Influence Maximization in Multi-Relational Social Networks

Wei Wang, Haili Yang, Yuanfu Lu*, Yuanhang Zou, Xu Zhang, Shuting Guo, Leyu Lin
WeChat Search Application Department, Tencent Inc. China
{unoywang,hailiyang,lucasyflu,yuanhangzou,xuonezhang,tinkleguo,goshawklin}@tencent.com

ABSTRACT

Influence maximization (IM) is a classic problem, which aims to find a set of k users (called seed set) in a social network such that the expected number of users influenced by the seed users is maximized. Existing IM algorithms mainly focus on one-by-one influence diffusion among users with friendships. However, in addition to *1-to-1* friendships, *1-to-N* group relations usually exist in real social platforms, which are seldom fully exploited by conventional methods.

In this paper, with the real-world datasets in WeChat, the largest online social platform in China, we first study the IM problem in multi-relational social networks consisting of friendships and group relations, and propose a novel Generate&Extend framework to find influential seed users for product promotion . Specifically, to achieve a trade-off between effectiveness and efficiency, we present a truncated meta-seed generator to select a small number of users, which are influential with consideration of both friendships and group relations. More importantly, a structural seed extender is put forward to extend the meta-seed set, so as to encode the differentiated propagation structures between friendships and group relations. Extensive online/offline experiments on three real-world datasets demonstrate that Generate&Extend significantly outperforms the state of the arts. Our Generate&Extend has been deployed at WeChat for mini-program promoting, and severing more than 200 million users.

CCS CONCEPTS

• **Information systems → Mobile information processing systems**; **Data mining**; • **Social and professional topics → User characteristics**.

KEYWORDS

social network analysis, influence maximization, social influence

ACM Reference Format:
Wei Wang, Haili Yang, Yuanfu Lu*, Yuanhang Zou, Xu Zhang, Shuting Guo, Leyu Lin. 2021. Influence Maximization in Multi-Relational Social Networks. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3459637.3481928

*The corresponding author.

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

ACM ISBN 978-1-4503-8446-9/21/11...$15.00
https://doi.org/10.1145/3459637.3481928

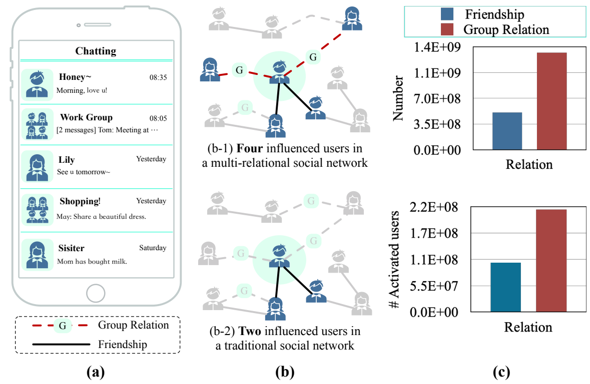


Figure 1: (a) An illustration of the chatting scenario on a social platform (e.g., WeChat), where users can chat with friends and other users in the same group (maybe non-friends). (b) Comparison of influence spread on a multi-relational social network and a conventional social network. (b-1) shows that the shaded user can influence 4 users via friendships and group relations, while there are only 2 reachable users in the conventional social network in (b-2). (c) Influence diffusion analysis of a product in WeChat within one month. The numbers of friendships and group relations indicate the number of edges involved in the sharing process of the product, and the activated users represent users who use the shared product.

1 INTRODUCTION

Online social networks (OSNs), such as Facebook with 1.5 billion and WeChat with 1.2 billion monthly active users, have come to play a vital role in numerous economical, social and political events [1, 43]. In online social networks, hundreds of millions of people actively interact with each other and share their opinions, which makes a piece of information could quickly become pervasive through the *word-of-mouth* propagation among social friends [16, 24]. Thus, seeking the most influential users in an online social network is practical and valuable for various applications, like product promotion and viral marketing [10, 25], which is well known as *influence maximization* (IM) problem [11, 22].

Formally, given an online social network $\mathcal{G}$, influence maximization asks for k users (a.k.a. *seed set*) in $\mathcal{G}$ that can influence the largest expected number of remaining users. The first seminal work presented by Kempe et al. [18] formulates the influence maximization as an optimization problem and shows that it is NP-hard in general. Along this line of research, a large body of work on influence maximization has been done in the past decade, detailed introductions can refer to [3, 6, 9–11, 18, 22, 22, 25, 27, 28, 34, 37, 38].

Although existing methods achieve either approximation guarantees or practical efficiency, they almost focus on influence diffusion between friends. In the real-world online social network, relations between two users are usually complex and diverse [32]. For instance, Fig. 1(a) illustrates a **multi-relational social network** in WeChat, an online social platform serving more than 1 billion users in China. Different from the conventional social network, the multi-relational social network consists of **friendships** and **group relations** between users, where the friendship represents that two users are in a *1-to-1* relation while the group relation means that two users are in the same group (i.e., a *1-to-N* relation) with similar interests but not necessarily friends. Intuitively, in Fig. 1(b), the shaded user can reach two more users and propagate wider influence with *1-to-N* group relations. Furthermore, in Fig. 1(c), influence diffusion analysis of one product in WeChat shows that much more group relations are involved in the influence diffusion process and most users are activated by group sharing. Thus, multiple social relations may significantly change the influence diffusion in an online social network, recasting the influence maximization problem in the need of considering not only *1-to-1* friendships but also *1-to-N* group relations.

To seek the most influential users in a multi-relational social network, one straightforward solution is to take all relations as friendships and then apply the state-of-the-art methods (e.g., SSA [28] or OPIM [36]). However, this simple idea suffers from two predominant problems: (1) *1-to-N* group relations lead to a high-degree and large-scale social network, which makes traditional IM algorithms inefficient and even infeasible. As shown in Fig. 1(b), the number of group relations is more than twice that of friendships in the multi-relational social network of WeChat. Moreover, in the multi-relational social network, a group usually comprises of hundreds of members and a user could join various groups, thus the degrees of nodes are much higher than that of nodes in a conventional social network only consisting of *1-to-1* friendships. (2) Influence diffusion in a multi-relational social network could be significantly changed. The basic assumption of traditional IM algorithms is that the more people a user can reach, the larger the influence of the user. However, this could not be strictly held in a multi-relational social network, since the *1-to-N* group relations show significant different structures from the *1-to-1* friendships, For instance, a seed user who joins many groups can reach much more people, but she/he always keeps silent and never share any content, which makes her/him a *fake seed*.

