Intertwined Viral Marketing in Social Networks

Jiawei Zhang[⋆], Senzhang Wang[†], Qianyi Zhan[‡], Philip S. Yu[⋆a]
[⋆] University of Illinois at Chicago, Chicago, IL, USA
[†] Nanjing University of Aeronautics and Astronautics, Nanjing, China
[‡] Nanjing University, Nanjing 210023, China
[a] Institute for Data Science, Tsinghua University, Beijing, China
jzhan9@uic.edu, szwang@nuaa.edu.cn, zhanqianyi@gmail.com, psyu@uic.edu

***Abstract*—Traditional viral marketing problems aim at selecting a subset of seed users for one single product to maximize its awareness in social networks. However, in real scenarios, multiple products can be promoted in social networks at the same time. At the product level, the relationships among these products can be quite intertwined, e.g., *competing*, *complementary* and *independent*. In this paper, we will study the "*interTwined Influence Maximization*" (i.e., TIM) problem for one product that we target on in online social networks, where multiple other competing/complementary/independent products are being promoted simultaneously. The TIM problem is very challenging to solve due to (1) few existing models can handle the *intertwined diffusion* procedure of multiple products concurrently, and (2) optimal seed user selection for the target product may depend on other products' marketing strategies a lot. To address the TIM problem, a unified greedy framework TIER (interTwined Influence EstimatoR) is proposed in this paper. Extensive experiments conducted on four different types of real-world social networks demonstrate that TIER can outperform all the comparison methods with significant advantages in solving the TIM problem.**

***Index Terms*—Intertwined Influence Maximization, Social Networks, Data Mining**

I. INTRODUCTION

Viral marketing (i.e., social influence maximization) first proposed in [14] has become a hot research problem in recent years and dozens of papers on this topic have been published so far [18], [19], [8], [7], [24], [16], [27], [11]. Traditional viral marketing problem aims at selecting the optimal set of seed users to maximize the awareness of ideas or products in social networks and has extensive concrete applications in the real world, e.g., product promotion [13], [20] and opinion spread [6]. In the traditional viral marketing setting [14], [18], only one product/idea is to be promoted. However, in the real scenarios, the promotions of multiple products can co-exist in the social networks at the same time. For example, in Figure 1, we show 4 different products to be promoted in an online social network and HP printer is our target product. At the product level, the relationships among these products can be quite intertwined:

- *independent*: promotion activities of some products (e.g., HP printer and Pepsi) can be *independent* of each other.
- *competing*: products having common functions will *compete* for the market share [1], [5] (e.g., HP printer and Canon printer). Users who have bought a HP printer are less likely to buy a Canon printer again.

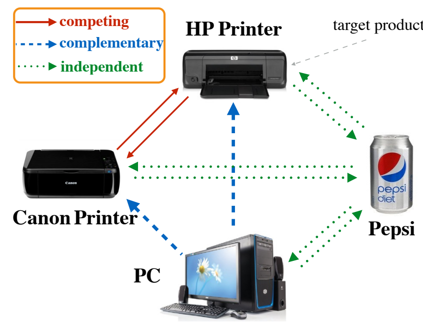


Fig. 1. Intertwined relationships among products.

- *complementary*: product cross-sell is also very common in marketing [20]. Users who have bought a certain product (e.g., PC) will be more likely to buy another product (e.g., HP printer) and the promotion of PC is said to be *complementary* to that of HP printer.

Problem: In this paper, we want to maximize the influence of one specific product that we target on in online social networks, where many other products are being promoted simultaneously. The relationships among these product can be obtained in advance via effective market research, which can be *independent*, *competitive* or *complementary*. Formally, we define this problem as the interTwined Influence Maximization (TIM) problem. A longer version of this paper is available at [29].

Before starting the promotions, companies need to design their *marketing strategies* carefully. *Marketing strategies* includes all basic and long-term activities in the field of marketing that can contribute to the goals of the company and its marketing objectives. However, in this paper, we are mainly concerned about the selected *seed users* who will spread the influence in social networks. Hence, for simplicity, we refer to the *marketing strategies* of products as the *seed users* selected for the products.

More specifically, depending on the promotional order of other products and the target product, the TIM problem can have two different variants (we don't care about the case that other products are promoted after the target product):

- C-TIM *problem*: In some cases, the other products have been promoted ahead of the target products, where their selected seed users are known and product information has already been propagated within the network. In such a case, the variant of TIM is defined as the Conditional interTwined Influence Maximization (C-TIM) problem.

- J-TIM *problem*: However, in some other cases, the promotion activities of multiple products occur simultaneously, where the *marketing strategies* of all these products are confidential to each other. Such a variant of TIM is defined as the Joint interTwined Influence Maximization (J-TIM) problem.

The TIM problem (both C-TIM and J-TIM) studied in this paper is a novel problem and totally different from existing works on viral marketing: *traditional single-product viral marketing problem* [18], *viral marketing for multiple independent products* [13], *viral marketing for competing products only* [1], [5], [3], and *viral marketing for cross-sell products only* [20]. More information of other related problems is available in Section V.

Despite its importance and novelty, the TIM problem is very challenging to solve due to the following reasons:

- *Lack of information diffusion model*: A new diffusion model which can handle the intertwined diffusion of these *independent*, *competing* and *complementary* products is the prerequisite for addressing the TIM problem.
- *Utilization of the known marketing strategies*: In the C-TIM problem, other products have been promoted in advance and their *marketing strategies* are public already. How to utilize these known *marketing strategies* to help identify the optimal seed user set for the target product is very challenging.
- *Unknown marketing strategies*: In the J-TIM problem, *marketing strategies* of other products are unknown. Inferring the potential *marketing strategies* of these products and developing the optimal *marketing strategies* for the target product based on the inference is still an open problem to this context so far.

