Information Systems 88 (2020) 101441

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

On optimal preference diffusion over social networks

Cheng Long [a,*], Anhua Chen [b], Pakawadee Pengcharoen [b], Raymond Chi-Wing Wong [b]

[a] School of Computer Science and Engineering, Nanyang Technological University[1], Singapore
[b] The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 30 July 2017
Received in revised form 24 November 2017
Accepted 20 September 2019
Available online 3 October 2019
Recommended by G. Vossen

Keywords:
Preference diffusion
Optimization

ABSTRACT

It was well observed that a user's preference over a product changes based on his/her friends' preferences, and this phenomenon is called "preference diffusion", and several models have been proposed for modeling the preference diffusion process. These models share an idea that the diffusion process involves many iterations, and in each iteration, each user has his/her preference affected by some other preferences (e.g., those of his/her friends). When computing users' preferences after a certain number of iterations, these models use users' preferences at the end of that iteration *only*, which we believe is not desirable since users' preferences at the end of other iterations should also have some effects on users' final preferences. Therefore, in this paper, we propose a new model for preference diffusion, which takes into consideration users' preferences at each iteration for computing users' final preferences. Under the new model, we study two problems for optimizing the preference diffusion process with respect to two different objectives. One is easy to solve for which we design an exact algorithm and the other is NP-hard for which we design a $(1 - 1/e)$-factor approximate algorithm. We conducted extensive experiments on real datasets which verified our proposed model and algorithms.

1. Introduction

It is well recognized that a user's *preference* over a product *changes* based on their friends' opinions over the same product. For example, people share with their friends how they like iPhone 6, people talk about their favorite movie stars with their friends, and people discuss with friends candidates for an election (e.g., the US President election). In these activities, people have their opinions over a product (e.g., iPhone 6, a movie star, or a candidate in an election) affected by those of their friends. The dynamics of people's preferences over a product, when considered on the top of a social network, is called *preference diffusion* (since a user's preference is affected by those of his/her friends which means that the preferences of his/her friends *diffuse* via friendship to him/her).

A related phenomenon is *information diffusion* which has been studied extensively [1–25]. Specifically, information diffusion refers to the phenomenon on a social network that a piece of information or *awareness* (of a product) spreads from some users to their friends and then further to their friends' friends and so on. Nevertheless, information diffusion is different from preference diffusion in quite a few aspects and as a result, exiting models for information diffusion cannot capture preference diffusion well, which we explain as follows. First, in information diffusion, a user's awareness of a product is *binary*, i.e., either yes or no, while in preference diffusion, a user's preference over a product is clearly continuous and corresponds to a *real number*. For example, in the most commonly used models of information diffusion, namely the *independent cascade* (IC) model [1] and the *linear threshold* (LT) model [2], each user is assumed to be either active or inactive which correspond to binary numbers. Second, in information diffusion, a user would usually keep an awareness of a product for the whole propagation process once he/she gets it, i.e., the propagation process is *irreversible*, while in preference diffusion, a user may have his/her preference over a product (e.g., a candidate in an election) *degraded* if some of his/her friends have very low preferences over the same product. Third, in information diffusion, the propagation of the awareness of a product usually happens *once*, i.e., once the awareness is propagated from a user to one of his/her friends, the same propagation process for the same product will not happen again in the future, while in preference diffusion, *multiple* iterations of propagation (e.g., multiple rounds of communications) can happen between two users, and during each iteration, users' preferences could be updated.

Several models have been proposed for capturing the preference diffusion process [26–29]. These models share the following

* Corresponding author.
E-mail address: c.long@ntu.edu.sg (C. Long).

[1] The majority of work was done when Cheng Long was a PhD student at The Hong Kong University of Science and Technology

https://doi.org/10.1016/j.is.2019.101441
0306-4379/

ideas: (1) users' preferences are captured by real numbers and (2) the diffusion process proceeds with many iterations and at each iteration, each user's preference is updated based on his/her own preference and his/her friends' preferences (specifically, it is updated to be a *linear combination* of these preferences). Different models use different strategies of updating a user's preference, e.g., in the model [26,29], it is updated to be a linear combination of his/her preference *at the current iteration* and his/her preferences (*at the current iteration*) and in the model [27,28], it is updated to be a linear combination of his/her preference *before the first iteration* and the *average* of his/her friends' preferences.

It is adopted by each of these models that when computing users' preferences after a certain number h of iterations, *only* users' preferences at the end of the hth iteration are used while those at the end of the $(h-1)$th iteration, those at the end of the $(h-2)$th iteration, ..., and those at the right beginning (or at the end of the 0th iteration) are not, which we argue is not desirable since users' preferences at the end of *each* iteration leave users some impression and hence they contribute to the users' final preferences. Motivated by this, in this paper, we propose a new model which accounts for users' preferences at the end of *each* iteration when computing users' final preferences. Besides, we incorporate the *decaying effects* from the social science literature [30] in our model such that users' preferences at earlier iterations are counted less while those at later iterations are counted more when computing users' preferences at the end of the preference diffusion process.

Our new model guarantees that after a certain of iterations of propagation, (1) users' preferences over the products stay unchanged/stable (or change only insignificantly), i.e., users' preferences *converge*, which is often the case in real life (For example, people usually get clear minds/preferences over the products after enough communication with their friends, which thus will not be changed) and (2) some users have exactly the same preferences over the products, i.e., a *consensus* is reached, which also happens often in real life (For example, a group of users who have similar interests might reach a consensus of the preferences).

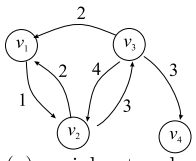
A common application scenario of preference diffusion is as follows. There are some existing products with a product type in the market, and now we want to promote a new product with the same product type. We are given some budget which allow us to *target* at most k users in the social network for the new product, where targeting a user for the new product means giving this user some incentives and consequently this user would have a certain degree of preference over the new product and we call a user that has been targeted as a *seed*.

Based on the above scenario, we study two problems. The first one is called the *preference maximization* (PM) problem which is to select k users as seeds such that at the end of the preference diffusion process, the *sum of users' preferences* over the new product is maximized. The second one is called the *adoption maximization* (AM) problem which is to select k users as seeds such that at the end of the preference diffusion process, the *sum of users' probabilities* to adopt the new product is maximized, where the probability that a user adopts the new product is defined as the *relative preference* over the product to his/her overall preferences over all products (including the existing products). The PM problem optimizes users' *absolute* preferences over the new product while the AM problem optimizes the users' *relative* preferences over the new product.

For the PM problem, we design an exact algorithm called *Top-k*. For the AM problem which is proved to be NP-hard, we design an approximate algorithm called *Greedy*. We prove that *Greedy* provides an $(1-1/e)$-factor approximation for the AM problem.

Contributions & Roadmap. Our contributions are summarized as follows. First, we propose a new model for the preference diffusion process, which is verified by our experiments to perform better than existing ones. Second, under the new model, we study two problems, namely the preference maximization (PM) problem and the adoption maximization (AM) problem, which optimize the preference diffusion process w.r.t. two different objectives by selecting proper sets of seeds. For the PM problem, we design an exact algorithm called *Top-k* and for the AM problem, we prove its NP-hardness and then design an $(1-1/e)$-factor approximate algorithm called *Greedy*. Third, we conducted extensive experiments on real datasets which verified our proposed model and algorithms.

The remaining of the paper is organized as follows. Section 2 presents our proposed preference diffusion model. Section 3 studies the PM problem and the AM problem. Section 4 reviews the related work and Section 5 gives the empirical study and Section 6 concludes the paper.



(a) social network

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 2 & 0 & 3 & 0 \\ 2 & 4 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) The adjacency matrix $\mathbf{A}$

$$\begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.2 & 0 & 0.8 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(c) The influenced-by matrix $\mathbf{M}_{ib}$

$$\begin{bmatrix} 0.5 & 0.25 & 0.25 & 0 \\ 0.1 & 0.5 & 0.4 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

(d) The matrix $\mathbf{M}$ $(\alpha = 0.5)$

Fig. 1. A running example used for illustration.

2. Preference diffusion model

We are given a social network which is represented by a weighted directed graph $G(V, E, W)$ where V is a set of n vertices each representing a user from a set $\{v_1, v_2, \ldots, v_n\}$, E is a set of edges each in the form of (v_i, v_j) meaning that v_i can influence v_j, and $W : E \rightarrow \mathbb{R}^+$ maps each edge $(v_i, v_j) \in E$ to a positive real number r meaning the strength of the influence, termed "influence strength", from v_i to v_j. For a vertex $v_i \in V$, we define its *in-degree*, denoted by $indeg(v_i)$, to be the overall influence strength from v_i's in-coming neighbors to v_i, i.e., $indeg(v_i) = \sum_{e \in E \text{ and } e \text{ is an edge to } v_i} W(e)$. For example, in the graph as shown in Fig. 1(a), the in-degrees of v_1, v_2, v_3 and v_4 are 4, 5, 3, and 3, respectively. We have an adjacency matrix corresponding to graph $G(V, E, W)$, denoted by $\mathbf{A}_{n \times n}$, which is defined as follows. The entry at the ith row and the jth column of $\mathbf{A}$, i.e., $\mathbf{A}[i][j]$, is equal to $W(e)$ where $e = (v_i, v_j)$ if e is an edge in E, and is equal to 0 otherwise. For example, Fig. 1(b) shows the adjacency matrix of the graph in Fig. 1(a).

Assume that we have m product brands (or simply products) under the same product type, namely, a_1, a_2, ..., and a_m. For example, Apple, Samsung and Blackberry are three product brands under the product type of smart phone. We denote by $p_{i,j}$ the *preference* of a user v_i $(1 \leq i \leq n)$ over a product a_j $(1 \leq j \leq m)$ where $p_{i,j} > 0$, and we assume that the larger $p_{i,j}$ is, the more v_i prefers a_j. Given a user v_i $(1 \leq i \leq n)$, we represent v_i's preferences over all products by a vector in the form of $(p_{i,1}, p_{i,2}, \ldots, p_{i,m})^T$ which we call the *preference vector* of user v_i and denote by P_i. Then, we represent the preference vectors of all users by a matrix in the form of $(P_1, P_2, \ldots, P_n)^T$ which we call the *preference matrix* and denote by $\mathbf{P}$. Note that $\mathbf{P}$ is an $n \times m$ matrix and $\mathbf{P}[i][j]$ corresponds to user v_i's preference over product a_j for any $1 \leq i \leq n$ and $1 \leq j \leq m$.

2.1. Preference propagation model

We first describe the preference propagation based on a single user v_i on a product a_j in Section 2.1.1. Then, we describe the preference propagation based on all users on all products in Section 2.1.2.

2.1.1. Single user on single product

In real life, a user v_i usually gets his/her preference over a product a_j *influenced by* his/her friends' preferences over the same product because of the communication among friends in the social network and also *persists* on his/her own preference to a certain extent. The former effect on the preference is called the *influenced-by effect* and the latter effect is called the *persistence effect*. Given a user v_i and a product a_j, we denote the influenced-by effect by $IE(v_i, a_j)$, indicating the preference of v_i on a_j influenced by his/her friends, and denote the persistence effect by $PE(v_i, a_j)$, indicating the preference of v_i on a_j based on his/her original personal preference. The details of $IE(v_i, a_j)$ and $PE(v_i, a_j)$ will be described later.

