Influence Maximization across Partially Aligned Heterogenous Social Networks

Qianyi Zhan[1], Jiawei Zhang[2], Senzhang Wang[3],

Philip S. Yu[2,4], and Junyuan Xie[1]

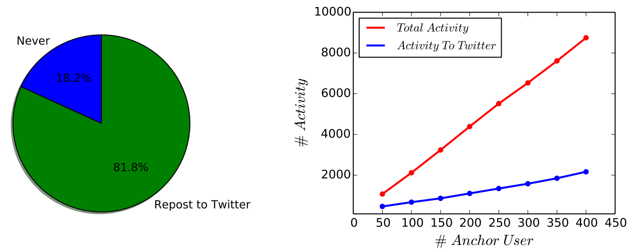
[1] National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
[2] University of Illinois at Chicago,Chicago, IL, USA
[3] Beihang University, Beijing, 100191, China
[4] Institute for Data Science, Tsinghua University, Beijing, 100084, China
zhanqianyi@gmail.com, jzhan9@uic.edu, szwang@buaa.edu.cn, psyu@uic.edu, jyxie@nju.edu.cn

Abstract. The influence maximization problem aims at finding a subset of seed users who can maximize the spread of influence in online social networks (OSNs). Existing works mostly focus on one single homogenous network. However, in the real world, OSNs (1) are usually heterogeneous, via which users can influence each others in multiple channels; and (2) share common users, via whom information could propagate across networks.

In this paper, for the first time we study the influence maximization problem in multiple partially aligned heterogenous OSNs. A new model, multi-aligned multi-relational network influence maximizer (M&M), is proposed to address this problem. M&M extracts multi-aligned multi-relational networks (MMNs) from aligned heterogeneous OSNs based on a set of inter and intra network social meta paths. Besides, M&M extends traditional linear threshold (LT) model to depict the information diffusion across MMNs. In addition, M&M, which selects seed users greedily, is proved to achieve a $(1 - \frac{1}{e})$-approximation of the optimal solution. Extensive experiments conducted on two real-world partially aligned heterogeneous OSNs demonstrate its effectiveness.

1 Introduction

Witnessing the rapid growth of online social networks, viral marketing (i.e., influence maximization) in social networks has attracted much attention of data mining community in the last decade [5, 7, 10]. Traditional viral marketing problem aims at selecting the set of seed users to maximize the awareness of ideas or products merely based on the *social connections* among users in *one single social network* [3, 8, 11]. However, in the real world, social networks usually contain heterogeneous information [18–20], e.g., various types of nodes and complex links, via which users are extensively connected and have multiple channels to influence each other [9].



(a) anchor users' reposting (b) cross-network reposted activities

Fig. 1: Cross-network information propagation analysis.

Meanwhile, as studied in [13, 20], users nowadays are usually involved in multiple social networks simultaneously to enjoy more social network services. The shared users across multiple social networks are named as *anchor users* [13]. *Anchor users* exist widely in the real world. Via these anchor users, influence can propagate not only within but also across social networks [16]. To support such a claim, we investigate the partially aligned network dataset studied in this paper (i.e., Twitter and Foursquare) and the results are given in Fig. 1. In Fig. 1(a), we randomly sample a subset of anchor users from Foursquare and observe that 409 out of 500 (i.e., 81.8%) sampled users have reposted their activities (e.g., tips, location checkins, etc.) to Twitter. Meanwhile, the activities reposted by these 409 anchor users only account for a small proportion of their total activities in Foursquare, as shown in Fig. 1(b).

In this paper, we study the influence maximization problem across multiple partially aligned heterogenous social networks simultaneously. This is formally defined as the *Aligned Heterogeneous network Influence maximization (AHI)* problem. The *AHI* problem studied in this paper is very important and has extensive concrete applications in real-world social networks, e.g., *cross-community* [1] even *cross-platform* [16] *product promotion* [17] and *opinion diffusion* [2].

