Influence Maximization Based on Dynamic Personal Perception in Knowledge Graph

Ya-Wen Teng[1], Yishuo Shi[2], Chih-Hua Tai[3], De-Nian Yang[1,4], Wang-Chien Lee[5], Ming-Syan Chen[6]
[1]*Research Center for Information Technology Innovation, Academia Sinica, Taiwan*
[2]*Department of Statistics and Information Science, College of Mathematics and Physics, Wenzhou University, China*
[3]*Department of Computer Science and Information Engineering, National Taipei University, Taiwan*
[4]*Institute of Information Science, Academia Sinica, Taiwan*
[5]*Department of Computer Science and Engineering, The Pennsylvania State University, USA*
[6]*Department of Electrical Engineering, National Taiwan University, Taiwan*
ywteng@citi.sinica.edu.tw, yishuo@wzu.edu.cn, hanatai@mail.ntpu.edu.tw
dnyang@iis.sinica.edu.tw, wlee@cse.psu.edu, mschen@ntu.edu.tw

2021 IEEE 37th International Conference on Data Engineering (ICDE) | 978-1-7281-9184-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICDE51399.2021.00132

***Abstract*—Viral marketing on social networks, also known as *Influence Maximization (IM)*, aims to select k users for the promotion of a target item by maximizing the total spread of their influence. However, most previous works on IM do not explore the dynamic user perception of promoted items in the process. In this paper, by exploiting the knowledge graph (KG) to capture dynamic user perception, we formulate the problem of *Influence Maximization based on Dynamic Personal Perception (IMDPP)* that considers user preferences and social influence reflecting the impact of relevant item adoptions. We prove the hardness of IMDPP and design an approximation algorithm, named *Dynamic perception for seeding in target markets (Dysim)*, by exploring the concepts of dynamic reachability, target markets, and substantial influence to select and promote a sequence of relevant items. We evaluate the performance of Dysim in comparison with the state-of-the-art approaches using real social networks with real KGs. The experimental results show that Dysim effectively achieves at least 6 times of influence spread in large datasets over the state-of-the-art approaches.**

***Index Terms*—influence maximization, multiple promotions, item relationships, dynamic personal perceptions**

I. Introduction

Social influence [1], [2], [3] refers to the impact of a social environment on people's behavior. By exploiting the social influence of users, a wide spectrum of applications (e.g., item promotion and viral marketing) have been formulated as various research problems, such as *influence maximization* (IM) [1], revenue maximization (RM) [2], and profit maximization (PM) [3]. Among them, IM selects k users as the seeds to promote *one* target item to maximize the number of influenced users. Nevertheless, in real life, companies often promote relevant items in *multiple* events, e.g., Apple Inc. usually promotes iPhones, AirPods, and iPads in September, followed by a series of subsequent promotions.[1] In this work, we address a new IM problem formulated for a sequence of promotions on relevant items.[2]

For multiple promotions, exploring the dynamic changes in personal perceptions on promoted items is important, since users' perceptions of item relationships may vary according to the changes in users' demand indicated by research in the marketing field [4]. First, the *complementary* and *substitutable* relationships between items affect users' preferences on items [4], [5], [6]. In economics, *cross elasticity of demand* [7] indicates that adopting complementary items of an item increases the preference for it, while adopting its substitutable items has the opposite effect. For example, users who own iPhones with no headphone jack may be interested in AirPods (due to its complementary relationship with iPhones), while users who have iPhones may have less interest in iPads (due to their substitutable relationship). Second, the association between items may trigger extra adoptions without promotions [8], [9]. For example, AirPods may be directly adopted together with iPhones due to their complementary relationship.

Third, the perceptions of these relationships between items are usually personal and dynamic [4], [10], [11], as the items got newly adopted usually bring fresh experiences to users. For example, users who care more about large screens than mobility may treat iPhones as substitutable items of iPads; when these iPad users start to care about the mobility, they may tend to regard iPhones as complementary items of iPads. In turn, the changes in personal perceptions of item relationships lead to changes in users' preferences. Fourth, the dynamic personal perceptions of item relationships also affect users' social influence strength over friends, since friends adopting similar items and sharing similar perceptions tend to become closer [12], [13]. To address IM in a sequence of promotions on relevant items, it is essential to carefully examine dynamic personal perceptions of item relationships, together with their ripple effect on personal preferences for items, social influence strength, and item associations.

Knowledge graph (KG) (along with weighted meta-graphs) to capture the relationships (e.g., the complementary and substitutable relationships) has been well-explored in recommendation systems [10]. As illustrated in Fig. 1, KG represents facts (e.g., ITEM iPhone and ITEM AirPods SUPPORT the FEATURE Bluetooth in Fig. 1(a)), while meta-graphs capture relationships in the KG (e.g., m_1 in Fig. 1(b) describes two

[1]https://www.apple.com/apple-events/.

[2]After the influence propagation of the seed group for the first promotion finishes, the second follows, and so on.

978-1-7281-9184-3/21/$31.00 ©2021 IEEE
DOI 10.1109/ICDE51399.2021.00132

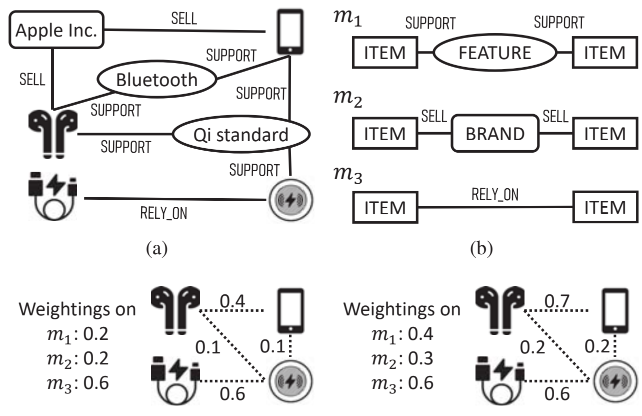


Fig. 1. (a) A tiny KG describing facts about the iPhone, AirPods, wireless charger, and charging cable. (b) Three meta-graphs specifying the complementary relationship. (c) Bob's initial personal item network, where a dotted edge denotes a complementary relationship. (d) Update of Bob's personal item network: after adopting iPhone and AirPods, Bob's weightings on m_1 and m_2 grow, which increases the relevance scores between iPhone, AirPods, and the wireless charger.

ITEMs SUPPORTing the same FEATURE are complementary). Note that these meta-graphs can be used to reflect the perception of item relationships, in forms of *personal item network*, for each individual. The personal weighting on each meta-graph describes the significance of this meta-graph to an individual (e.g., the values next to $m_1, \ldots, m_3$ in Fig. 1(c)), while the relevance scores between items describe the strength of their relationships in the mind of this individual [10], [11] (e.g., the values on dotted edges in Fig. 1(c)). By adjusting the weightings on meta-graphs according to previous adoptions [10], [11], dynamic personal perceptions of item relationships in individual users can be updated (in Fig. 1(d)). In this paper, we aim to leverage dynamic personal item networks for a sequence of IM promotions.

Following up the example in Fig. 1, Fig. 2 illustrates the IM process considering dynamic personal perceptions of item relationships, personal preferences for items, social influence strength, and item associations. As shown, the number of hearts indicates Bob's preference for a not-yet-adopted item, and a solid arrow represents the social influence between users (thickness implies strength). After Bob is promoted iPhone by Alice, Bob's purchase decision depends not only on the influence strength from Alice but also on his own preference for iPhone (in Fig. 2(a)). Meanwhile, item associations usually trigger extra adoptions of relevant items, such as AirPods, according to Bob's item network (in Fig. 1(c)). After Bob purchases iPhone and AirPods, his perception of the complementary relationship changes (i.e., he becomes to regard items supporting common features or belonging to the same brand as complementary), which increases the relevance between iPhone, AirPods, and the wireless charger (in Fig. 1(d)). After that, as Bob has adopted iPhone and AirPods, and their relevance to the wireless charger increases, Bob's preference

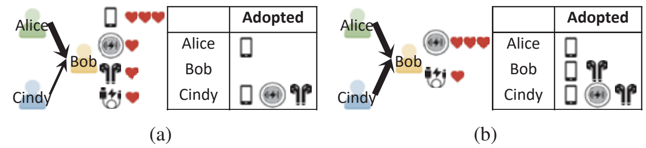


Fig. 2. Illustration of the IMDPP problem. (a) The states before Bob adopts iPhone and AirPods. (b) The states after Bob adopts iPhone and AirPods.

for the wireless charger grows accordingly.[3] Moreover, if Cindy acts as a seed to promote the wireless charger, as Bob and Cindy have similar adopted items (in Fig. 2(b)) (indicating Bob shares a similar perception of item relationships with Cindy and tends to behave similarly with Cindy), the influence strength from Cindy to Bob thus becomes stronger. It is easier for Cindy to promote the wireless charger to Bob now, since both Cindy's influence strength to Bob and Bob's preference for the wireless charger increase.