To solve all the above challenges, we propose a unified *greedy* framework interTwined Influence EstimatoR (TIER) in this paper. The TIER method also has two variants: (1) C-TIER (Conditional TIER) for the C-TIM problem, and (2) J-TIER (Joint TIER) for the J-TIM problem. TIER is based on a novel information diffusion model interTwined Linear Threshold (TLT) introduced in this paper. TLT quantifies the *impacts* among products with the *intertwined threshold updating strategy* and can handle the intertwined diffusions of these products at the same time. To solve the C-TIM problem, C-TIER will select seed users greedily and is proved to achieve a $(1-\frac{1}{e})$-approximation to the optimal result. For the J-TIM problem, we show that the theoretical influence upper and lower bounds calculation is *NP-hard*. Alternatively, we formulate the J-TIM problem as a game among different products and propose to infer the potential *marketing strategies* of other products. The *step-wise greedy* method J-TIER can achieve promising results by selecting seed users wisely according to the inferred marketing strategies of other products.

The rest of this paper is organized as follows. In Section II, we give the concept and problem definitions. In Section III, the TLT diffusion model and TIER method are introduced in details, which will be evaluated in Section IV. Finally, we give the related works in Section V and conclude the paper in Section VI.

II. Problem Formulation

In this section, we will define some important concepts and give the formulation of the TIM problem.

A. Concept Definitions

Definition 1 (Social Network): An online *social network* can be represented as $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of users and $\mathcal{E}$ contains the interactions among users in $\mathcal{V}$. The set of n different products to be promoted in network G can be represented as $\mathcal{P} = \{p^1, p^2, \cdots, p^n\}$.

Definition 2 (User Status Vector): For a given product $p^j \in \mathcal{P}$, users who are influenced to buy p^j are defined to be "*active*" to p^j, while the remaining users who have not bought p^j are defined to be "*inactive*" to p^j. User u_i's status towards all the products in $\mathcal{P}$ can be represented as "*user status vector*" $\mathbf{s}_i = (s_i^1, s_i^2, \cdots, s_i^n)$, where s_i^j is u_i's status to product p^j. Users can be activated by multiple products at the same time (even competing products), i.e., multiple entries in *status vector* $\mathbf{s}_i$ can be "*active*" concurrently.

Definition 3 (Independent, Competing and Complementary Products): Let $P(s_i^j = 1)$ (or $P(s_i^j)$ for simplicity) denote the probability that u_i is activated by product p^j and $P(s_i^j|s_i^k)$ be the conditional probability given that u_i has been activated by p^k already. For products $p^j, p^k \in \mathcal{P}$, the promotion of p^k is defined to be (1) *independent* to that of p^j if $\forall u_i \in \mathcal{V}$, $P(s_i^j|s_i^k) = P(s_i^j)$, (2) *competing* to that of p^j if $\forall u_i \in \mathcal{V}$, $P(s_i^j|s_i^k) < P(s_i^j)$, and (3) *complementary* to that of p^j if $\forall u_i \in \mathcal{V}$, $P(s_i^j|s_i^k) > P(s_i^j)$.

Definition 4 (Marketing Strategy): In this paper, we are mainly concerned about the *seed user* selection problem. For simplicity, we refer to the *marketing strategy* of product $p^j \in \mathcal{P}$ as the *seed user set* $\mathcal{S}^j$ selected for p^j. And the *marketing strategies* of all products in $\mathcal{P}$ can be represented as seed user set list $\mathcal{S} = (\mathcal{S}^1, \mathcal{S}^2, \cdots, \mathcal{S}^n)$.

B. Problem Definition

In traditional single-product viral marketing problems, the selected *seed users* will propagate the influence of the target product in the network and the number of users get activated can be obtained with the *influence function* $I : \mathcal{S} \to \mathbb{R}$, which maps the selected seed users to the number of influenced users.

Traditional one single product viral marketing problem aims at selecting the optimal seed users $\bar{\mathcal{S}}$ for the target product, who can achieve the maximum influence:

$$\bar{\mathcal{S}} = \arg_{\mathcal{S}} \max I(\mathcal{S}).$$

However, in the TIM problem, promotions of multiple products in $\mathcal{P}$ co-exist simultaneously. The influence function of the target product $p^j \in \mathcal{P}$ depends on not only the seed user set $\mathcal{S}^j$ selected for itself but also the seed users of other products in $\mathcal{P} \setminus \{p^j\}$. Based on such a intuition, we formally define the *conditional intertwined influence function*, *joint intertwined influence function* and give the formulation of C-TIER, J-TIER problems as follows.

Definition 5 (Conditional Intertwined Influence Function): Let $\mathcal{S}^{-j} = (\mathcal{S}^1, \cdots, \mathcal{S}^{j-1}, \mathcal{S}^{j+1}, \cdots, \mathcal{S}^n)$ be the known seed user sets selected for all products in $\mathcal{P} \setminus \{p^j\}$, the *influence function* of the target product p^j given the known *seed user sets* $\mathcal{S}^{-j}$ is defined as the *conditional intertwined influence function*: $I(\mathcal{S}^j | \mathcal{S}^{-j})$.

C-TIM Problem: C-TIM problem aims at selecting the optimal *marketing strategy* $\bar{\mathcal{S}}^j$ to maximize the *conditional intertwined influence function* of p^j in the network, i.e.,

$$\bar{\mathcal{S}}^j = \arg_{\mathcal{S}^j} \max I(\mathcal{S}^j | \mathcal{S}^{-j}).$$

Definition 6 (Joint Intertwined Influence Function): When the seed user sets of products $\mathcal{P} \setminus \{p^j\}$ are unknown, i.e., $\mathcal{S}^{-j}$ is not given, the *influence function* of product p^j together with other products in $\mathcal{P} \setminus \{p^j\}$ is defined as the *joint intertwined influence function*: $I(\mathcal{S}^j; \mathcal{S}^{-j})$.

J-TIM Problem: J-TIM problem aims at choosing the optimal *marketing strategy* $\bar{\mathcal{S}}^j$ to maximize the *joint intertwined influence function* of p^j in the network, i.e.,

$$\bar{\mathcal{S}}^j = \arg_{\mathcal{S}^j} \max I(\mathcal{S}^j; \mathcal{S}^{-j}),$$

where set $\mathcal{S}^{-j}$ can take any possible value.

III. PROPOSED METHOD

In this section, we will introduce the TIER framework in details. We will first propose a new diffusion model TLT to deal with the intertwined diffusion of multiple products. Based on TLT, we will analyze the C-TIM problem and show that the proposed greedy method C-TIER can achieve a $(1 - \frac{1}{e})$-approximation of the optimal results. Finally, we will study the J-TIM problem and propose a new approach J-TIER, which formulates the J-TIM problem as a game among multiple products.