To help illustrate the *AHI* problem, we give an example in Fig. 2, where Fig. 2-A shows the two partially aligned heterogeneous input networks. To conduct viral marketing in the input networks and solve the *AHI* problem, we first extract multiple influence channels (i.e., multi-relations) among users with the heterogeneous information (e.g., traditional follow links, retweet, location checkins, as well as anchor links, etc.) and then select the optimal seed user set based on the constructed multi-relational network, as shown in Fig. 2-B.

The AHI problem is a novel problem and totally different from conventional works on information diffusion and influence maximization, including:(1) traditional viral marketing problems in one single homogeneous social network [6, 12, 17], like the Twitter network shown in Fig. 2-C; (2) topic diffusion in heterogeneous information networks [9], which explores information diffusion in one single multi-relational network (e.g., the Twitter network in Fig. 2-D); and (3) influence maximization in multiplex social networks [16], which studies infor-

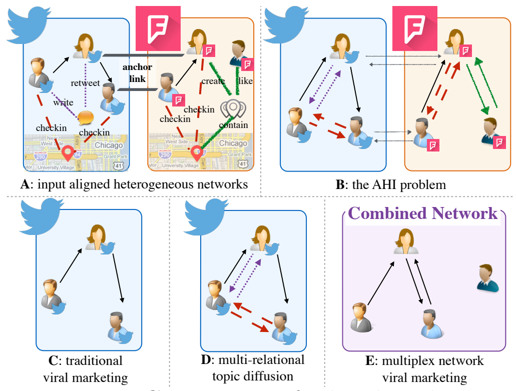


Fig. 2: Comparison of related works

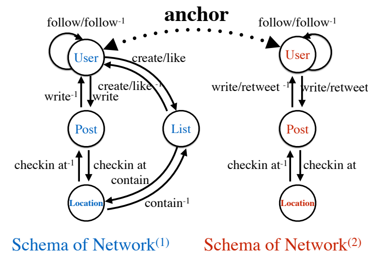


Fig. 3: Network schema

mation maximization problem across multiple homogeneous social networks by simply combining multiple networks into one single homogeneous network (e.g., the network shown in Fig. 2-E). In paticular, [16] assumes that the shared users will propagate all the information reaching them to the other network, which is unrealistic and severely violates our observation in Fig. 1(b). Different from all these related works, in the *AHI* problem: (1) the social networks are heterogeneous [18]; (2) multiple social networks [20] are studied simultaneously, where the different heterogeneous networks may have different structures or network schema as shown in Fig. 3; and (3) social networks studied in this paper are partially aligned by *anchor links* [20] instead of being simply merged together.

Addressing the *AHI* problem is very difficult due to the following challenges:

– *Information Diffusion in Heterogeneous Networks*: Users in heterogeneous networks are extensively connected with each other by different types of links and information can diffuse among users via different channels. Modeling information diffusion in heterogeneous social networks is very challenging.
– *Cross-Network Information Propagation*: Via the anchor links, information can propagate across networks. Modeling inter-network information diffusion remains an open problem.
– *NP-hard*: The *AHI* problem is proved to be *NP-hard*, which cannot be solved in polynomial time.

To address the above challenges, a new model Multi-aligned Multi-relational network influence maximizer (M&M) is proposed in this paper. M&M first extracts multi-aligned multi-relational networks with the heterogeneous information across the input OSN based on a set of inter and intra network social meta paths [18, 20]. M&M extends the traditional Linear Threshold (LT) model to depict the information propagation within and across these multi-aligned multi-relational networks. Based on the extended diffusion model, the influence function which maps seed user set to the number of activated users is proved to be both *monotone* and *submodular*. Thus the greedy algorithm used in M&M,

which selects seed users greedily at each step, is proved to achieve a $(1 - \frac{1}{e})$-approximation of the optimal result.

The remaining parts of this paper are organized as follows. We formulate the studied problem in Section 2. In Sections 3-4, we introduce the proposed M&M method. Experiments are given in Section 5. Finally, we introduce the related works in Section 6 and conclude the paper in Section 7.