To incorporate factors depicted in the example above, several new challenges arise. (i) *Propagation of item impact (i.e., impact due to item adoption)*: Item adoptions change users' personal perceptions of item relationships, their preferences for other items, their strength of social influence among friends, and the item associations. In other words, the promotion of an item may affect the adoptions of subsequent items and thereby the planning for the next promotions. The order of items being promoted matters. (ii) *Antagonism of the substitutable relationship*: Promoting an item after adopting a substitutable item is not beneficial when the first item has met the users' needs. It is thus vital to avoid promoting substitutable items to the same users in consecutive promotions. (iii) *Determination of promotional timing*: As the promotions are dependent on previous ones, a seed in early promotions should facilitate subsequent promotions, while a seed in later promotions should focus on potential adoptions benefited from previous promotions. Therefore, determining the proper promotional timing for each seed is essential.

In this work, we formulate a new problem, named *Influence Maximization based on Dynamic Personal Perception* (referred to as IMDPP). In contrast to most previous works [14] focusing on one item, given the social network, KG, and meta-graphs for different item relationships, IMDPP targets on multiple promotions to maximize the overall spread of influence by choosing items and selecting seed users for promotion at proper timings under a total budget, where users have different costs as seeds [3], and each promotion allows multiple items to be promoted. We exploit personal item networks to capture dynamic personal perceptions of complementary and substitutable relationships between items. The adoptions of items dynamically adjust users' weightings on meta-graphs, reflecting dynamic personal perceptions and updating personal item networks. Also, users' preferences for other items, their social influence strength over friends, and the item associations change accordingly, in turn affecting other users' adoptions, their personal weightings on meta-graphs, their preferences

[3] A real example is in https://amzn.to/3fW7JLC.

for other items, their social influence strength, and the item associations in a ripple effect. For the ease of understanding, we first present the fundamental problem of IMDPP, referred to as *Simple IMDPP (SIMDPP)*, by focusing on the important factors of dynamic personal perceptions of item relationships and their fundamental ripple effect on dynamic preferences for items, i.e., neglecting dynamic social influence strength and item associations.

We prove that SIMDPP and IMDPP are NP-hard and inapproximable within $O(\frac{1}{|V|^{1-\epsilon}})$, where $|V|$ is the number of users and ϵ is an arbitrarily small constant. We design an approximation algorithm, named *Dynamic perception for seeding in target markets (Dysim)*, to tackle the above challenges of IMDPP. For the first challenge in the propagation of item impacts, Dysim introduces *dynamic reachability* to evaluate the impacts from previously promoted items on the currently chosen item, as well as the potential impact from the current item on any candidate item in subsequent promotions. For the second challenge in the antagonism of the substitutable relationship, Dysim identifies *target markets* to promote complementary items to socially close users in consecutive promotions. For the third challenge in determining the promotional timing, Dysim introduces *substantial influence* to evaluate both immediate and subsequent adoptions under the impact of a candidate seed (assigned at some promotional timing). For SIMDPP, where dynamic social influence strength and item associations are not considered, we develop an approximation algorithm, namely *Simple Dysim (SDysim)*, in this paper as well. We evaluate the performance of SDysim and Dysim on real social networks with KGs, i.e., Amazon, Yelp, Douban, and Gowalla. Due to the space constraint, the details of IMDPP and Dysim are presented in the full-length version [15] of this paper. The contributions of this work include:

- To the best of our knowledge, IMDPP is the first attempt to study the IM problem under a sequence of promotions on relevant items, where the personal perceptions of item relationships are dynamically captured from users' previously adopted items by KG and meta-graphs, and the changes in preferences for items, social influence strength, and item associations are considered as a ripple effect in the diffusion process.
- We prove that SIMDPP and IMDPP are inapproximable within $O(\frac{1}{|V|^{1-\epsilon}})$ even for a simple case with only the complementary relationship and only one promotion.
- We design an approximation algorithm Dysim, which plans a distinct effective promotional strategy for each target market to avoid antagonism between substitutable items. Dysim carefully examines the dynamic reachability of items to prioritize the promotion of relevant items, and evaluates the substantial influence of candidate seeds to properly determine the promotional timing.
- Via real social networks and real KGs, experimental results demonstrate that SDysim and Dysim effectively achieve at least 6 times of the influence spread over the state-of-the-arts.

II. Related work

Influence maximization (IM) aims at maximizing the number of influenced users by selecting seed users. It was first formulated as a discrete optimization problem and proved as NP-hard by Kempe et al. [1]. Since then, various issues in IM have been actively studied. To address the inefficiency in computing influence spread, some exploit the submodular property and certain heuristics [16]. Recent works further introduce the reverse influence sampling to approximate the influence with guarantees [17].[4] Recently, Huang et al. [14] point out that users' adopting probabilities of the promoted item should depend on users' previously adopted complementary and substitutable items (which is modeled as dynamic preferences for items in IMDPP). However, [14] targets only on a specified item in only one promotion with fixed item relationships, whereas IMDPP explores multiple promotions on relevant items and carefully examines the dynamic user perceptions of item relationships. Although various issues, e.g., target audience, scalability, and complementary/substitutable items, are studied, previous works [1], [3], [14], [16], [17] promote only one target item in only a single promotion, instead of multiple target items in multiple promotions, and the phenomenon of dynamic personal perceptions of item relationships together with its ripple effect are not considered. By contrast, IMDPP aims at a sequence of promotions on *relevant items* modeled by Knowledge Graph, where the item relationships and promotional timings are able to alter users adoption decisions.

Some studies investigate IM on promoting multiple target items, e.g., making exclusive adoption among items [18], avoiding spamming seeds by overwhelming promotions [19], learning diffusion probabilities of different items [20], and maximizing utility-based adoption among desired items [21]. However, they focus on a single promotion and do not consider multiple promotions to promote a sequence of relevant items modeled by KG and meta-graphs. Moreover, they study the problems under simpler scenarios without capturing the dynamic changes in personal perceptions of item relationships, personal preferences for items, social influence strength, and item associations. By contrast, in IMDPP, the adoption of items dynamically changes the personal perceptions of complementary and substitutable relationships between items. The changed perceptions of item relationships affect users' preferences for items and users' social influence strength over time, in turn affecting other users' adoptions, their preferences for other items, and their social influence strength as a ripple effect. Therefore, the above works have limitations to IMDPP, since the promotional timing is critical as users' perceptions of item relationships, preferences for items, and influence strength on friends are dynamic.

Research on adaptive IM [22], [23], [24] aims to select the seeds adaptively based on the adoptions in the previous influence diffusion. However, although multiple promotional timings are considered, they consider only one item in the IM

[4] More introduction on other IM problems is presented in [15].

problem and ignore multiple target items, item relationships, and dynamic preference for items. Moreover, adaptive IM requires a predefined budget allocation to different promotions, and it does not have the adaptive monotonicity and the adaptive submodularity (or even the adaptive bounded weak-submodularity).[5] By contrast, IMDPP does not require a predefined budget allocation to promotions and can be solved by Dysim with an approximation guarantee (detailed in [15]).

Knowledge graph (KG) is employed to describe facts in a wide spectrum of applications, e.g., relevance measures and search [11], [25], and recommendation [10]. Shi et al. [10] present a new similarity measure through personal weighted meta-paths to include different semantics of similarity. Users' own preferences can thus be derived from these meta-paths. Gu et al. [11] point out that a user may have different perceptions of similarity due to the change in her interests. They propose to automatically pick up meta-paths to best characterize the similarity by user-provided examples. Huang et al. [25] further extend meta-paths to meta-graphs to measure similarity with more complex connections. Note that these works focus on predicting the ratings of unknown items for users, which is essentially different from the IM problem. Inspired by the above research, we first attempt to incorporate the above relevance measurements and adopt the above meta-graphs with dynamic personal weightings in influence diffusion of multiple relevant items.

III. Fundamental Problem Formulation

To study various issues in multiple promotions of relevant items, we first introduce two important factors, which can be easily incorporated into existing diffusion models, e.g., triggering models [1], by extending the diffusion process, to consider dynamic changes in personal perceptions of item relationships [4], [10], [11] and their fundamental ripple effect on personal preferences for items [4], [5], [6], [7].[6] (1) *Relevance measurement*: KG is leveraged to measure the relevance between two items and find personal item networks, by learning the personal weightings on meta-graphs from users' previously adopted items.[7] (2) *Preference estimation*: Users' preferences for not-yet-adopted items are derived and updated based on previously adopted items and personal item networks. The dependency of these factors is illustrated in Fig. 3, while the discussion of deriving and updating them is detailed in [15]. For ease of discussion, we summarize the notations in Table I.

[5]The adaptive monotonicity is a property that conditional expected marginal adoptions of any item are non-negative. The adaptive submodularity (or the adaptive bounded weak-submodularity) is the property that conditional expected marginal adoptions of any fixed item do not increase (or boundedly increase) as more items are selected and their states are observed.

[6]The complete problem that considers the comprehensive ripple effect on social influence strength [12], [13], [26] and item associations [8], [9] is presented in the full version [15].

[7]Instead of KG, some lightweight alternatives, such as Tagging algorithm [27], Sceptre [28], PMSC [29], and DecGCN [30] can also be adopted to learn item relationships. Since the above works derive the item relationships according to all users' adoption history, the item network is no longer personalized. When they are adopted in our problem, all personal item networks will be identical.

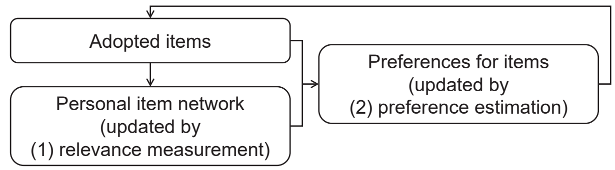


Fig. 3. Illustration of the two dependent factors.