A. Intertwined Information Diffusion

To depict the intertwined diffusions of multiple independent/competing/complementary products, we propose a new information diffusion model TLT in this paper. In the TLT model, given the promotions of multiple products $\mathcal{P}$, user $u_i \in \mathcal{V}$ can influence his neighbor $u_k \in \Gamma(u_i)$ in promoting product $p^j \in \mathcal{P}$ according to weight $w^j_{i,k} \geq 0$ ($w^j_{i,k} = 0$ if link (u_k, u_i) doesn't exist or user u_i is *inactive* to p^j), where $\Gamma(u_i)$ represents the set of users following u_i (i.e., users that u_i can influence). Each user, e.g., u_i, is associated with a *static threshold* θ^j uniformly chosen at random from interval $[0, 1]$ towards each product $p^j \in \mathcal{P}$, which represents the minimal required influence for u_i to become *active* to p^j. Initially, only users in the seed user set $\mathcal{S}^j$ of product p^j are active to p^j and their influence will propagate within the network in discrete steps. At step t, all *active* users to p^j in step $t-1$ remain *active* and *inactive* users, e.g., u_i, will be activated by their neighbors to buy product p^j if $\sum_{u_l \in \Gamma_{out}(u_i)} w^j_{l,i} \geq \theta^j_i$, where $\Gamma_{out}(u_i)$ represents the set of users that u_i follows (i.e., people who can influence u_i).

What's more, in TLT, users in online social networks can be activated by multiple products at the same time, which can be either *independent*, *competing* or *complementary*. As shown in Figure 1, we observe the probabilities for users' to buy the HP printer will be (1) unchanged given that they have bought Pepsi (i.e., the *independent* product of HP printer), (2) increased if they own PCs (i.e., the *complementary* product of HP printer), and (3) decreased if they already have the Canon printer (i.e., the *competing* product of HP printer). To model such a phenomenon in TLT, we introduce the following *intertwined threshold updating strategy*, where users' *thresholds* to different products will change *dynamically* as the influence of other products propagates in the network.

Definition 7 (Intertwined Threshold Updating Strategy): Assuming that user u_i has been activated by m products p^{τ_1}, p^{τ_2}, $\cdots$, $p^{\tau_m} \in \mathcal{P} \setminus \{p^j\}$ in a sequence, then u_i's *threshold* towards product p^j will be updated as follows:

$$(\theta^j_i)^{\tau_1} = \theta^j_i \frac{P(s^j_i)}{P(s^j_i | s^{\tau_1}_i)}, (\theta^j_i)^{\tau_2} = (\theta^j_i)^{\tau_1} \frac{P(s^j_i | s^{\tau_1}_i)}{P(s^j_i | s^{\tau_1}_i, s^{\tau_2}_i)}, \cdots$$
$$(\theta^j_i)^{\tau_m} = (\theta^j_i)^{\tau_{m-1}} \frac{P(s^j_i | s^{\tau_1}_i, \cdots, s^{\tau_{m-1}}_i)}{P(s^j_i | s^{\tau_1}_i, \cdots, s^{\tau_{m-1}}_i, s^{\tau_m}_i)},$$

where $(\theta^j_i)^{\tau_k}$ denotes u_i's threshold to p^j after he has been activated by p^{τ_1}, p^{τ_2}, $\cdots$, p^{τ_k}, $k \in \{1, 2, \cdots, m\}$.

In this paper, we do not focus on the order of products that activate users [6] and to simplify the calculation of the *threshold updating strategy*, we assume only the most recent activation has an effect on updating current thresholds, i.e.,

$$\frac{P(s^j_i | s^{\tau_1}_i, \cdots, s^{\tau_{m-1}}_i)}{P(s^j_i | s^{\tau_1}_i, \cdots, s^{\tau_{m-1}}_i, s^{\tau_m}_i)} \approx \frac{P(s^j_i)}{P(s^j_i | s^{\tau_m}_i)} = \phi^{\tau_m \to j}_i.$$

Definition 8 (Threshold Updating Coefficient): Term $\phi^{l \to j}_i = \frac{P(s^j_i)}{P(s^j_i | s^l_i)}$ is formally defined as the "*threshold updating coefficient*" of product p^l to product p^j for user u_i, where

$$\phi^{l \to j}_i \begin{cases} < 1, & \text{if } p^l \text{ is } \textit{complementary} \text{ to } p^j, \\ = 1, & \text{if } p^l \text{ is } \textit{independent} \text{ to } p^j, \\ > 1, & \text{if } p^l \text{ is } \textit{competing} \text{ to } p^j. \end{cases}$$

The *intertwined threshold updating strategy* can be rewritten based on the *threshold updating coefficients* as follows:

$$(\theta^j_i)^{\tau_m} \approx \theta^j_i \cdot \phi^{\tau_1 \to j}_i \cdot \phi^{\tau_2 \to j}_i \cdots \phi^{\tau_m \to j}_i.$$

B. Conditional Intertwined Influence Maximization

In the C-TIM problem, the promotion activities of other products have been done before we start to promote our target product. Subject to the TLT diffusion model, users' thresholds to the target product can be updated with the *threshold updating strategy* after the promotions of other products. Based on the updated network, the C-TIM can be mapped to the *tradition single-product viral marketing*, which has been proved to be *NP-hard* already.

Theorem 1: The C-TIM problem is *NP-hard* based on the TLT diffusion model.

The proof of Theorem 1 is omitted due to limited space.

Algorithm 1 The C-TIER Algorithm

Input: input social network $G = (\mathcal{V}, \mathcal{P}, \mathcal{E})$
target product: p^j
known seed user sets of $\mathcal{P} - \{p^j\}$: $\mathcal{S}^{-j}$
conditional influence function of p^j: $I(\mathcal{S}^j|\mathcal{S}^{-j})$
seed user set size of p^j: k^j
Output: selected seed user set $\mathcal{S}^j$ of size k^j
initialize seed user set $\mathcal{S}^j = \emptyset$
propagate influence of products $\mathcal{P} - \{p^j\}$ with $\mathcal{S}^{-j}$ and update users' thresholds with intertwined threshold updating strategy
while $\mathcal{V} \setminus \mathcal{S}^j \neq \emptyset \wedge |\mathcal{S}^j| \neq k^j$ **do**
  pick a user $u \in \mathcal{V} - \mathcal{S}^j$ according to equation $\arg\max_{u \in \mathcal{V}} I(\mathcal{S}^j \cup \{u\}|\mathcal{S}^{-j}) - I(\mathcal{S}^j|\mathcal{S}^{-j})$
  $\mathcal{S}^j = \mathcal{S}^j \cup \{u\}$
end while
return $\mathcal{S}^j$.

