existence of two authors collaborating on a paper. The structure of the four types of networks is shown in Table 2.

4.1.2 Experimental Model

The goal of our algorithm is to perform validation in the Independent Cascade (IC) model, so we use the following two models to generate non-uniform information propagation probabilities:

- UIC: i.e., Uniform Independent Cascade Model (UIC) On each edge (v, w), we uniformly choose the probability at random in the set 0.1,0.01, which corresponds to the level of influence;
- WIC:i.e., Weighted Independent Cascade Model (WIC), in which the probability of influence on each edge (v, w) is $1/d_w$, where d_w is the number of degrees of entry of node w. However, the model can generate asymmetric, non-uniform propagation probabilities even if the original graph is undirected.

4.1.3 Comparison Method

- Random: As a basic comparison algorithm, k nodes are randomly selected in graph G. The graph is referred to as Rand;
- MaxDegree: as a comparison algorithm, one that selects k nodes with a maximum degree based on their topology, abbreviated as HeuMD in the figure, with time complexity of $O(N)$;
- DegreeDiscount: proposed in the literature [6], a degree discount heuristic, abbreviated as HeuDD in the figure, with a running time of $O(klogN + M)$;
- BetweennessCentrality: proposed in [3], a heuristic based on the betweenness centrality of nodes, abbreviated as HeuBet in the figure, and the optimal running time is $O(MN)$;
- CelfGreedy: proposed in [13] a greedy algorithm optimization scheme based on submodular properties, abbreviated as Celf in the figure, and the optimal running time is $O(RMN)$.
- The method of this paper: The greedy algorithm based on two rounds of filtration proposed in this paper, referred to as TRFM in the figure, has an optimal running time of $O(zRNC)$;

4.1.4 Evaluation Indicators

Influence range: To better obtain the influence propagation scope of these algorithms, for each node seed set, we got a more stable propagation scope by Monte Carlo simulations 2000 times on UIC and WIC models, respectively. The larger the influence propagation value, the better the algorithms perform.

Runtime: We also compare the runtime of influence propagation for the set of k=50 nodes. The smaller the running time, the better the algorithm performs.

4.2 Analysis of Results

4.2.1 Experimental Results Demonstration

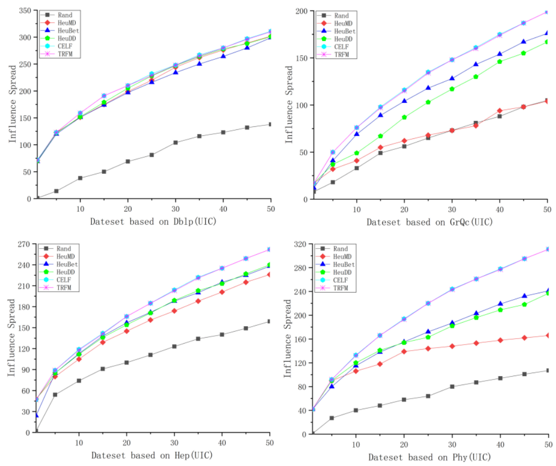

Figure 1: Experimental comparison of algorithms based on different data sets under the UIC model

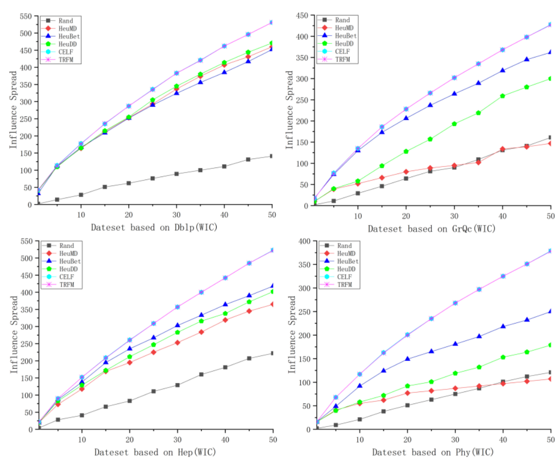

Figure 2: Experimental comparison of algorithms based on different data sets under the WIC model

Influence propagation: Figure 1 (based on the UIC model) and Figure 2 (based on the WIC model) show the scope of influence propagation based on various algorithms on four different datasets. For straightforward reading, in all the influence propagation legends, the legends rank the algorithms from the direction down according to the scope of influence propagation (k=50). Figure 3 shows the running time comparison of the Celf algorithm and TRFM algorithm when the k=50 seed set.

4.2.2 Analysis of Propagation Scope

First, the influence propagation scope based on the UIC model and the WIC model is shown in Figure 1 and Figure 2, where the CELF algorithm shown in cyan as the optimal coverage guarantees an approximate optimal solution over the four data sets, and we used as the target for the benchmark test and marked as 100%.