Accordingly, we elaborate on the *diffusion process* as follows. A campaign includes T promotions. The t-th promotion contains multiple steps $\zeta_t = 0, 1, \ldots$, where each step represents an influence propagation from users adopting items to their friends that haven't adopted those items yet.[8] As a promotion depends on previous promotions, the initial state of a user in the t-th promotion (i.e., adopted item, perceptions, and preferences at $\zeta_t = 0$) is the same as the state at the end of the $(t-1)$-th promotion, while the seeded users in the t-th promotion newly adopt the promoted items at $\zeta_t = 0$. When the diffusion starts at step $\zeta_t \geq 1$, a user u may be promoted x by any friend u' who newly adopted x at $\zeta_t - 1$ only if u has not adopted x yet. The probability that u will adopt x is derived according to the social influence strength from u' (denoted as $p_{u',u}$) and u's preference for x (denoted as $P_{\text{pref}}(u, x, \zeta_t - 1)$), i.e., $p_{u',u} \times P_{\text{pref}}(u, x, \zeta_t - 1)$. Then, at the end of this step, u's personal perceptions of item relationships (i.e., personal item network) are updated by *(1) relevance measurement* (detailed in [15]) if u newly adopts any item, while her preferences for not-yet-adopted items also change accordingly by *(2) preference estimation* (detailed in [15]). If there is any new adoption at ζ_t, the next step $\zeta_t + 1$ starts with users having new adoptions at ζ_t to promote their newly adopted items to their friends (who have not yet adopted those items).[9] In other words, the diffusion of t-th promotion stops when no new adoptions happen since users cannot be promoted the adopted items again. Thus, the diffusion of the $(t+1)$-th promotion follows.

Based on the above diffusion process for relevant items in multiple promotions, we aim to choose a number of items, seed suitable users, and decide the proper timing, such that the influence spread (defined below) is maximized. Formally, $S = \{(u, x, t)\}$ is a seed group, where a seed (u, x, t) indicates that an item x is chosen for promotion starting at a seeded user u in the decided t-th promotion.[10] Let $S_t \subseteq S$ denote a subgroup of seeds chosen for the t-th promotion. We first define the influence spread and then formulate the problem as follows.

Definition 1 (Influence function). *Let T denote the number of promotions. For a seed group S, the influence spread in the social network $G_{SN} = (V, E)$, denoted as*

[8]Note that $\zeta_t - 1 \equiv \zeta_{t-1}^{\text{last}}$ if $\zeta_t = 0$, where $\zeta_{t-1}^{\text{last}}$ is the last step of the $(t-1)$-th promotion.

[9]Following other IM problems [1], [16], these users promote items to their friends without costs.

[10]An item x can be assigned to multiple seeded users at multiple promotions; each promotion can promote multiple chosen items by multiple seeded users.

TABLE I
SUMMARY OF NOTATIONS IN SECTIONS III-IV.

Notation	Description
ζ_t	Step ζ of the t-th promotion
$P_{\text{pref}}(u, y, \zeta_t)$	u's preference for y updated at ζ_t
$p_{u,v}$	u's influence strength on v
(u, x, t); (u, x)	Seed; nominee
$S = \{(u, x, t)\}$; S_t	Seed group; subgroup of seeds in the t-th promotion
$\sigma^{G_{\text{SN}}}(S)$; T	Influence function; number of promotions
$\{m^{\mathsf{C}}\}$ / $\{m^{\mathsf{S}}\}$	Sets of meta-graphs for describing complementary/substitutable relationships
τ; $\mathcal{G}$	Target market; set of target markets with common users
$\bar{r}^{\mathsf{C}}_{x,y}$ / $\bar{r}^{\mathsf{S}}_{x,y}$	Average complementary/substitutable relevance between x and y per user (the timing is specified from context)
$DR^{\mathcal{W},\tau}(S^{\mathcal{G}}, x)$	x's dynamic reachability of τ's users given $S^{\mathcal{G}}$ and $\mathcal{W}$
$\hat{t}$	Latest promotional timing in $S^{\mathcal{G}}$

$\sigma^{G_{SN}}(S)$, is the expected adoptions in all T promotions, i.e., $\sigma^{G_{SN}}(S) = \sum_{t=1}^{T} \sigma_t^{G_{SN}}(S_t \mid S_1, S_2, \ldots, S_{t-1}) = \sum_{t=1}^{T} \sum_{x \in I} n_x(S_t \mid S_1, S_2, \ldots, S_{t-1})$, where $n_x(S_t \mid S_1, S_2, \ldots, S_{t-1})$ is the expected new adoptions of x for S_t in the t-th promotion conditioned on $S_1, \ldots, S_{t-1}$ in previous promotions.[11] *(When G_{SN} is clear from context, we write $\sigma(S)$ for short.)*

Definition 2 (Simple IMDPP (SIMDPP))**.** *Let m^{C} and m^{S} denote the meta-graphs for describing the complementary and substitutable relationships between items, respectively. Based on the diffusion process described earlier, given a social network $G_{SN} = (V, E)$ with the influence strength $p_{u,v}$ for all $u, v \in V$, a KG $G_{KG} = (\mathcal{V}, \mathcal{E}, \Phi, \Psi)$, two sets of meta-graphs $\{m^{\mathsf{C}}\}$ and $\{m^{\mathsf{S}}\}$, a target item set $I = \{x\}$, the cost $c_{u,x}$ of hiring a user $u \in V$ to promote an item $x \in I$, the budget b, and the total number of promotions T, the SIMDPP problem is to find the seed group $S = \bigcup_{t=1}^{T} S_t$ such that the influence spread $\sigma(S)$ is maximized within the budget constraint b, i.e., $\sum_{t=1}^{T} \sum_{(u,x,t) \in S_t} c_{u,x} \leq b$.*[12]

Theorem 1. *SIMDPP cannot be approximated within $O(\frac{1}{|V|^{1-\epsilon}})$ in polynomial time unless $P = NP$, even with only the complementary relationship and $P_{ext} \equiv 0$ in only one promotion.*

Proof Sketch. We prove the theorem with the gap-introducing reduction from the decision problem of Set Cover. Given a set cover instance, by constructing a corresponding special case of SIMDPP, we prove that if there is a set cover solution with at most k sets, there is a feasible solution of SIMDPP with the total influence at least $|U|^c + 2|U| + 2$, where $|U|$ is the size of the ground set of set cover instance, and c is a large constant. Otherwise, if there exists no set cover solution with at most k sets, the optimal value of SIMDPP is at most $2|U| + k + 2$. Then, we assign c suitably to satisfy $|V|^{1-\epsilon} \leq \frac{|U|^c + 2|U| + 2}{2|U| + k + 2}$, where $|V|$ and ϵ are related to $|U|$ and c. Consequently, if there is a $|V|^{1-\epsilon}$ approximation algorithm of SIMDPP, we can solve the decision problem of set cover in polynomial time, implying $P = NP$, which is a contradiction. For more details, please refer to [15]. □

IV. APPROXIMATION ALGORITHM

A. Algorithm Overview

To efficiently solve SIMDPP, we design an approximation algorithm, namely Simple Dysim (SDysim), which embodies a number of ideas. (i) To tackle the challenge in the propagation of item impacts, SDysim introduces *Dynamic Reachability (DR)* to measure the impact made by an item promotion and the impacts resulted from the promotions of other items based on users' dynamic perceptions of item relationships (detailed later in Eq. (1)). Specifically, DR evaluates both *proactive* and *reactive impacts* for each item. For an item, the proactive impact is the probability for this item to result in an increase of users' preferences *on other items*. The reactive impact is the probability to increase users' preferences *on this item* resulted from other items promoted previously. The item with the highest DR is prioritized for promotion. Previous works [18], [20], [21] select users only and do not consider the items in IM.

(ii) To avoid antagonism between substitutable items, SDysim identifies *target markets*, each of which consists of socially close users to promote complementary items in consecutive promotions. Specifically, it identifies some *nominees* (where a nominee is a user-item pair (u, x)) as candidate seeds, denoted by (u, x, t), for an incoming promotion at time t (decided later). Note that a target market targets on a cluster of nominees in order to promote complementary items to socially close users. Since different target markets may share some common users, it is important to avoid promoting substitutable items to them. Accordingly, SDysim prioritizes the target market promoting items with the least substitutable relevance to items in the overlapping target markets (i.e., the target markets sharing many common users). By contrast, prior works [18], [20], [21] consider only one relationship and thereby may promote substitutable items to the same users.

Equipped with the above strategies, SDysim includes two phases: Target Market Identification (TMI) and Dynamic Reachability Evaluation (DRE).[13] Since users in social networks usually have different needs and diverse purchase intentions, a promotional strategy is planned more sophisticatedly if the target users are identified first. Intuitively, intensively promoting a few items within a short period can better draw users' attention. Hence, SDysim first exploits TMI to identify target markets and then leverages DRE to plan the

[11]Following [1], [16], σ is estimated by the Monte Carlo method, which simulates the influence diffusion of seeds according to the probabilities.