Meanwhile, based on the TLT diffusion model, the *conditional influence function* of the target product $I(\mathcal{S}^j|\mathcal{S}^{-j})$ are observed to be both *monotone* and *submodular*.

Theorem 2: For the TLT diffusion model, the *conditional influence function* is *monotone* and *submodular*.

Proof:

(1) monotone: Given the existing seed user sets $\mathcal{S}^{-j}$ for existing products $\mathcal{P} - \{p^j\}$ in the market, let $\mathcal{T}$ be a seed user set of product p^j. Users in the network who are not involved in $\mathcal{T}$ can be represented as $\mathcal{V} - \mathcal{T}$. For the given seed user set $\mathcal{T}$ and the fixed seed users set $\mathcal{S}^{-j}$ of other products, adding a new seed user, e.g., $u \in \mathcal{V} - \mathcal{T}$, to the seed user set $\mathcal{T}$ will not decrease the number of influenced users, i.e., $I(\mathcal{T} \cup \{u\}|\mathcal{S}^{-j}) \geq I(\mathcal{T}|\mathcal{S}^{-j})$.

(2) submodular: After the diffusion process of the existing products in $\mathcal{P} - \{p^j\}$, users the thresholds towards product p^j will be updated. Based on the updated network, for two given seed user sets $\mathcal{R}$ and $\mathcal{T}$, where $\mathcal{R} \subseteq \mathcal{T} \subseteq \mathcal{V}$, it is easy to show that $I(\mathcal{R} \cup \{v\}|\mathcal{S}^{-j}) - I(\mathcal{R}|\mathcal{S}^{-j}) \geq I(\mathcal{T} \cup \{v\}|\mathcal{S}^{-j}) - I(\mathcal{T}|\mathcal{S}^{-j})$ with the "*live-edge path*" [18].

According to the above analysis, a greedy algorithm C-TIER is proposed to solve the problem C-TIM in this paper, whose pseudo code is available in Algorithm 1. In C-TIER, we select the user u who can lead to the maximum increase of the conditional influence function $I(\mathcal{S}^j \cup \{u\}|\mathcal{S}^{-j})$ at each step as the new seed user. This process repeats until either no potential seed user is available or all the k^j required seed users have been selected. The time complexity of C-TIER is $O(k^j|\mathcal{V}|(|\mathcal{V}|+|\mathcal{E}|))$. Since the *conditional influence function* is *monotone* and *submodular* based on the TLT diffusion model, then the *step-wise greedy* algorithms C-TIER, 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 for the target product.

C. Joint Intertwined Influence Maximization

C-TIM studies a common case in real-world viral marketing, where different companies have different schedules to release the promote their products and some can be conducted ahead of the target product. Meanwhile, in this section, we will study a more challenging case: J-TIM, where other products are being promoted at the same time as our target product and the marketing strategies of different products are totally confidential.

Algorithm 2 The J-TIER Algorithm

Input: input social network $G = (\mathcal{V}, \mathcal{P}, \mathcal{E})$
target product: p^j
set of other products: $\mathcal{P} - \{p^j\}$
joint influence function of p^j: $I(\mathcal{S}^j; \mathcal{S}^{-j})$
seed user set size of products in $\mathcal{P}$:$k^1, k^2, \cdots, k^j, \cdots, k^n$
Output: selected seed user sets $\{\mathcal{S}^1, \mathcal{S}^2, \cdots, \mathcal{S}^n\}$ of products in $\mathcal{P}$ respectively
initialize seed user set $\mathcal{S}^1, \mathcal{S}^2, \cdots, \mathcal{S}^n = \emptyset$
while $(\mathcal{V} \setminus \mathcal{S}^1 \neq \emptyset \vee \cdots \vee \mathcal{V} \setminus \mathcal{S}^n \neq \emptyset) \wedge (|\mathcal{S}^1| \neq k^1 \vee \cdots \vee |\mathcal{S}^n| \neq k^n)$ **do**
  for random $i \in \{1, 2, \cdots, n\}$ (p^i has not selected seeds in the round yet) **do**
    if $\mathcal{V} \setminus \mathcal{S}^i \neq \emptyset \wedge |\mathcal{S}^i| \neq k^i$ **then**
      p^i infers the seed user sets $\bar{\mathcal{S}}^{-i}$ of other products
      p^i selects its seed user $u^i \in \mathcal{V} - \mathcal{S}^i$, who can maximize $I(\mathcal{S}^i \cup \{u^i\}; \bar{\mathcal{S}}^{-i}) - I(\mathcal{S}^i; \bar{\mathcal{S}}^{-i})$
      $\mathcal{S}^i = \mathcal{S}^i \cup \{u^i\}$
      propagate influence of u in G and update influenced users' thresholds to products in $\mathcal{P}$ with the *intertwined threshold updating strategy*.
    end if
  end for
end while
return $\mathcal{S}^1, \mathcal{S}^2, \cdots, \mathcal{S}^n$.

1) The J-TIM Problem: When the *marketing strategies* of other products are unknown, the *influence function* of the target product and other products co-exist in the network is defined as the *joint influence function*: $I(\mathcal{S}^j; \mathcal{S}^{-j})$. Meanwhile, by setting $\mathcal{S}^1 = \cdots = \mathcal{S}^{j-1} = \mathcal{S}^{j+1} = \cdots = \mathcal{S}^n = \emptyset$, the J-TIM problem can be mapped to the traditional *single-product influence maximization* problem in polynomial time, which is an NP-hard problem.

Theorem 3: The J-TIM problem is *NP-hard* based on the TLT diffusion model.