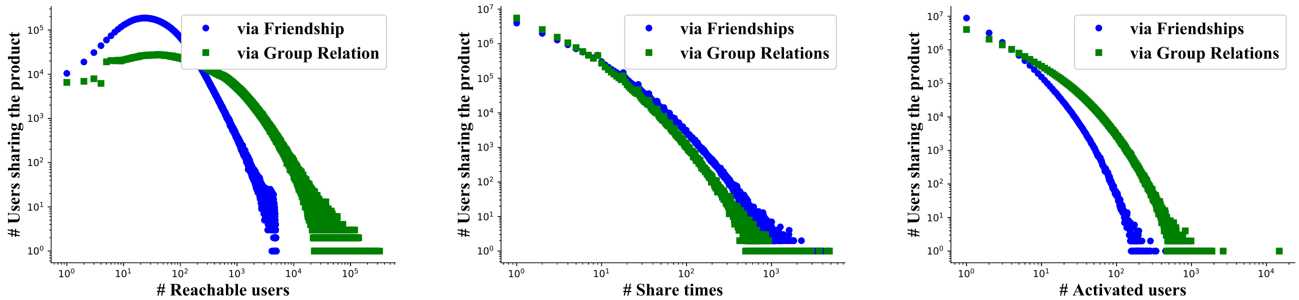
The above issues trigger us to investigate the influence maximization problem in multi-relational social networks. Motivated by the extensive analysis of a multi-relational social network in WeChat, we propose a novel **Generate&Extend** framework. The basic idea of Generate&Extend is to first find a small set of seed users, called meta-seed set, by constructing truncated reachable neighborhood sets, and then extend the meta-seed set through selecting the structure similar users for influence maximization in the multi-relational social network. More specifically, the proposed Generate&Extend consists of two components: **truncated meta-seed generator** and **structural seed extender**. Towards a trade-off between effectiveness and efficiency, a truncated meta-seed generator is put forward to address the large-scale and high-degree challenges proposed by the *1-to-N* group relations. Taking inspiration from the reverse influence sampling [7] and the observations on real data of WeChat, the truncated meta-seed generator truncates the reachable neighborhood structures, which serves for the selection of a small number of influential users, i.e., meta-seed set. On the other hand, we present a structural seed extender to cope with the different influence propagation structures of friendships and group relations in the multi-relational social network. The structural seed extender first encodes the differentiated structures into low-dimensional representations, and then takes the meta-seed set and the structure-embedded representations as input to extend the meta seeds in terms of structural similarity. Finally, the proposed Generate&Extend framework derives a seed set for influence maximization in a multi-relational social network.

To summarize, this work makes the following major contributions.

- To the best of our knowledge, this is the first attempt to investigate influence maximization in a multi-relational social network, which not only contains *1-to-1* friendships, but more importantly includes *1-to-N* group relations.
- We conduct extensive analysis on real social influence diffusion in WeChat for the first time, which motivates us to propose a novel Generate&Extend framework to find the most influential users in industrial scenarios.
- Our proposed Generate&Extend leverages a truncated meta-seed generator to select meta-seed users and a structural seed extender to generate the final seed set with the consideration of differentiated friendships and group relation structures.
- We conduct extensive empirical studies (including online and offline) on three real-world datasets in WeChat, and demonstrate that Generate&Extend consistently and significantly outperforms various state of the arts.

2 RELATED WORK

Influence Maximization. Influence maximization is first modeled as an algorithmic problem in [18], which proposes a greedy framework that returns $(1 - 1/e - \epsilon)$-approximation for several influence diffusion models. Subsequently, a plethora of greedy-based heuristics has been developed to achieve approximate solutions in the literature [6, 22]. Broadly, based on how to evaluate the influence spread, existing IM algorithms can be classified into simulation-based, proxy-based, and sketch-based approaches. The simulation-based algorithms perform Monte-Carlo (MC) simulation for evaluating influence spread, which can incorporate differernt diffusion models (e.g., IC [13] or LT [15]) and have a good theoretical property. However, the MC simulation incurs significant computational overheads [14]. Instead of performing heavy MC simulations, the proxy-based algorithms devise proxy models to approximate influence spread, which makes it more practical and efficient on large-scale networks at the expense of theoretical guarantees [11, 23]. To overcome the drawbacks of the aforementioned algorithms, the sketch-based approaches construct theoretically grounded sketches based on the diffusion model and then evaluate influence spread via the constructed sketches, which ensures that the algorithm achieves theoretical efficiency while preserving the approximation guarantee [28, 37]. Besides, there exist some works



(a) Distribution of reachable users.

(b) Distribution of share times.

(c) Distribution of activated users.

Figure 2: Data observations of a multi-relational social network in WeChat.

that attempt to incorporate content information, such as location and time [12, 21, 42].

Unfortunately, all the above-mentioned approaches are designed for social networks with only simple *1-to-1* friendship structures, therefore they cannot be directly applied for influence maximization in multi-relational social networks that are ubiquitous in real applications. In this paper, we firstly investigate both *1-to-1* friendships and *1-to-N* group relations in social networks for influence maximization problem.

Social Influence Analysis. In recent decades, the booming of various social platforms has promoted the research on social influence analysis [34, 35]. For instance, Bakshy et al. [5] investigate the attributes and relative influence of *Twitter* users by tracking diffusion events that took place on the *Twitter*. Similarly, on *Twitter* network, Azaza et al. [4] propose an influence assessment approach for *Twitter* users by considering three actions: retweet, mention and reply. Moreover, [2] and [40] attempt to analyze how social influence on *Facebook* affects the news we read and the product we purchase. While social influence analysis is much practical and valuable in real-world applications, few efforts has been made for analyzing the influence of group relations in social networks, especially for IM problem. Thus, we make a preliminary attempt to maximize influence in a multi-relational social network.

3 PRELIMINARIES

In this section, we first present the formal definition of the multi-relational social network and then formalize the problem of influence maximization in multi-relational social networks.

Definition 1. ***Multi-Relational Social Network (MRSN).*** A multi-relational social network is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, consisting of a user set $\mathcal{V}$ and a relation set $\mathcal{E} = \mathcal{E}_S \cup \mathcal{E}_G$. Here $\mathcal{E}_S$ and $\mathcal{E}_G$ are the sets of ***1-to-1* friendships** and ***1-to-N* group relations** between users. In fact, group relations formalize the social relations between two users, and two users connected via group relations usually share similar interests.

Example 1. Fig. 1(a) shows the chatting scenario of a user in WeChat, which derives a multi-relational social network. There are friendships between the user and *Honey*, *Lily* and *Sisiter*, as well as group relations with *Tom* and *May*. Intuitively, influence spreads in a group is wider and quicker than that between *1-to-1* friendships.

Definition 2. ***Influence Maximization in MRSNs.*** Given a MRSN $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a positive integer $1 \leq k \leq |\mathcal{V}|$ and a diffusion model $\mathcal{M}$, IM aims to select a seed set $\mathcal{S}^* \subset \mathcal{V}$ of k users that maximizes the influence propagation under the given diffusion model $\mathcal{M}$, i.e., the expected number of influenced users via both friendships and group relations.

4 DATA OBSERVATIONS

In this section, we look into the unique characteristics of multi-relational social networks with a real-world dataset of WeChat[1], and draw some observations motivating our proposed framework.

