Influence Maximization in Partially Observable Mobile Social Networks

Zhenyu Xu[1,2], Yifan Li[1,2], Xiaolin Li[1,2], Xinxin Zhang[3], and Li Xu[1,2](✉)

[1] College of Computer and Cyber Security, Fujian Normal University, Fuzhou, Fujian, China

[2] Fujian Provincial Key Laboratory of Network Security and Cryptology, Fujian Normal University, Fuzhou, Fujian, China

xuli@fjnu.edu.cn

[3] School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, China

Abstract. Most of the existing influence maximization problems assume that k user promotion targets are selected to the entire mobile social networks (MSN) under the complete network structure. However, in reality, it is unrealistic to acquire the complete network structure. Therefore, it is our motivation to maximizing influence under partially observable networks. Firstly, we propose a new model named Variational Graph Auto-Encoder with Network Gravity (VGAE-WNG) which combined VGAE with a new effective decoder to obtain the link structure that was not presented before. Secondly, we propose a novel Similarity Decreasing Transfer Algorithm (SDTA) to evaluates the reachability of a node's influence on other nodes, by according to the transfer of similarity between nodes and the distance of information spread on the path between nodes. Finally, we performed experiments on three different scale networks. The results show that our model outperforms other algorithms by about 2% in link prediction, and our method achieves similar or even better propagation performance in the absence of partial network structures than state-of-the-art algorithms with full network structures.

Keywords: Influence maximization · Mobile social network · Link prediction · Node similarity

1 Introduction

The development of mobile device technology has made it possible for almost everyone to connect to the Internet for communication at any time, MSN have

This work was supported by the National Natural Science Foundation of China Nos. (62471139,U1905211). The authors would like to thank the National Natural Science Foundation of China No.U1905211, Science and technology projects in Fujian Province NOs. (2022G02003, 2021L3032), Fujian Provincial Department of Education Middle and Young People's Program No.JAT220814, Enterprise industry-university-research project DH-1565.

J. Wang (Ed.): MobiMedia 2023, LNICST 576, pp. 238–255, 2025.
https://doi.org/10.1007/978-3-031-60347-1_20

emerged as a result of the combination of social networks and mobile Internet. Through MSN, people can communicate anytime and anywhere even if they are separated by thousands of miles [1]. The convenience of communication in MSN enables information to be quickly and widely disseminated through important nodes in the network in a "word-of-mouth" way. Along with this phenomenon, issues such as viral marketing [2], rumor control [3] and community detection [4] have been raised and extensively studied, and how to protect private information from being leaked in the face of such rapid information diffusion is also an important research direction [5]. Researchers put forward the influence maximization (IM) problem to study this kind of communication phenomenon, which has attracted widespread attention [6]. The goal of maximizing influence is to select some high-influential seed nodes to form a node set in the network, and spread the message through the node set in order to achieve the maximum spread in the network. Kempe *et al.*[7] analyzed that maximizing influence is an NP-hard problem.

Most of the existing research on influence maximization problem assumes that the complete network topology can be obtained to analyze and solve the problem. However, in fact this hypothetical scenario is unreasonable. However, this hypothetical scenario is unreasonable. In fact, a complete network structure is difficult to obtain, and the agent can only observe part of the structure of the entire network [8]. How to maximize the influence under the partially observable network is an urgent problem to be solved, we propose Network Gravity-Similarity Decreasing Transfer Algorithm (NG-SDTA) to solve this proplem by extending variational graph auto-encoders (VGAE) [9] and proposing a novel influence maximization algorithm.

The main contributions of this paper are summarized below:

- We propose a novel model, Variational Graph Auto-Encoder with Network Gravity (VGAE-WNG), which combines variational graph auto-encoder with a new and effective decoder to predict and supplement the non-observable part of a network. To our knowledge, we are the first to incorporate link prediction in a partially observable network to aid in solving the influence maximization problem.
- We design a novel algorithm named Similarity Decreasing Transfer Algorithm (SDTA). When evaluating the reachability of a node's influence on other nodes, it takes into account the influence of the similarity on the transmission of information and the probability of successful information transmission decreases with the increase of link length.
- The experimental results of different scale networks shows that our method can achieve almost the same impact maximization effect when it has only partial network structure and other algorithms have complete network structure.

The organization structure of the remaining content is as follows. In Sect. 2, We exhibit notations used in this article, and some research work related to IM problem and link prediction are introduced. In Sect. 3, we explain the concepts of influence maximization and variational graph auto-encoders. Section 4 presents our method for addressing the IM problem. In Sect. 5, we have experimented on multiple networks and verified that our proposed methods can effectively solve the influence maximization problem under a partially observable network. Finally, we conclude and look forward to this paper in Section 6.

2 Preliminaries and Related Work

2.1 Notations and Preliminaries

All notations and meanings used are displayed in Table 1.