Meanwhile, if all the products in $\mathcal{P} \setminus \{p^j\}$ are *independent* to p^j, the *joint influence function* $I(\mathcal{S}^j; \mathcal{S}^{-j})$ will be both *monotone* and *submodular*.

Theorem 4: Based on the TLT diffusion model, the *joint influence function* is *monotone* and *submodular* if all the other products are *independent* to p^j.

However, when there exist products in $\mathcal{P} \setminus \{p^j\}$ to be either *competing* or *complementary* to p^j, the *joint influence function* $I(\mathcal{S}^j; \mathcal{S}^{-j})$ will be neither *monotone* nor *submodular*.

Theorem 5: Based on the TLT diffusion model, the *joint influence function* is not *monotone* nor *submodular* if there exist products which are either *competing* or *complementary* to the target product p^j.

The proofs of Theorems 3-5 is omitted due to the limited space.

2) Challenges in J-TIM: When all the other products are *independent* to p^j, the *joint influence function* of p^j will be *monotone* and *submodular*, which is solvable with the *traditional greedy algorithm* proposed [18] and can achieve $(1 - \frac{1}{e})$-approximation of the optimal results. However, when there exist at least one product which is either *competing* or *complementary* to p^j, the *joint influence function* will be no longer *monotone* or *submodular*. In such a case, the J-TIM will be very hard to solve and no promising optimality bounds of the results are available.

By borrowing ideas from the game theory studies [21], [2], for product p^j, the lower-bound and upper-bound of influence the J-TIM problem can be achieved by selecting seed users of size k can be represented as

$$\max_{\mathcal{S}^j} \min_{\mathcal{S}^{-j}} I(\mathcal{S}^j; \mathcal{S}^{-j}), \quad \max_{\mathcal{S}^j} \max_{\mathcal{S}^{-j}} I(\mathcal{S}^j; \mathcal{S}^{-j})$$

respectively, which denotes the maximum influence p^j can achieve in the worst (and the best) cases where all the remaining products work together to make p^j's influence as low (and high) as possible. The *seed user set* selected by p^j when achieving the lower-bound and upper-bound of influence can be represented as

$$\hat{\mathcal{S}}^j_{low} = \arg\max_{\mathcal{S}^j} \min_{\mathcal{S}^{-j}} I(\mathcal{S}^j; \mathcal{S}^{-j}), \;\; \hat{\mathcal{S}}^j_{up} = \arg\max_{\mathcal{S}^j} \max_{\mathcal{S}^{-j}} I(\mathcal{S}^j; \mathcal{S}^{-j}).$$

However, the lower and upper bounds of the optimal results of the J-TIM problem is hard to calculate mathematically.

Theorem 6: Computing the Max-Min for 3 or more player games is NP-hard.

Proof: As proposed in [2], the problem of finding any (approximate) Nash equilibrium for a three-player game is computationally intractable and it is NP-hard to approximate the min-max payoff value for each of the player [2], [12], [9], [10].

3) The J-TIER *Algorithm:* In addition, in the real world, the other products will not co-operate together in designing their marketing strategies to create the worst or the best situations for the target product p^j, i.e., choosing the *marketing strategies* $\mathcal{S}^{-j}$ such that the *joint influence function* $I(\mathcal{S}^j; \mathcal{S}^{-j})$ is minimized or maximized. To address the J-TIM problem, in this part, we propose the J-TIER algorithm to simulate the intertwined round-wise greedy seed user selection process of all the products.

In J-TIER, all products are assumed to be *selfish* and wants to maximize their own influence when selecting seed users based on the "*current*" situation created by all the products. J-TIER will infer the next potential *marketing strategies* of other products round by round and select the *optimal* seed users for each product based on the inference.

In algorithm J-TIER, we let all products in $\mathcal{P}$ choose their optimal *seed users* randomly at each round. For example, let $(\mathcal{S})^{\tau-1}$ be the seed users selected by products in $\mathcal{P}$ at round $\tau - 1$. At round τ, a random product p^i can select one seed user. To achieve the largest influence, product p^i will infer the next potential seed users to be selected by other products based on the assumption that they are all selfish. For example, based p^i's inference, the next seed user to be selected by p^j can be represented as $\bar{u}^j$, i.e.,

$$\arg\max_{u \in \mathcal{V} - (\mathcal{S}^j)^{\tau-1}} [I\left((\mathcal{S}^j)^{\tau-1} \cup \{u\}; (\mathcal{S}^{-j})^{\tau-1}\right) - I\left((\mathcal{S}^j)^{\tau-1}; (\mathcal{S}^{-j})^{\tau-1}\right)].$$

Similarly, p^i can further infer the potential seed users to be selected next by products in $\mathcal{P} \setminus \{p^i, p^j\}$, who can be represented as $\{\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_{i-1}, \bar{u}_{i+1}, \cdots, \bar{u}_{j-1}, \bar{u}_{j+1}, \cdots, \bar{u}_n\}$ respectively. Based on such inference, p^i knows who are the next seed users to be selected by other products and will make use of the "prior knowledge" to select its own seed user $\hat{u}^i$ in round τ:

$$\hat{u}^i = \arg\max_{u \in \mathcal{V} - (\mathcal{S}^i)^{\tau-1}} [I\left((\mathcal{S}^i)^{\tau-1} \cup \{u\}; \bar{\mathcal{S}}^{-i}\right) - I\left((\mathcal{S}^i)^{\tau-1}; \bar{\mathcal{S}}^{-i}\right)].$$

where $\bar{\mathcal{S}}^{-i}$ is the "inferred" seed user sets of other products inferred by p^i based on current situation by "adding" these inferred potential seed users to their seed user sets.

TABLE I
PROPERTIES OF THE DIFFERENT NETWORKS

network	# nodes	# links	link type
Facebook	4,039	88,234	undirected
Wikipedia	7,115	103,689	directed
arXiv	5,242	14,496	undirected
Epinions	7,725	82,861	directed

The selected $(\hat{u}^i)^\tau$ will be added to the seed user set of product p^i, i.e.,

$$(\mathcal{S}^i)^\tau = (\mathcal{S}^i)^{\tau-1} \cup \{(\hat{u}^i)^\tau\}.$$

And the "*current*" seed user sets of all the products, i.e., $\mathcal{S}$, is updated as follows:

$$\mathcal{S} = ((\mathcal{S}^1)^\tau, (\mathcal{S}^2)^{\tau-1}, \cdots, (\mathcal{S}^n)^{\tau-1}).$$

The selected $(\hat{u}^i)^\tau$ will propagate his influence in the network and all the users just activated to product p^i will update their thresholds to other products in $\mathcal{P} \setminus \{p^i\}$.