2 Problem Formulation

In this paper, we will follow the definitions of concepts "*anchor user*", "'*heterogeneous networks*", "*aligned networks*", ect., proposed in [20]. Based on the definitions of these terminologies, the *AHI* problem can be formulated as follows:

AHI: Given two partially aligned networks [20] $G^{(1)}$ and $G^{(2)}$ together with the undirected anchor link set $\mathcal{A}$ [13] between $G^{(1)}$ and $G^{(2)}$, the user sets of $G^{(1)}$ and $G^{(2)}$ can be represented as $\mathcal{U}^{(1)}$ and $\mathcal{U}^{(2)}$ respectively. Let $\sigma(\cdot) : \mathcal{Z} \to \mathbb{R}, \mathcal{Z} \subset \mathcal{U}^{(1)} \cup \mathcal{U}^{(2)}$ be the *influence function* [12] which maps the seed user set $\mathcal{Z}$ to the number of users influenced by users in $\mathcal{Z}$. The *AHI* problem aims at selecting the optimal set $\mathcal{Z}^*$ which contains d seed users to maximize the propagation of information across the networks, i.e., $Z^* = \arg\max_{\mathcal{Z} \subseteq \mathcal{U}^{(1)} \cup \mathcal{U}^{(2)}} \sigma(\mathcal{Z})$.

3 Proposed Model

In this section, we will introduce the method M&M in details. M&M can extract multi-aligned multi-relation networks (MMNs) based on a set of inter and intra network social meta paths. The traditional LT model is extended in M&M to depict the information propagation across MMNs.

3.1 Multi-aligned Multi-relational Networks Extraction

We utilize the meta paths [18,20] defined based on the *network schema* to extract multi-aligned multi-relational networks with the heterogeneous information in aligned networks.

Definition 1 ***Network Schema:*** *For the given network G, its network schema can be defined as $S_G = (O, R)$ with O and R denoting the set of node types and link types in G.*

For the partially aligned input networks shown in Fig. 2-A. We note that the network schemas of the two networks are different, so the heterogeneous networks cannot be simply merged together as in the homogeneous case [16]. Based on the network schema, we can represent the diffusion channels as a set of intra and inter network social meta paths that are defined as follows.

Definition 2 ***Intra-network Social Meta Path:*** *An intra-network social meta path $\mathcal{P}$, based on the given network schema $S_G = (O, R)$, is denoted as $\mathcal{P} = O_1 \xrightarrow{R_1} O_2 \xrightarrow{R_2} \cdots \xrightarrow{R_{k-1}} O_k (k > 1)$ where $O_i \in O, i \in \{1, 2, \cdots, k\}$ and*

$R_i \in R, i \in \{1, 2, \cdots, k-1\}$. *In addition,* $O_1 \cdots O_k = User \in O$ *as we are mainly concerned about meta paths connecting users, i.e., social meta paths [20].*

Definition 3 ***Inter-network Social Meta Path:*** *Given two partially aligned heterogenous networks* $G^{(1)}$ *and* $G^{(2)}$ *with network schemas* $S_{G^{(1)}} = (O^{(1)}, R^{(1)})$ *and* $S_{G^{(2)}} = (O^{(2)}, R^{(2)})$*,* $\mathcal{Q} = O_1 \xrightarrow{R_1} O_2 \xrightarrow{R_2} \cdots \xrightarrow{R_{k-1}} O_k (k > 1)$ *can be defined to be an inter-network social meta path between* $G^{(1)}$ *and* $G^{(2)}$*, where* $O_i \in O^{(1)} \cup O^{(2)}, i \in \{1, 2, \cdots, k\}$*,* $R_i \in R^{(1)} \cup R^{(2)} \cup \{Anchor\}, i \in \{1, 2, \cdots, k-1\}$ *and Anchor is the anchor link type. Furthermore,* $O_1 = User \in O^{(1)}$*,* $O_k = User \in O^{(2)}$*, and* $\exists m \in \{1, 2, \cdots, k-1\}$ *such that* $R_m = \{Anchor\}$*.*

In both Foursquare and Twitter, users can follow other users and check-in at locations, forming two intra-network influence channels among users. Meanwhile, (1) in Foursquare, users can create/like lists containing a set of locations; (2) while in Twitter, users can retweet other users' tweets, both of which will form an intra-network influence channel among users in Foursquare and Twitter respectively. The set of intra network social meta paths considered in this paper as well as their physical meanings are listed as follows:

intra-network social meta paths in Foursquare

(1) *follow*: User $\xrightarrow{follow^{-1}}$ User

(2) *co-location checkins*: User $\xrightarrow{checkin}$ Location $\xrightarrow{checkin^{-1}}$ User