Table 1. Notations

Notation	Definition
$G(V, E)$	Donote a graph
V	The set of all nodes
E	The set of all edges
p	The propagation probability of each edge
S	Influence Seed Node Set
$\sigma(S)$	The size of the set of nodes that S can influence
A	Adjacency matrix
X	Initial eigenvector
Z	The latent matrix
z_i	Eigenvector of Node i
W_i	The i-th layer weight matrix
$\boldsymbol{\mu}$	Matrix of mean vectors
$\log \sigma^2$	Matrix of variance vectors

Influence Maximization Problem Compared with traditional social networks, a MSN covers various types of networks and has become a popular social communication platform that people often use [10]. In the MSN, users participate in virtual social networks through mobile devices, and users with similar interests communicate and interact on this platform [11]. Therefore, the social behavior and relationships composed of users' own behavior and the communication with other users can be used to define the structure and interaction of users with their related organizations [12].

In MSN, the structure is typically modeled using $G = (V, E)$. Each node represents an individual participating in the network, and V is a collection of these nodes. Edges represents a certain relationship between two nodes, and E represent the set of edges.

The way to solve the influence maximization problem is to find the seed node set S, so that the number of nodes influenced by S under a specific diffusion model and its rules is maximized. The expected value of the quantity activated by S is denoted by $\sigma(S)$. Above all, the following defines the IM problem. The following fomula expresses this optimization problem:

$$S = \arg\max \sigma(S)$$

Given a graph $G = (V, E)$ and seed node set size k, select k nodes to join S to maximize their influence spread under a specific diffusion model, that is, maximize $\sigma(S)$.

In the diffusion models, selects a set of initial nodes for activation, and the remaining nodes in the network are activated through the propagation information of these nodes. Many diffusion models are designed to study the influence maximization problems, of which two models are widely used: Independent Cascade Model (IC) and Linear Threshold Model (LT) [7].

In the IC model, the node activated at time t only has one chance to try to activate the unwanted neighbor borrowing points at $t+1$ time, the probability of successful activation is usually used p to represent, this behavior is independent. Each node in the LT model has multiple opportunities to activate neighbor nodes. The change of the node from unactivated to the activation state is related to activated in-degree nodes. The difficulty of each node activation is randomly generated, each node has an activation threshold $\theta \in (0, 1)$, and the behavior of the activated node is not independent. For an inactive node v, when the sum of the in-degree node's influence on it is greater than θ, node v becomes the active state. Our algorithm focuses on influence maximization under the IC model.

Variational Graph Auto-Encoders Graph auto-encoder (GAE) [9] is an unsupervised model that applies autoencoders to graph data. VGAE combines variational Bayes and GAE to generate a new model. A typical VGAE consists of two parts:

• A graph convolution neural network (GCN) [13] encoder: the inference model constructed by two GCNs is used to design as a encoder, the encoder is used to capture potential low-latitude vectors for each node in the network. which is used to:

$q(\mathbf{Z} \mid \mathbf{X}, \mathbf{A}) = \prod_{i=1}^{N} q(\mathbf{z}_i \mid \mathbf{X}, \mathbf{A})$, with $q(\mathbf{z}_i \mid \mathbf{X}, \mathbf{A}) = \mathcal{N}\big(\mathbf{z}_i \mid \boldsymbol{\mu}_i, \operatorname{diag}\left(\boldsymbol{\sigma}_i^2\right)\big)$

• An decoder: generate a new adjacency matrix A based on the vector obtained by the encoder:

$p(\mathbf{A} \mid \mathbf{Z}) = \prod_{i=1}^{N} \prod_{j=1}^{N} p(A_{ij} \mid \mathbf{z}_i, \mathbf{z}_j)$, with $p(A_{ij} = 1 \mid \mathbf{z}_i, \mathbf{z}_j) = \sigma\left(\mathbf{z}_i^\top \mathbf{z}_j\right)$

Whether there is an edge between two nodes is measured by the cosine similarity of the vector between i and j.

2.2 Related Work

Influence Maximization IM hopes to identify a nodes set that can maximize their ultimate impact under a specific network propagation model. Kempe *et al.* [7] proposed a greedy algorithm: SimpleGreedy to solve IM problem, and proved that the error between this algorithm and the optimal solution is approximately $\left(1 - \frac{1}{e} - \varepsilon\right)$. Shang *et al.* [14] designed IMPC, an IM framework built by mining the potential among neighbors in community networks. Many heuristic algorithms, such as degree centrality [15], closeness centrality [15], betweenness centrality [16], etc., measure the influence of nodes according to centrality, and select the largest top k as seed nodes. Zhang *et al.* [17] integrated ITÖ algorithm into PSO algorithm. They also proposed a modal-based influence evaluation algorithm that combined subgraphs and social attributes [18]. Li *et al.* [19] proposed the layered gravity bridge algortihm (LGB) to solve IM problem. Evaluate the influence of nodes by introducing the gravity equation into the community detection algorithm by mining the local structure information of the network. Chatterjee *et al.* [20] proposed a framework to solve the influence maximization problem by combining the community algorithm with the Shuffled Frog Leaping algorithm, so as to maximize the influence propagation of two distances under the IC model. In order to avoid the problem of overlapping influence between nodes, Liao *et al.* [21] considers the multi-hop coverage of nodes, and introduces the statistical physics method combined with the search for influential seed nodes.