With the terms $IE(v_i, a_j)$ and $PE(v_i, a_j)$, we define the *overall effect* of user v_i's preference over product a_j, denoted by $OE(v_i, a_j)$, to be

$$\alpha \cdot IE(v_i, a_j) + (1 - \alpha) \cdot PE(v_i, a_j) \tag{1}$$

where α is a parameter $\in [0, 1]$ controlling the trade-off between the influenced-by effect and the persistence effect. As could be noticed clearly, when $\alpha = 0$, the overall effect correspond to the persistence effect only which means that user v_i persists on his/her own preference without change while when $\alpha = 1$, the overall effect correspond to the influenced-by effect only which means that user v_i changes his/her preferences over a product totally based on his/her friends' preferences over the same product.

Next, we discuss these two types of effect in detail.

Consider a user v_i and his/her preference over a product a_j, i.e., $p_{i,j}$. Recall that the *influence strengths* of v_i's in-coming friends to v_i correspond to the elements in the ith column of the adjacency matrix $\mathbf{A}$, i.e., $\mathbf{A}[h][i]$ for $h = 1, 2, \ldots, n$, and the *preferences* of v_i's in-coming friends over a_j correspond to the elements in the jth column of the preference matrix $\mathbf{P}$, i.e., $\mathbf{P}[h][j]$ for $h = 1, 2, \ldots, n$.

- **Influenced-by effect:** We use the linear combination of v_i's friends' preferences over a_j weighted by their *relative* influence strengths on v_i for capturing the influenced-by effect. That is, the influenced-by effect $IE(v_i, a_j)$ is measured by $\sum_{1 \le h \le n} \frac{\mathbf{A}[h][i]}{indeg(v_i)} \cdot \mathbf{P}[h][j]$. Note that $\mathbf{A}[h][i]$ represents the influence strength from v_h to v_i.
- **Persistence effect:** We capture the persistence effect by v_i's *current* preference over a_j. That is, the persistence effect $PE(v_i, a_j)$ is measured by $\mathbf{P}[i][j]$.

2.1.2. All users on all products

Next, we describe the preference propagation of all users on all products.

Influenced-by effect: Let $\mathcal{IE}$ be the $n \times m$ matrix where the entry at the ith row and the jth column in this matrix is $IE(v_i, a_j)$ where $i \in [1, n]$ and $j \in [1, m]$. Note that $\mathcal{IE}$ contains the influenced-by effects on all users' influences over all products. In the following, we show that $\mathcal{IE}$ can be computed easily by a simple matrix multiplication operation between two matrices. We define the *influenced-by matrix* denoted by $\mathbf{M}_{ib}$ to be $\mathbf{NA}^T$ where $\mathbf{N}$ is an $n \times n$ diagonal matrix with $\mathbf{N}[h][h] = \frac{1}{indeg(v_h)}$ for $h = 1, 2, \ldots, n$ and $\mathbf{A}^T$ is the transposed matrix of the adjacency matrix $\mathbf{A}$. Note that $\mathbf{M}_{ib}$ is a *right stochastic* matrix[2]. For example, Fig. 1(c) shows the influenced-by matrix based on the graph shown in Fig. 1(a).

Lemma 1. $\mathcal{IE} = \mathbf{M}_{ib}\mathbf{P}$

Persistence effect: Let $\mathcal{PE}$ be the $n \times m$ matrix where the entry at the ith row and the jth column in this matrix is $PE(v_i, a_j)$ where $i \in [1, n]$ and $j \in [1, m]$. Note that $\mathcal{PE}$ contains the persistence effects on all users' influences over all products. In the following, we show that $\mathcal{PE}$ can be computed easily by a simple matrix multiplication operation between two matrices. We define the *persistence matrix* denoted by $\mathbf{M}_p$ to be an $n \times n$ identity matrix. Note that $\mathbf{M}_p$ is also a right stochastic matrix.

Lemma 2. $\mathcal{PE} = \mathbf{M}_p\mathbf{P}$

Overall effect: Let $\mathcal{OE}$ be the $n \times m$ matrix where the entry at the ith row and the jth column in this matrix is $OE(v_i, a_j)$ where $i \in [1, n]$ and $j \in [1, m]$. Since $\mathcal{OE} = \alpha \cdot \mathcal{IE} + (1 - \alpha) \cdot \mathcal{PE}$ (by Eq. (1)), with the help of the above lemmas, $\mathcal{OE}$ is equal to

$$\alpha \cdot \mathbf{M}_{ib}\mathbf{P} + (1 - \alpha) \cdot \mathbf{M}_p\mathbf{P} = (\alpha \cdot \mathbf{M}_{ib} + (1 - \alpha) \cdot \mathbf{M}_p)\mathbf{P} \tag{2}$$

Note that $(\alpha \cdot \mathbf{M}_{ib} + (1 - \alpha) \cdot \mathbf{M}_p)$ is a right stochastic matrix.

Let $\mathbf{P}_0$ be the preference matrix corresponding to users' preferences over all products at the beginning (i.e., $\mathbf{P}_0$ is the initial preference matrix). Let

$$\mathbf{M} = \alpha \cdot \mathbf{M}_{ib} + (1 - \alpha) \cdot \mathbf{M}_p \tag{3}$$

Note that $\mathbf{M}$ is a right stochastic matrix. For example, the matrix $\mathbf{M}$ based on the graph shown in Fig. 1(a) with the setting of $\alpha = 0.5$ is shown in Fig. 1(d).

We denote by $\mathbf{P}(n)$ the preference matrix at the end of nth propagation step. By modeling each propagation step as a DeGroot's like iteration, the preference matrix at the current step is equal to the matrix $\mathbf{M}$ times that at the previous step. It follows that the preference matrix at the end of nth step, i.e., $\mathbf{P}(n)$, is equal to M^n times that before the first step which is $\mathbf{P}_0$, i.e., $\mathbf{P}(n) = \mathbf{M}^n \cdot \mathbf{P}_0$. As could be noticed, $\mathbf{P}(n)$ corresponds to a matrix which we call the *update matrix* and denote by $\mathbf{K}(n)$ times $\mathbf{P}_0$, i.e., $\mathbf{P}(n) = \mathbf{K}(n) \cdot \mathbf{P}_0$ where $\mathbf{K}(n) = \mathbf{M}^n$. In this paper, motivated by the phenomenon that the preference matrix after a propagation step may be affected by those after *each* previous propagation step (but not that after the last propagation step only), we model update matrix $\mathbf{K}(n)$ as follows.

$$\mathbf{K}(n) = \begin{cases} \mathbf{M}^n + \delta \cdot \mathbf{K}(n-1) & n \ge 1 \\ [[1, 0, \ldots, 0], \ldots, [0, 0, \ldots, 1]] & n = 0 \end{cases} \tag{4}$$

where the part $\mathbf{M}^n$ (in the case of $n \ge 1$) captures the update rule of a DeGroot's like iteration as adopted in existing studies and the part $\delta \cdot \mathbf{K}(n-1)$ (in the case of $n \ge 1$) which involves a decaying factor $\delta \in [0, 1]$ and also a recursion $\mathbf{K}(n-1)$ captures the update rule that all previous preference matrices that affected $\mathbf{P}(n-1)$ would affect $\mathbf{P}(n)$ with a decaying factor of δ. The decaying model which corresponds to an exponentially weighted function is borrowed from some social science research [30] which captures that the more recent the preference matrices are, the more important they are. We note that the same decaying effect has been adopted in some existing models such as the *Voter Model* [31], but to the best of our knowledge, it is the first time that this decaying effect is adopted in a preference diffusion model.

[2] A right stochastic matrix is a square matrix of non-negative real numbers, with each row summing to 1.

2.2. Convergence and consensus

In this part, we present two interesting features of our preference propagation model, namely *convergence* and *consensus*.

Convergence means that after many rounds of communication/propagation, users' preferences over products stay unchanged (or change only insignificantly), and this is often the case in real life. For example, people usually get clear minds/preferences over the products after enough communication with their friends, which thus will not be changed (i.e., converged). Our model guarantees that users' preferences over products converge at the end of the propagation, and we present this result in the following lemma.

Lemma 3. *If* $\alpha \neq 1$, $\lim_{N\to\infty} \mathbf{P}(N)$ *exists.*

We note here that the convergence result in this paper cannot be trivially derived from the existing results (e.g., given that $\mathbf{P}_N$ converges which is the result in [26], it is not obvious whether $\mathbf{P}(N)$ converges).

Consensus of preferences among a group of users means that at the end of the propagation process, all users in the group have exactly the same preferences over the products, and this also happens often in real life. For example, a group of users who have similar interests might reach a consensus of the preferences over the products. Within our model, when the preference matrix, i.e., $\mathbf{P}(N)$ converges, a consensus is reached within each *closed* SCC (a SCC is said to be *closed* if there exist no incoming edges from vertices outside the SCC), i.e., all users in a closed SCC have exactly the same preference. We present this consensus result in the following Lemma 4.

Lemma 4. *When the preference matrix, i.e.,* $\mathbf{P}(N)$ *converges, within each closed SCC (strongly connected component), all users have the same preferences over the products.*

3. Preference diffusion optimization

Suppose that besides the m existing products (i.e., a_1, a_2, …, a_m), we have a new product which we denote by a_{m+1}. Also, we assume that the propagation process based on users' preferences over the existing products (i.e., a_1, a_2, …, a_m) has been finished. As a result, we only need to consider the propagation process on users' preferences over the new product a_{m+1}.

Since product a_{m+1} is new, at the very beginning, users' preferences over this product are all 0's. Suppose that we have a budget which allows us to target k users for the product a_{m+1}. Here, a user is targeted for a_{m+1} means that the user has her preference over a_{m+1} changed from 0 to 1, which follows the strategy used in [32]. That is, in the case that we select a set S of seeds for a_{m+1}, we have that the preferences of these seeds over a_{m+1} are all 1's while the preferences of all other users (non-seeds) are all 0's. Specifically, Suppose $S = \{v_{s_1}, v_{s_2}, \ldots, v_{s_k}\}$ $(k \leq n)$, and let $\mathbf{P}_0^S$ denotes the initial preference matrix representing users' initial preferences over a_{m+1} with the seed set as S. Note that $\mathbf{P}_0^S$ has the size of $n \times 1$ and hence we use it as a vector in the following. Then, we have $\mathbf{P}_0^S[i] = 1$ for $i \in \{s_1, s_2, \ldots, s_k\}$ and $\mathbf{P}_0^S[i] = 0$ for those $1 \leq i \leq n$ but not in $\{s_1, s_2, \ldots, s_k\}$.

In the following, we study the *preference maximization* (PM) problem in Section 3.1 and the *adoption maximization* (AM) problem in Section 3.2.

Algorithm 1 *Top-k*

Input: Social network: $G(V, E, W)$; an integer $k \leq n$
Output: a seed set S_{pm}
1: compute $\mathbf{M}_c$
2: **for** $j : 1 \to n$ **do**
3: $\quad sum_j \leftarrow \sum_{1\leq i\leq n} \mathbf{M}_c[i][j]$
4: $S_{pm} \leftarrow \{v_{h'} | sum_{h'}$ is among the top-k in $\{sum_h | 1 \leq h \leq n\}\}$
5: **return** S_{pm}

3.1. Preference maximization

The preference maximization (PM) problem is to target k users (or selecting k users as seeds) for a_{m+1} such that after the propagation process (i.e., when the preference matrix converges), the sum of users' preferences over a_{m+1} is maximized. We formalize the problem as follows.