(3) *co-location via shared lists*: User $\xrightarrow{create/like}$ List $\xrightarrow{contain}$ Location $\xrightarrow{contain^{-1}}$ List $\xrightarrow{create/like^{-1}}$ User

intra-network social meta paths in Twitter

(1) *follow*: User $\xrightarrow{follow^{-1}}$ User

(2) *co-location checkins*: User $\xrightarrow{checkin}$ Location $\xrightarrow{checkin^{-1}}$ User

(3) *contact via tweet*: User $\xrightarrow{write}$ Tweet $\xrightarrow{retweet}$ Tweet $\xrightarrow{write^{-1}}$ User

Users can diffuse information across networks via the anchor links formed by anchor users. This can be abstracted as *inter-network social meta path*: *User* $\xrightarrow{Anchor}$ *User*. By taking the inter-network meta paths into account, the studied problem becomes even more complex due to the fact that non-anchor users in both networks can also be connected via intra- and inter-network meta paths. As a result, the number of social meta path instances grows mightily.

Each meta path defines an influence propagation channel among linked users. If linked users u, v are connected by only intra-network meta path, we say u has *intra-network relation* to v, otherwise there is *inter-network relation* between them. Based on these relations, we can construct multi-aligned multi-relational networks (e.g., the network shown in Fig. 2-B) for the aligned heterogeneous networks (e.g., the networks shown in Fig. 2-A). The formal definition of multi-aligned multi-relational networks is given as follows:

Definition 4 ***Multi-Aligned Multi-Relational Networks:*** *For two given heterogenous networks* $G^{(1)}$ *and* $G^{(2)}$*, we can define the multi-aligned multi-relational network constructed based on the above intra and inter network social* *meta paths as $M = (U, E, R)$, where $U = U^{(1)} \cup U^{(2)}$ denote the user nodes in the MMNs M. Set E is the set of links among nodes in U and element $e \in E$ can be represented as $e = (u, v, r)$ denoting that there exists at least one link (u, v) of link type $r \in R = R^{(1)} \cup R^{(2)} \cup \{Anchor\}$, where $R^{(1)}$, $R^{(2)}$ are the intra-network link types of networks $G^{(1)}$, $G^{(2)}$ and the inter-network Anchor link between $G^{(1)}$ and $G^{(2)}$ respectively.*

3.2 Influence Propagation in Multi-Aligned Multi-Relational Networks

In this subsection, we will extend the traditional *linear threshold* (LT) model to handle the information diffusion across the multi-aligned multi-relational networks (MMNs).

In traditional *linear threshold* (LT) model for single homogeneous network $G = (V, E)$, user $u_i \in V$ can influence his neighbor $u_k \in \Gamma_{in}(u_i) \subseteq V$ according to weight $w_{i,k} \geq 0$ ($w_{i,k} = 0$ if u_i is *inactive*), where $\Gamma_{in}(u_i)$ represents the users following u_i (i.e., set of users that u_i can influence) and $\sum_{u_k \in \Gamma_{in}(u_i)} w_{i,k} \leq 1$. Each user, e.g., u_i, is associated with a *static threshold* θ_i, which represents the minimal required influence for u_i to become *active*.

Meanwhile, based on the MMNs $M = (U, E, R)$, the weight of each pair of users with different diffusion relations is estimated by pathsim [18]. Formally, the intra-network (inter-network) diffusion weight between user u and v with relation $i(j)$ is defined as:

$$\phi^i_{(u,v)} = \frac{2|P^i_{(u,v)}|}{|P^i_{(u,)}| + |P^i_{(,v)}|},\ \psi^j_{(u,v)} = \frac{2|Q^j_{(u,v)}|}{|Q^j_{(u,)}| + |Q^j_{(,v)}|},$$

where $P^i_{(u,v)}(Q^j_{(u,v)})$ is the set of intra-network (inter-network) diffusion meta paths instances, starting from u and ending at v with relation $i(j)$. $|\cdot|$ denotes the size of the set. Thus, $P^i_{(u,)}(Q^j_{(u,)})$ and $P^i_{(,v)}(Q^j_{(,v)})$ means the number of meta path instances with users u, v as the starting and ending users, respectively.