All the above work is focused on the assumption of a complete network structure. However, in fact, it is difficult and costly to obtain this topology, so these works are inconsistent with the real network, more and more scholars are attracted by the problem of maximizing influence in the absence of network structure. Eshghi *et al.* [22] shown that it is NP hard to solve the IM problem in the case of lack of network structure, and proposed an agent's new computational efficient seed selection approximation algorithm, an analytical guarantee was provided for the algorithm's performance. Stein *et al.* [23] thought that influential people usually want to spread their information outside the known network, and proposed some heuristic algorithms to select nodes located at the network border to maximize the impact of the entire network.

Link Prediction Link prediction is a technique used to understand the relationships between nodes in a network and determine whether unconnected nodes will form connections in the future. This allows us to analyze network topology, predict unobserved structure of the network, and dig deeper into the network as a whole. Link prediction methods can be roughly summarized into two categories: network-based topological similarity and characteristic vector similarity, which involve comparing the network structure and feature vectors of the nodes.

The similarity of network neighbors has been widely used to solve the problem of link prediction because of its low computing complexity and high interpretation. [24]. Ghorbanzadeh *et al.* [25] extended the method, combined neighbor similarity with authority and hub, proposed a new prediction method according to the structure of co-neighbors between nodes and the motif where the node is located. Aghabozorgi and Khayyambashi [26] obtained triad similarity by calculating the number of motifs jointly participated by nodes, and proposed a new method by combining similarity of neighbor and triad similarity between nodes.

The network representation learning method aggregates and extracts the network topology information and the attribute information of the nodes, so as to obtain the vector representation of the nodes in the low-latitude space, where link generation is determined based on the vector distance between nodes. DeepWalk [27] uses random wandering to obtain the local neighborhood structure of a node, abstracts the wandering path as a phrase in the article, and obtains the feature vector of the node. Graph Auto-Encoder (GAE) [9] and Variational Graph Auto-Encoder (VGAE) [9] are both Graph Neural Network (GNN) based methods. Both models use GCN as the encoder to aggregate network topology information and attribute information. Xiao *et al.* [28] builds a new link prediction model for solving graph publishing privacy problems.

While there have been numerous studies on influence maximization, there is a lack of research on this topic in the absence of a network. The research work in this paper aims to maximize the influence of partially observable networks by exploring the underlying structure of the network.

3 Constructions of NG-SDTA

In this section, we introduce our proposed method, named Network Gravity-Similarity Decreasing Transfer Algorithm (NG-SDTA), which comprises two main parts. We first introduce a new general decoder: the network gravity equation, which is used to improve variational graph autoencoders to form a new model. Secondly, we develop a new algorithm to assess the influence of nodes. An overview of our proposed method is depicted in Fig. 1.

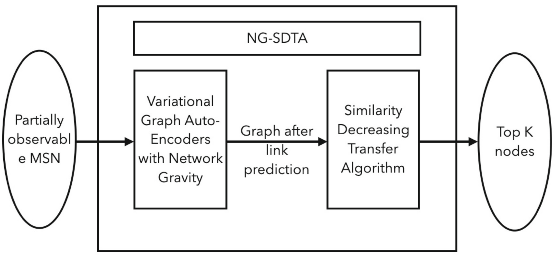


Fig. 1. Overview of the proposed methods.

As shown in Fig. 1, firstly, input partially observable MSN into the variational graph autoencoders wtih network graph to obtain the predicted graph through the link. Then, use similarity decreasing transfer algorithm to calculate the predicted graph to obtain the final Top k nodes form a seed node set.

3.1 Variational Graph Auto-Encoders Wtih Network Gravity

According to the universal law of gravitation [29], all particles in the universe attract each other due to a force known as gravity. The larger the quality of the two particles, the greater the gravity of each other, and the farther the distance between the particles will cause the gravity between particles to weaken. The formula of gravitation is as follows [29]:

$$F_{m_1 m_2} = G\frac{m_1 m_2}{r^2} \tag{1}$$

m_1 and m_2 represent the mass of the object, r represents the distance between the objects, and G represents the gravitational constant. Theoretically, the greater the gravitational force between two objects, the closer the distance between them will be. This theory also applies to MSNs. Baek and Kim [30] found that the more similar two people are, they will communicate with the other more comfortable. Therefore, in MSNs, the higher the similarity between two users, the higher the possibility of communication. We abstract the similarity between users as the product of the mass of objects in the gravitational equation, and use the number of hops between two users as the distance to formulate Network Gravity Equation:

$$F_{ij} = G_n\frac{z_i z_j}{r_{ij}^2} \tag{2}$$

The G_n represents the gravitational constant of the network itself. Different networks have their unique properties, we set G_n as a parameter that can be learned and obtained through training. Based on VAGE, we use the Network Gravity Equation as a decoder to form a new model Variational Graph Auto-Encoders with Network Gravity (VGAE-WNG).