[12]As stated in Sec. II, adaptive IM problems may incur an unbounded cost and require a predefined budget allocation. By contrast, SIMDPP does not require a predefined budget allocation and can be solved (by our proposed algorithm) with a limited approximation ratio (as stated in Theorem 2 later).

[13]The complete algorithm Dysim to tackle all challenges is presented in [15]. Besides, our proposed algorithm can deal with adaptive IM (even without a predefined budget allocation to different promotions), detailed in [15].

Algorithm 1: SDysim

Input: Social network $G_{\text{SN}} = (V, E)$; knowledge graph $G_{\text{KG}} = (\mathcal{V}, \mathcal{E}, \Phi, \Psi)$; item set I; total budget b; total number of promotions T

Output: Seed group

/* TMI phase */

$U \leftarrow \{(u, x) \mid u \in V, x \in I\}$

$N \leftarrow$ `selectNominees` (U, b)

$\{N^\tau\} \leftarrow$ `clusterNominees` (N)

for *each* N^τ **do**

  Identify the target market τ by N^τ

$CG \leftarrow$ `prioritizeTargetMarket` $(\{\tau\})$

for *each* $\mathcal{G}$ *in* CG **do**

  $S^{\mathcal{G}} \leftarrow \emptyset$

  for *each* $\tau_k \in \mathcal{G}$, *where* $k = 1, 2, \ldots$ **do**

    /* DRE phase */

    $N^{\tau_k} \leftarrow$ nominees in τ_k

    $I^{\tau_k} \leftarrow \{x \mid (u, x) \in N^{\tau_k}\}$

    while $I^{\tau_k} \neq \emptyset$ **do**

      $x_{\text{p}} \leftarrow \text{argmax}_{x \in I^{\tau_k}} DR^{\tau_k}(S^{\mathcal{G}}, x)$

      $I^{\tau_k} \leftarrow I^{\tau_k} \setminus \{x_{\text{p}}\}$

      if $S^{\mathcal{G}} \neq \emptyset$ **then**

        $\hat{t} \leftarrow \max\{t \mid (u, x, t) \in S^{\mathcal{G}}\}$

        $t_{\text{p}} \leftarrow \min\{\hat{t} + 1, T\}$

      else

        $t_{\text{p}} \leftarrow 1$

      for *each* $(u, x_p) \in N^{\tau_k}$ **do**

        $S^{\mathcal{G}} \leftarrow S^{\mathcal{G}} \cup \{(u, x_{\text{p}}, t_{\text{p}})\}$

return $\bigcup_{\mathcal{G}} S^{\mathcal{G}}$

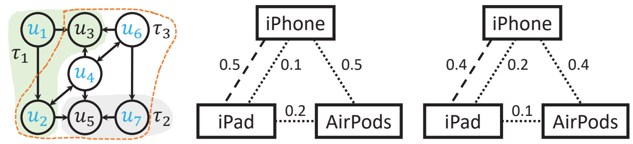
distinct effective promotional strategy for each target market. Specifically, TMI selects and clusters nominees to promote complementary items to each target market and prioritizes target markets with fewer substitutable items to the nominees in the overlapping target markets. For each target market, DRE finds the item with the highest DR to exploit item impacts and decides the promotional timings for the corresponding nominees to be seeds. Algorithm 1 presents the pseudo-code of SDysim.[14]

B. Algorithm Description

1. Target Market Identification (TMI): TMI selects the nominees that exert large influence spread, clusters select nominees to identify each target market for promoting complementary items to socially close users, and prioritizes the target market with fewer substitutable items to the nominees in the overlapping target markets.

For nominee selection, TMI carefully examines the marginal gain of influence for each nominee. It's crucial to select a cost-effective nominee due to different costs of nominees and a limited budget. Therefore, we propose *marginal cost-performance ratio (MCP)* to jointly consider the above factors and ensure the approximation ratio of SDysim in Theorem 2. Specifically, given a set N of selected nominees, MCP of a nominee (u, x) is $\frac{f(N \cup \{(u,x)\}) - f(N)}{c_{u,x}}$, where f is the influence spread σ with the nominees placed in the first promotion

[14]Please refer to [15] for more pseudo-codes.



(a) (b) (c)

Fig. 4. An example of TMI. (a) A social network. (b) Average relevance over all users in the whole social network. (c) Average relevance in τ_3.

as the seeds and P_{pref} assigned at the beginning of this promotion. For the nominees with the costs satisfying $c_{u,x} < b - \sum_{(u',x') \in N} c_{u',x'}$, TMI iteratively extracts the one with the highest MCP into N.

Afterward, TMI identifies the target markets by clustering the nominees. To promote complementary items to the users in a target market, TMI first clusters the nominees in N (e.g., by clustering methods POT [31] and FGCC [32])[15] according to the social distances between the nominees and the relevance between their promoting items, i.e., $\bar{r}^{\mathsf{C}}_{x,y} - \bar{r}^{\mathsf{S}}_{x,y}$, where $\bar{r}^{\mathsf{C}}_{x,y}$ and $\bar{r}^{\mathsf{S}}_{x,y}$ are the average complementary and substitutable relevance between x and y over all users, respectively.[16] Larger complementary and smaller substitutable relevance are encouraged. For each cluster, a target market τ is identified by exploring the influenced users from the nominees N^τ (e.g., by MIOA [16]).[17] Note that with TMI, the budget allocation of SDysim is realistic since a larger target market is inclined to have a larger budget to promote items. In TMI, the target markets are identified by the influence of nominees, where more nominees and influential nominees lead to a larger target market. As more nominees and influential nominees usually incur a larger cost [3], SDysim allocates larger budgets to those target markets accordingly.

Afterward, TMI prioritizes the target market with fewer substitutable items to the nominees in the overlapping target markets. Let $\mathcal{G}$ denote a set of target markets with common users. A target market τ_i is in $\mathcal{G}$ if there is another target market $\tau_j \in \mathcal{G}$ with the common user number above a threshold θ.[18] TMI arranges the promoting order for the target markets in each $\mathcal{G}$ by deriving *Antagonistic Extent (AE)* of each target market τ_i according to the substitutable relationship between every promoting item x and the items of other target market τ_j, i.e., $AE(\tau_i) = \sum_{x \in \tau_i, y \in \tau_j} \bar{r}^{\mathsf{S}}_{x,y}$, where $\tau_i, \tau_j \in \mathcal{G}, i \neq j$. The target market (and the items in the corresponding nominees) with a smaller AE is promoted earlier in $\mathcal{G}$.[19]

Example 1. Figs. 4(a) and 4(b) present an example of TMI with a social network and the average relevance over all users in the whole social network, where the dotted and dashed edges are the complementary and substitutable relationships, respectively. The number beside each edge is the relevance.

[15]More details of POT and FGCC are presented in [15].

[16]The derivation of relevance is described in [15].

[17]More details of MIOA are presented in [15].

[18]The sensitivity of SDysim to θ is evaluated in [15].

[19]Alternatively, according to research in the marketing field, the profitability [33] of a target market is also a good metric to prioritize target markets. The comparison of different marketing orders is presented in Sec. V-D.

Assume $N = \{(u_1,\text{ iPad}),(u_2,\text{ AirPods}),(u_4,\text{ iPhone}),(u_6,\text{ AirPods}),(u_7,\text{ iPad})\}$ by TMI according to MCP. Then, TMI finds three clusters $N^{\tau_1} = \{(u_1,\text{ iPad})\}$, $N^{\tau_2} = \{(u_7,\text{ iPad})\}$, and $N^{\tau_3} = \{(u_2,\text{ AirPods}),(u_4,\text{iPhone}),(u_6,\text{ AirPods})\}$ from Figs. 4(a) and 4(b), and identifies τ_1, τ_2, and τ_3 accordingly, as shown in Fig. 4(a). Assume $\theta = 1$. Then, τ_1, τ_2, and τ_3 belong to the same $\mathcal{G}$ since τ_1 and τ_3 have two common users, and τ_2 and τ_3 have two common users. After that, according to the substitutable relevance in Fig. 4(b), $AE(\tau_1) = 0.5$ since iPad promoted in τ_1 is substitutable to iPhone promoted in τ_3. Similarly, $AE(\tau_2) = 0.5$ and $AE(\tau_3) = 0.5 + 0.5 = 1$. TMI thereby promotes τ_1, τ_2, and τ_3 sequentially. ■

2. Dynamic Reachability Evaluation (DRE): For each target market $\tau_k \in \mathcal{G}$ selected by TMI, DRE evaluates *Dynamic Reachability (DR)* of each item in τ_k, and the nominees (in N^{τ_k}) promoting the item with the highest DR serve as the candidate seeds. In other words, after TMI has identified target markets with socially close users to promote complementary items and has prioritized the target markets using AE, DRE allows each target market to prioritize its promoting items differently and lets the nominees promoting items with higher DR be the seeds earlier. Specifically, let d^{τ_k} denote the diameter of the target market τ_k, and $S^{\mathcal{G}}$ is the seed group determined so far for all the target markets in $\mathcal{G}$. Let I^{τ_k} denote the items that have not yet been promoted in τ_k. DR of an item $x \in I^{\tau_k}$ is

$$DR^{\tau_k}(S^{\mathcal{G}}, x) = PI^{\tau_k}(S^{\mathcal{G}}, x, d^{\tau_k}) + RI^{\tau_k}(S^{\mathcal{G}}, x, d^{\tau_k}). \quad (1)$$