4.1 Background

WeChat is the largest online social communication service in China, with more than 1.2 billion monthly active users (MAUs), where users are allowed to send and receive multimedia messages, articles and mini-programs[2] (i.e., "mini-applications" running within WeChat) in real-time via Internet. One important feature in WeChat is that users can not only chat with their friends but also with several other users in the same chat group at the same time. For example, a user can share a mini-program to his/her friends or groups, and the persons who use the shared mini-program at the next moment are regarded as *activated users*. Thus, in WeChat, information can be propagated one by one through friendships, and more importantly, can also be widely diffused among numerous users in a chat group through group relations.

It is worth mentioning that there also exist other applications like WhatsApp[3] and LINE[4] that provide the similar chat group feature. Here we mainly study the influence diffusion in WeChat, and it is flexible to apply some conclusions and our Generate&Extend framework to their scenarios.

[1]https://en.wikipedia.org/wiki/WeChat

[2]https://wechatwiki.com/wechat-resources/wechat-mini-program-epic-tutorial-guide/

[3]https://en.wikipedia.org/wiki/WhatsApp

[4]https://en.wikipedia.org/wiki/Line_(software)

4.2 Data Observations

The data for this study comes from anonymized logs of a mini-program (i.e., the product to be promoted) sharing data in WeChat. We collect the sharing logs ranging from 2020/07/01 to 2020/07/31, which derives a large scale multi-relational social network consisting of 25.59 million nodes (i.e., user) and 1.8 billion edges (i.e., friendships and group relations). Note that we only have access to records of a share action but no detailed content due to user privacy. Here we conduct an analysis on the influence diffusion of the product (i.e., the mini-program) within one week, and showcase the results in Fig. 2. We have the following observations:

(1) *More users can be reached through group relationships than through friendships.* Fig. 2(a) presents the distribution of the number of reachable users w.r.t users sharing the product in our collected data. Overall, the number of users reachable via group relations (i.e., $\sum_{\textit{for each group the user joins}}$ (*#members in a group*)) is larger than the number of users reachable via friendships (i.e., the number of friends). Only a few users can reach more users via friendships since they have many friends. It inspires us to incorporate not only *1-to-1* friendships but also *1-to-N* group relations for influence maximization problem.

(2) *Influence spread through group relations is much wider and more powerful than the influence spread through friendships.* Observing from Fig. 2(b) and (c), we find that although users generally prefer to share the product through friendships, fewer sharing via group relations yields more activated users. This observation motivates us to carefully distinguish the influence diffusion of differentiated relations in a multi-relational social network.

5 METHODOLOGY

As motivated by the aforementioned data observations, we next elaborate our key designs of the Generate&Extend framework, which effectively selects a small number of influential users as the meta-seed set and extends the meta-seeds with the consideration of the differentiated diffusion structures of friendships and group relations.

5.1 Overview of Generate&Extend

As illustrated in Fig. 3, the Generate&Extend framework consists of two components: truncated meta-seed generator in Fig. 3(a) and structural seed extender in Fig. 3(b).

First, existing IM algorithms mainly focus on social networks with only *1-to-1* friendships that are small and low-degree [19, 28]. However, the *1-to-N* group relations make the social network high-degree and large-scale since a group usually contains hundreds of members. Thus, taking inspiration from reverse influence sampling (RIS) [7] for IM, we design a truncated meta-seed generator to select a small number of influential users as the meta-seed set, which only takes h-hop neighborhoods into consideration, so as to achieve a trade-off between effectiveness and efficiency for large-scale networks.

Second, since the meta-seed users are generated with a limited number for a large-scale network, it is crucial to extend the seed users with the consideration of the significant different propagation structures between *1-to-1* and *1-to-N* relations. Thus, based on a relation-aware structure similarity, our structural seed extender encodes *1-to-1* and *1-to-N* structures into low-dimensional representations with graph neural networks [17], and then extends the seed users with similar structures.

5.2 Truncated Meta-seed Generator

As motivated, towards handling the high-degree and large-scale MRSN caused by *1-to-N* group relations, it is important to balance the effectiveness and efficiency of the seed selection. Given a MRSN $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we take inspiration from [7] and define the truncated reachable set of source user v as follows:

$$\mathcal{T}_v = \{u | u \in \mathcal{N}_h(v), v \in \mathcal{V}\} \cup \{v\}, \tag{1}$$

where $\mathcal{N}_h(v)$ is the h-hop reachable neighbors of the user v via friendships or/and group relations. For example, Fig. 3(a) shows the 2-hop truncated reachable sets for read users. In essence, the truncated reachable set is a subgraph of the MRSN $\mathcal{G}$, in which the social influence is capable of spreading among each other.

Likewise, we can construct multiple truncated reachable sets as $\mathcal{T} = \{\mathcal{T}_1, \cdots, \mathcal{T}_m\}$. Intuitively, influential users are likely to appear frequently in the truncated reachable sets. In other words, a user who cover most of the truncated reachable sets is likely to be selected as a seed user and achieves the maximum influence. Thus, we can select top-k_0 seed users that cover the maximum number of truncated reachable sets as

$$\mathcal{S}_0 = \textsc{Rank}(\{v : \sum_{i=1}^{m} \min(|\{v\} \cap \mathcal{T}_i|, 1) | v \in \mathcal{V}\}, k_0), \tag{2}$$

where $\mathcal{S}_0$ is the k_0-size seed set, and $\textsc{Rank}(key : value, k_0)$ is a ranking function that outputs the top-k_0 keys based on the corresponding $value$s.

Until now, we can directly solve Eq. (2) based on the greedy algorithm for max-coverage problem [20]. However, as mentioned before, the *1-to-N* group relations in a multi-relational social network make it inefficient and even impractical in the real application scenario. Thus, we only generate a small number of users as a meta-seed set $\mathcal{S}_0$ (i.e., k_0 is small), and then present a structural seed extender to derive the final seed set that maximizes influence in a multi-relational social network.

5.3 Structural Seed Extender

With the generated meta-seed set $\mathcal{S}_0$, next we extend $\mathcal{S}_0$ to achieve the influence maximization in the multi-relational social network.

Relation-aware Structural Distance. As motivated, the extended seed set is expected to be similar to the meta seeds in structures, rather than in features, so as to avoid the influence overlap problem [41]. However, there are significant differences in the propagation structures of friendships and group relations, as analysis in Fig. 1(b) and Section 4. Thus, it is necessary to differentiate *1-to-1* and *1-to-N* relations in the measurement of the structural distance in the multi-relational social network.