3.2 Similarity Decreasing Transfer Algorithm

In the process of transmission, the influence presents a decaying trend. Christakis and Fowler [31] found that the spread of influence was limited to a certain range, and believed that the distance of propagation would generally not exceed three hops. When two users have similar mutual friends, the two users may establish a good relationship [24]. Considering the impact of node similarity on influence propagation, and the decreasing influence in propagation, we designed a novel algorithm named Similarity Decreasing Transfer Algorithm (SDTA).

The similarity calculation between nodes is obtained by the following formula:

$$W_{v_i v_j} = \frac{Nbr_{v_i} \cap Nbr_{v_j}}{Nbr_{v_i} \cup Nbr_{v_j}} \tag{3}$$

where Nbr_i represents the first-order neighbor node set of node i, W_{ij} represents the first-order neighbor similarity between nodes i and j.

Communication between two nodes can be achieved through multiple paths, as illustrated in Fig. 2:

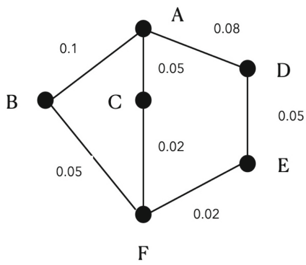


Fig. 2. Paths between two nodes.

As shown in Fig. 2, We can observe that there are 3 different paths between A and F, the probability that A affects F from each path is different, the possibility of all path propagation should be considered. From the Fig. 2, we can observe that the probability of affecting the F node through the ABF path is 0.005, the probability of successfully influencing the F node through the ACF path is 0.001, and the probability of successfully influencing the F node through the ADEF path is only 0.00008, which is almost impossible to activate the F node through this path. Therefore, such low probability paths should be ignored. At the same time, since the influence shows an attenuation trend in the process of transmission, and whether it can influence success is is positively correlated with the similarity between nodes. Above all, we designed SDTA to calculate the influence of nodes.

$$Inf_{P_1 v_i v_j} = W_{v_i v_a} * W_{v_a v_b} * \ldots * W_{v_k v_j} \tag{4}$$

$$Inf_{v_i v_j} = \sum_{k}^{n} Inf_{P_k v_i v_j} \tag{5}$$

$$STDA_{v_i} = \sum_{j \in V} (Inf_{v_i v_j} > \theta ? 1 : 0) \tag{6}$$

The Eq. 4 is used to calculate the propagation probability of node v_i to node v_j on the path P_1. The propagation probability decreases along the similarity product between nodes in the path. Add up all path propagation probabilities to get the sum of node v_i to node v_j propagation probabilities as in Eq. 5. In Eq. 6, if the influence probability of node v_i on a certain node is greater than θ, the node is treated as a true node that can be influenced by v_i, and the influence of v_i is increased by 1. On this basis, the algorithm computes the propagation probability of v_i to other nodes and counts it as the sum of the influence of v_i. The complete algorithm is presented below:

The algorithm first evaluates the similarity between two connected nodes, and then calculates the nodes' influence according to the formula. After that, all

Algorithm 1. Similarity Decreasing Transfer Algorithm.

Require: graph $G(V, E)$, nodes N
Ensure: influence node set S
for each edge e **do**
  For nodes v_i and v_j on edge e, $W_{v_i v_j} = \frac{Nbr_{v_i} \cap Nbr_{v_j}}{Nbr_{v_i} \cup Nbr_{v_j}}$
end for
for $v_i \in N$ **do**
  $STDA_{v_i} = \sum_{j \in V} (Inf_{v_i v_j} > \theta ? 1 : 0)$
end for
Sort all nodes in descending order according to SDTA
pick the nodes with the highest rank as a seed node set S
return S

nodes are arranged in descending order of influence, and the top-ranked nodes form the influence seed node set.

The algorithm can be divided into three parts based on its time complexity. The first part is to calculate the propagation probability between nodes, and the cost is directly proportional to the number of edges. Therefore, the time complexity is $O(e)$. The second part is to calculate the number of possible propagation paths from a node to any third-order neighbor node. The time cost is related to the number of third-order neighbor nodes of each node, which can be represented as $O(k_r)$, where k_r represents the number of third-order neighbors of node r. Finally, the influence of all nodes is sorted with a time complexity of $O(n)$.

4 Experimental Evaluation and Results

Through experiments on different networks to prove the performance of the method we put forward. The experiment is divided into two parts. The first part is used to verify that VGAE-WNG has better prediction performance. The second part is the influence diffusion experiment, which proves that our method can solve the influence maximization problem under partially observable network.

4.1 Link Prediction

We used two publicly available real-world datasets (Cora[32] and Citeseer[32]) for link prediction experiment, both of them are citation networks composed of mutual citations between scientists. The relevant information of the datasets is shown in Table 2.