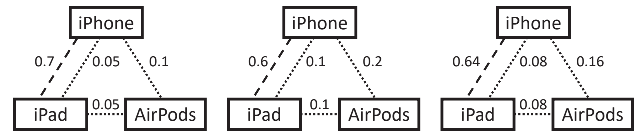
The proactive impact $PI^{\tau_k}(S^{\mathcal{G}}, x, d^{\tau_k})$ is the probability of x to increase the preferences of users in τ_k *for other items*. The reactive impact $RI^{\tau_k}(S^{\mathcal{G}}, x, d^{\tau_k})$ is the probability to increase the preferences of users in τ_k *for* x under the impact from other items in $S^{\mathcal{G}}$.[20] The adoption of x increases (decreases) the preferences for the items complementary (substitutable) to x [7]. Given $S^{\mathcal{G}}$, the likelihood of regarding x and y as complementary (substitutable) for each user is proportional to the complementary (substitutable) relevance between x and y, i.e., $\mathcal{L}^{\mathsf{C},\tau_k}(x, y, S^{\mathcal{G}}) = \frac{\bar{r}^{\mathsf{C}}_{x,y}}{\bar{r}^{\mathsf{C}}_{x,y}+\bar{r}^{\mathsf{S}}_{x,y}}$ and $\mathcal{L}^{\mathsf{S},\tau_k}(x, y, S^{\mathcal{G}}) = \frac{\bar{r}^{\mathsf{S}}_{x,y}}{\bar{r}^{\mathsf{C}}_{x,y}+\bar{r}^{\mathsf{S}}_{x,y}}$, where $\bar{r}^{\mathsf{C}}_{x,y}$ and $\bar{r}^{\mathsf{S}}_{x,y}$ are the average complementary and substitutable relevance between x and y over all users in τ_k after the promotion of $S^{\mathcal{G}}$, respectively.[21] Therefore, PI is recursively formulated as follows.

$$\begin{aligned} &PI^{\tau_k}(S^{\mathcal{G}}, x, d) \qquad (2)\\ &= \sum_y \Big(\mathcal{L}^{\mathsf{C},\tau_k}(x, y, S^{\mathcal{G}})\bar{r}^{\mathsf{C}}_{x,y} - \mathcal{L}^{\mathsf{S},\tau_k}(x, y, S^{\mathcal{G}})\bar{r}^{\mathsf{S}}_{x,y} \\ &\qquad + PI^{\tau_k}(S^{\mathcal{G}}, y, d-1)\Big), \end{aligned}$$

where y represents each item relevant to x. The first two terms are the likelihood to increase and decrease the preferences of the users in τ_k *for* y (weighted by the corresponding relevance between x and y). The last term $PI^{\tau_k}(S^{\mathcal{G}}, y, d-1)$ recursively captures the likelihood to increase or decrease the preferences (of users in τ_k) *for other items* via item impact propagation from y, where $PI^{\tau_k}(S^{\mathcal{G}}, y, 0) = 0$.[22]

[20] d^{τ_k} appears in PI and RI to restrict the item impact propagation to the users at most d^{τ_k} away in τ_k.

[21] The update of relevance is described in [15].



(a) $t = 0$ (b) $t = 1$ (c) $t = 1$

Fig. 5. An example of DRE: an illustration of the dynamics in u_5's personal item network. (a) Initially. (b) After u_5 adopts iPad. (c) Expectation.

Similarly, RI evaluates the item impact propagation from any promoted item z to x according to $\bar{r}^{\mathsf{C}}_{z,x}$ and $\bar{r}^{\mathsf{S}}_{z,x}$ as follows.

$$\begin{aligned} &RI^{\tau_k}(S^{\mathcal{G}}, x, d) \qquad (3)\\ &= \sum_z \Big(\mathcal{L}^{\mathsf{C},\tau_k}(z, x, S^{\mathcal{G}})\bar{r}^{\mathsf{C}}_{z,x} - \mathcal{L}^{\mathsf{S},\tau_k}(z, x, S^{\mathcal{G}})\bar{r}^{\mathsf{S}}_{z,x} \\ &\qquad + RI^{\tau_k}(S^{\mathcal{G}}, z, d-1)\Big), \end{aligned}$$

where z is each item relevant to x and $RI^{\tau_k}(S^{\mathcal{G}}, y, 0) = 0$.

Consequently, for each τ_k selected by TMI, DRE extracts the nominees $\{(u, x_{\text{p}}) \mid x_{\text{p}} = \operatorname*{argmax}_{x\in I^{\tau_k}} DR^{\tau_k}(S^{\mathcal{G}}, x), (u, x_{\text{p}}) \in N^{\tau_k}\}$ with the highest DR as the candidate seeds iteratively, and this property (i.e., the highest DR) is important to approximate the optimal solution in Theorem 2.

Example 2. Following Example 1, this example shows how the DRE of SDysim works based on the diffusion model to solve SIMDPP. Assume that the seed group becomes $S^{\mathcal{G}} = \{(u_1,\text{ iPad}, 1)\}$ after τ_1 is promoted. To update the complementary and substitutable relevance in each user's dynamic perception, SDysim employs the Monte Carlo method to generate different cases of users' adoption decisions according to their preferences for items and the influence strength. For example, suppose that, due to the seed u_1's promotion at $t = 1$, u_2 adopts iPad and promotes to u_5 by her social influence. For the case that u_5 adopts iPad, the update of her adopting item (referring Fig. 3) in turn changes her personal item network from Fig. 5(a) to Fig. 5(b). Meanwhile, u_5's preference for AirPods increases due to the adopted item {iPad} and her changed personal item network (Fig. 5(b)), where the complementary relevance between iPad and AirPods increases from 0.05 to 0.1. By contrast, for the case that u_5 does not adopt iPad at $t = 1$, her personal item network and her preference for AirPods remains unchanged. Based on the number of times these cases being observed, SDysim computes the expectation of u_5' personal item network, as shown in Fig. 5(c).

After that, assume that $S^{\mathcal{G}} = \{(u_1,\text{ iPad}, 1),(u_7,\text{ iPad}, 2)\}$ after τ_2 is also promoted. SDysim now concentrates on τ_3, where $N^{\tau_3} = \{(u_2,\text{ AirPods}),(u_4,\text{ iPhone}),(u_6,\text{ AirPods})\}$,

[22] Here it is $d-1$ because item impact has propagated 1-hop from x to y.

$d^{\tau_3} = 3$, and $I^{\tau_3} = \{\text{iPhone, AirPods}\}$ (due to the items not yet promoted by the nominees in N^{τ_3}). DRE calculates the DR for iPhone and AirPods, i.e., $DR^{\tau_3}(S^{\mathcal{G}}, \text{iPhone})$ and $DR^{\tau_3}(S^{\mathcal{G}}, \text{AirPods})$, respectively, according to the updated (same as above) personal item networks.

$$\begin{aligned}
&DR^{\tau_3}(S^{\mathcal{G}}, \text{iPhone})\\
&= \Big(\frac{0.2 \cdot 0.2}{0.4+0.2} - \frac{0.4 \cdot 0.4}{0.4+0.2} + PI^{\tau_3}(S^{\mathcal{G}}, \text{iPad}, 2)\Big)\\
&\quad + \big(1 \cdot 0.4 + PI^{\tau_3}(S^{\mathcal{G}}, \text{AirPods}, 2)\big)\\
&\quad + \Big(\frac{0.2 \cdot 0.2}{0.4+0.2} - \frac{0.4 \cdot 0.4}{0.4+0.2} + RI^{\tau_3}(S^{\mathcal{G}}, \text{iPad}, 2)\Big)\\
&= 1.4 + (-0.2) = 1.2.
\end{aligned}$$

Since $DR^{\tau_3}(S^{\mathcal{G}}, \text{AirPods}) = 1.8 > DR^{\tau_3}(S^{\mathcal{G}}, \text{iPhone}) = 1.2$, DRE then extracts $\{(u_2, \text{AirPods}), (u_6, \text{AirPods})\}$ for promotion first. ■

After a set of nominees $\{(u, x_{\text{p}})\}$ are extracted by DRE, SDysim finds the promotional timing for them. Since the target markets have been arranged in a promoting order (by TMI) and the items with higher DR have been promoted with higher priority (by DRE), the promotional timing t_{p} for $\{(u, x_{\text{p}})\}$ is assigned right after the latest promotion in $S^{\mathcal{G}}$ to ensure the influence spread of the seeds in $S^{\mathcal{G}}$ is not reduced. Let $\hat{t}$ denote the latest promotion in $S^{\mathcal{G}}$, i.e., $\hat{t} = \max\{t \mid (u, x, t) \in S^{\mathcal{G}}\}$. SDysim assigns $\{(u, x_{\text{p}})\}$ in the promotion $t_{\text{p}} = \min\{\hat{t}+1, T\}$ to be the seeds, i.e., $S^{\mathcal{G}} \cup \{(u, x_{\text{p}}, t_{\text{p}})\}$. Then, SDysim selects the next item with DRE and determines the timing for the corresponding nominees. After all nominees in τ_k are assigned their promotional timings as the seeds, TMI moves on to the next target market $\tau_{k+1} \in \mathcal{G}$. It returns the seed group $S = \bigcup_{\mathcal{G}} S^{\mathcal{G}}$ as the solution after all target markets are examined.[23]