Problem 1 (*Preference Maximization (PM)*)**.** Given a social network $G(V, E, W)$ and a positive integer $k \leq n$, the **preference maximization** (PM) problem is to find a set S of k seeds such that at the end of the propagation process based on the initial preference matrix $\mathbf{P}_0^S$, the sum of users' preferences over the product is maximized. □

Suppose that S is the set of seeds. Then, the preference matrix at the end of the propagation process which we denote by $\mathbf{P}^S$ corresponds to $\lim_{N\to\infty} \mathbf{K}(N) \cdot \mathbf{P}_0^S = \lim_{N\to\infty} \sum_{0\leq h\leq N} \delta^h \cdot \mathbf{M}^{N-h} \cdot \mathbf{P}_0^S$ (Eq. (4)). According to Lemma 3, we know that $\lim_{N\to\infty} \sum_{0\leq h\leq N} \delta^h \cdot \mathbf{M}^{N-h} \cdot \mathbf{P}_0^S$ exists. It follows that $\lim_{N\to\infty} \sum_{0\leq h\leq N} \delta^h \cdot \mathbf{M}^{N-h}$ also exists (by contradiction). Therefore, we let $\mathbf{M}_c$ be $\lim_{N\to\infty} \sum_{0\leq h\leq N} \delta^h \cdot \mathbf{M}^{N-h}$, and thus we have

$$\mathbf{P}^S = \mathbf{M}_c \mathbf{P}_0^S \tag{5}$$

Let $S = \{v_{s_1}, v_{s_2}, \ldots, v_{s_k}\}$ $(k \leq n)$ be a set of k seeds. According to Eq. (5), we know that

$$\begin{aligned} \sum_{1\leq i\leq n} \mathbf{P}^S[i] &= \sum_{1\leq i\leq n} \sum_{1\leq j\leq n} \mathbf{M}_c[i][j] \cdot \mathbf{P}_0^S[j] \\ &= \sum_{1\leq j\leq n} \mathbf{P}_0^S[j] \cdot \Big(\sum_{1\leq i\leq n} \mathbf{M}_c[i][j]\Big) \\ &= \sum_{j\in\{s_1,s_2,\ldots,s_k\}} \sum_{1\leq i\leq n} \mathbf{M}_c[i][j] \end{aligned} \tag{6}$$

According to Eq. (6), we can select a set S_{pm} of seeds as follows which maximizes the resulting sum of users' preferences over the product. We first set S_{pm} to be $\emptyset$ and then add v_j into S_{pm} if jth column of $\mathbf{M}_c$ is among the top-k in terms of the sum of entries in the column (note that $\sum_{1\leq i\leq n} \mathbf{M}_c[i][j]$ is exactly equal to the sum of the entries in jth column of $\mathbf{M}_c$). We call this algorithm *Top-k* and present it in Algorithm 1.

3.2. Adoption maximization

The PM problem maximizes users' *absolute* preferences over the product a_{m+1}, which, however, does not optimize the users' adoption behaviors directly for the new product a_{m+1}. This is because a user who has a larger preference over a_{m+1} does not necessarily mean that she would adopt a_{m+1} with a larger probability since she might also have large preference over the other products. A better way to capture the probability that a user adopts a product is to use her *relative* preference over a_{m+1} to her overall preferences over all products (including both the

existing products and a_{m+1}). To illustrate, consider that there is one case where a user has his/her preference over a product to target equal to 1 and her preferences over other products equal to 1 as well and another case where the user has his/her preference over the product equal to 0.8 and those over other products equal to 0.2. If the objective of the PM problem is used, the former case would be favored by a company since the preference over the target product is larger, which, however, does not capture the real picture since in the latter case, the user has the majority of his/her preference on the target product which naturally implies that he/she would adopt the target product with a higher probability than that in the former case. Motivated by this, in this paper, we model the probability that a user adopts a product to be the ratio between her preference over a target product and her overall preferences over all products (under the same product type). Specifically, let p_i be the sum of the user v_i's preferences over the existing products and Pr_i^S be the probability that user v_i adopts the new product a_{m+1} at the end of the propagation process on users' preferences over a_{m+1} with S as the seed set. Then, we have

$$Pr_i^S = \mathbf{P}^S[i]/(p_i + \mathbf{P}^S[i]) \tag{7}$$

For example, if a user v_i has his/her preference over a_{m+1} equal to 0.2 after the diffusion process of the preference over a_{m+1} (i.e., $\mathbf{P}^S[i] = 0.2$) and the sum of his/her preferences over all existing products equal to 1.8 (i.e., $p_i = 1.8$), the probability that user v_i would adopt the new product is equal to $0.2/(1.8+0.2) = 0.1$.

With the view of the above discussion, we propose a problem called *adoption maximization* (AM) which is to find a set S of k seeds such that at the end of propagation process of users' preferences over the new product a_{m+1}, the sum of users' probabilities to adopt a_{m+1} is maximized. Note that using different sets of seeds, we have different users' preferences over a_{m+1} and thus different sums of users' probabilities to adopt a_{m+1}. In this paper, we define a function $\sigma(\cdot)$ such that it takes a set S of users as input and returns the sum of users' probabilities to adopt a_{m+1} at the end of the propagation process on users' preferences over a_{m+1} when the users in S are targeted as the seeds for a_{m+1}, i.e.,

$$\sigma(S) = \sum_{i=1}^{n} Pr_i^S \tag{8}$$

The formal definition of the AM problem is provided in Problem 2.

Problem 2 (*Adoption Maximization (AM)*)**.** Given a social network $G(V, E, W)$, the sum of user v_i's preferences over the existing m products, p_i, for $i = 1, 2, \ldots, n$, and an integer k, the **adoption maximization** (AM) problem is to find a set S of k seeds for a new product a_{m+1} such that at the end of the propagation process on users' preferences over a_{m+1}, the sum of all users' probabilities to adopt a_{m+1}, i.e., $\sigma(S)$, is maximized. □

Compared to the PM problem which allows a simple exact solution, the AM problem is more difficult to tackle. In this paper, we prove that with the matrix $\mathbf{M}_c$ (recall that matrix $\mathbf{M}_c$ is defined as $\lim_{N\to\infty} \sum_{0\le h\le N} \delta^h \cdot \mathbf{M}^{N-h}$) given arbitrarily, the AM problem is NP-hard.

Lemma 5 (*NP-hardness*)**.** *The AM problem with the matrix* $\mathbf{M}_c$ *given arbitrarily is NP-hard.*

Motivated by the NP-hardness result as presented in Lemma 5, in this paper, we design an approximate algorithm called *Greedy* for the AM problem, which we describe as follows. *Greedy* is a greedy algorithm involving k steps.

Algorithm 2 *Greedy*

Input: Social network: $G(V, E, W)$; user v_i's sum of preferences over the existing products, p_i for $i = 1, 2, \ldots, n$; an integer $k \le n$

Output: a seed set S_{am}

1: $S_{am} \leftarrow \emptyset$
2: **for** $i : 1 \to k$ **do**
3: $v^* \leftarrow \arg\max_{v\in V\setminus S_{am}} g(v|S_{am})$ where $g(v|S_{am})$ means the marginal gain of adoption when adding v into S_{am}, i.e., $g(v|S_{am}) = \sigma(S_{am} \cup \{v\}) - \sigma(S_{am})$
4: $S_{am} \leftarrow S_{am} \cup \{v^*\}$
5: **return** S_{am}

Let S_{am} be the set of seeds outputted by *Greedy*. *Greedy* first initializes S_{am} to be $\emptyset$ and then proceeds with k steps. At each step, it selects the user which incurs the greatest marginal gain in terms of the $\sigma(\cdot)$ function when selected as a new seed and includes it into S_{am}. At the end, it outputs S_{am}. As could be noticed, *Greedy* is a greedy algorithm w.r.t. the $\sigma(\cdot)$ function. The pseudo-code of *Greedy* is presented in Algorithm 2.

Interestingly, *Greedy* could provide a $(1 - 1/e)$-factor approximation for the AM problem where e is the natural logarithmic base.

Lemma 6. *Greedy provides a* $(1-1/e)$*-factor approximation for the AM problem.*

The approximation results here are based on the fact that function $\sigma(\cdot)$ is *submodular*.[3] A stronger property is that function $\sigma(\cdot)$ is submodular even the diffusion process stops after a certain number of iterations of propagation without reaching a convergence. Details of proof could be found in the Appendix.

4. Related work

In the literature, several models have been proposed for capturing the preference diffusion process [26–29]. These models share the following ideas: (1) users' preferences are captured by real numbers and (2) the diffusion process proceeds with many iterations and at each iteration, each user's preference is updated based on his/her own preference and his/her friends' preferences (more specifically, it is updated to be a *linear combination* of these preferences with appropriate weights). Different models use different strategies of updating a user's preference, e.g., in the model [26] which is called the *DeGroot* model, it is updated to be a linear combination of his/her preference *at the current iteration* and his/her friends' preferences (*at the current iteration*), in the model [27,28] which is called the *opinion formation* (OF) model, it is updated to be a linear combination of his/her preference *before the first iteration* and the *average* of his/her friends' preferences (*at the current iteration*), and in the model [29] which is called the *Lou's model* (LM), it is updated exactly the same as in [26] but with an additional normalization procedure. The model proposed in this paper differs from these models by adopting a new strategy for computing users' final preferences, i.e., instead of using users' preferences at the *last* iteration only, it uses users' preferences at *all* iterations with the *decaying effects* employed.

Under the OF model, an optimization problem called *CAMPAIGN* was studied [32], which is to find k individuals in the social network for a given integer k such that once the values

[3] Let U be a universe set. Function $f : 2^U \to \mathbb{R}$ is said to be *submodular* iff given any two sets X and Y where $Y \subseteq X \subset U$, $\forall e \in U - Y, f(X \cup \{e\}) - f(X) \le f(Y \cup \{e\}) - f(Y)$ [33].

of the expressed opinion of these individuals are set to be the maximum, the average of the values of the expressed opinion of all other individuals is maximized. Note that the PM problem, which is to maximize the *absolute* preferences, is very similar to the CAMPAIGN problem though, they are based on different models, and as a consequence, they have different tractability results (i.e., CAMPAIGN is NP-hard [32] and PM is polynomially-solvable). The AM problem, which is to maximizes the *relative* preferences, is totally new and more interesting (since a user has a large (absolute) preference over a product does not imply that the user would adopt the product with a high probability (due to several products with high absolute preferences), and instead, the probability is better captured by the user' relative preference over the product).

A closely related work is *information diffusion* which has been studied extensively [1–25]. Several models have been proposed for capturing the information diffusion process, and among them, the *independent cascade* (IC) model [1] and the *linear threshold* (LT) model [2] are most commonly used. Two types of optimization problems of information diffusion, namely *influence maximization* (IM) [4,6,9–19,25,34,35] and *seed minimization* (SM) (also called the *target set selection* problem) [20,36–41] were studied. Both problems are usually NP-hard, and many approximate algorithms have been developed.