Based on the traditional LT model, influence propagates in discrete steps in the network. In step t, all *active* users remain *active* and *inactive* user can be *activated* if the received influence exceeds his threshold. Only activated users at step t can influence their neighbors at step $t+1$ and the activation probability for user v in one network (e.g., $G^{(1)}$) with intra-network relation i and inter-network relation j can be represented as $g^{(1)}_{v,i}(t+1)$ and $h^{(1)}_{v,j}(t+1)$ respectively:

$$g^{(1)}_{v,i}(t+1) = \frac{\sum_{u \in \Gamma_{in}(v,i)} \phi^i_{(u,v)} \varphi(u,t)}{\sum_{u \in \Gamma_{in}(v,i)} \phi^i_{(u,v)}},\ h^{(1)}_{v,j}(t+1) = \frac{\sum_{u \in \Gamma_{in}(v,j)} \phi^j_{(u,v)} \varphi(u,t)}{\sum_{u \in \Gamma_{in}(v,j)} \phi^j_{(u,v)}}$$

where $\Gamma_{in}(v,i), \Gamma_{in}(v,j)$ are the neighbor sets of user v in relations i and j respectively and $\varphi(u,t)$ denotes if user u is activated at timestamp t. Note that anchor user $v^{(1)}$ is activated does not mean that $v^{(2)}$ is activated at the same time, but $v^{(2)}$ will get influence from $v^{(1)}$ via anchor link.

Algorithm 1 M&M Greedy Algorithm for AHI problem

Input: $G^{(1)}$, $G^{(2)}$, anchor user matrix $A_{n_{(1)} \times n_{(2)}}$, d

Output: seed set Z

```
initialize Z =, seed index i = 0;
get network schema S_G^(1) and S_G^(2), get user set U = U^(1) ∪ U^(2);
for v = 0 to |U| do
    extract intra and inter network diffusion meta paths of v;
end for
calculate relations' diffusion strength φ_(u,v) and ψ_(u,v);
define activation probability vector P^(1), P^(2) and calculate their initial value;
while i < d do
    for u ∈ U \ Z do
        using Monte Carlo method to estimate u's marginal gain M_u = σ(Z ∪ {u}) − σ(Z) based
        on users' activation probability;
    end for
    select z = arg max_{u ∈ U \ Z} M_u
    Z = Z ∪ {z}
    update users' activation probability in P^(1), P^(2) and i = i + 1.
end while
```

By aggregating all kinds of intra-network and inter-network relations, we can obtain the integrated activation probability of $v^{(1)}$ [9]. Here logistic function is used as the aggregation function.

$$p_v^{(1)}(t+1) = \frac{e^{\sum_{(i)} \rho_i^{(1)} g_{v,i}^{(1)}(t+1) + \sum_{(j)} \omega_j^{(1)} h_{v,j}^{(1)}(t+1)}}{1 + e^{\sum_{(i)} \rho_i^{(1)} g_{v,i}^{(1)}(t+1) + \sum_{(j)} \omega_j^{(1)} h_{v,j}^{(1)}(t+1)}},$$

where $\rho_i^{(1)}$ and $\omega_j^{(1)}$ denote the weights of each relation in diffusion process, whose value satisfy $\sum_{(i)} \rho_i^{(1)} + \sum_{(j)} \omega_j^{(1)} = 1,\ \rho_i^{(1)} \geq 0,\ \omega_j^{(1)} \geq 0$. Similarly, we can get activation probability of a user $v^{(2)}$ in $G^{(2)}$.

4 Influence Maximization Problem in M&M model

In this section, we will first analyze the influence maximization problem based on M&M model, and then provide M&M Greedy algorithm for seed users selection.