Formally, given a multi-relational social network $\mathcal{G} = (\mathcal{V}, \mathcal{E} = \mathcal{E}_S \cup \mathcal{E}_G)$, a meta-seed set $\mathcal{S}_0$, we denote $\mathcal{N}_l^S(\cdot)$ and $\mathcal{N}_l^G(\cdot)$ as the l-hop neighbors of nodes via friendships and group relations, respectively. In particular, based on the l-hop neighborhood structures (i.e., $\mathcal{N}_l^S(\cdot)$ and $\mathcal{N}_l^G(\cdot)$), the relation-aware structural distance

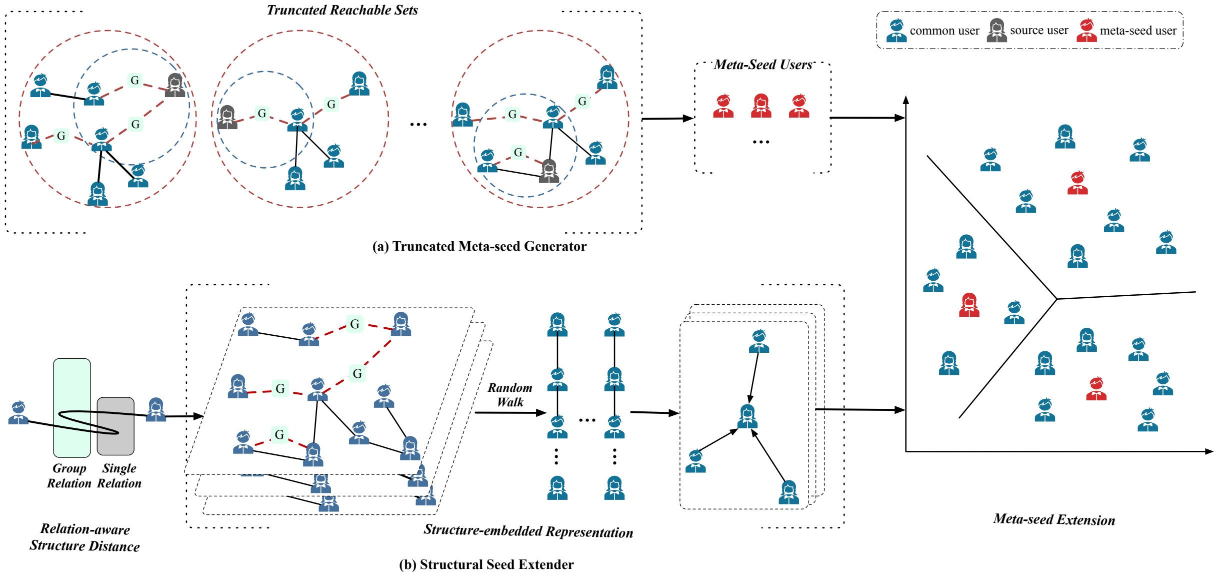


Figure 3: Illustration of the proposed Generate&Extend framework. (a) Truncated meta-seed generator, which generates multiple truncated reachable sets given the source nodes (in grey), and selects a small set of influential users (in red) that cover most truncated reachable sets as the meta-seed set. (b) Structural seed extender encodes different structures of friendships and group relations into low-dimensional representations, which are fed to a cluster to gather users with structure similar to the meta seeds as the final seed set.

between user u and user v is defined as follows:

$$\mathcal{D}(u,v,l) = \alpha * d(\mathcal{N}_l^S(u), \mathcal{N}_l^S(v)) + (1-\alpha) * d(\mathcal{N}_l^G(u), \mathcal{N}_l^G(v)), \tag{3}$$

where the hyper-parameter α is the trade-off factor between the *1-to-1* friendships and *1-to-N* group relations. The function $d(\cdot)$ measures the distance between two node sequences, which can be implemented with various distance measure methods, such as degree-based dynamic time warping [30] adopted in our work.

As neighborhood structures (i.e., $\mathcal{N}_l^S(\cdot)$ and $\mathcal{N}_l^G(\cdot)$) imply the role of nodes in the graph and social influence propagates along neighbors, it is natural that the defined structural distance $\mathcal{D}(u,v,l)$ copes with distinguishing structures between friendships and group relations and captures more refined structural similarity.

Structure-embedded Representation. By definition $\mathcal{D}(u,v,l)$ is based on the l-hop neighborhood structures, we can flexibly measure the similarity between two users in terms of different neighborhood structures. Towards fully exploring the diffusion of social influence among users, we construct a multi-layer weight graph that hierarchically encodes the structural similarity, then perform the random walk on the multi-layer weight graph to derive the user sequences, and finally learn the structure-embedded representations for users, as shown in Fig. 3(b).

Formally, given the structural distance $\mathcal{D}(u,v,l)$ based on the l-hop neighborhood structures, we hierarchically formulate the structural similarity in the l-th layer of the multi-layer graph as:

$$\mathcal{R}(u,v,l) = -\sum_{i=1}^{l} \mathcal{D}(u,v,i). \tag{4}$$

Intuitively, the above-defined hierarchical similarity determines the similarity between users base on their neighborhood structures on the bottom of the hierarchy, while the similarity depends on the entire graph at the top of the hierarchy. In essence, the structural similarity allows the social influence in the network to spread among users' neighbors and the entire graph with different weights.

Thus, in the l-th layer of the multi-layer network, the edge weight between two nodes is defined as

$$w(u,v,l) = e^{\mathcal{R}(u,v,l)}. \tag{5}$$

Naturally, the larger the distance between two users with respect to neighborhood structures, the more similar the two users, as well as the larger the edge weight between the two users. Thus, the graph in layer l can be formulated as $\mathcal{G}^l = (\mathcal{V}, \mathcal{E}^l)$, where $|\mathcal{E}^l| = \binom{|\mathcal{V}|}{2}$.

In terms of the connection of two layers, inspired by [31], we link each node to its corresponding node in the layer above and below. Thus, the edge weight between layers are defined as follows:

$$\begin{aligned} &w(u_{l-1}, u_l) = 1, \\ &w(u_l, u_{l+1}) = \log(\sum_{v \in \mathcal{V}} \mathbf{1}(w(u,v,l) > \overline{w(l)}) + e), \end{aligned} \tag{6}$$

where $w(u_{l-1}, u_l)$ and $w(u_l, u_{l+1})$ are the edge weights of the node u in the above and below layers. The term $\overline{w(l)} = \sum_{u,v \in \mathcal{V}} w(u,v,l) / \binom{|\mathcal{V}|}{2}$ is the average edge weight of the complete graph in layer l, and the log function reduces the magnitude of the potentially large number of nodes that are similar to u in a given layer.

Now, given the number of layers, denoted as l, we can formulate the multi-layer graph $\mathcal{H}$ as

$$\mathcal{H} = \mathcal{G}^1 \star \mathcal{G}^2 \star \cdots \star \mathcal{G}^l, \tag{7}$$

where the operation $\star$ represents the connection between two layers of networks. In fact, the multi-layer graph $\mathcal{H}$ can cope with the different structures between the *1-to-1* friendships and *1-to-N* group relations in the multi-relational social network $\mathcal{G}$, since the weights between nodes in each layer of $\mathcal{H}$ completely lie on the relation-aware structural distance defined before.