Some variants of the IM problem have also been studied which include [42] where each user in the social network has a location and the problem is to select k seeds such that after the information diffusion process based on these seeds, the total amount of the awareness gained by the users located in a given spatial query region is maximized, [43] where each user has a set of attributes/keywords and the problem is to select k seeds such that after the information diffusion process based on these seeds, the total amount of the awareness gained by the users containing those given attributes/keywords is maximized, [44] where each user has a skill set and the problem is to select k seeds such that the seeds together cover a given set of required skills and after the information diffusion process based on these seeds, the total amount of the awareness gained by the users is maximized, and [25] where the budget for seeding could be allocated *fractionally* to different users and a user could be influenced *partially* and the problem is to allocate limited budget of seeding to some users such that the incurred influence is maximized.

Some other related work is reviewed as follows. In [23], the authors studied the cascading process of "socware" (e.g., malware and spam), in [24], the authors aimed to answer the question of whether the information cascade/diffusion process could be predicted and showed some positive evidence, in [45], the authors studied the problem of recommending connections among nodes in order to boost the information diffusion process, and in [21], the authors studied the role of social networks in the information diffusion process.

5. Empirical study

5.1. Experimental setup

We conducted experiments on a 3 GHz machine with 32 GB memory under a Linux platform. All algorithms were implemented in C++.

5.1.1. Datasets

Following [29,32], we used two real datasets in our experiments, namely HEP-TH (in the period of 1992–2003) and DBLP (in the period of 2002–2014). The rationale of using scientific citation networks for our experiments is as follows. First, in academia, citations of papers imply some of authors' interests/preferences on which venues they tend to read/cite papers from and/or publish their papers at. Second, collaborations among researchers correspond naturally to a phenomenon of communication among users, as a result of which, authors' preferences are diffused from one to another. For the HEP-TH dataset, we constructed a corresponding social network as follows. For each author who has at least 4 publications either in the period of 1993–1996 or in the period of 1997–2000, we generated a node in the social network. For each pair of (remaining) authors who co-authored in the period of 1993–2003, we created an edge between the two corresponding nodes in the social network. Totally, there are 1516 nodes and 10,738 edges. For the DBLP dataset, we constructed a corresponding social network as follows. For each author who has the number of publications from 15 to 45 and at least 10 co-authors, we generated a node in social network. For each pair of two co-authors, we created an edge. Totally, there are 2001 nodes and 14,486 edges. Each publisher/venue was regarded as a product in our problem setting. We chose the top m publishers which were cited the most in the period of 1993–1996 for HEP-TH (and 2009–2011 for DBLP).

For the HEP-TH dataset, each author's preference on a publisher x is initialized to be the total number of times that his/her papers that were published in 1993–1996 cited papers published by publisher x divided by the total number of times that his/her papers that were published in 1993–1996 cited a paper published by any publisher. For the DBLP dataset, each author's preference on a publisher x is initialized to be the total number of his/her publications that were published by publisher x divided by the total number of his/her publications published by any publisher in 2009–2011.

5.1.2. Types of experiments

There are three parts in the experiments. The first part is to study the quality of the proposed preference diffusion model proposed in this paper by predicting what is the probability an author will cite a paper from a given publication venue, given their past citation preferences and the co-authorship social network. We consider two existing preference diffusion models, namely the opinion formation (OF) model [27] and the Lou's Model (LM) [29]. Note that we did not consider the DeGroot model since the DeGroot model is abstract in the sense that the weights used for the linear combination are not specified and the Lou's model is one instance of the DeGroot model. Following [29], we also consider four other models, namely the Independent Cascade (IC) model [1,46], the Linear Threshold (LT) model [2,3], the PageRank model [47], the DiffusionRank model [48]. For the IC model and the LT model, the diffusion process stops when there is no additional node activated. For both the IC model and the LT model, we conducted experiments 20 times and then used the *fraction* that a user is activated as the user's preference over a product. For the PageRank method and the DiffusionRank method, we set the parameter values as in [48]. The LM model [29] can be considered as the state-of-the-art model for preference diffusion. There is one parameter in the LM model [29], namely the *susceptible ratio* parameter $\in [0, 1]$. We varied this parameter value from 0 to 1 and set this value to the one with the best performance.

Following [29], we evaluated the "goodness" of each diffusion model with two measurements, namely Jaccard coefficient and Kendall's Tau coefficient. Each of these two measurements is an indicator measuring the similarity between the users' preferences on products generated by a diffusion model and the *ground-truth* users' preferences on products. In this paper, the information in the period of 1997–1999 for HEP-TH (2012–2014 for DBLP) are regarded as ground truth. Specifically, for each author in HEP-TH, his/her *ground-truth* preference on a publisher x in year $y \in$

{1997, 1998, 1999} is equal to the total number of times his/her papers that were published in year y cited papers published by publisher x divided by the total number of times his/her papers that were published in year y cited a paper published by any publisher. For each author in DBLP, his/her *ground-truth* preference on a publisher x in year $y \in \{2012, 2013, 2014\}$ is equal to the total number of his/her publishers that were published by publisher x in year y divided by the total number of his/her publishers that were published in year y by any publisher.

Let p_{ij} be the preference of user i on product j generated by a diffusion model and g_{ij} be the ground-truth preference of user i on product j. Jaccard coefficient is defined to be the proportion of the user-product pairs (i, j) such that $|p_{ij} - g_{ij}| \leq 0.05$ where $i \in [1, n]$ and $j \in [1, m]$. Note that the larger the Jaccard coefficient is, the better the model is. In Kendall's Tau coefficient, initially, a variable is initialized to 0. Then, for any pair of entries (x, y) in the preference matrix returned by the diffusion model, we can find the corresponding pair of entries (x', y') in the ground-truth matrix. The variable is updated based on different cases. If either ($x < y$ and $x' < y'$) or ($x > y$ and $x' > y'$), then the variable is incremented by 1. If either ($x < y$ and $x' > y'$) or ($x > y$ and $x' < y'$), then the variable is decremented by 1. The variable is updated for each possible pair of entries in the preference matrix. Finally, the Kendall's Tau coefficient is equal to the variable divided by the total number of possible pairs in the matrix. Same as the Jaccard coefficient, the larger the Kendall's Tau coefficient is, the better the model is.

Ideally, the iterations of the diffusion model should be tied with some temporal information such as the publication dates of papers, e.g., some propagation should be incurred (locally) when a new paper is published. Nevertheless, in the datasets we used for experiments, the temporal information is not available quite often in its raw form, and because of this, existing studies [29] assumed that the iterations happen all-at-once for each year. To make the comparison against existing studies more consistent, we adopt a similar strategy but with a small difference in this paper. Specifically, we define a parameter r such that in each year, r iterations of propagations are performed, to take into account that iterations may happen multiple times during a one-year time interval.

The second part is to study the efficiency of the proposed *Top-k* algorithm (i.e., Algorithm 1) for the preference maximization problem by comparing it against some baseline methods in terms of preference and running time. Following existing studies [4], we compared our proposed method with three baseline methods, namely *Degree-heuristic*, *Centrality-heuristic* and *Random*. Specifically, we use each of methods including our *Top-k* algorithm and also the baseline methods to select a set of seeds and perform the preference diffusion process starting from the set of seeds based on the proposed diffusion model in this paper, and then measure the resulting preference spread and also the time used to select the set of seeds.

The third part is to study the effectiveness and also the efficiency of the proposed *Greedy* algorithm (i.e., Algorithm 2) for the adoption maximization problem by comparing it against some baseline methods in terms of adoption and running time, respectively. Similar to the case of the second part, we use each of methods including our *Greedy* algorithm and also the baseline methods to select a set of seeds and perform the preference diffusion process starting from the set of seeds based on the proposed diffusion model in this paper, and then measure the resulting adoption and also the time used to select the set of seeds.

5.1.3. Parameters

We conducted experiments by varying some parameter values. We varied the number of publishers (m) which was chosen from {10, 12, 14, 16}. We varied the parameter r from 4, 6, 8, 10 for HEP-TH (and 1, 2, 3, 4 for DBLP). Note that the initial preferences of users on products were constructed in the period of 1993–1996, and the ground-truth preferences were constructed in one of the years in {1997, 1998, 1999} for HEP-TH (and {2012, 2013, 2014} for DBLP). When we compared the preferences generated by our model with the ground-truth preferences in year $x \in \{1997, 1998, 1999\}$ for HEP-TH (and $x \in \{2012, 2013, 2014\}$ for DBLP), the number of iterations involved in our model is equal to $r \cdot (x - 1996)$ for HEP-TH (and $r \cdot (x - 2011)$ for DBLP). There are two parameters in our proposed model, namely α and δ. In our experiments, we varied parameter α from 0 to 1 and varied parameter δ from 0 to 1.

We conducted experiments with the following default parameter values: $m = 16$, $r = 5$, $\alpha = 0.5$ and $\delta = 0.25$. By default, we adopt the preference matrix $\mathbf{P}(N)$ such that the sum of the difference between the entries of $\mathbf{P}(N)$ and the corresponding entries of $\mathbf{P}(N-1)$ is smaller than 10^{-3}. In our experiments, N is equal to 105.

5.2. Experimental results

In this part, we present the results for the aforementioned three parts of experiment (in this part, we show the results on the HEP-TH dataset only, and results on the DBLP show similar clues and could be found in the appendix).

5.2.1. Part 1: Diffusion model comparison

We study the performance of our diffusion model in this section.

Diffusion model comparison: Fig. 2 shows that our proposed diffusion model gives the greatest Jaccard coefficient and the LM model gives the second greatest one. The worst model is the LT model. This could probably be explained by the fact that our model captures several phenomena like "continuous preference", "reversible propagation", "repetitive propagation" and "memory-based decay" which are not captured all by any of the competitors of our model. We notice that the Jaccard coefficient of each model increases slightly with r. This may be explained by the fact that with more products (i.e., more information), users/authors need more iterations (or communications) to decide to cite a paper published by a publisher, resulting in more accurate results. When n increases, the Jaccard coefficient of each model increases in general since there is more preference information captured, resulting in more accurate results. The results when Kendalls's Tau coefficient is used are similar (Fig. 3).

Varying forgetting coefficient (δ): We vary the forgetting coefficient where the default value of m used is 10. Fig. 4(a) shows that the greatest Jaccard coefficient of the proposed model with $\alpha = 0.25$ is obtained when δ is equal to 0.75. Fig. 4(b) shows that the greatest Kendall's Tau coefficient of the proposed model with $\alpha = 0.75$ is obtained when δ is equal to 0.75.

Varying persistence effects (α): We vary the persistence effect where the default value of m used is 10. Fig. 5(a) shows that the greatest Jaccard coefficient of the proposed model with $\delta = 0.25, 0.5$ or 1 is obtained when α is in the range of 0.85–0.95. Fig. 5(b) shows that the greatest Kendall's Tau coefficient of the proposed model with $\delta = 0.25, 0.5$ or 0.75 is obtained when α is in the range of 0.75–0.95.