We conducted a comparative analysis between the proposed VAGE-WNG method and two classic graph neural network models, GAE and VGAE, which are widely used for link prediction due to their excellent prediction performance. We randomly selected 10% of the dataset as the validation set and another 5% as the test set. The three models were trained on the remaining incomplete

Table 2. Network Datesets for Link Prediction.

Dateset	#Nodes(|V|)	#Edges(|E|)	Type
Cora	2708	5429	Citation
Citeseer	3327	4732	Citation

data, without the use of node attributes and initialized with weights. The Adam optimization algorithm [33] was employed for iterative training, and the number of training iterations was set to 200 with a learning rate of 0.01. The experimental results are presented in Table 3.

Table 3. Link Prediction on Networks.

	Cora		Citeseer	
Method	AUC	AP	AUC	AP
GAE	84.3 ± 0.02	88.1 ± 0.01	78.7 ± 0.02	84.1 ± 0.02
VGAE	84.0 ± 0.02	87.7 ± 0.01	78.9 ± 0.03	84.1 ± 0.02
VGAE-WNG	85.2 ± 0.02	87.5 ± 0.02	81.7 ± 0.02	85.3 ± 0.02

The results of the link prediction task on the datasets show that VGAE-WNG consistently outperforms other models on various metrics. It shows that compared with the simple cosine similarity calculation, it is more reasonable to consider the distance of users in the network in our scheme, and we describe a certain property that may exist in the network by mining the potential gravity of the network, which also helps to predict the relationship between users does it exist. Specifically, VGAE-WNG achieved the highest performance on all metrics, demonstrating the effectiveness of our proposed model for link prediction.

4.2 Influence Maximization

To evaluate the performance of our proposed influence evaluation scheme, we compared our algorithm with four other algorithms on three different datasets: Facebook [34], formed by collecting information about users using Facebook; LastFM Asia [35], which was collected from the public API and contains a network of relationships between Asian users; and ca-HepTh [36], a cooperation network from the ARXIV electronic magazine collection that includes cooperation between author papers studying the theoretical category of high-energy physics. The statistical information for these datasets is shown in the (Table 4).

Table 4. Network Datesets for influence maximization.

Dateset	#Nodes(|V|)	#Edges(|E|)	Type
Facebook	4039	88234	Social Network
LastFM Asia	7624	27806	Social Network
ca-HepTh	9877	25998	Collaboration Network

We compare with 4 algorithms to verify the effectiveness of SDTA in evaluating node influence, each algorithm selects 50 seed nodes to form the node set. The relevant information of the algorithms are as follows:

- Degree Centrality(DC) [15] : Degree centrality measures the sum of how well a node is connected to other nodes in the network.
- Eigenvector Centrality(EC) [37] : Eigenvector centrality determines the influence of nodes based on the number of their neighbors and the importance of those neighbors in the network. If a node has neighbors with high importance in the network, its own importance will also increase.
- Maximum Influence Arborescence(MIA) [2] : MIA considers the influence propagated through local arborescences in a network.
- Reversed Node Ranking (RNR) [38] : RNR ranks the importance of nodes and selects seed nodes through ranking information. Each time choose nodes, continue to delete the currently selected nodes and its neighbors in the network, and continue to select seed nodes at the remaining network nodes.

Since the propagation of information in the network often does not exceed three degrees [31], we only consider the influence of a certain node in its two-hop neighbor node set, thereby reducing computational overhead. The propagation probability between two nodes is proportional to the similarity, so the propagation probability, which is set to W_{ij} (calculated by Eq. 3). The propagation probability of edges in the IC model is generally set to 0.1 [2], and the influence of nodes can typically spread to four hops. Therefore, we set θ to the fourth power of 0.1 (0.0001).

Partially Observable Network In this part, we compared the influence diffusion effect that SDTA and other algorithms can achieve under some observable networks, and the experimental data set retains 90% of the structure. The results are shown in the figures.