Next, we let another random product (which has not selected seed users yet) to infer the next seed users to be selected by other products and choose its seed user based on the inferred situation. In each round, each product will have a chance to select one seed user and the user selection order of different products in each round is totally random. Such a process will stop when all the products either have selected the required number of *seed users* or no users are available to be chosen. With the J-TIER model, we simulate an alternative seed user selection procedure of multiple products in viral marketing and the pseudo-code J-TIER method is given in Algorithm 2. The time complexity of the J-TIER algorithm is $O((\sum_i k_i \cdot n)|\mathcal{V}|(|\mathcal{V}| + |\mathcal{E}|)$, where $k_i = |\mathcal{S}^i|$ is the number of seed users to be selected for product p^i.

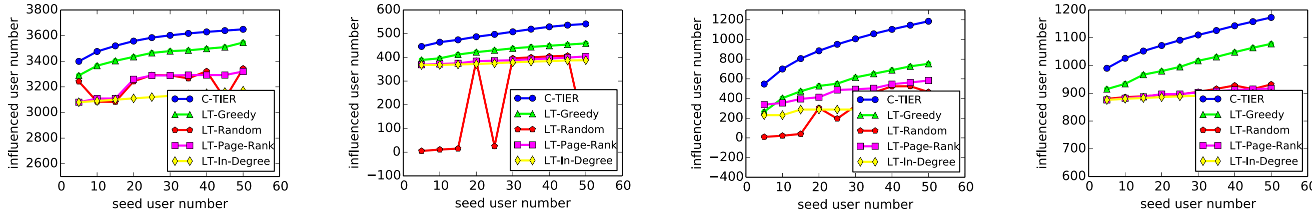
IV. EXPERIMENTS

Considering that real-world social networks with multiple competing, complementary and independent products being promoted simultaneously is extremely difficult to obtain. To test the effectiveness of TIER in addressing the TIM problem, we will conduct extensive experiments on 4 real-world social network datasets, where 4 generated products with intertwined relationships will be promoted simultaneously. This section contains 5 parts: (1) dataset descriptions, (2) experiment setting of the C-TIM problem, (3) experiment results of the C-TIM problem, (4) experiment setting of the J-TIM problem, and (5) experiment results of the J-TIM problem.

A. Dataset Description

The datasets used in the experiment include (1) Facebook social network[1], (2) Wikipedia administrator vote network[2], (3) arXiv collaboration network[3], and (4) Epinions e-commerce trust network[4]. These 4 different network datasets are all public and of different categories, which include the widely used social networks Facebook (where various social

[1]http://snap.stanford.edu/data/egonets-Facebook.html
[2]http://snap.stanford.edu/data/wiki-Vote.html
[3]http://snap.stanford.edu/data/ca-GrQc.html
[4]http://www.public.asu.edu/ jtang20/datasetcode/truststudy.htm



(a) Facebook (b) Wikipedia (c) arXiv (d) Epinions

Fig. 2. Experiment results of the C-TIM problem.

influence can diffuse among users), vote network (where voters' opinions about candidates could diffuse), academic co-author network (where academic ideas can propagate among researchers), and e-commerce network (where customers' reviews of products can influence other customers). Some statistical information about these 4 datasets is given in Tables I. More detailed information about these datasets is available at their corresponding webpages.

B. Experiment Setting of the C-TIM *Problem*

In this subsection, we will introduce comparison methods and experiment setups of the C-TIM problem.

1) Comparison Methods: In the C-TIM problem, the marketing strategies of all the other products are known in advance. "*Utilizing these known marketing strategies to select seed users for the target product can help achieve larger social influence in the social network.*" To demonstrate such a claim, different methods are compared in the experiments, which can be divided into two categories:

Methods using the known strategies

- C-TIER: C-TIER based on the TLT diffusion model is the method proposed in this paper. Other products' known marketing strategies are used to update users thresholds towards the target product dynamically with the *intertwined threshold updating strategy*. In each step, C-TIER selects the user who can lead to the maximum influence as the seed user.

Methods without using the known strategies

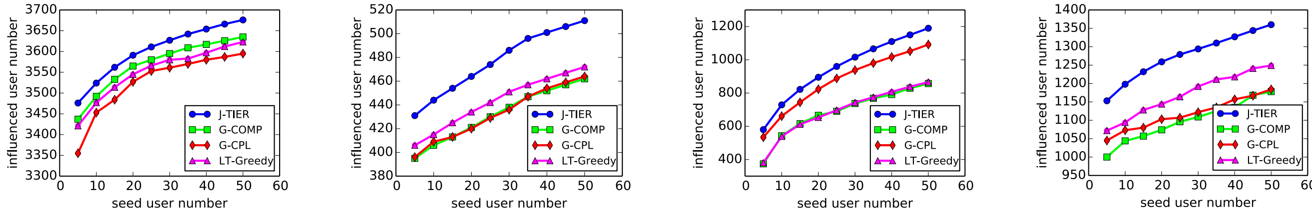
- LT-GREEDY: LT-GREEDY is the *greedy* seed user selection method based on the traditional LT diffusion model. LT-GREEDY ignores the existence of other products in seed user selection [18].
- LT-PAGE RANK: LT-PAGE RANK is based on the traditional LT diffusion model and doesn't use the know marketing strategies of other products. LT-PAGE RANK is a heuristics-based method and chooses users with the top K page rank scores as the final seed users [4].
- LT-IN DEGREE: LT-IN DEGREE is quite similar to LT-PAGE RANK: (1) it is a heuristics-based method, (2) it is based on traditional LT diffusion model, and (3) it doesn't use the known marketing strategies of other products. LT-IN DEGREE chooses users with the top K in degrees (i.e., # followers) as the seed users [8].
- LT-RANDOM: LT-RANDOM chooses K seed users from the network randomly from the network.