Studies on convergence: We studied the convergence of our proposed model. Fig. 6 shows that the difference between the

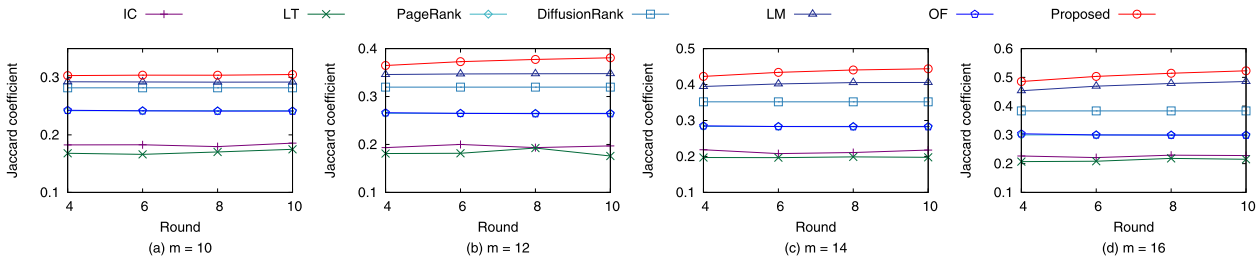


Fig. 2. The average of Jaccard coefficient for 1997, 1998 and 1999 against the number of iterations per year (r).

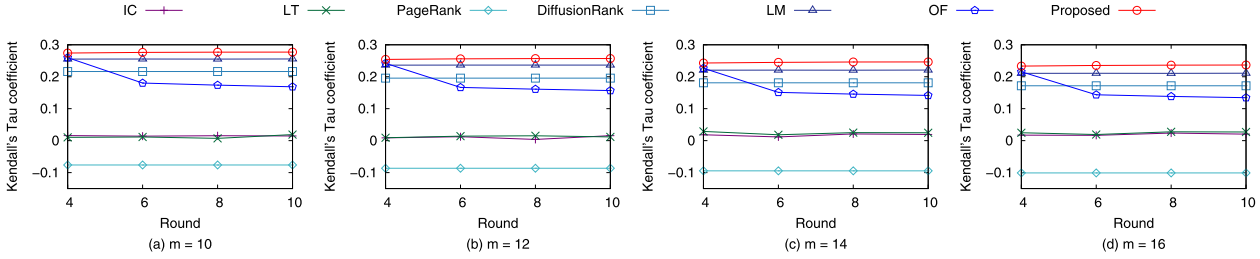


Fig. 3. The average of Kendall's Tau coefficient for 1997, 1998 and 1999 against the number of iterations per year (r).

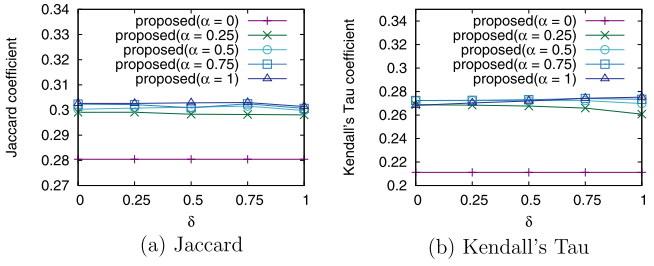


Fig. 4. The average scores for 1997, 1998 and 1999 against δ.

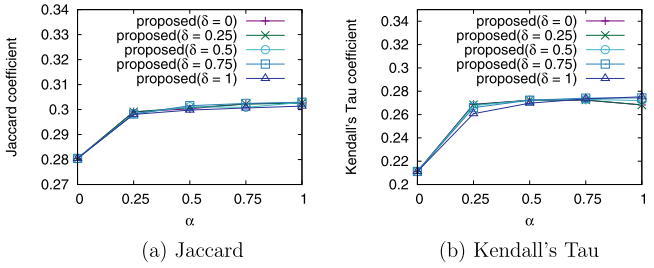


Fig. 5. The average scores for 1997, 1998 and 1999 against α.

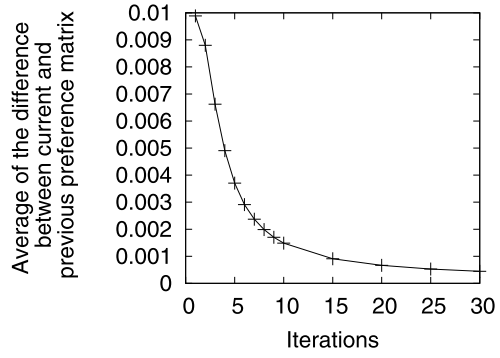


Fig. 6. Convergence (Mean Absolute Error (MAE) is used).

preference matrix in the current iteration and the preference matrix in the previous iteration decreases dramatically for the first few iterations and then slowly for the following iterations. Besides, it could be noted that the convergence speed of the model is high, e.g., it converges within 30 iterations.

Studies on consensus: We studied the consensus property in our proposed model. Fig. 7 shows a case study from the dataset. In this case study, we show a SCC with 3 authors. The author name is shown in each node in the figure. There are 3 sub-figures in Fig. 7 showing 3 different instances of preference diffusion at 3 different time stages. It shows that the users' probabilities to adopt a publisher reach consensus within a closed SCC after 14 iterations. We note here that the consensus results presented here are based on a case where many iterations of propagation were performed on a network assumed to stay unchanged. In practice, networks are always dynamic (e.g., papers are published every year and new collaborations are being formed in each year), which would prevent it from reaching a consensus within a closed SCC easily.

5.2.2. Part 2: Preference maximization

As shown in Fig. 8, our *Top-k* algorithm achieves the best results in terms of the sum of preferences. Besides, *Top-k* runs faster than *Centrality-heuristic*, but slower than *Degree-heuristic* and *Random*.

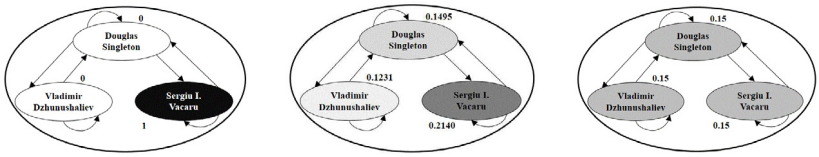
5.2.3. Part 3: Adoption maximization

As shown in Fig. 9, our *Greedy* algorithm returns a seed set with the best quality, but it also runs the slowest.

We note here that as the first attempt of tackling the problem, our algorithms have some efficiency issues this is similar to the case of influence diffusion where the first attempt is a slow algorithm [4] and then faster algorithms were developed (e.g., the algorithms in [9–11]). In the future, we plan to develop more efficient algorithms, and one possible way is to use approximate algorithms instead of exact ones for the matrix manipulations.

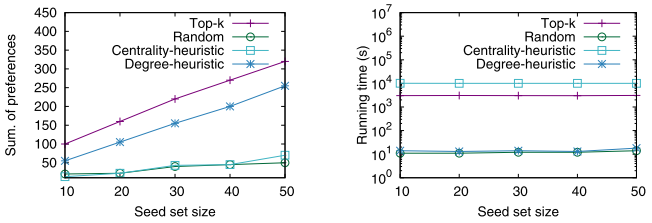
6. Conclusion

In this paper, we proposed a new model for modeling the preference diffusion process on social networks, which guarantees



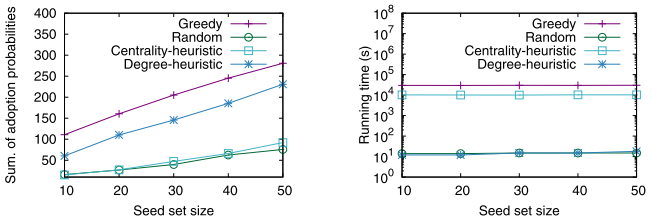
(a) $i = 0$ (b) $i = 4$ (c) $i = 14$

Fig. 7. The users' probabilities to adopt a publisher at ith iteration which are (a) at the beginning (b) at the middle stage and (c) at the end of the diffusion process.



(a) Preference (b) Running time

Fig. 8. Preference maximization (the preference of a user on a product was computed by simulating the proposed diffusion model in this paper with an sufficient number of iterations of propagation).



(a) Adoption (b) Running time

Fig. 9. Adoption maximization (the adoption of a user on a product corresponds to his/her relative preference over the product where the preference was computed by simulating the proposed diffusion model in this paper with an enough number of iterations of propagation).

convergence and gives some consensus. Based on the proposed model, we studied two problems, namely PM and AM, which optimizes the preference diffusion process w.r.t. two different objectives by selecting proper seed sets. We developed an exact algorithm *Top-k* for the PM problem, proved the NP-hardness of the AM problem and designed an $(1 - 1/e)$-factor approximate algorithm *Greedy* for the AM problem. Extensive experiments on real datasets were conducted which verified our proposed model and algorithms. One interesting research direction is to study the problem of selecting a smallest set of seeds such that the sum of users' preferences over the new product (and/or probabilities to adopt the new product) is at least a pre-set goal.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research of Raymond Chi-Wing Wong is supported by IRS17EG25.

Appendix A. Proof of Lemma 1

It could be verified that the entry at the ith row and the jth column of $\mathbf{M}_{ib}\mathbf{P}$ which is an $n \times m$ matrix is equal to $\sum_{1\le h\le n} \mathbf{M}_{ib}[i][h]\cdot\mathbf{P}[h][j] = \frac{1}{indeg(v_i)}\cdot\sum_{1\le h\le n}\mathbf{A}^T[i][h]\cdot\mathbf{P}[h][j] = \sum_{1\le h\le n}\frac{\mathbf{A}[h][i]}{indeg(v_i)}\cdot\mathbf{P}[h][j]$, and thus it captures the influenced-by effects on v_i's preference over product a_j.

Appendix B. Proof of Lemma 2

Clearly, the entry at the ith row and the jth column of $\mathbf{M}_p\mathbf{P}$ which is an $n \times m$ matrix is equal to $1 \cdot \mathbf{P}[i][j] = \mathbf{P}[i][j]$, and thus it captures the persistence effect on v_i's preference over a_j.

Appendix C. Proof of Lemma 3

Before we proceed, we introduce two concepts first, namely *closeness* (of a set of vertices in a graph) and *aperiodicity* (of a graph or the corresponding adjacency matrix). Given a graph $G(V, E, W)$, a set $S \subseteq V$ of vertices in G is said to be *closed* if there exists no such an edge which goes from a vertex in $V\backslash S$ to a vertex in S. In other words, a set is closed if it does not have any incoming edges from outside to this set. We say that a set which is closed is a closed set. For example, in the graph shown in Fig. 1(a), set $S = \{v_1, v_2, v_3\}$ is closed since there exist no incoming edges from outside S to S. A sub-structure of a graph (or its corresponding adjacency matrix) is said to be *aperiodic* if the greatest common divisor of the lengths of the simple cycles in this sub-structure is equal to 1. For example, in the graph as shown in Fig. 1(a), the sub-graph reduced by the set $S = \{v_1, v_2, v_3\}$ is aperiodic since all cycles in this sub-graph have the lengths of 2 or 3 which have the greatest common divisor equal to 1.