With the multi-layer graph $\mathcal{H}$, we next learn the structure-embedded representations for users in the MRSN $\mathcal{G}$. Following [31], we conduct a random walk on the multi-layer graph $\mathcal{H}$ to generate the context sequences that are likely to include structurally similar nodes, denoted as C. Further, inspired by recently emerging graph convolutional networks [17], we learn the structure-embedded representations for users by optimizing the co-occurrence probability:

$$\mathcal{L} = -\log \sigma(\mathbf{x}_u^\top \mathbf{x}_v) - K \cdot \mathbb{E}_{v' \sim P} \log \sigma(-\mathbf{x}_u^\top \mathbf{x}_{v'}), \tag{8}$$

where $\mathbf{x}_u \in \mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times d}$ is the structure-embedded representation for the user u in the MRSN $\mathcal{G}$, and σ means the *sigmoid* function. Here K is the number of negative sample and P is the negative sample distribution. Since v in Eq. (8) is a node that co-occurs near u in the random sequence C traversing the multi-layer graph $\mathcal{H}$, the learned representations for users is capable of encoding different structures of friendships and group relations in the original multi-relational social network $\mathcal{G}$.

Meta-seed Extension. Recall that our goal is to extend the meta-seed set considering the different propagation structures of the *1-to-1* and *1-to-N* relations, next we cluster the seeds of similar structure with the learned structure-embedded representations.

In specific, we take each seed user in the meta-seed set as a cluster center, and then input the learned structure-embedded representations $\mathbf{X}$ into a cluster to gather users with structures similar to the meta seed as the extended seeds. In particular, we have

$$\mathcal{S} = \textsc{Cluster}(\mathcal{S}_0, \mathbf{X}), \tag{9}$$

where $\mathcal{S}$ is the final seed set for influence maximization in the multi-relational social network $\mathcal{G}$. The function $\textsc{Cluster}$ can be implemented with clustering algorithms, such as K-nearest neighbor [29] adopted in our work. In fact, the structure-embedded user representations ensure that the extended seeds are similar to the meta seeds in structure, which effectively tackle the challenges of different structures and the large-scale network cased by the *1-to-N* group relations.

6 EXPERIMENTS

In this section, we conduct extensive online and offline experiments to answer three research questions: (**RQ1**) In terms of the online number of activated users by the seeds, how does Generate&Extend perform compared to the state of the arts? (**RQ2**) What is the offline consumption of the promoted product by seed users w.r.t. different methods? (**RQ3**) How does Generate&Extend benefit from the structural seed extender for both online and offline performance? Besides, a case study is provided to intuitively demonstrate the influence diffusion in a multi-relational social network.

Table 1: Statistics of the three datasets.

Dataset	7-days	15-days	31-days
#Users	179,080,550	190,932,127	255,878,282
#Friendships	117,028,138	249,286,424	508,216,853
#Group Relations	300,675,112	643,033,468	1,322,769,827

6.1 Experimental Setup

Datasets. As described in Section 4, we further extract three subsets of the real-world dataset in WeChat for evaluation, named **7-days dataset** (contains 17.91 million users ranging from 2020/07/01 to 2020/07/07), **15-days dataset** (contains 19.09 million users ranging from 2020/07/01 to 2020/07/15), and **31-days dataset** (contains 25.59 million users ranging from 2020/07/01 to 2020/07/31). Their statistics are summarized in Table 1.

Baselines. We compare our proposed Generate&Extend with several state-of-art methods in the literature, including the classic **Random-based** and **DegreeDiscountIC** [11] methods, as well as the recent **SSA** [28] and **OPIM** [36]. Please note that we have also tried to compare our Generate&Extend to some classic methods (e.g., IMM [37] and TIM [38], etc.), however, they either fail to obtain results due to time or memory complexity, or perform much worse.

- Random-based method simply selects k random nodes in the multi-relational social network.
- DegreeDiscountIC [11] is a degree discount heuristic designed for the uniform IC model [13] with a propagation probability of $p = 0.01$, as suggested in [11].
- SSA [28] is also a reverse influence sampling based algorithm that achieves the $(1 - 1/e - \epsilon)$ theoretical guarantee.
- OPIM [36] is also based on reverse influence sampling algorithm that achieve the $(1 - 1/e - \epsilon)$ theoretical guarantee.

Evaluation Metrics. We evaluate the comparison performance by predicting the next week's influence propagation of seed users. We adopt a variety of metrics widely used in the industry to measure the online and offline performance of influence maximization. For the online experiment, we mainly use the number of activated users by seed users $\mathcal{S}$ (i.e., the expected influence as in [28]) and the activation rate. Besides, to show the quality of the seed set, as mentioned before, we also measure the true influential seeds and the fake seed rate. Formally,

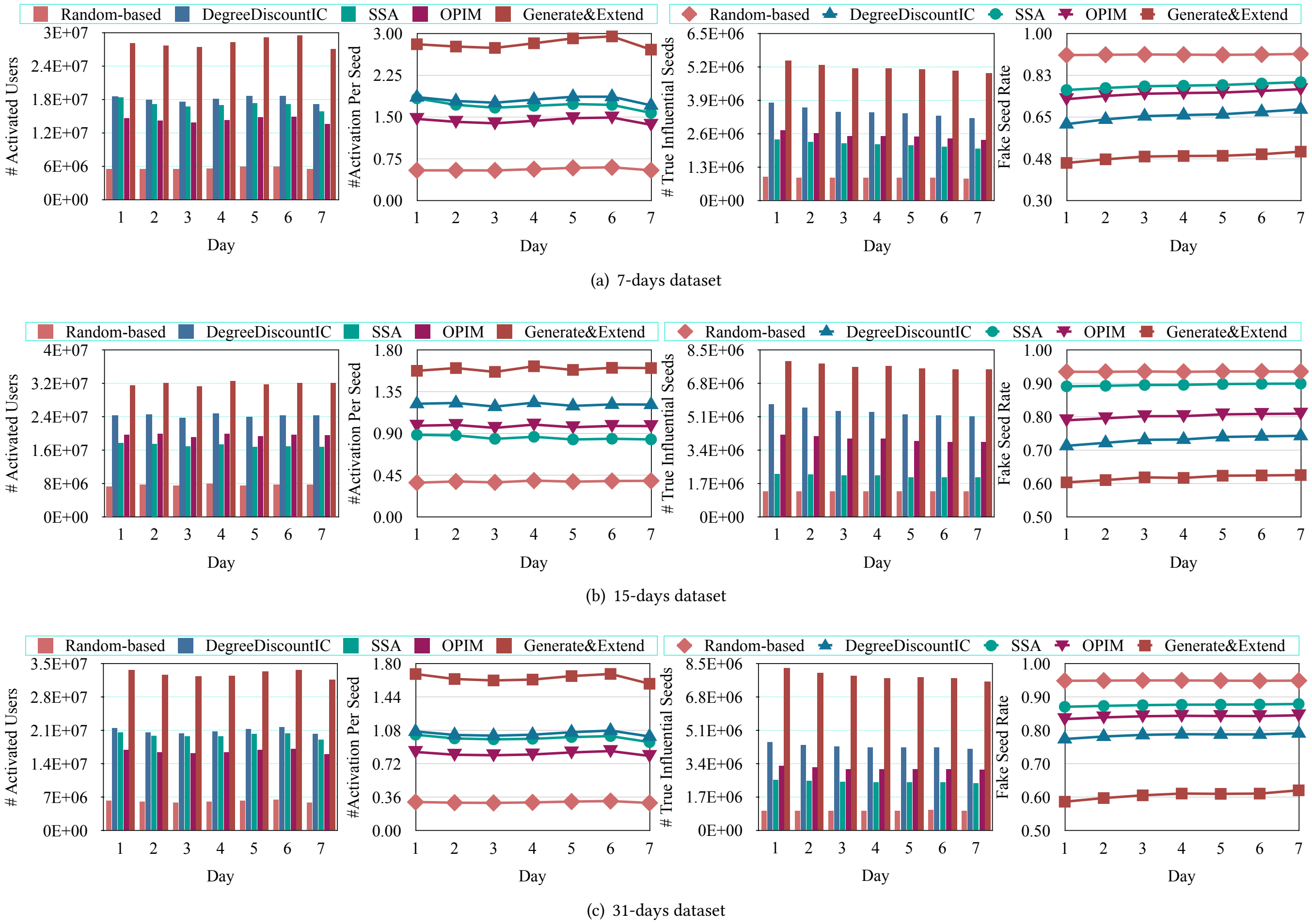
$$\#\textit{Activated Users} = \#\textit{Users that use the product shared by } \mathcal{S}$$