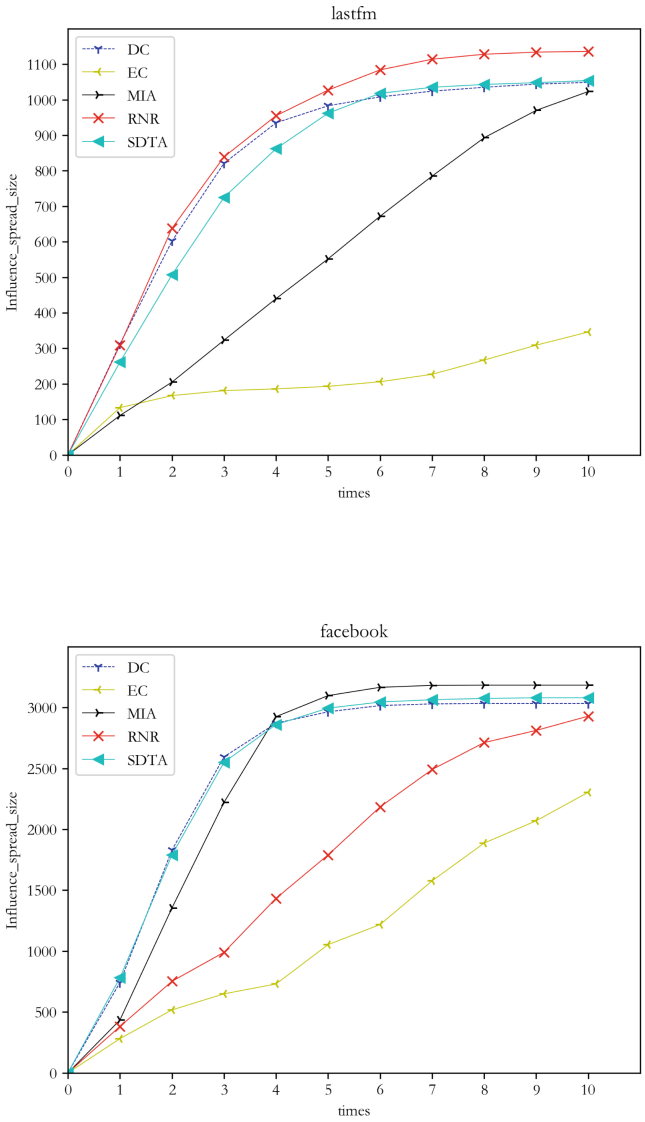
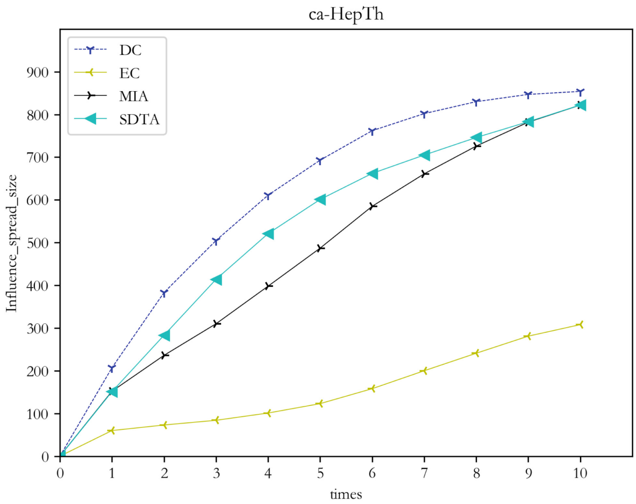


Fig. 3. Comparison of influence spread of the five algorithms on partially observable network.

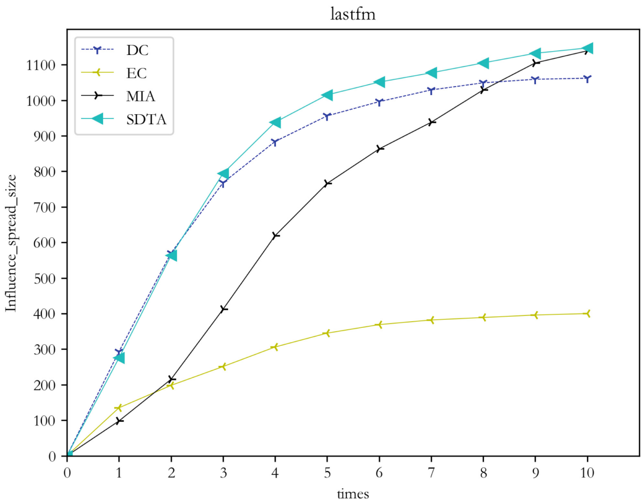


(c) Comparison of influence spread on ca-HepTh

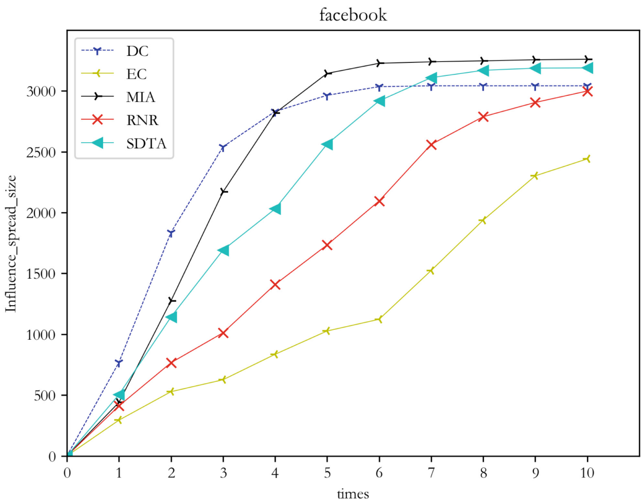
Fig. 3. (*continued*)

In figures, we can observe that SDTA had good performance in all three networks. In the LastFM Asia network, SDTA is obviously slightly inferior to DC and RNR in terms of propagation speed due to MIA and EC, but the final propagation range is comparable to DC, and RNR performs the best. Both SDTA and DC have excellent performance in the early stage in facebook, and the final spread range of both SDTA is slightly better, while MIA shows the best spread effect and RNR and EC perform poorly. In the ca-HepTh network, the RNR algorithm cannot obtain enough seed node sets due to the inability to converge, and DC has the best performance in the propagation performance, SDTA is the second, MIA has the same propagation range as SDTA at the end, and EC has the worst performance. Both MIA and DC only start from the point of view of node degree, ignoring the influence of the distance of information in propagation, while RNR considers the problem of overlapping influence, but the given convergence conditions are unreasonable and cannot converge. To sum up, SDTA has excellent and stable performance in the three networks, indicating that it is more reasonable for us to consider multiple paths to evaluate the possibility of information dissemination.