Theorem 2. *SDysim is a* $(1 - \frac{1}{\sqrt{e}} - \epsilon)(\min\{P^c_{\textit{minpref}}, P^c_{\textit{minext}}\})$ *approximation algorithm for SIMDPP in* $O(M\,|V|\,|I|\,k_{\max})$ *time, where* $P_{\textit{minpref}} > 0$ *and* $P_{\textit{minext}} > 0$ *are the minimum preference and extra adoption probability, respectively.* c *is the maximum hop of influence propagation,* M *is the time to evaluate* σ *depending on the evaluation error* $\epsilon > 0$,[24] *and* $k_{\max}$ *is the maximum size of a feasible solution.*

Proof. For more details, please refer to [15]. □

V. Experiments

A. Experiment Setup

The experiment includes four datasets, where each one consists of a KG and a social network: i) *Douban* [20], ii) *Gowalla*, iii) *Yelp*, and iv) *Amazon*.[25] Since there are no social relationships in *Amazon*, we supplement it with Pokec[26] according to the user profiles. To capture the complementary and substitutable relationships between items, the meta-graphs are generated according to [28], and the relevance of a certain relationship regarding a meta-graph is derived according to [25]. For the diffusion models, the two factors, relevance measurement (including the learning of personal weightings on meta-graphs and the constructions of personal item networks) and preference estimation are learned and updated based on [10] and [34], respectively. The statistics of the datasets are listed in Table II. Following [3], the costs of hiring users to promote items are set proportional to users' out-degree and their preferences for items, since users who are more influential and who prefer the item less may need more incentive to be seeds. In the implementation of Dysim, we exploit the submodularity to speed up the nominee selection, and follow [31] and [16] to cluster nominees and explore influenced users, respectively, in TMI.

TABLE II
THE STATISTICS OF DATASETS.

Dataset	*Douban*	*Gowalla*	*Yelp*	*Amazon*
# of node types	3	3	6	6
# of nodes	7.6M	3.2M	251K	260K
# of users	5.5M	407K	17K	1.6M
# of items	2.1M	2.8M	22K	20K
# of edge types	3	3	6	6
# of edges	100M	42M	1.6M	1.4M
# of friendships	86M	4.4M	140K	30.6M
Directed friendship?	No	No	No	Yes
Avg. initial influence strength	0.011	0.092	0.121	0.050

We compare SDysim with OPT (derived from a brute-force approach) and four state-of-the-art approaches: BGRD [21], HAG [20], PS [18], and DRHGA [14] as the baselines.[27] We extend [14], [18], [20], [21] to consider different costs of selecting a user to promote an item by selecting from the user-item pairs or the users that satisfy the remaining budget. Furthermore, since they cannot be directly applied to our problem, we augment [14], [18], [20], [21] with CR-Greedy [22] to support multiple promotions and determine the promotion timings of the user-item pairs as the seeds in each baseline. The performance metrics include the 1) influence spread σ (Def. 1) and 2) execution time. We perform a series of sensitivity tests in terms of the budget b and the number of promotions T. To verify our algorithm, we further conduct an empirical study on course promotion in viral marketing for the course selection system. The complete experiments for IMDPP (including the case study on *Amazon*) are shown in [15] due to the space constraint. We conduct all experiments on an HP DL580 server with an Intel 2.10GHz CPU and 1TB RAM. Each simulation result is averaged over 100 samples.

B. Performance Comparison

First, we compare all approaches and OPT on small datasets sampled from *Amazon* with 100 users. Fig. 6(a) shows the influence under different budgets. SDysim has the closest

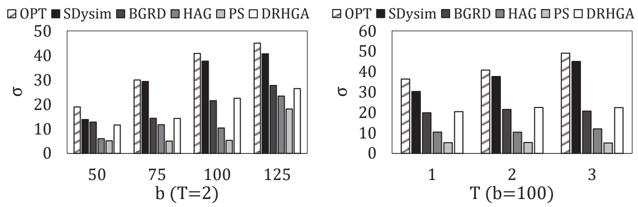
[23] $S^{\mathcal{G}}$ of different $\mathcal{G}$ can be derived in parallel due to the independency of different $\mathcal{G}$.

[24] Note that the technique of reverse influence sampling cannot support multiple promotions since the dependency among different promotions makes positive propagation irreversible.

[25] *Gowalla*, *Yelp*, and *Amazon* are from https://www.yongliu.org/datasets, https://www.yelp.com/dataset, and https://jmcauley.ucsd.edu/data/amazon, respectively. KGs are HINs in the datasets, where the HINs contain diverse node types like items, categories, brands, etc.

[26] https://snap.stanford.edu/data/soc-Pokec.html.

[27] Codes and datasets are available on https://tinyurl.com/y26fx2mp.



(a) Diff. budgets. (b) Diff. numbers of promotions.

Fig. 6. Comparisons with optimal solutions.

performance to OPT, and outperforms BGRD, HAG, PS, and DRHGA, because TMI of SDysim carefully selects influential nominees by MCP, and DRE of SDysim then prioritizes nominees based on dynamic perceptions of item relationships. In contrast, the baselines neglect the changes in item relationships and do not promote items beneficial to each other over time. Fig. 6(b) compares the influence under various numbers of promotions. SDysim creates a larger influence spread as T increases because TMI avoids promoting substitutable items to the same users in near promotions. All baselines do not incorporate the item impact propagation to achieve a larger influence spread as T grows even a sophisticated algorithm based on CR-Greedy [22] is employed to schedule promotions at different timings.

Figs. 7(a)-7(c) compare the influence in large datasets under different budgets.[28] For all datasets, SDysim achieves the largest influence spread, followed by DRHGA, BGRD, HAG, and PS, because SDysim is able to exploit the changes in users' preferences. PS fails to obtain a large influence spread because it only estimates the influence of a seed alone and cannot utilize the impact of items from other promotions to find seeds. BGRD usually achieves smaller than half of the influence compared with SDysim, because it neglects the substitutable relationship and regards all items as a bundle to be promoted. Although DRHGA also promotes all items, it is usually better than BGRD since DRHGA is able to select appropriate users to promote each item, instead of regarding all items as a bundle in BGRD. However, as DRHGA does not choose items to be promoted, it still generate a smaller influence spread compared with SDysim. HAG outperforms BGRD in *Yelp* with low budgets and in *Amazon* when the budget is relatively low to the social network size. This is because HAG greedily selects the most influential combination of user-item pairs as the seeds, instead of the most influential user to promote a bundle of items, making the solutions of HAG more cost-effective. BGRD fails to achieve a large influence spread for a large b in *Douban* since items (e.g., songs and books) in *Douban* are usually complementary, but BGRD still allocates the budget to the same users to promote a bundle of complementary items.

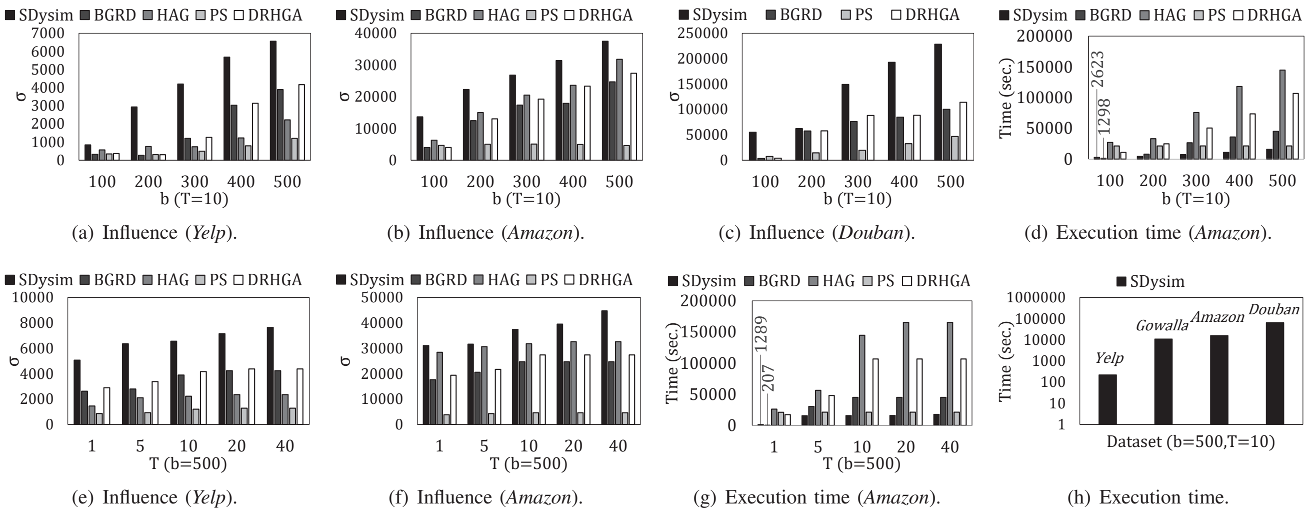
Figs. 7(e)-7(f) present the influence in large datasets under different numbers of promotions with the maximal T as 40 (following [22]). SDysim achieves the largest influence spread for all T with significant increments as T grows, because TMI of SDysim first arranges the promoting order of target markets, and SDysim then exploits DR to prioritize items to be promoted for each target market. In contrast, the influence spreads grow slowly for the baselines, especially when $T \geq 20$, because they cannot arrange the promoting order holistically and fail to utilize more promotions to properly gain more adoptions.