4.1 Analysis of Influence Maximization Problem

Kempe et al. [12] proved traditional influence maximization problem is a NP-hard for LT model, but the objective function of influence $\sigma(\mathcal{Z})$ is *monotone* and *submodular*. Based on these properties, the greedy approximation algorithms can achieve an approximation ratio of $1 - 1/e$.

With the above background knowledge, we will show that the influence maximization problem under the M&M model is also NP-hard and prove the influence spread function $\sigma(\mathcal{Z})$ is monotone and submodular.

Theorem 1 *Influence Maximization Problem across Partially Aligned Heterogenous Social Networks(AHI) is NP-hard.*

Table 1: Properties of the Heterogeneous Social Networks

	property	network Twitter	Foursquare
# node	user	500	500
	post	741,529	7,504
	location	34,413	6,300
# link	friend/follow	5,341	2,934
	write	741,529	7,504
	locate	40,203	7,504

Proof: The AHI problem can be easily mapped to "Vertex Cover" problem which is NP-complete. Thus AHI problem is NP-hard.

Theorem 2 *For the M&M model, the influence function $\sigma(\mathcal{Z})$ is monotone.*
Proof: Given the existing seed user sets $\mathcal{Z}$, let z be a seed user selected in this round. Since the weights of multi-relation are nonnegative, adding a new seed user z will not decrease the number of influenced users, i.e.,$\sigma(\mathcal{Z} + z) \geq \sigma(\mathcal{Z})$. Therefore the influence spread function is monotone for the given M&M model.

Theorem 3 *For the M&M model, the influence function $\sigma(\mathcal{Z})$ is submodular.*
Proof: It can be proved with the live-edge path method proposed in [12] very easily. The detail is omitted due to space limitation.

4.2 Greedy Algorithm for AHI problem

Since the influence function is monotone and submodular based on the M&M model, step-wise greedy algorithms which select the users who can lead to the maximum increase of influence can achieve a $(1-\frac{1}{e})$-approximation of the optimal result. Algorithm 1 is a greedy algorithm to solve the AHI problem based on M&M model.

5 Experiment

5.1 Experiment Preparation

In this part, we will introduce the dataset and baselines used in the experiments.

Dataset Description: The partially aligned heterogeneous network dataset used in the experiment are Foursquare and Twitter, The statistics of the two datasets are given in Table 1. For more detailed information about the dataset as well as its crawling methods, please refer to [13].

Baselines: We use following methods as baselines:

- *The* M&M *method* (M&M): M&M is the method proposed in this paper, which can select seed users greedily from the extracted MMNs. Depending on from which network to select the seed users, different variants of M&M

are compared: (1) M&M (which selects seed users from both Foursquare and Twitter), (2) M&M-Foursquare (selecting only from Foursquare), and (3) M&M-Twitter (selecting only from Twitter).

– *Lossless method for multiplex networks (LCI)*: Method LCI is the influence maximization method proposed for multiplex networks in [16], which selects seed users from the merged network as shown in Fig. 2-E.
– *Greedy method for single heterogenous network (Greedy)*: Based on a multi-relational network (as shown in Fig. 2-D), method Greedy selects seed users who can lead to the maximum influence gain within one single network. Similar to M&M, Greedy also has two variants: Greedy-Foursquare and Greedy-Twitter.
– *Seed Selection method based on traditional LT model(LT)*: Based on one single homogeneous network (e.g., Fig. 2-C), LT selects seed users who can lead to the maximum influence gain. Two variants of method LT, LT-Foursquare and LT-Twitter, are compared in the experiments.

5.2 Experiment Setup

Based on the input aligned heterogeneous networks, the MMNs are extracted based on a set of intra and inter network social meta paths. The influence score among users in each relation is used to calculate the aggregated activation probability with the logistic function. For simplicity, the weights of all relations (both intra and inter network) are set to be equal (i.e., 0.25 in this paper). The thresholds of users are randomly select from the uniform distribution within range [0,1]. The number of selected seed users is selected from $\{5, 10, \cdots, 50\}$. To simulate different partially aligned networks, we randomly sample the anchor links from the networks with different anchor ratios: $\{0.3, 0.6\}$, where 0.3 denotes that 30% anchor links are preserved while the remaining 70% are removed.