First, we prove that $\mathbf{M}^N$ converges. According to [49], $\mathbf{M}$ converges if and only if the following condition is satisfied: each SCC (strongly connected component) in the graph with the adjacency matrix as $\mathbf{M}$ that is closed is aperiodic. In our case, the above condition is satisfied which we explain as follows. Recall that $\mathbf{M} = \alpha \cdot \mathbf{M}_{ib} + (1 - \alpha) \cdot \mathbf{M}_p$. Thus, in the graph which has the adjacency matrix the same as $\mathbf{M}$, there is a self-loop on each vertex (this is because $\mathbf{M}_p$ is an identity matrix and $\alpha \neq 1$), and this further implies that each SCC that is closed is aperiodic (this is because each SCC has at least a self-loop as a simple cycle of length equal to 1). We denote the entry at the ith row and the jth column of $\lim_{N\to\infty}\mathbf{M}^N$ by $\pi_{i,j}$, i.e., $\pi_{i,j} = (\lim_{N\to\infty}\mathbf{M}^N)[i][j]$.

Second, we present three limits as follows which will be used in our proof.

- **Limit 1:** $(\lim_{N\to\infty}\mathbf{M}^N)[i][j] = \pi_{i,j}$ for $1 \le i, j \le n$ (this is based on the definition of $\pi_{i,j}$). That is, $\forall\epsilon' > 0$, $\exists T_1 \in \mathbb{Z}^+$ such that for any $N > T_1$, we have $\max_{1\le i,j\le n}|\mathbf{M}^N[i][j] - \pi_{i,j}| < \epsilon'$.
- **Limit 2:** $\lim_{N\to\infty}\sum_{h=0}^{N}\delta^h = \frac{1}{1-\delta}$ (this is simply because $\sum_{h=0}^{N}\delta^h = \frac{1-\delta^{N+1}}{1-\delta}$). That is, $\forall\epsilon' > 0$, $\exists\, T_2 \in \mathbb{Z}^+$ such that for any $N > T_2$, we have $|\sum_{h=0}^{N}\delta^h - \frac{1}{1-\delta}| < \epsilon'$.

- **Limit 3:** $\lim_{N\to\infty}\sum_{h=N}^{\infty}\delta^h = 0$ (this is simply because $\sum_{h=N}^{\infty}\delta^h = \delta^N\cdot\sum_{h=0}^{\infty}\delta^h = \frac{\delta^N}{1-\delta}$). That is, $\forall\epsilon' > 0,\ \exists T_3 \in \mathbb{Z}^+$ such that for any $N > T_3$, we have $\sum_{h=N}^{\infty}\delta^h < \epsilon'$.

Third, we further prove that $\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j]$ converges to $\frac{\pi_{i,j}}{1-\delta}$ as follows. Let $T = \max\{T_1 + T_3, T_2\}$. Consider the difference between $\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j]$ and $\frac{\pi_{i,j}}{1-\delta}$, i.e., $|\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j] - \frac{\pi_{i,j}}{1-\delta}|$, for any $1 \le i, j \le n$. For any $N > T$, we have

$$
\begin{aligned}
&|\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j] - \frac{\pi_{i,j}}{1-\delta}| \\
&= |\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j] - \sum_{h=0}^{N}\delta^h\pi_{i,j} + \sum_{h=0}^{N}\delta^h\pi_{i,j} - \frac{\pi_{i,j}}{1-\delta}| \\
&\le \sum_{h=0}^{N}\delta^h|(\mathbf{M}^{N-h}[i][j] - \pi_{i,j})| + \pi_{i,j}|\sum_{h=0}^{N}\delta^h - \frac{1}{1-\delta}| \\
&\le \sum_{h=0}^{T_3}\delta^h|(\mathbf{M}^{N-h}[i][j] - \pi_{i,j})| + \sum_{h=T_3+1}^{N}\delta^h|(\mathbf{M}^{N-h}[i][j] - \pi_{i,j})| \\
&+ \pi_{i,j}|\sum_{h=0}^{N}\delta^h - \frac{1}{1-\delta}| \qquad \text{(C.1)} \\
&\le \sum_{h=0}^{T_3}\delta^h\epsilon' + \sum_{h=T_3+1}^{N}\delta^h + \pi_{i,j}\epsilon' \qquad \text{(C.2)} \\
&\le \frac{1}{1-\delta}\epsilon' + \epsilon' + \pi_{i,j}\epsilon' = (\frac{1}{1-\delta} + 1 + \pi_{i,j})\epsilon' \qquad \text{(C.3)}
\end{aligned}
$$

The deduction from (C.1) to (C.2) is because (a) $|\mathbf{M}^{N-h}[i][j] - \pi_{i,j}| \le \epsilon'$ based on the results of Limit 1 (note that $N - h$'s for $h = 0, 1, \ldots, T_3$ are all larger than T_1 since $N - h > T - h \ge T_1 + T_3 - h \ge T_1$), (b) $|(\mathbf{M}^{N-h}[i][j] - \pi_{i,j})|$ is bounded by 1 (note that $\mathbf{M}^{N-h}[i][j] \in [0, 1]$ and $\pi_{i,j} = (\lim_{N\to\infty}\mathbf{M}^N)[i][j] \in [0, 1]$ since $\mathbf{M}$ is right-stochastic as discussed previously), and (c) $|\sum_{h=0}^{N}\delta^h - \frac{1}{1-\delta}| \le \epsilon'$ based on the results of Limit 2. The deduction from (C.2) to (C.3) is because $\sum_{h=0}^{T_3}\delta^h < \sum_{h=0}^{\infty}\delta^h = \frac{1}{1-\delta}$ and $\sum_{h=T_3}^{N}\delta^h < \epsilon'$ based on the results of Limit 3.

For any $\epsilon > 0$, we let $\epsilon' = \frac{\epsilon}{(\frac{1}{1-\delta}+1+\pi_{i,j})}$. According to Eq. (C.3), for any $N > T$, we have $|\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j] - \frac{\pi_{i,j}}{1-\delta}| \le \epsilon$ for any $1 \le i, j \le n$ which implies that

$$\lim_{N\to\infty}\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}[i][j] = \frac{\pi_{i,j}}{1-\delta} \qquad \text{(C.4)}$$

Fourth, based on the previous result, we know that $\lim_{N\to\infty}\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}$ exists and

$$\lim_{N\to\infty}\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h} = \frac{\lim_{N\to\infty}\mathbf{M}^N}{1-\delta} \qquad \text{(C.5)}$$

Fifth, we prove that $\lim_{N\to\infty}\mathbf{P}(N)$ exists which is simply because $\lim_{N\to\infty}\mathbf{P}(N) = \lim_{N\to\infty}\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}\mathbf{P}_0$ and $\lim_{N\to\infty}\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}$ exists.

Appendix D. Proof of Lemma 4

First, according to [50], in $\lim_{N\to\infty}\mathbf{M}^N\mathbf{P}_0$, a SCC that is closed would reach a consensus among all users in the SCC. Second, according to Eq. (C.5) in the proof of Lemma 3, we know that $\lim_{N\to\infty}\mathbf{P}(N) = \lim_{N\to\infty}\sum_{h=0}^{N}\delta^h\mathbf{M}^{N-h}\mathbf{P}_0 = \frac{\lim_{N\to\infty}\mathbf{M}^N}{1-\delta}\mathbf{P}_0 = \frac{\lim_{N\to\infty}\mathbf{M}^N\mathbf{P}_0}{1-\delta}$, and thus each closed SCC reaches a consensus of the preferences over the products among all users in the SCC.

Appendix E. Proof of Lemma 5

We prove the lemma by transforming the traditional *Exact Cover by 3-Sets* (EC3S) problem which is NP-hard to our AM problem.

First, we give the decision problem of EC3S, also denoted by EC3S for simplicity, as follows. Given a universe set $U = \{e_1, e_2, \ldots, e_{3q}\}$ (q is a positive integer) and a collection $\mathcal{C} = \{E_1, E_2, \ldots, E_l\}$ where $E_i \subseteq U$ and $|E_i| = 3$ (E_i is called a 3-set) for $i = 1, 2, \ldots, l$ (l is a positive integer), the problem is to decide whether there exists a subset $\mathcal{T}$ of $\mathcal{C}$ such that $\mathcal{T}$ is an exact cover over U (i.e., $|\mathcal{T}| = q$ and $U \subseteq \cup_{E\in\mathcal{T}}E$).

Second, we give the decision problem of the AM problem (with the matrix $\mathbf{M}_c$ already known) as follows. Given a social network involving n users $(v_1, v_2, \ldots, v_n)$, the sum of user v_i's preferences over existing products p_i for $i = 1, 2, \ldots, n$, a matrix $\mathbf{M}_c$ of size $n \times n$, an integer k and a real value P, the problem is to decide whether there exists a set S of k seeds such that $\sigma(S) \ge P$.

Third, we describe the process of transforming an arbitrary EC3S problem instance to an AM problem instance as follows. We set n to be $\max\{3q, l\}$ and construct n users, namely v_1, v_2, ..., and v_n. We set p_i to be 1 for $i = 1, 2, \ldots, n$, k to be q, and P to be $3q/2$. We construct $\mathbf{M}_c$ as a $n \times n$ matrix and set its entries as follows. First, we initialize all entries of $\mathbf{M}_c$ to be 0's. Second, for each $i \in [1, 3q]$ and each $j \in [1, l]$ satisfying $e_i \in E_j$, we update $\mathbf{M}_c[i][j]$ to be 1. Note that in the case of $l > 3q$ (thus, $n = l$), the last $l - 3q$ rows contain all 0's.

Fourth, we show the equivalence between the EC3S problem instance and its corresponding AM problem instance by two cases: $l \le 3q$ (Case 1) and $l > 3q$ (Case 2). Before we proceed, we introduce a property first.

Property 1. *Let $\mathcal{T} = \{E_{s_1}, E_{s_2}, \ldots, E_{s_d}\}$ ($d \in [1, l]$) be a subset of $\mathcal{C}$ in the EX3S problem instance and S be the set containing the s_1th user, the s_2th user, ..., and the s_dth user in the AM problem instance, i.e., $S = \{v_{s_1}, v_{s_2}, \ldots, v_{s_d}\}$. Then, $\mathbf{P}^S[i] = c$ if and only if element e_i is covered by $\mathcal{T}$ c times.* □

Proof. The property could be easily verified as follows.

$$
\begin{aligned}
\mathbf{P}^S[i] &= \sum_{j=1}^{n}\mathbf{M}_c[i][j]\cdot\mathbf{P}_0^S[j] = \sum_{j:e_i\in E_j}1\cdot\mathbf{P}_0^S[j] \\
&= \sum_{j:e_i\in E_j\wedge v_j\in S}1\cdot 1 = \sum_{j:e_i\in E_j\wedge E_j\in\mathcal{T}}1 = c \qquad \text{(E.1)}
\end{aligned}
$$

The first equation is by definition of $\mathbf{P}^S[i]$, the second equation is because $\mathbf{M}_c[i][j] = 1$ for those j's satisfying $e_i \in E_j$ and $\mathbf{M}_c[i][j] = 0$ for all other j's, the third equation is because $\mathbf{P}^S[j] = 1$ for those j's satisfying $v_j \in S$ and $\mathbf{P}^S[j] = 0$ for all other j's, and the fourth equation is because $v_j \in S$ is equivalent to $E_j \in \mathcal{T}$, and the fifth equation is by definition of c.