Figs. 7(d) and 7(g)-7(h) compare the execution time under different budgets and different numbers of promotions, respectively. As shown in Fig. 7(d), when b varies, SDysim requires the least execution time for most cases. HAG suffers from finding numerous combinations of seeds for a large budget. PS requires much time to search for maximum influence paths to evaluate the influence of a user. Although DRHGA only selects users, it takes more time than BGRD since the selection process is repeated for each item. As b becomes larger, the execution time of SDysim only slightly increases since TMI quickly selects influential nominees by MCP according to the cost and increment on influence for each candidate nominee. PS is less sensitive to b since it employs a discounting strategy to estimate a seed's influence under the impact of selected seeds. On the other hand, as shown in Fig. 7(g), SDysim requires a low overhead to find promotional timings since it assigns the promotions by TMI and DRE, which are less sensitive to T, whereas the baselines greedily assigning the promotional timings tend to suffer from larger T. To show the scalability of SDysim, Fig. 7(h) compares the execution time of SDysim on different datasets (in the order of the number of users in the social network). The time increases not only as the number of users increases but also as that of items increases (e.g., so the time on *Gowalla* and *Amazon* are similar) due to the propagation of item impact.

C. Ablation Study

Fig. 8 compares SDysim, SDysim without target markets (i.e., w/o TM), and SDysim without item priority (i.e., w/o IP). We have the following three observations. First, the influence spread is smaller when target markets are not identified, since the selected nominees may promote substitutable items to the same users in consecutive promotions, which detracts from users' preferences for the posterior items to be promoted. By contrast, SDysim effectively avoids the antagonism of the substitutable relationship by identifying and prioritizing the target markets. Second, the influence spread of SDysim without item priority is also smaller than that of SDysim, because all items in a target market are promoted simultaneously, and therefore the promotion of an item is hardly facilitated by promoting its complementary items first. In contrast, SDysim determines the item priority by exploiting DR, which carefully measures the impact from previously promoted items on an item and also the potential impact from this item on other items in subsequent promotions. Third, as T increases, the gaps between SDysim and SDysim w/o TM/IP increase. This is because the number of promotions in SDysim w/o TM/IP is limited, i.e., at most the number of items/target markets, implying that more promotions are not beneficial

[28]Fig. 7(c) doesn't include HAG due to execution time longer than 12 hours.



(a) Influence (*Yelp*). (b) Influence (*Amazon*). (c) Influence (*Douban*). (d) Execution time (*Amazon*).

(e) Influence (*Yelp*). (f) Influence (*Amazon*). (g) Execution time (*Amazon*). (h) Execution time.

Fig. 7. Comparisons on large datasets.

for a larger influence spread. By contrast, SDysim effectively schedules the promotional timings of different target markets and different items to exploit the propagation of item impacts.

D. Comparison of Different Market Orders

To compare with Antagonistic Extent (AE), we leverage the following metrics to evaluate additional promotional orders of target markets: profitability (PF) [33], size of the market (SZ) [33], relative market share (RMS) [35], and random (RD). PF and SZ are two of the most common criteria to prioritize target markets in the marketing research field. PF is the expected adoptions under the promotion from the corresponding nominees minus the cost of the nominees. SZ is the number of customers in the target market. A target market with a larger PF or SZ is preferred to be promoted earlier. RMS is widely used to assess the value of a firm's item in the product management field. RMS of an item x is defined as the ratio of x's market share to the largest market share of its substitutable item, where the market share is evaluated by the number of users preferring the item most. The target market that promotes items with a higher RMS is prioritized.

Fig. 9 manifests that AE and PF usually achieve the largest influence spread, followed by SZ, RMS, and RD. AE and PF outperform the others since AE prioritizes the target markets that have less substitutable relationship on the subsequent target markets, while PF prioritizes the target markets with more profits to ensure their influence spread. When there exist excessively large target markets (e.g., identified by plenty of nominees), PF is suggested as the ordering metric, since PF can accurately prioritize these large target markets to maximize the influence. In general cases, AE is usually a better metric to prioritize the target markets since the impact from the substitutable items promoted by prior target markets is minimized. By contrast, SZ, RMS, and RD, without carefully examining the relationships of items promoted in other target markets, cannot avoid the antagonism of the substitutable relationship. The results manifest that promoting target markets with a smaller AE or a larger PF earlier in TMI is beneficial to achieve a larger influence spread.

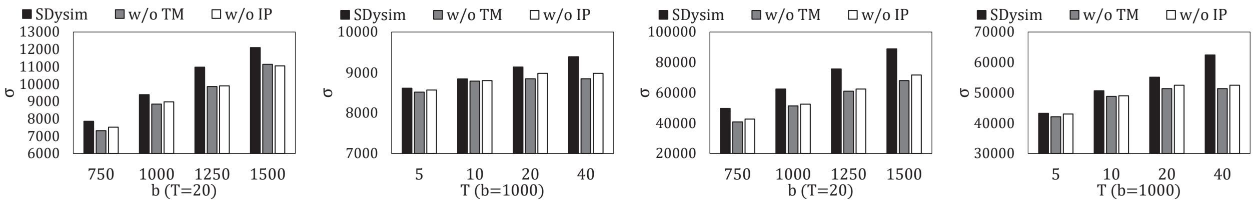
E. Empirical Study

In this study, we have recruited five classes for promoting courses by viral marketing to evaluate the effectiveness of SDysim in real-world settings. There were 30 elective courses for computer science college students, including artificial intelligence (AI), objective-oriented programming (OOP), and big data, to name a few. The goal of the campaigns is to encourage the students in Taiwan University to select those courses, i.e., maximizing the total number of students selecting the elective courses. The statistics of all classes are presented in Table III.

To construct KG of these courses, we crawled their syllabuses from Taiwan University, and extracted keywords of courses, related compulsory courses, and research fields of teachers. The meta-graphs were defined according to the curriculum guidelines in Taiwan.[29] Following [3], the costs of hiring users to promote courses are set as users' out-degree over their initial preferences for courses, since users who are more influential and who prefer the course less may need more incentive to be the seeds.

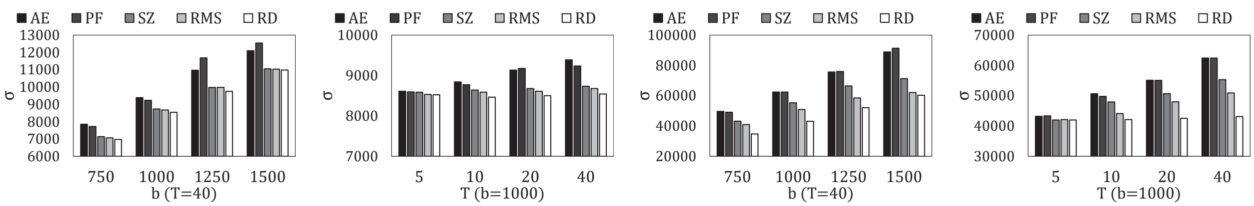
To evaluate the effectiveness of different approaches, we have launched campaigns based on the following approaches: 1) SDysim, 2) BGRD [21], 3) HAG [20], and 4) PS [18]. In this study, the budget and the number of promotions were set to 50 and 3, respectively. For SDysim, relevance measurement (including the learning of personal weightings on meta-graphs and the constructions of personal item networks) and preference estimation are updated according to [10] and [34], respectively. TMI of SDysim follows [31] and [16] to cluster nominees and explore influenced users, respectively.

[29]https://cirn.moe.edu.tw/Upload/file/32077/83646.pdf.



(a) Diff. b (*Yelp.*) (b) Diff. T (*Yelp*). (c) Diff. b (*Amazon.*) (d) Diff. T (*Amazon*).

Fig. 8. The ablation study.



(a) Diff. b (*Yelp.*) (b) Diff. T (*Yelp*). (c) Diff. b (*Amazon.*) (d) Diff. T (*Amazon*).

Fig. 9. Comparisons of different market orders.

Fig. 10 reports the total number of students selecting the elective courses for different approaches in each class. For all classes, SDysim induces the most students who selected those courses, followed by BGRD, HAG, and PS. These results validate that SDysim is able to encourage students to select those courses by carefully evaluating the dynamic changes in the relationships between courses. For instance, we observe that a student in Class A initially regarded the complementary relevance between AI and software design for cloud computing (SDCC) as 0.1. After he selected AI and big data, the complementary relevance between AI and SDCC increased to 0.6 (derived according to [10]). He then selected SDCC accordingly. In Class D, another student initially reported that the influence from one of her classmates is 0.2. During the promotions, both of them selected cloud computing and IoT, which increased the classmate's influence to this student to 0.7 (derived according to [36]). Then, this student selected big data after being informed that the classmate selected big data as well. By contrast, BGRD, HAG, and PS do not capture the dynamic changes in the relationships between courses and the ripple effect, resulting in fewer students selecting the elective courses in the end.

Besides, although BGRD is able to select influential students in each class, all courses are promoted as a bundle without considering their relationships. For example, in Class B, BGRD selects a student to promote python and C++ in a bundle, but the two courses were usually regarded as substitutable for most students (i.e., the average substitutable relevance between python and C++ was 0.7). We observe that more than two-thirds of the students who selected python did not select C++ when they were promoted C++ by their classmates. Similar to BGRD, HAG does not examine the substitutable relationship when promoting courses. In Class B,

TABLE III
THE STATISTICS OF RECRUITED CLASSES.