Entire Network In this section, we combined VAGE-WNG with SDTA to carry out influence propagation experiments under the condition that only 90% of the structure of the network is partially observable, and compared with other algorithms under the condition of complete network structure. We first performed complete network structure prediction by VAGE-WNG with only 90% of the network, and later used the SDTA algorithm to select the seed node set in the predicted network for influence propagation experiments.

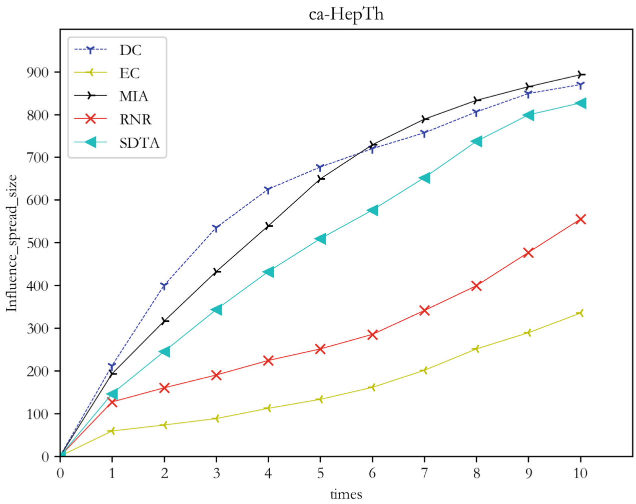


(a) Comparison of influence spread on LastFM Asia



(b) Comparison of influence spread on Facebook

Fig. 4. Comparison of influence spread between missing network structure and other algorithms without missing network structure.



(c) Comparison of influence spread on ca-HepTh

Fig. 4. (*continued*)

In LastFM Asia network, RNR is also unable to obtain the complete set of seed nodes due to non-convergence. MIA has a slower propagation speed in the early stage but achieves the same propagation range as SDTA, DC is similar to SDTA in terms of propagation speed and inferior to SDTA in terms of final propagation range. SDTA has a slightly lower propagation speed than DC and MIA in the Facebook network, but is comparable to MIA in terms of final propagation range. In the ca-HepTh network, DC and MIA have the best propagation speed and the final propagation range is similar, SDTA is weaker than the first two, and RNR is worse. EC performs the worst in all three networks.

Experimental results showed that our scheme is able to achieve comparable influence propagation with a partially observable network structure as other algorithms that have a complete network structure.

5 Conclusion

In this paper, we abstracted it as a combination of the link prediction problem and the node influence evaluation problem to address the problem of how to maximize influence in real-life situations where the complete structure of the network cannot be observed. We first develop a new decoder for VGAE to achieve better link prediction. Then a new node influence assessment method is proposed to obtain the seed node set. Experimental results show that our scheme has comparable performance to the best influence propagation effect achieved by other algorithms with a complete network structure.

In future work, how to optimize the model to achieve better link prediction results is the next focus of our research. Besides, the study of how to achieve the problem of maximizing the influence of unknown network structure will also be our research goal afterwards.

References

1. Qie, T., Chen, B., Sangaiah, A.K., Ma, J., Huang, R.: A survey of MSNs: applications, social characteristics, and challenges. IEEE Syst. J. **12**(4), 3932–3947 (2018)
2. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1029–1038 (2010)
3. Budak, C., Agrawal, D., Abbadi, A.E.: Limiting the spread of misinformation in social networks. In: Proceedings of the 20th international conference on World wide web, pp. 665-674 (2011)
4. Chunaev, P.: Community detection in node-attributed social networks: a survey. Comput. Sci. Rev. **37**, 100286 (2020)
5. Youliang, T., Zhiying, Z., Jinbo, X., Lei, C., Jianfeng, M., Changgen, P.: Achieving graph clustering privacy preservation based on structure entropy in social IoT. IEEE Int. Things J. **9**(4), 2761–2777 (2022)
6. Li, Y., Wang, Y., Tan, K.L.: Influence maximization on social graphs: a survey. IEEE Trans. Knowl. Data Eng. **30**(10), 1852–1872 (2018)
7. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network In:Proceedings of the 9th International Conference on Knowledge Discovery and Data Mining, pp. 137-146 (2003)
8. Yadav, A., Chan, H., Xin Jiang, A., Xu, H., Rice, E., Tambe, M.: Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty, In: Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems, pp. 740-748 (2016)
9. Kipf, T.N., Welling, M.: Variational graph auto-encoders, arXiv preprint arXiv:1611.07308 (2016)
10. Mao, Z., Jiang, Y., Min, G., Leng, S., Jin, X., Yang, K.: MSNs: Design requirements, architecture, and state-of-the-art technology. Comput. Commun. **100**, 1–19 (2017)
11. Boyd, D.M., Ellison, N.B.: Social network sites: definition, history, and scholarship. J. Comput.-Mediat. Commun. **13**(1), 210–230 (2007)
12. Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the social network of scientific collaborations. Phys. A **311**(3), 590–614 (2002)
13. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907 (2016)
14. Shang, J., Wu, H., Zhou, S., Zhong, J., Feng, Y., Qiang, B.: IMPC: influence maximization based on multi-neighbor potential in community networks. Phys. A **512**, 1085–1103 (2018)
15. Freeman, L.C.: Centrality in social networks conceptual clarification. Soc. Netw. **1**(3), 215–239 (1978)