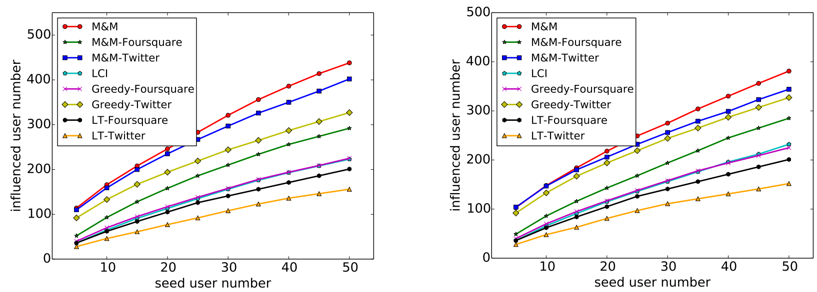
To evaluate the performance of all comparison methods, the number of finally activated users by the seed users is counted as the evaluation metric in the experiments, where anchor users are counted at most once. For example, for an anchor user u (whose accounts in Foursquare and Twitter are $u^{(1)}$ and $u^{(2)}$ respectively), if neither $u^{(1)}$ nor $u^{(2)}$ is activated, then u will not be counted as the activated user (i.e., 0); otherwise u will be counted as one activated user finally (i.e., 1).

5.3 Experiment Results

The experiment results are given in Fig.4, where the anchor ratios of (a) and (b) are 30% and 60% respectively.

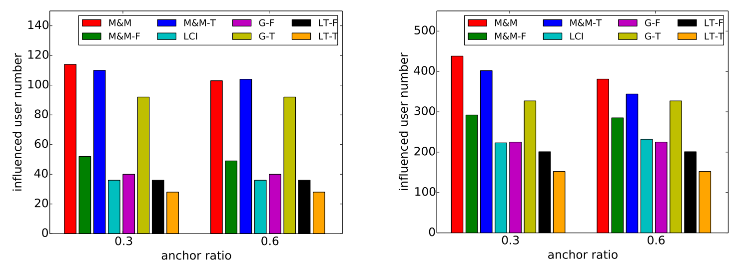
As shown in both figures, the number of influenced users will increase as more users are added as the seed users. M&M outperforms all the baselines consistently.

By comparing M&M with M&M-Foursquare and M&M-Twitter, we observe that M&M can perform better than both M&M-Foursquare and M&M-Twitter in both Fig. 4(a)- 4(b). It demonstrates that selecting seed users globally (i.e.,



(a) 30% anchor user (b) 60% anchor user

Fig. 4: Performance of different comparison methods.



(a) 5 seed users (b) 50 seed users

Fig. 5: Influence diffusion range with different anchor user ratio

both of the networks) can achieve better results than the method selecting seed users locally (i.e., either Foursquare or Twitter).

Compared with LCI, M&M can outperform LCI with significant advantages. For example, in Fig. 4(a) with seed user set size 20, seed users selected by M&M can activate 246 users, which is 117% larger than the 113 users activated by the seed users selected by LCI. Similar results can be observed for other seed user set sizes in both Fig. 4(a) and Fig. 4(b). As a result, M&M which selects seed users from MMNs can perform better than LCI which selects seed users from combined multiplex networks.

Furthermore, by comparing M&M with Greedy methods (both Greedy-Foursquare and Greedy-Twitter) and LT methods (both LT-Foursquare and LT-Twitter), M&M can always achieve better performance for different seed user set sizes and anchor ratios in Fig. 4(a)-4(b). In summary, selecting seed users based on cross-network information propagation model can select better seed user sets than those merely based on intra-network information propagation models.

5.4 Parameter Analysis

To study the effects of anchor ratio parameter, we compare the performance of all these comparison methods achieved at anchor ratio 0.3 and 0.6, whose results are shown in Fig. 5, where Fig. 5(a)-5(b) correspond to the seed user set sizes 5 and 50 respectively. We abbreviate M&M-Twitter and M&M-Foursquare as M&M-T and M&M-F, while Greedy is abbreviated as G.