2) Experiment Setup: The connections among users in some networks are undirected, e.g., Facebook and arXiv, but in some others are directed, e.g., Wikipedia and Epinions. To unify different kinds of networks in our model, we replace undirected links, e.g., $u_i - u_j$, with two directed links $u_i \rightarrow u_j$, $u_i \leftarrow u_j$, and links among users in our model are all directed. In the TLT diffusion model, each user can influence his neighbors with certain influence weights and has a threshold denoting the minimal required influence to be activated by other users. The weight of directed social link $(u_j \rightarrow u_i)$ (u_j follows u_i or u_i influences u_j) quantifies the influence propagated from u_i to u_j. In the experiments, the influence weight of link (u_i, u_j) is quantified as $JC(u_i \rightarrow u_j) = \frac{|\Gamma(u_i) \cap \Gamma_{out}(u_j)|}{|\Gamma(u_i) \cup \Gamma_{out}(u_j)|}$, which is widely used in existing works [25] and depends on not only the shared users between u_i and u_j but also the degrees of u_i and u_j respectively. Considering that there exist multiple products to be promoted in the network, for simplicity, the influence weights of link $(u_i \rightarrow u_j)$ in promoting different products are all set as $JC(u_i \rightarrow u_j)$. Meanwhile, users will have multiple thresholds towards all these products, which can be represented as $\theta_i = (\theta_i^1, \cdots, \theta_i^n)$ and θ_i^j is the threshold of user u_i towards product p^j. The thresholds are randomly selected from uniform distribution within range $[0, 1]$. In the experiment, we consider 4 different products shown in Figure 1, where "HP printer" is the target product and "Canon printer", "PC" and "Pepsi Diet" are *competing*, *complementary* and *independent* respectively to "HP printer". The *threshold updating coefficient* between (1) *independent* products is set as 1.0; (2) *competing* products is randomly selected from $[1, 2]$, and (3) *complementary* products is randomly chosen from range $[0, 1]$.

The number of selected seed user for "HP printer" changes in range $\{5, 10, 15, \cdots, 45, 50\}$. For methods without utilizing the known strategies, we can just select seed users for "HP printer" based on the traditional LT model with methods LT-GREEDY, LT-PAGE RANK, LT-IN DEGREE and LT-RANDOM without considering the other products, which is exactly how these methods work in traditional single-product problem settings.

Meanwhile, C-TIER will update the network with the *intertwined threshold updating strategy* to use the known strategies of other products. The known seed users of products "Canon printer", "PC" and "Pepsi Diet" are selected with the LT-GREEDY algorithm from the network, whose sizes are all 50. The selected seed users of these products will propagate their influence in the network. Thresholds of users who get activated the products will be updated according to the *threshold updating strategy*. Based on the updated network, we apply C-TIER to select seed users for "HP printer".

To evaluate the performance of all these methods, we will



(a) Facebook (b) Wikipedia (c) arXiv (d) Epinions

Fig. 3. Experiment results of the J-TIM problem.

calculate the number of users influenced by the seed users based on the updated network.

C. *Experiment Results of the* C-TIM *Problem*

The experiment results of different comparison methods are given in Figure 2, where Subfigures 2(a)- 2(d) correspond to Facebook, Wikipedia, arXiv and Epinions datasets respectively.

Based on the results in Subfigures 2(a)- 2(d), the number of users who get influenced generally increases as more seed users are selected for most methods except LT-RANDOM. LT-RANDOM selects seed users randomly and the number of influenced users achieved by which can vary dramatically.

By comparing C-TIER with LT-GREEDY, we observe that C-TIER can perform better than LT-GREEDY consistently for different seed user set sizes in all these 4 datasets. For example, in the arXiv dataset when seed user set size is 50, the number of users get influenced by TIER is $1,185$, which is over 50% larger than the 753 influenced users achieved by LT-GREEDY. Experiments on other datasets show the similar results with various sizes of the seed users. It demonstrates that (1) the TLT diffusion model with *threshold updating strategy* works better than the traditional LT model, and (2) utilizing the known *marketing strategies* of other products can help lead to greater influence.

D. *Experiment Setting of the* J-TIM *Problem*

In this subsection, we will introduce comparison methods and experiment setups of the J-TIM problem.

1) Comparison Methods: In J-TIM problem, the marketing strategies of other products are unknown and we consider the seed user selection process as a game among all the products. All products are assumed to be *selfish* and want to choose users who can maximize their influence in the network. "*Meanwhile, in the seed user selection process, incorporating all the other products into the game can lead to better results*." To demonstrate such a claim, depending on the opponents incorporated in the game, the comparison methods used to address the J-TIM problem can be divided into 2 categories:

Methods with Complete Game Opponents

- J-TIER: In seed user selection process, all the products (i.e., independent, competing and complementary products) are involved in the game. This is the J-TIER method proposed in this paper.

Methods with Partial Game Opponents

- G-COMP: Enlightened by the analysis in [26], we propose G-COMP (Game among COMPeting products) as a potential comparison method, which can select seed nodes by only considering the competing products as the game opponents but ignoring the other two types of products.
- G-CPL: Method G-CPL (Game among ComPLementary products) extends the B-IMCP model proposed in [20], which can select seed nodes by only considering complementary products as the game opponents.
- G-INDEP/LT-GREEDY: Method G-INDEP (Game among INDEPendent products) ignores the *competing* and *complementary* products and only considers the *independent products* as the potential game opponents. Considering that *independent products* will not change users' thresholds towards the target product, method G-INDEP is identical to the traditional *step-wise greedy method* LT-GREEDY, which ignores all the other products in the network [18].

2) Experiment Setup: The experiment setup of the J-TIM problem is similar to that of the C-TIM problem. For different comparison methods, specific types of products are involved in the game and the selected seed users at each step are recorded. In evaluation, we simulate the game among different products again, where seed users of other products are those selected by J-TIER but seed users of the target product are replaced with those selected by different comparison methods. In the simulation, each product choose its seed users by turns and the influence of the seed users will propagate within the network and update users' thresholds right after it is selected. We calculate the number of users get influenced by the seed users of the target product to evaluate the comparison methods' performance.