16. Newman, M.J.: A measure of betweenness centrality based on random walks. Soc. Netw. **27**(1), 39–54 (2005)
17. Zhang, X., Xu, L., Gao, M.: An efficient influence maximization algorithm based on social relationship priority in msns, In: Security and Privacy in Social Networks and Big Data, PP. 164–177 (2020)
18. Zhang, X., Xu, L., Xu, Z.: Influence maximization based on network motifs in MSNs. IEEE Trans. Netw. Sci. Eng. **9**(4), 2353–2363 (2022)
19. Li, Q., Cheng, L., Li, X., Li, S., Zhu, P.: Influence maximization through exploring structural information. Appl. Math. Comput. **442**, 127721 (2023)
20. Chatterjee, B., Bhattacharyya, T., Ghosh, K.K., Chatterjee, A,. Sarkar R.: A novel meta-heuristic approach for influence maximization in social networks, Expert Systems, **4**(40) (2023) https://doi.org/10.1111/exsy.12676
21. Liao, F., Hayashi, Y.: Identify multiple seeds for influence maximization by statistical physics approach and multi-hop coverage. Appl. Netw. Sci. **7**(1), 52 (2022)
22. Eshghi, S., Maghsudi, S., Restocchi, V., Stein, S., Tassiulas, L.: Efficient influence maximization under network uncertainty, In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 365–371 (2019)
23. Stein, S., Eshghi, S., Maghsudi, S., Tassiulas, L., Bellamy, R.K.E., Jennings, N.R.: Heuristic algorithms for influence maximization in partially observable social networks. In: Proceedings of the 3rd International Workshop on Social Influence Analysis, pp. 20–32 (2017)
24. Newman, M.E.: Clustering and preferential attachment in growing networks. Phys. Rev. E **64**(2), 025102 (2001)
25. Ghorbanzadeh, H., Sheikhahmadi, A., Jalili, M., Sulaimany, S.: A hybrid method of link prediction in directed graphs. Expert Syst. Appl. **165**, 113896 (2021)
26. Aghabozorgi, F., Khayyambashi, M.R.: A new similarity measure for link prediction based on local structures in social networks. Phys. A **501**, 12–23 (2018)
27. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 701–710 (2014)
28. Li, X., Xu, L., Zhang, H., Xu, Q.: Differential privacy preservation for graph auto encoders: a novel anonymous graph publishing model. Neurocomputing **521**, 113–125 (2022)
29. Wieslaw, A.K.: Issac Newton's "Philosophiae Naturalis Principia Mathematica", World Scientific Publishing Company (1687)
30. Seungsoo, B., Seungjoo, K.: Trust-based access control model from sociological approach in dynamic online social network environment. The Sci. World J. 936319 (2014)
31. Christakis, N.A., Fowler, J.H.: Connected: the surprising power of our social networks and how they shape our lives. Little, Brown (2009)
32. Sen, P., Namata, G.M., Bilgic, M., Getoor, L., Gallagher, B., Eliassi-Rad, T.: Collective classification in network data. AI Mag. **29**(3), 93–106 (2008)
33. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2015)
34. McAuley, J., Leskovec, J.: Learning to discover social circles in ego networks, Adv. Neural Inf. Proc. Syst. **25** (2012)
35. Rozemberczki, B., Sarkar, R.: Characteristic functions on graphs: birds of a feather,from statistical descriptors to parametric models. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management, pp. 1325–1334 (2020)

36. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: densification and shrinking diameters. ACM Trans. Knowl. Discovery from Data, **1**(1), (2007)
37. Bonacich, P.: Some unique properties of eigenvector centrality. Soc. Netw. **29**(4), 555–564 (2007)
38. Rui, X., Meng, F., Wang, Z., Yuan, G.: A reversed node ranking approach for influence maximization in social networks. Appl. Intell. **49**(7), 2684–2698 (2019)