Consider Case 1. In this case, $n = 3q$. Consider "EC3S ⇒ AM". Suppose $\mathcal{T} = \{E_{s_1}, E_{s_2}, \ldots, E_{s_q}\}$ is an exact cover of U. Consider the set S containing the s_1th user, the s_2th user, ..., and the s_qth user, i.e., $S = \{v_{s_1}, v_{s_2}, \ldots, v_{s_q}\}$, as the seed set. According to Property 1, we know that $\mathbf{P}^S[i] = 1$ for $i = 1, 2, \ldots, n$ (note that $n = 3q$). As a result, we know that the probability that user v_i adopts the new product, i.e., Pr_i^S, is equal to $\mathbf{P}^S[i]/(\mathbf{P}^S[i] + p_i) = 1/(1 + 1) = 1/2$. Therefore, the sum of the probabilities that the users would adopt the new product, i.e., $\sigma(S)$, is equal to $\sum_{i=1}^{n}Pr_i^S = n/2 = 3q/2$.

Consider "AM ⇒ EC3S". Suppose $S = \{v_{s_1}, v_{s_2}, \ldots, v_{s_q}\}$ is a set of seeds such that $\sigma(S) \ge 3q/2$. Consider the sum of users' preferences over the new product. According to Eq. (7), we know

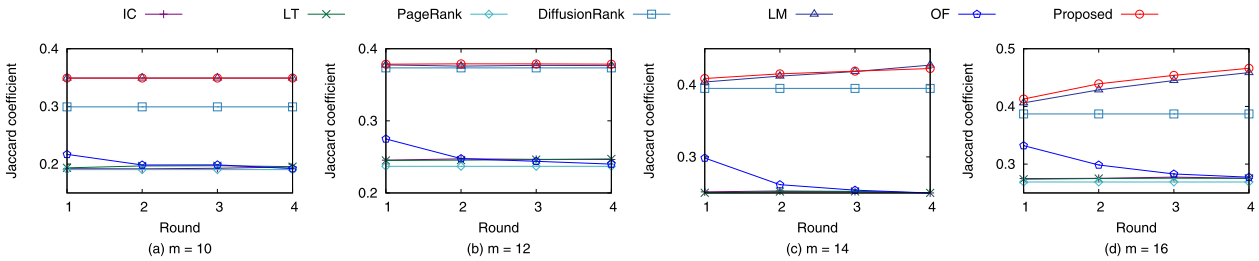


Fig. G.10. The average of Jaccard coefficient for 2012, 2013 and 2014 against the number of iterations per year (r).

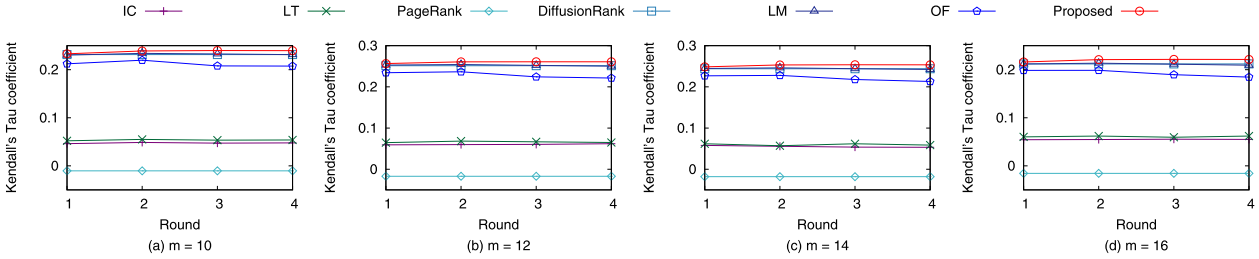


Fig. G.11. The average of Kendall's Tau coefficient for 2012, 2013 and 2014 against the number of iterations per year (r).

$$\sum_{1\le i\le n}\mathbf{P}^S[i] = \sum_{j\in\{s_1,s_2,\ldots,s_q\}}\sum_{1\le i\le n}\mathbf{M}_c[i][j] = \sum_{j\in\{s_1,s_2,\ldots,s_q\}} 3 = 3q \tag{E.2}$$

Then, we deduce that $\mathbf{P}^S[i] = 1$ for $i = 1, 2, \ldots, n$ since otherwise $\sigma(S) \ge 3q/2$ and $\sum_{1\le i\le n}\mathbf{P}^S[i] = 3q$ cannot be true simultaneously (this could be verified by considering the problem of allocating the sum of preferences equal to $3q$ to the n users such that the sum of the probabilities that the users adopt the new product is at least $3q/2$, and due to page limit, we omit the details here). Consider $\mathcal{T}$ which is $\{E_{s_1}, E_{s_2}, \ldots, E_{s_q}\}$. According to Property 1, we know that $\mathcal{T}$ corresponds to an exact cover over U.

Consider Case 2. In this case, $n = l$. We can show the equivalence similarly as we did for Case 1 except that $\mathbf{P}^S[i] = 0$ for $i \in [3q+1, l]$.

Appendix F. Proof of Lemma 6

We first show that function $\sigma(\cdot)$ is *submodular* as follows.

Let $T \subset V$ be a set of users and S be a subset of T, i.e., we have $S \subseteq T \subset V$. Let v_h be a user that is not in T, i.e., $v_h \in V\backslash T$. We also let $T' = T \cup \{v_h\}$ and $S' = S \cup \{v_h\}$. Consider $\mathbf{P}^S[i]$. According to Eq. (5), we have

$$\begin{cases}\mathbf{P}^S[i] = \sum_{1\le j\le n}\mathbf{M}_c[i][j]\cdot\mathbf{P}_0^S[j] = \sum_{j:v_j\in S}\mathbf{M}_c[i][j]\\ \mathbf{P}^{S'}[i] = \sum_{1\le j\le n}\mathbf{M}_c[i][j]\cdot\mathbf{P}_0^{S'}[j] = \sum_{j:v_j\in S'}\mathbf{M}_c[i][j]\end{cases}$$

As a result, we have

$$\begin{cases}\mathbf{P}^{S'}[i] - \mathbf{P}^S[i] = \sum_{j:v_j\in S'}\mathbf{M}_c[i][j] - \sum_{j:v_j\in S}\mathbf{M}_c[i][j]\\ = \sum_{j:v_j\in S'\backslash S}\mathbf{M}_c[i][j] = \mathbf{M}_c[i][h] \quad (a)\\ \mathbf{P}^{T'}[i] - \mathbf{P}^T[i] = \sum_{j:v_j\in T'}\mathbf{M}_c[i][j] - \sum_{j:v_j\in T}\mathbf{M}_c[i][j]\\ = \sum_{j:v_j\in T'\backslash T}\mathbf{M}_c[i][j] = \mathbf{M}_c[i][h] \quad (b)\end{cases} \tag{F.1}$$

Let $b_i = \mathbf{P}^{S'}[i]$, $c_i = \mathbf{P}^S[i]$, and $d_i = \mathbf{M}_c[i][h]$ for $i = 1, 2, .., n$. According to Eq. (F.1(a)), we know that $b_i - c_i = d_i$ for $i = 1, 2, \ldots, n$. Then, we have

$$\begin{aligned}&\sigma(S\cup\{v_h\}) - \sigma(S) = \sigma(S') - \sigma(S)\\ &= \sum_{1\le i\le n}\frac{b_i}{p_i+b_i} - \sum_{1\le i\le n}\frac{c_i}{p_i+c_i} = \sum_{1\le i\le n}\left(\frac{b_i}{p_i+b_i} - \frac{c_i}{p_i+c_i}\right)\\ &= \sum_{1\le i\le n}\left(\frac{c_i+d_i}{p_i+c_i+d_i} - \frac{c_i}{p_i+c_i}\right) = \sum_{1\le i\le n}\frac{d_i\cdot p_i}{(p_i+c_i+d_i)\cdot(p_i+c_i)}\end{aligned}$$

Similarly, let $b_i' = \mathbf{P}^{T'}[i]$ and $c_i' = \mathbf{P}^T[i]$ for $i = 1, 2, .., n$. According to Eq. (F.1(b)), we know that $b_i' - c_i' = d_i$ for $i = 1, 2, \ldots, n$. Then, we have

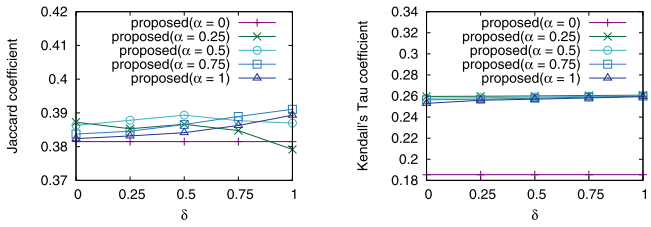
$$\sigma(T\cup\{v_h\}) - \sigma(T) = \sum_{1\le i\le n}\frac{d_i\cdot p_i}{(p_i+c_i'+d_i)\cdot(p_i+c_i')}$$

As a result, we have

$$\begin{aligned}&\sigma(T\cup\{v_h\}) - \sigma(T) - (\sigma(S\cup\{v_h\}) - \sigma(S))\\ &= \sum_{1\le i\le n}\frac{d_i\cdot p_i}{(p_i+c_i'+d_i)\cdot(p_i+c_i')} - \sum_{1\le i\le n}\frac{d_i\cdot p_i}{(p_i+c_i+d_i)\cdot(p_i+c_i)}\\ &= \sum_{1\le i\le n}\left(\frac{d_i\cdot p_i}{(p_i+c_i'+d_i)\cdot(p_i+c_i')} - \frac{d_i\cdot p_i}{(p_i+c_i+d_i)\cdot(p_i+c_i)}\right)\\ &\le 0\end{aligned} \tag{F.2}$$

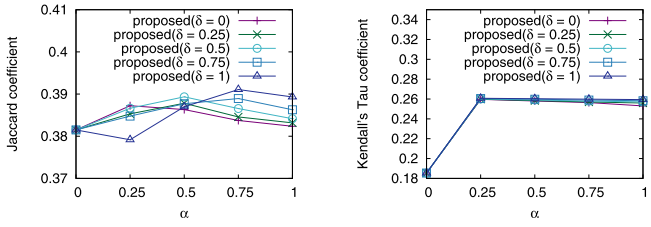
Note that the last inequality above is true since $\mathbf{P}^T[i] \ge \mathbf{P}^S[i]$, i.e., $c_i' \ge c_i$ (this is simply because $S \subseteq T$). Eq. (F.2) implies that function $\sigma(\cdot)$ is submodular. It is also worth mentioning that the proof above is valid for an arbitrary matrix $\mathbf{M}_c$, which further implies that function $\sigma(\cdot)$ is submodular even if the diffusion process stops after a certain number of iterations of propagation without reaching a convergence.

By using that fact that the *Greedy* algorithm is simply a greedy algorithm based on the function $\sigma(\cdot)$ under a cardinality constraint (i.e., $|S| \le k$) and the fact that a simply greedy algorithm provides a $(1-1/e)$-factor approximation for maximizing a submodular function under a cardinality constraint [33], we know that *Greedy* provides a $(1-1/e)$-factor approximation for the AM problem.



(a) Jaccard (b) Kendall's Tau

Fig. G.12. The average scores for 2012, 2013 and 2014 against δ.



(a) Jaccard (b) Kendall's Tau

Fig. G.13. The average scores for 2012, 2013 and 2014 against α.