$$\#\textit{Activation Per Seed} = \frac{\#\textit{Activated Users}}{|\mathcal{S}|}$$

$$\#\textit{True Influential Seeds} = \#\textit{Seeds that successfully activate others}$$

$$\textit{Fake Seed Rate} = \frac{\#\textit{Seeds that unsuccessfully activate others}}{|\mathcal{S}|}$$

For the offline experiment, we adopt two widely used metrics to evaluate the offline product consumption of the selected seed users, including the sharing rate and the number of sharing times per



(a) 7-days dataset

(b) 15-days dataset

(c) 31-days dataset

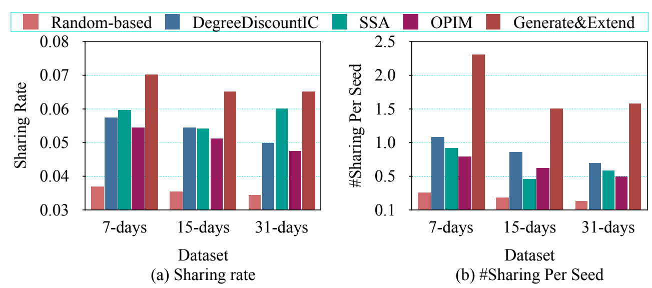
Figure 4: Online experimental results on three datasets with respect to each day in the next week.

seed user. In specific,

$$\textit{Sharing Rate} = \frac{\#\textit{Seeds who share the product}}{\#\textit{Seeds who use the product}}$$

$$\#\textit{Sharing Per Seed} = \frac{\#\textit{Total sharing times}}{|\mathcal{S}|}$$

Implementation Details. We randomly initialize our model parameters with an xavier initialization and set the optimizer as Adam. Moreover, we set the sizes of meta-seed set to 1,000,000, 2,000,000 and 2,000,000, and the sizes of the final seed set is set to 10,000,000, 20,000,000 and 20,000,000 for three datasets, respectively. The number of layers of the multi-layer graph is $l = 2$, the trade-off factor α is set to 0.4, and the dimension of structure-embedded representations for users is set to 128. We set the number of negative samples to $K = 5$ and the learning rate to $1e-5$. The hops of the truncated reachable set and structure distance are set to $h = 6$ and $l = 3$, respectively. For the baselines, we set the same seed set size as our framework and optimize other parameters according to literatures. Since the baselines are not designed for multi-relational social networks, we take all relations as the same type.



(a) Sharing rate

(b) #Sharing Per Seed

Figure 5: Offline product consumption of the selected seed users on three datasets.

6.2 Online Evaluation (RQ1)

To evaluate the online performance of methods, we mainly measure the number of activated users by seeds (i.e., the expected influence as in [28]) and the number of activation per seed. Besides, to demonstrate the quality of the seed set, we further measure the number of true influential seeds and the fake rate of seeds.

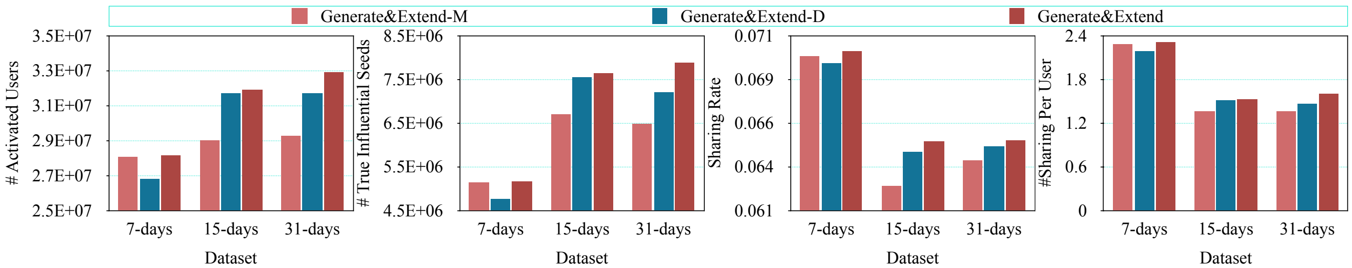
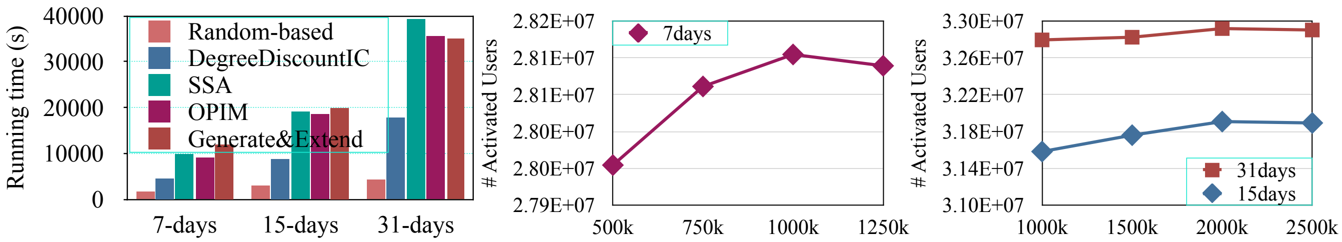


Figure 6: Analysis of Generate&Extend using various ablated models.



(a) Running time analysis.

(b) Impact of the size of meta-seed set.

Figure 7: Analysis of running time and parameter.

We report the comparison results w.r.t. every day in the next week in Fig. 4, from which we have the following findings: (1) Overall, our Generate&Extend consistently yields the highest quality seeds among all methods on three datasets. Specifically, seed users selected by our Generate&Extend significantly improve the number of activated users and the activation rate, which implies that our Generate&Extend is capable of modeling user influence in the MRSN and thus finding the most influential seeds. (2) In terms of the quality of the seed set, the number of true influential seeds in the seed set selected by our Generate&Extend is obviously larger than that of the baselines. Besides, our Generate&Extend achieves the lowest fake seed rate, i.e., the proportion of users who do not share the promoted product in the seed set, which benefits from the extension of the meta-seed with the consideration of structures of multiple relations. (3) Among different baselines, random-based method has a much worse influence spread, indicating that a careful seed selection is indeed important to effective product promotion. DegreeDiscountIC and SSA improve the quality of the seed set, but the performance is still unsatisfactory since they lose the different structures of friendships and group relations. (4) From the perspective of online influence in a week, we obverse that the number of activated users is generally stable on different days, since the number of daily active users of WeChat is stable.