By comparing the performance of all the comparison methods achieved with different anchor ratios in Fig. 5(a)-5(b), we observe that Greedy-Foursquare, Greedy-Twitter, LT-Foursquare and LT-Twitter can perform exactly the same with different ratios, as these comparison methods are all based on intra-network information propagation models.

However, the M&M methods can influence more users in aligned networks with lower anchor ratio, e.g., 0.3. With lower anchor ratio, less information can propagate across networks. However, with lower anchor ratio, more users will be non-anchor users. According to the evaluation metric introduced in Subsection 5.2, anchor users' accounts in multiple aligned networks will be counted at most once in the results, which is the reason why M&M can perform a little better for networks with anchor ratio 0.3 than those with anchor ratio 0.6.

6 Related Work

Influence maximization problem as a popular research topic recent years was first proposed by Domingos et al. [6]. It was first formulated as an optimization problem in [12].Since then a considerable number of work focused on speeding up the seed selection algorithms. CELF in [14] is faster 700 times than original Greedy method, and Chen designed heuristic algorithms for both IC model [4] and LT model [5]. Some other papers extended information diffusion models and provided efficient algorithms [3]. However almost all existing work studied influence maximization problem only for one single network. Nguyen et al. [16] studied the least cost influence problem across multiplex networks.

As to another related topic, information diffusion study, heterogenous and multi-relational networks became an increasingly hot topic [18, 19]. Tang et al. [15] proposed a generative graphical model to mine topic-level influence strength with both link and textual information. Gui et al. [9] proposed models by considering weighted combination of different types of relations. While all these work focused on one network.

7 Conclusion

In this paper, we study the novel problem of influence maximization across partially aligned heterogeneous social networks. To solve this problem, we propose multi-aligned multi-relation network based on intra and inter network meta paths to model information diffusion process. Greedy algorithm is proposed to select seed users in multiple heterogenous networks. Extensive experiments conducted on two real OSNs verify the effectiveness of the proposed algorithm. We believe that our work will not only advance the research on influence maximization problem, but also benefit many real-world applications.

8 Acknowledgement

This work is supported in part by NSFC(61375069, 61403156), NSF through grants CNS-1115234, and OISE-1129076, Google Research Award, and the Pinnacle Lab at Singapore Management University.

References

1. V. Belak, S. Lam, and C. Hayes. Towards maximising cross-community information diffusion. In *ASONAM*, 2012.
2. W. Chen, A. Collins, and R. Cummings et al. Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate - Microsoft Research. In *SDM*, 2011.
3. W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.
4. W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, 2009.
5. W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, 2010.
6. P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, 2001.
7. N. Du, L. Song, M. Gomez-Rodriguez, and H. Zha. Scalable influence estimation in continuous-time diffusion networks. In *NIPS*, 2013.
8. A. Goyal, W. Lu, and L. Lakshmanan. Celf++: optimizing the greedy algorithm for influence maximization in social networks. In *WWW*, 2011.
9. H. Gui, Y. Sun, J. Han, and G. Brova. Modeling topic diffusion in multi-relational bibliographic information networks. In *CIKM*, 2014.
10. X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, 2012.
11. Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, and K. Xie. Simulated annealing based influence maximization in social networks. In *AAAI*, 2011.
12. D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.
13. X. Kong, J. Zhang, and P. Yu. Inferring anchor links across multiple heterogeneous social networks. In *CIKM*, 2013.
14. J. Leskovec, A. Krause, C. Guestrin, and et al. Cost-effective outbreak detection in networks. In *KDD*, 2007.
15. L. Liu, J. Tang, J. Han, M. Jiang, and S. Yang. Mining topic-level influence in heterogeneous networks. In *CIKM*, 2010.
16. D. Nguyen, H. Zhang, S. Das, M. Thai, and T. Dinh. Least cost influence in multiplex social networks: Model representation and analysis. In *ICDM*, 2013.
17. M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD*, 2002.
18. Y. Sun, J. Han, X. Yan, P. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In *VLDB*, 2011.
19. J. Tang, T. Lou, and J. Kleinberg. Inferring social ties across heterogenous networks. In *WSDM*, 2012.
20. J. Zhang, P. Yu, and Z. Zhou. Meta-path based multi-network collective link prediction. In *KDD*, 2014.