Secondly, the random strategy shown in black shows the practical significance of the strategy selection that must be employed. The maximum degree strategy is shown in red; However, it offers a specific propagation scope on the dblp dataset and Hep dataset; due to the generative characteristics of the scale-free network, the nodes with more significant degrees are often linked together preferentially, which quickly causes repeated coverage of the propagation scope and cancels out part of the propagation effect; therefore it has to perform poorly on the GrQc dataset, Phy dataset, even inferior to the random strategy.

Then, the betweenness centrality shown in blue and the degree discounting algorithm shown in green, either based on the UIC or WIC models, reflect a better and more stable propagation scope as heuristic strategies. However, there is still some distance to improve the propagation scope compared to the optimal one.

Finally, the CELF algorithm under the greedy strategy shown in cyan, and the TRFM algorithm based on two rounds of filtration proposed in this paper shown in pink, maintain excellent propagation scope on the four data sets, and the performance can remain stable on both the UIC model and the WIC model. The TRFM algorithm proposed in this paper can significantly approximate the optimal solution of the CELF algorithm. Benchmark comparisons of the propagation scope of different algorithms are shown in Table 3 and Table 4.

4.2.3 Running Time Analysis

Current algorithms: As shown in Figure 3 and Figure 4, the CELF algorithm corresponds to the cyan histogram, and the TRFM algorithm corresponds to the pink histogram. From the comparison of the data based on the UIC model (as shown in Figure 3), compared with CELF, the TRFM algorithm saves 96%, 89.3%, 92.9%, and 93.5% of the running time; from the comparison of the data based on WIC model (as shown in Figure 4), compared with CELF, TRFM algorithm saves 97.1%, 91.5%, 92.2%, 95.8% of the running time; the TRFM algorithm proposed in this paper substantially reduces the running time and improves the running time by about 10 times to 30 times compared with the CELF algorithm of greedy strategy and maintains good stability.

Experiments on four publicly available datasets show that the TRFM method proposed in this paper can obtain a propagation scope that can approach the optimal solution of CELF, whether based on the UIC model or the WIC model, and, at the same time, substantially reduces the computation time and maintains good stability.

Table 3: Comparison of the propagation scope of different algorithms based on the UIC model

Data\Algorithm	Rand	HeuMD	HeuMD	HeuDD	TRFM	CRLF
Dblp	44.40%	97.10%	96.10%	96.80%	99.70%	100%
GrQc	52.80%	52.30%	88.40%	83.90%	100%	100%
Hep	60.70%	86.30%	90.80%	91.60%	100%	100%
Phy	34.40%	52.40%	77.50%	76.20%	100%	100%

Table 4: Comparison of the propagation scope of different algorithms based on the WIC model

Data\Algorithm	Rand	HeuMD	HeuMD	HeuDD	TRFM	CRLF
Dblp	26.60%	86.60%	85.10%	88.70%	100%	100%
GrQc	37.60%	34.30%	84.60%	70.10%	99.80%	100%
Hep	42.40%	69.80%	79.90%	76.90%	99.80%	100%
Phy	31.90%	28.20%	66.00%	47.20%	99.70%	100%

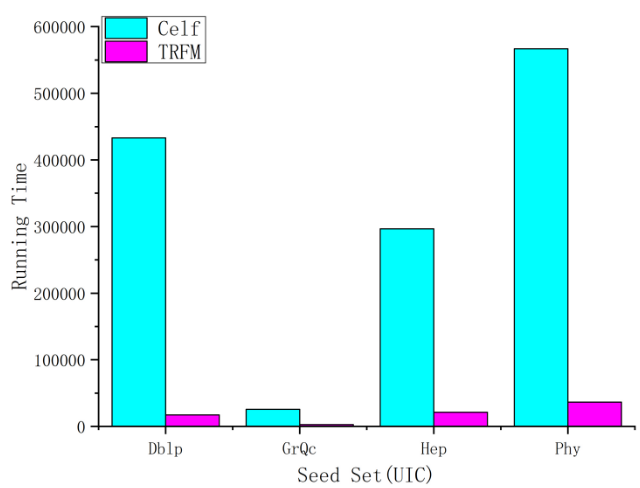

Figure 3: Comparison of the algorithm running time for different data sets based on the UIC model

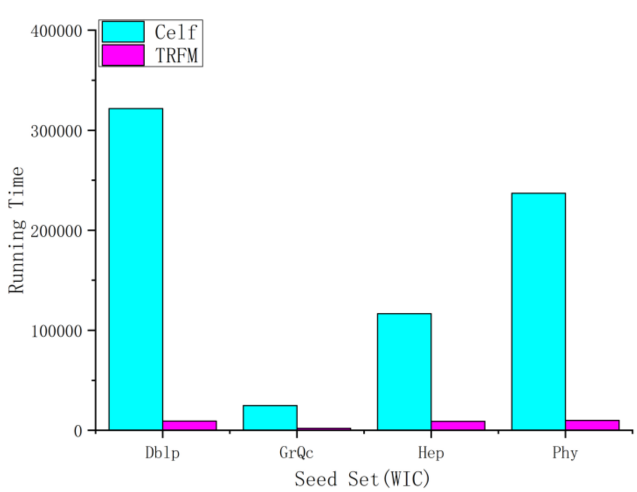

Figure 4: Comparison of the algorithm running time for different data sets based on the WIC model

5 Conclusion and Outlook

As the problem of maximizing the influence of social networks is a hot research topic, this paper proposes the TRFM algorithm with two rounds of filtration metrics, which significantly reduces the time complexity compared with current methods and approaches the optimal propagation scope on four different datasets and has stable performance. Nevertheless, there is still much room for future research work; for example, solving the influence maximization problem based on "topic semantic modeling," "large-scale social networks," "dynamic online computing," etc. may provide ideas for the application of social networks.