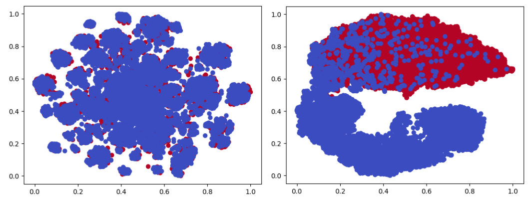
6.3 Offline Evaluation (RQ2)

We investigate the offline consumption of the promoted product by seed users: the product sharing rate, and the number of sharing times per user. Since similar trends are observed every day in next week, here we report the average performance of the entire week.

As presented in Fig. 5, the higher sharing rate indicates that more seed users share the promoted product offline, while the more sharing times per user means that the seed users are more active. Compared to the baselines, we observe that our Generate&Extend achieves the largest sharing rate, which implies that much more seed users of Generate&Extend are willing to share the product offline after using the product. We believe the reason is that seed uses selected by Generate&Extend have the similar structures encoding *1-to-N* group relations in the MRSN, which is in line with the before analysis that more activated users are brought by group relations. On the other hand, it is obvious that the seed users selected by the proposed Generate&Extend have the strongest sharing and consumption power, achieving the largest sharing times per users. As the *1-to-N* group relations significantly affect a user's sharing behavior, Generate&Extend effectively captures the different propagation structures of friendship and group relations,

6.4 Model Analysis (RQ3)

Ablation Analysis. To investigate the underlying mechanism of our Generate&Extend, we conduct an ablation analysis with two variant models: **Generate&Extend-M** without the multi-layer network, as well as **Generate&Extend-D** without differentiating friendships and group relations. As presented in Fig. 6, our Generate&Extend is consistently superior to all variant models, in terms of the number of online activated user and the offline consumption of the seed users. Among different ablated models, Generate&Extend-M is least competitive in most cases, which makes sense as it optimizes node embedding based on feature homophily (i.e., two connected node should be similar) and loses the hierarchical structure information in the multi-relational social network. Compared with Generate&Extend-M, Generate&Extend-D achieves better performance but still underperforms Generate&Extend, illustrating that



(a) Generate&Extend-M. (b) Generate&Extend.

Figure 8: Visualization of node embeddings The meta-seed nodes are in red, while the others are in blue.

the structures-embedded representation is a crucial component of Generate&Extend and differentiating friendships and group relations is vital for IM.

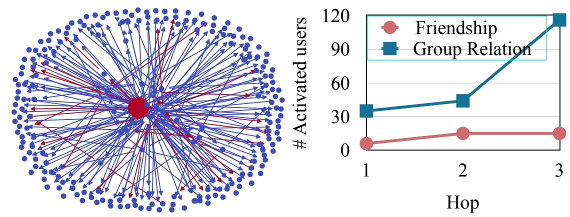
Moreover, in order to intuitively demonstrate the capability of structure-embedded representations, we comparatively visualize the low-dimensional node embeddings learned by our Generate&Extend and Generate&Extend-M. In specific, we map the low-dimensional node embeddings to a 2-D space using the t-SNE [39] package, and color the meta-seed nodes as red and the others as blue in Fig. 8. It is obvious that our Generate&Extend can better separate the meta-seed users from other users, showing that the selected meta-seed users are similar in structure rather than feature homophily. Besides, this observation shows that it is critical to extend the meta-seed users with the consideration of the structure similarly, so as to avoid the influence overlap problem [26].

Running Time Analysis. To verify the efficiency of our proposed Generate&Extend, we compare the running time of Generate&Extend to the baselines on three datasets. We run all methods on the same machine with 12 Intel(R) Xeon(R) CPUs. As presented in Fig. 7(a), as the scale of dataset increases, the advantages of our Generate&Extend become more obvious, with less running time than the comparison method. Among all baselines, the random-based and DegreeDiscountIC achieve the less running time since they do not need to operate reverse influence sampling as in other baselines and our Generate&Extend.

Parameter Analysis. Lastly, we study the impact of the size of meta-seed set (i.e., k_0) on model performance, and report the number of activated users w.r.t different meta-seed sizes in Fig. 7(b). We observe that as the number of meta seeds increases, the number of activated users also increases at the beginning, but it gradually stabilizes, indicating that Generate&Extend is generally robust w.r.t. the size of meta-seed set. Since Generate&Extend selects meta seeds without differentiating friendships and group relations in consideration of efficiency, a larger meta-seed set will include more fake seed, leading to a gradually stable influence of the entire meta seeds. The results imply the importance of structural seed extender to distinguish different structures of friendships and group relations. As we select enough meta-seeds, the method achieves the largest influence and the performance tends to be stable.

6.5 Case Study

To intuitively demonstrate the capability of modeling influence diffusion in a multi-relational network, here we conduct a case



(a) Influence propagation graph. (b) Statistics of activated users.

Figure 9: Case study of the influence diffusion of a user (the red node) through friendships (red lines) and group relations (blue lines).

study to visualize the influence propagation graph of a seed user. In specific, we randomly select a seed user in the seed set generated by our Generate&Extend, and then we enumerate the activated users influenced by the seed user during the $\{1, 2, 3\}$-hop propagation.

As showcased in Fig. 9, the blue edges occupy almost the entire graph, which indicates that the spread of influence via *1-to-N* group relations are much wider than the spread of influence via *1-to-1* friendships. Besides, we can observe that the number of activated users influenced through group relations is much larger than the number of users influenced by friendships. This case study is also in line with the data observations in Section 4, which demonstrates the effectiveness of Generate&Extend in capturing multiple relational structures in a multi-relational network.

7 CONCLUSION

In this paper, we conducted analysis on social influence diffusion in WeChat, and first studied the problem of influence maximization in multi-relational social networks. To simultaneously cope with the *1-to-1* friendships and *1-to-N* group relations, we proposed a novel Generate&Extend framework consisting of a truncated meta-seed generator and a structural seed extender. Specifically, the truncated meta-seed generator selected a small number of influential users as the meta-seed set, which achieved a trade-off between effectiveness and efficiency. More importantly, a structural seed extender was put forward to differentiate influence propagation structures of friendships and group relations, and then derived the final seed set for influence maximization in a multi-relational social network. Extensive experiments on three datasets of WeChat demonstrated that Generate&Extend significantly outperforms the baselines in online and offline scenarios.

8 ETHICS STATEMENT

While our proposed Generate&Extend is evaluated on datasets of WeChat, it sheds an interesting insight for influence maximization in multi-relational social networks, which is adaptable to various scenarios in online social platforms. As we all know, machine learning in general runs the risk of violating privacy due to the collection of user data, especially in the filed of social network analysis [8, 33]. Thus, to avoid any leakage of personal privacy information, in our work, we desensitize the collected data and only use meaningless ids to denote users. We state that the datasets are strictly anonymous and have no access to the detailed user profiles.