E. *Experiment Results of the* J-TIM *Problem*

The results of different comparison methods in addressing the J-TIM problem on different datasets are available in Figure 3, where Subfigures 3(a)- 3(d) correspond to the Facebook, Wikipedia, arXiv and Epinions networks respectively.

Based on Subfigures 3(a)- 3(d), the results achieved by J-TIER is much better than those obtained by other methods. It shows that for the target product, when selecting seed users, considering all the existing products as game opponents (including competing, complementary and independent products) can help make better choices. For example, in Epinions network when seed user set size is 50, the influenced user numbers achieved by J-TIER, G-COMP, G-CPL and LT-GREEDY are 1390, 1178, 1184 and 1249 respectively. The results achieved by considering all the products in the game is (1) 11.2% better than that achieved by only considering *competing products* in the game, (2) 17.3% better than that gained by considering *complementary products* only, and (3)

7.28% better than that obtained by considering independent products only.

V. Related Work

Viral marketing (i.e., influence maximization) problem in customer networks first proposed by Domingos et al. [14] has been a hot research topic. Richardson et al. [23] study the viral marketing based on knowledge-sharing sites and propose a new model which needs less the computational cost than the model proposed in [14]. Kempe et al. propose to study the influence maximization problem through a social network [18] and propose to different diffusion models: Independent Cascade (IC) model and Linear Threshold (LT) model, which have been widely used in later influence maximization papers. Zhan et al. propose to extend the traditional single-network viral marketing problem to multiple aligned networks in [28].

Meanwhile, the promotions of multiple products can exist in social networks simultaneously, which can be independent, competing or complementary. Datta et al. [13] study the viral marketing for multiple independent products at the same time and aim at selecting seed users for each products to maximize the overall influence. Pathak et al. [22] propose a generalized linear threshold model for multiple cascades. Bharathi et al. [1] propose to study the competitive influence maximization in social networks, where multiple competing products are to be promoted. He et al. [17] propose to study the influence blocking maximization problem in social networks with the competitive linear threshold model. Carnes et al. [5] study the influence maximization problem in a competitive social network from a follower's perspective and Chen et al. [6] study the influence maximization in social networks when negative opinions can emerge and propagate. Multiple threshold models for competitive influence in social networks are proposed in [3], whose submodularity and monotonicity are studied in details. A nash equilibrium based model is proposed by Dubey et al. [15] to compete for customer in online social networks. Meanwhile, Narayanam et al. [20] study the viral marketing for product cross-sell through social networks to maximize the revenue, where products can have promotion cost, benefits and promotion budgets.

VI. Conclusion

In this paper, we have studied the TIM problem in online social networks. A novel unified framework TIER has been proposed to address the TIM problem. TIER is based on a novel diffusion model TLT, which can update users' thresholds dynamically. For the C-TIM problem, *greedy* method C-TIER selects the optimal seed users at each step and can achieve a $(1-\frac{1}{e})$-approximation to the optimal results. For the J-TIM problem, J-TIER formulates the seed user selection process of multiple products as a game and selects the optimal seed users step by step based on the inferred *marketing strategies* of other products. Extensive experiments on 4 real-world social network datasets demonstrate the superior performance of C-TIER and J-TIER in addressing the C-TIM and J-TIM problems.

VII. Acknowledgement

This work is supported in part by NSF through grants III-1526499.

References

[1] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *WINE*, 2007.

[2] C. Borgs, J. Chayes, N. Immorlica, A. Kalai, V. Mirrokni, and C. Papadimitriou. The myth of the folk theorem. *Games and Economic Behavior*, 2010.

[3] A. Borodin, Y. Filmus, and J. Oren. Threshold models for competitive influence in social networks. In *WINE*, 2010.

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *WWW*, 1998.

[5] T. Carnes, R. Nagarajan, S. Wild, and A. Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *ICEC*, 2007.

[6] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan. Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate - Microsoft Research. In *SDM*, 2011.

[7] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.

[8] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, 2009.

[9] X. Chen, X. Deng, and S. Teng. Computing nash equilibria: Approximation and smoothed complexity. In *FOCS*, 2006.

[10] X. Chen, S. Teng, and P. Valiant. The approximation complexity of win-lose games. In *SODA*, 2007.

[11] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what?: Item-level social influence prediction for users and posts ranking. In *SIGIR*, 2011.

[12] C. Daskalakis, P. Goldberg, and C. Papadimitriou. The complexity of computing a nash equilibrium. In *STOC*, 2006.

[13] S. Datta, A. Majumder, and N. Shrivastava. Viral marketing for multiple products. In *ICDM*, 2010.

[14] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, 2001.

[15] P. Dubey, R. Garg, and B. De Meyer. Competing for customers in a social network: The quasi-linear case. In P. Spirakis, M. Mavronicolas, and S. Kontogiannis, editors, *Internet and Network Economics*. Springer Berlin Heidelberg, 2006.

[16] A. Goyal, F. Bonchi, and L. Lakshmanan. Discovering leaders from community actions. In *CIKM*, 2008.

[17] X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, 2012.

[18] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[19] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.

[20] R. Narayanam and A. Nanavati. Viral marketing for product cross-sell through social networks. In *ECML PKDD*, 2012.

[21] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[22] N. Pathak, A. Banerjee, and J. Srivastava. A generalized linear threshold model for multiple cascades. In *ICDM*, 2010.

[23] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD*, 2002.

[24] X. Song, Y. Chi, K. Hino, and B. Tseng. Identifying opinion leaders in the blogosphere. In *CIKM*, 2007.

[25] J. Sun and J. Tang. A survey of models and algorithms for social influence analysis. In C. Aggarwal, editor, *Social Network Data Analytics*. Springer US, 2011.

[26] V. Tzoumas, C. Amanatidis, and E. Markakis. A game-theoretic analysis of a competitive diffusion process over social networks. In *WINE*, 2012.

[27] L. Yu, P. Cui, F. Wang, C. Song, and S. Yang. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. *CoRR*, abs/1505.07193, 2015.

[28] Q. Zhan, J. Zhang, S. Wang, P. Yu, and J. Xie. *Influence Maximization Across Partially Aligned Heterogenous Social Networks*. 2015.

[29] J. Zhang, S. Wang, Q. Zhan, and P. Yu. Intertwined viral marketing through online social networks. *CoRR*, 2016.