Acknowledgments

The research of this paper was supported by a grant from the National Social Science Foundation of China under the General Project (19BXW107). The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] "High energy physics [eb/ol]," in *Available online: http://www.arxiv.org/archive/hep*, accessed on 2003.

[2] L. Aliahmadipour and E. Valipour, "A new fuzzy propagation model for influence maximization in social networks," *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, vol. 30, pp. 279–292, 2022.

[3] M. Barthelemy, "Betweenness centrality in large complex networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 163–168, 2004.

[4] 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 (SIAM)*, pp. 946–957, 2014.

[5] M. Chen, Q. P. Zheng, V. Boginski, and E. L. Pasiliao, "Influence maximization in social media networks concerning dynamic user behaviors via reinforcement learning," *Comput Soc Netw*, pp. 8–9, 2021.

[6] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (ACM SIGKDD)*, 2009.

[7] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM)*, pp. 88–97, 2010.

[8] E. Cohen, D. Delling, T. Pajor, and R. F. Werneck, "Sketch-based influence maximization and computation: Scaling up with guarantees," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (ACM)*, pp. 629–638, 2014.

[9] G. Cordasco, L. Gargano, and A. A. Rescigno, "Active spreading in networks," in *Proceedings of the ICTCS*, pp. 149–162, 2016.

[10] G. Cordasco, L. Gargano, and A. A. Rescigno, "Influence propagation over large scale social networks," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 ,ACM*, pp. 1531–1538, 2015.

[11] 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, ACM*, pp. 35–36, 2015.

[12] A. Goyal, W. Lu, and L. V. Lakshmanan, "Simpath: An efficient algorithm for influence maximization under the linear threshold model," in *Proceedings of the 2011 IEEE 11th International Conference on Data Mining (ICDM)*, 2011.

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

[14] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, and K. Xie, "Simulated annealing based influence maximization in social networks," in *Proceedings of the 35th AAAI Conference on Artificial Intelligencet (AAAI)*, 2011.

[15] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," in *arXiv, 2011, preprint arXiv:1111.4795*, 2011.

[16] K. Jung, W. Heo, and W. Chen, "Irie: Scalable and robust influence maximization in social networks," in *2012 IEEE 12th International Conference on Data Mining (ICDM)*, pp. 918–923, 2012.

[17] D. Kempe, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (ACM SIGKDD)*, 2003.

[18] A. Kermani, R. Ghesmati, and M. S. Pishvaee, "A robust optimization model for influence maximization in social network with heterogeneous nodes," *Comput Soc Netw*, pp. 8–17, 2021.

[19] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," *Knowledge Discovery in Databases: PKDD 2006*, pp. Springer: 259–271, 2006.

[20] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[21] 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 (ACM SIGKDD)*, 2007.

[22] R. Narayanam and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation Science and Engineering(99)*, pp. 1–18, 2010.

[23] 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 (ACM SIGMOD)*, pp. 75–86, 2014.

[24] J. Tong, L. Shi, L. Liu, J. Panneerselvam, and Z. Han, "A novel influence maximization algorithm for a competitive environment based on social media data analytics," *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 130–139, 2022.

[25] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detec community structures in large-scale networks," *Physical Review*, vol. 76, no. 3, p. 036106, 2007.

[26] W. Wang, H. Yang, Y. Lu, Y. Zou, X. Zhang, S. Guo, and L. Lin, "Influence maximization in multi-relational social networks," in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 4193–4202, 2021.

[27] W. Wang and W. N. Street, "Modeling and maximizing influence diffusion in social networks for viral marketing," *Applied Network Science*, pp. 3–6, 2018.

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

[29] W. Yang, Y. Zhang, and D. Z. Du, "Influence maximization problem: Properties and algorithms," *J. Comb. Optim.*, vol. 40, p. 907–928, nov 2020.

[30] C. Zhou, P. Zhang, W. Zang, and L. Guo, "On the upper bounds of spread for greedy algorithms in social network influence maximization," *2015 IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2770–2783, 2015.

Biography

Yang Li is currently a senior engineer at the State Information Center. He graduated with a doctor's degree. His main research interests are social networks, information security and data mining.

Zhiqiang Wang received the Ph.D. degree in information security from Xidian University. He is currently an Assistant Professor with the Department of Computer Science and Technology. His research interests include system security and network security.