REFERENCES

[1] Eva Anduiza, Camilo Cristancho, and José M Sabucedo. 2014. Mobilization through online social networks: the political protest of the indignados in Spain. *Information, Communication & Society* 17, 6 (2014), 750–764.

[2] Nicolas M Anspach. 2017. The new personal influence: How our Facebook friends influence the news we read. *Political Communication* 34, 4 (2017), 590–606.

[3] Akhil Arora, Sainyam Galhotra, and Sayan Ranu. 2017. Debunking the myths of influence maximization: An in-depth benchmarking study. In *Proceedings of SIGMOD*. 651–666.

[4] Lobna Azaza, Sergey Kirgizov, Marinette Savonnet, Eric Leclercq, and Rim Faiz. 2015. Influence assessment in twitter multi-relational network. In *Proceedings of SITIS*. 436–443.

[5] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of WSDM*. 65–74.

[6] Suman Banerjee, Mamata Jenamani, and Dilip Kumar Pratihar. 2020. A survey on influence maximization in a social network. *Knowledge and Information Systems* (2020), 1–39.

[7] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of SIAM*. 946–957.

[8] Salvatore A Catanese, Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. Crawling facebook for social network analysis purposes. In *Proceedings of WIMS*. 1–8.

[9] Wei Chen, Wei Lu, and Ning Zhang. 2012. Time-Critical Influence Maximization in Social Networks with Time-Delayed Diffusion Process. In *Proceedings of AAAI*.

[10] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of KDD*. 1029–1038.

[11] Wei Chen, Yajun Wang, and Siyu Yang. 2009. Efficient influence maximization in social networks. In *Proceedings of KDD*. 199–208.

[12] Nan Du, Le Song, Manuel Gomez Rodriguez, and Hongyuan Zha. 2013. Scalable influence estimation in continuous-time diffusion networks. In *Proceedings of NeurIPS*. 3147–3155.

[13] Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters* 12, 3 (2001), 211–223.

[14] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. 2011. Celf++ optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of WWW*. 47–48.

[15] Mark Granovetter. 1978. Threshold models of collective behavior. *American journal of sociology* 83, 6 (1978), 1420–1443.

[16] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. 2013. Information diffusion in online social networks: A survey. *ACM Sigmod Record* 42, 2 (2013), 17–28.

[17] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Proceedings of NeurIPS*. 1024–1034.

[18] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of KDD*. 137–146.

[19] Moein Khajehnejad, Ahmad Asgharian Rezaei, Mahmoudreza Babaei, Jessica Hoffmann, Mahdi Jalili, and Adrian Weller. 2020. Adversarial Graph Embeddings for Fair Influence Maximization over Social Networks. In *Proceedings of IJCAI*. 4306–4312.

[20] Samir Khuller, Anna Moss, and Joseph Seffi Naor. 1999. The budgeted maximum coverage problem. *Information processing letters* 70, 1 (1999), 39–45.

[21] Guoliang Li, Shuo Chen, Jianhua Feng, Kian-lee Tan, and Wen-syan Li. 2014. Efficient location-aware influence maximization. In *Proceedings of SIGMOD*. 87–98.

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

[23] Qi Liu, Biao Xiang, Enhong Chen, Hui Xiong, Fangshuang Tang, and Jeffrey Xu Yu. 2014. Influence maximization over large-scale social networks: A bounded linear approach. In *Proceedings of CIKM*. 171–180.

[24] Yuanfu Lu, Ruobing Xie, Chuan Shi, Yuan Fang, Wei Wang, Xu Zhang, and Leyu Lin. 2020. Social Influence Attentive Neural Network for Friend-Enhanced Recommendation. In *Proceedings of ECML-PKDD*, Vol. 12460. 3–18.

[25] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. 2008. Mining social networks using heat diffusion processes for marketing candidates selection. In *Proceedings of CIKM*. 233–242.

[26] Peter V Marsden and Noah E Friedkin. 1993. Network studies of social influence. *Sociological Methods & Research* 22, 1 (1993), 127–151.

[27] Akash Mittal, Anuj Dhawan, Sahil Manchanda, Sourav Medya, Sayan Ranu, and Ambuj Singh. 2019. Learning heuristics over large graphs via deep reinforcement learning. *arXiv preprint arXiv:1903.03332* (2019).

[28] Hung T Nguyen, My T Thai, and Thang N Dinh. 2016. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In *Proceedings of SIGMOD*. 695–710.

[29] Leif E Peterson. 2009. K-nearest neighbor. *Scholarpedia* 4, 2 (2009), 1883.

[30] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2013. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data* 7, 3 (2013), 1–31.

[31] Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. 2017. struc2vec: Learning node representations from structural identity. In *Proceedings of KDD*. 385–394.

[32] Marko A Rodriguez and Joshua Shinavier. 2010. Exposing multi-relational networks to single-relational network analysis algorithms. *Journal of Informetrics* 4, 1 (2010), 29–41.

[33] Michael Seufert, Tobias Hoßfeld, Anika Schwind, Valentin Burger, and Phuoc Tran-Gia. 2016. Group-based communication in WhatsApp. In *2016 IFIP networking conference (IFIP networking) and workshops*. IEEE, 536–541.

[34] Jimeng Sun and Jie Tang. 2011. A survey of models and algorithms for social influence analysis. In *Social network data analytics*. Springer, 177–214.

[35] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. 2009. Social influence analysis in large-scale networks. In *Proceedings of KDD*. 807–816.

[36] Jing Tang, Xueyan Tang, Xiaokui Xiao, and Junsong Yuan. 2018. Online processing algorithms for influence maximization. In *Proceedings of SIGMOD*. 991–1005.

[37] Youze Tang, Yanchen Shi, and Xiaokui Xiao. 2015. Influence maximization in near-linear time: A martingale approach. In *Proceedings of SIGMOD*. 1539–1554.

[38] Youze Tang, Xiaokui Xiao, and Yanchen Shi. 2014. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of SIGMOD*. 75–86.

[39] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[40] Jyun-Cheng Wang and Ching-Hui Chang. 2013. How online social ties and product-related risks influence purchase intentions: A Facebook experiment. *Electronic Commerce Research and Applications* 12, 5 (2013), 337–346.

[41] Ning Wang, Zi-Yi Wang, Jian-Guo Liu, and Jing-Ti Han. 2019. Maximizing spreading influence via measuring influence overlap for social networks. *arXiv preprint arXiv:1903.00248* (2019).

[42] Xiaoyang Wang, Ying Zhang, Wenjie Zhang, and Xuemin Lin. 2016. Efficient distance-aware influence maximization in geo-social networks. *IEEE Transactions on Knowledge and Data Engineering* 29, 3 (2016), 599–612.

[43] Christine B Williams and Girish J'Jeff' Gulati. 2013. Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008. *New Media & Society* 15, 1 (2013), 52–71.