Class ID	A	B	C	D	E
# of users	33	26	22	20	20
# of edges	293	420	387	227	308

HAG also promoted OOP and C++ to the same set of students. However, more than half of the students selected only one of OOP and C++, indicating the waste of simultaneous promotions for substitutable items. PS induces the fewest students to select the elective courses, since it does not facilitate students to promote multiple courses and cannot properly utilize the course promotion from other seeds. For example, in Class C, PS selected a student to promote deep learning (DL) to a set of students who were very interested in DL (i.e., their average initial preference for DL was 0.9). As DL and natural language processing (NLP) were regarded as highly complementary for this set of students (i.e., the average complementary relevance between DL and NLP was 0.75), a good strategy is to let the students selected by PS to promote NLP as well. However, PS did not promote any other course to this set of students in Class C. The above results lead to conclusions consistent with the experiments in Sec. V-B, indicating that exploring the dynamic personal perceptions of item relationships, dynamic preference for items, dynamic social influence strength, and item associations is the cornerstone of influence maximization under a sequence of promotions on relevant items.

VI. CONCLUSION

To the best of our knowledge, this paper makes the first attempt to study the problem of influence maximization under a sequence of promotions for multiple relevant items. By exploring KG and meta-graphs to capture dynamic personal perceptions of item relationships, we formulate a new problem,

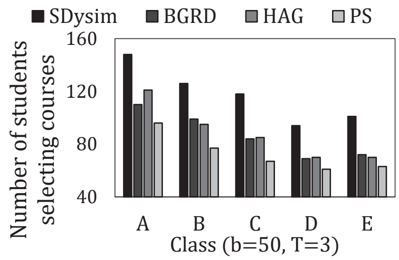


Fig. 10. The empirical study: the total number of students selecting the elective courses for each classes.

named IMDPP, and its fundamental problem, namely SIMDPP, to choose items and select seed users for promotions at proper timings. We prove the hardness of SIMDPP and IMDPP and design an approximation algorithm Dysim to solve IMDPP. Dysim first identifies nominees and target markets to promote complementary items to socially close users in consecutive promotions. For each target market, Dysim prioritizes the items to be promoted by dynamic reachability of items. Then, Dysim determines proper promotional timings with the highest substantial influence for each nominee. We also design an approximation algorithm SDysim for SIMDPP. Experiments on real social networks and KGs demonstrate that SDysim and Dysim can effectively achieve at least 6 times of the influence spread. Furthermore, the empirical study validates that exploring the dynamic personal perceptions of item relationships, dynamic preference for items, dynamic social influence strength, and item associations is crucial for influence maximization under a sequence of promotions on relevant items.

Acknowledgement

This work is supported in part by the National Science Foundation under Grant No. IIS-1717084, by MOST through grants 108-2221-E-002-022-MY3, 109-2221-E-001-017-MY2, 107-2221-E-001-011-MY3, 109-2221-E-305-013-MY3, and 108-2221-E-305-010-, and by the Institute for Information Industry under Contract 109-1134 and 108-1602.

References

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *ACM SIGKDD*, 2003.
[2] K. Han, C. Xu, F. Gui, S. Tang, H. Huang, and J. Luo, "Discount allocation for revenue maximization in online social networks," in *ACM Mobihoc*, 2018.
[3] H. T. Nguyen, T. N. Dinh, and M. T. Thai, "Cost-aware targeted viral marketing in billion-scale networks," in *IEEE INFOCOM*, 2016.
[4] A. D. Shocker, B. L. Bayus, and N. Kim, "Product complements and substitutes in the real world: The relevance of "other products"," *JM*, 2004.
[5] R. Venkatesh and W. Kamakura, "Optimal bundling and pricing under a monopoly: Contrasting complements and substitutes from independently valued products," *J. Bus*, 2003.
[6] J. Gerlach, R. M. Stock, and P. Buxmann, "Never forget where you're coming from: The role of existing products in adoptions of substituting technologies," *JPIM*, 2014.
[7] R. H. Frank and E. Cartwright, *Microeconomics and behavior*. McGraw-Hill New York, 1991.
[8] M. Zhang and J. Bockstedt, "Complements and substitutes in online product recommendations: The differential effects on consumers' willingness to pay," *Inf. Manag.*, 2020.
[9] N. T. Koukova, P. Kannan, and A. Kirmani, "Multiformat digital products: how design attributes interact with usage situations to determine choice," *JMR*, 2012.
[10] C. Shi, Z. Zhang, Y. Ji, W. Wang, P. S. Yu, and Z. Shi, "Semrec: a personalized semantic recommendation method based on weighted heterogeneous information networks," *W3J*, 2019.
[11] Y. Gu, T. Zhou, G. Cheng, Z. Li, J. Z. Pan, and Y. Qu, "Relevance search over schema-rich knowledge graphs," in *ACM WSDM*, 2019.
[12] L. Ma, R. Krishnan, and A. L. Montgomery, "Latent homophily or social influence? an empirical analysis of purchase within a social network," *Manag. Sci.*, 2015.
[13] M. Yavaş and G. Yücel, "Impact of homophily on diffusion dynamics over social networks," *SSCR*, 2014.
[14] H. Huang, Z. Meng, and H. Shen, "Competitive and complementary influence maximization in social network: A follower's perspective," *Knowl.-Based Syst.*, 2020.
[15] Y.-W. Teng, S. Shi, C.-H. Tai, D.-N. Yang, W.-C. Lee, and M.-S. Chen, "Influence maximization based on dynamic personal perception in knowledge graph," arXiv:2010.07125, 2020.
[16] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *ACM SIGKDD*, 2010.
[17] Q. Guo, S. Wang, Z. Wei, and M. Chen, "Influence maximization revisited: Efficient reverse reachable set generation with bound tightened," in *ACM SIGMOD*, 2020.
[18] Y.-W. Teng, C.-H. Tai, P. S. Yu, and M.-S. Chen, "Revenue maximization on the multi-grade product," in *SDM*, 2018.
[19] S. Datta, A. Majumder, and N. Shrivastava, "Viral marketing for multiple products," in *IEEE ICDM*, 2010.
[20] H.-J. Hung, H.-H. Shuai, D.-N. Yang, L.-H. Huang, W.-C. Lee, J. Pei, and M.-S. Chen, "When social influence meets item inference," in *ACM SIGKDD*, 2016.
[21] P. Banerjee, W. Chen, and L. V. S. Lakshmanan, "Maximizing welfare in social networks under a utility driven influence diffusion model," in *ACM SIGMOD*, 2019.
[22] L. Sun, W. Huang, P. S. Yu, and W. Chen, "Multi-round influence maximization," in *ACM SIGKDD*, 2018.
[23] K. Han, K. Huang, X. Xiao, J. Tang, A. Sun, and X. Tang, "Efficient algorithms for adaptive influence maximization," *VLDB*, 2018.
[24] B. Peng and W. Chen, "Adaptive influence maximization with myopic feedback," in *NeurIPS*, 2019.
[25] Z. Huang, Y. Zheng, R. Cheng, Y. Sun, N. Mamoulis, and X. Li, "Meta structure: Computing relevance in large heterogeneous information networks," in *ACM SIGKDD*, 2016.
[26] Y. Timmor and T. Katz-Navon, "Being the same and different: A model explaining new product adoption," *JCB*, 2008.
[27] J. Zheng, X. Wu, J. Niu, and A. Bolivar, "Substitutes or complements: another step forward in recommendations," in *ACM EC*, 2009.
[28] J. McAuley, R. Pandey, and J. Leskovec, "Inferring networks of substitutable and complementary products," in *ACM SIGKDD*, 2015.
[29] Z. Wang, Z. Jiang, Z. Ren, J. Tang, and D. Yin, "A path-constrained framework for discriminating substitutable and complementary products in e-commerce," in *ACM WSDM*, 2018.
[30] Y. Liu, Y. Gu, Z. Ding, J. Gao, Z. Guo, Y. Bao, and W. Yan, "Decoupled graph convolution network for inferring substitutable and complementary items," in *ACM CIKM*, 2020.
[31] H. Chen, H. Yin, X. Li, M. Wang, W. Chen, and T. Chen, "People opinion topic model: opinion based user clustering in social networks," in *WWW*, 2017.
[32] F. Wang, G. Wang, S. Lin, and P. S. Yu, "Concurrent goal-oriented co-clustering generation in social networks," in *IEEE ICSC*, 2015.
[33] L. Simkin and S. Dibb, "Prioritising target markets," *MIP*, 1998.
[34] T. Zhao, J. McAuley, M. Li, and I. King, "Improving recommendation accuracy using networks of substitutable and complementary products," in *IEEE IJCNN*, 2017.
[35] P. W. Farris, N. Bendle, P. E. Pfeifer, and D. Reibstein, *Marketing metrics: The definitive guide to measuring marketing performance*. Pearson Education, 2010.
[36] Z. Zhang, W. Zhao, J. Yang, C. Paris, and S. Nepal, "Learning influence probabilities and modelling influence diffusion in twitter," in *WWW*, 2019.