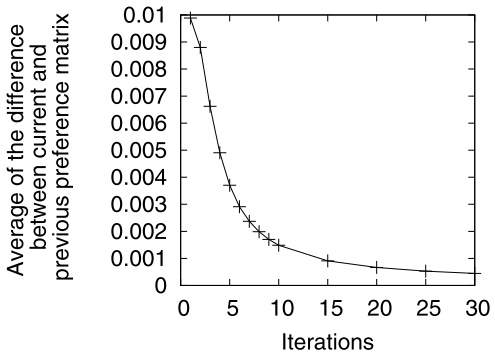


Fig. G.14. Convergence (Mean Absolute Error (MAE) is used).

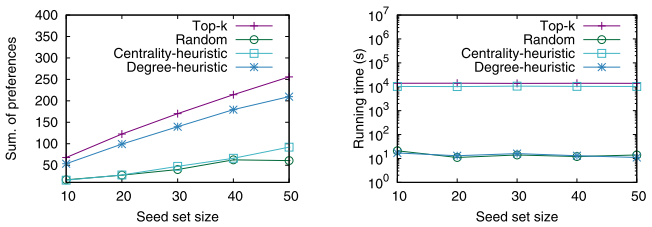
Appendix G. Experimental results on the DBLP datasets

In this part, we present the results for the aforementioned three parts of experiment on the DBLP dataset. We used the following default parameters: m = 12, r = 2, α = 0.25 and δ = 0.25.

G.1. Part 1: Diffusion model comparison

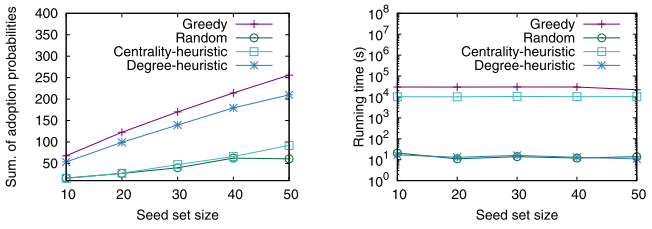
We study the performance of our diffusion model in this section.

Diffusion model comparison: Fig. G.10 shows that our proposed diffusion model gives the greatest Jaccard coefficient. The results



(a) Preference (b) Running time

Fig. G.16. Preference maximization (the preference of a user on a product was computed by simulating the proposed diffusion model in this paper with an enough number of iterations of propagation).



(a) Adoption (b) Running time

Fig. G.17. Adoption maximization (the adoption of a user on a product corresponds to his/her relative preference over the product where the preference was computed by simulating the proposed diffusion model in this paper with an enough number of iterations of propagation).

when Kendalls's Tau coefficient is used are similar and could be found in Fig. G.11.

Varying forgetting coefficient (δ): We vary the forgetting coefficient where the default value of m used is 12. Fig. G.12(a) shows that the greatest Jaccard coefficient of the proposed model with $\alpha = 0.25$ is obtained when δ is equal to 0.5. Fig. G.12(b) shows that the greatest Kendall's Tau coefficient of the proposed model with $\alpha = 0.25$, 0.5 or 0.75 is obtained when δ is equal to 0.75.

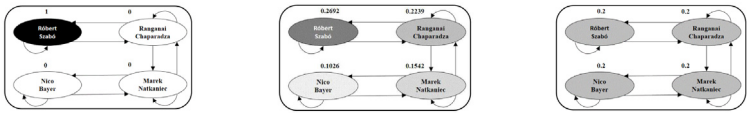
Varying persistence effects (α): We vary the persistence effect where the default value of m used is 12. Fig. G.13(a) shows that the greatest Jaccard coefficient of the proposed model with $\delta = 0.25$ or 0.5 is obtained when α is equal to 0.5. Fig. G.13(b) shows that the greatest Kendall's Tau coefficient of the proposed model with $\delta = 0.25$, 0.5 or 0.75 is obtained when α is equal to 0.25.

Studies on convergence:

The results are shown in Fig. G.14, and same as the results on the HEP-TH dataset, the results show that the difference between the preference matrix in the current iteration and the preference matrix in the previous iteration decreases dramatically for the first few iterations and then slowly for the following iterations.

Studies on consensus:

Fig. G.15 shows a case study from the dataset of DBLP. In this case study, we show a SCC with 4 authors. The author name is shown in each node in the figure. There are 3 sub-figures in



(a) Before Preference Diffusion (b) Just after 10 iterations (c) Just after 90 iterations

Fig. G.15. The users' probabilities to adopt a publisher at 3 different stages which are (a) at the beginning (b) at the middle stage and (c) at the end of the diffusion process.

Fig. G.15 showing 3 different instances of preference diffusion at 3 different time stages. It shows that the users' probabilities to adopt a publisher reach consensus within a closed SCC after 90 iterations.

G.2. Part 2: Preference maximization

The results are shown in Fig. G.16 which are similar to those on the HEP-TH dataset.

G.3. Part 3: Adoption maximization

The results are shown in Fig. G.17 which are similar to those on the HEP-TH dataset.

References

[1] J. Goldenberg, B. Libai, E. Muller, Talk of the network: A complex systems look at the underlying process of word-of-mouth, Mark. Lett. 12 (3) (2001) 211–223.
[2] M. Granovetter, Threshold models of collective behavior, Am. J. Sociol. 83 (6) (1978) 1420–1443.
[3] T.C. Schelling, Micromotives and Macrobehavior, WW Norton and Company, 2006.
[4] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 137–146, 2003.
[5] D. Gruhl, R. Guha, D. Liben-Nowell, A. Tomkins, Information diffusion through blogspace, in: Proceedings of the 13th International Conference on World Wide Web, ACM, 491–501, 2004.
[6] M. Kimura, K. Saito, Tractable models for information diffusion in social networks, PKDD (2006) 259–271.
[7] H. Ma, H. Yang, M.R. Lyu, I. King, Mining social networks using heat diffusion processes for marketing candidates selection, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, ACM, 233–242, 2008.
[8] B. Ryan, N.C. Gross, The diffusion of hybrid seed corn in two iowa communities, Rural Sociol. 8 (1) (1943) 15–24.
[9] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: SIGKDD, 2009, pp. 199–208.
[10] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 1029–1038, 2010.
[11] W. Chen, Y. Yuan, L. Zhang, Scalable influence maximization in social networks under the linear threshold model, in: ICDM, IEEE, 88–97, 2010.
[12] Y. Wang, G. Cong, G. Song, K. Xie, Community-based greedy algorithm for mining top-k influential nodes in mobile social networks, in: SIGKDD, ACM, 1039–1048, 2010.
[13] R. Narayanam, Y. Narahari, A shapley value-based approach to discover influential nodes in social networks, IEEE Trans. Autom. Sci. Eng. 8 (1) (2011) 130–147.
[14] A. Goyal, W. Lu, L.V. Lakshmanan, Simpath: An efficient algorithm for influence maximization under the linear threshold model, in: 2011 IEEE 11th International Conference on Data Mining (ICDM), IEEE, 211–220, 2011.
[15] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, K. Xie, Simulated annealing based influence maximization in social networks, in: AAAI, 2011, pp. 127–132.
[16] A. Goyal, F. Bonchi, L.V. Lakshmanan, A data-based approach to social influence maximization, VLDB (2011) 73–84.
[17] K. Jung, W. Heo, W. Chen, Irie: Scalable and robust influence maximization in social networks, in: ICDM, IEEE, 918–923, 2012.
[18] Y. Li, W. Chen, Y. Wang, Z.-L. Zhang, Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships, in: WSDM, ACM, 2013, pp. 657–666.
[19] C. Borgs, M. Brautbar, J. Chayes, B. Lucier, Influence maximization in social networks: Towards an optimal algorithmic solution, in: arXiv preprint arXiv:1212.0884, 2012.
[20] N. Chen, On the approximability of influence in social networks, in: SODA, 2008, pp. 1029–1037.
[21] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, The role of social networks in information diffusion, in: WWW, ACM, 2012, pp. 519–528.
[22] W. Chen, L.V. Lakshmanan, C. Castillo, Information and influence propagation in social networks, Synth. Lect. Data Manage. 5 (4) (2013) 1–177.
[23] T.-K. Huang, M.S. Rahman, H.V. Madhyastha, M. Faloutsos, B. Ribeiro, An analysis of socware cascades in online social networks, in: WWW, 2013, pp. 619–630.
[24] J. Cheng, L. Adamic, P.A. Dow, J.M. Kleinberg, J. Leskovec, Can cascades be predicted?, in: WWW, 2014, pp. 925–936.
[25] E.D. Demaine, M. Hajiaghayi, H. Mahini, D.L. Malec, S. Raghavan, A. Sawant, M. Zadimoghadam, How to influence people with partial incentives, in: WWW, 2014, pp. 937–948.
[26] M.H. DeGroot, Reaching a consensus, J. Amer. Statist. Assoc. 69 (345) (1974) 118–121.
[27] N.E. Friedkin, E.C. Johnsen, Social influence and opinions, J. Math. Sociol. 15 (3–4) (1990) 193–206.
[28] N.E. Friedkin, E.C. Johnsen, Social influence networks and opinion change, Adv. Group Process. 16 (1) (1999) 1–29.
[29] J.-K. Lou, F.-M. Wang, C.-H. Tsai, S.-C. Hung, P.-H. Kung, S.-D. Lin, Modeling the diffusion of preferences on social networks, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM, 605–613, 2013.
[30] H. Ebbinghaus, Memory: A Contribution to Experimental Psychology, No. 3, University Microfilms, 1913.
[31] M. Kimura, K. Saito, K. Ohara, H. Motoda, Opinion formation by voter model with temporal decay dynamics, Mach. Learn. Knowl. Discov. Databases (2012) 565–580.
[32] A. Gionis, E. Terzi, P. Tsaparas, Opinion maximization in social networks, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM, 387–395, 2013.
[33] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions-I, Math. Program. 14 (1) (1978) 265–294.
[34] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: SIGKDD, 2007, pp. 420–429.
[35] Y. Tang, X. Xiao, Y. Shi, Influence maximization: Near-optimal time complexity meets practical efficiency, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 75–86.
[36] O. Ben-Zwi, D. Hermelin, D. Lokshtanov, I. Newman, An exact almost optimal algorithm for target set selection in social networks, in: Proceedings of the 10th ACM Conference on Electronic Commerce, ACM, 2009, pp. 355–362.
[37] D. Reichman, New bounds for contagious sets, in: Discrete Mathematics, 2012.
[38] P. Shakarian, D. Paulo, Large social networks can be targeted for viral marketing with small seed sets, in: arXiv preprint arXiv:1205.4431, 2012.
[39] A. Goyal, F. Bonchi, L.V.S. Lakshmanan, S. Venkatasubramanian, On minimizing budget and time in influence propagation over social networks, Soc. Netw. Anal. Min. (2012) 1–14.
[40] C. Long, R.C.W. Wong, Minimizing seed set for viral marketing, in: ICDM, IEEE, 2011, pp. 427–436.
[41] P. Zhang, W. Chen, X. Sun, Y. Wang, J. Zhang, Minimizing seed set selection with probabilistic coverage guarantee in a social network, in: arXiv preprint arXiv:1402.5516, 2014.
[42] G. Li, S. Chen, J. Feng, K.-l. Tan, W.-s. Li, Efficient location-aware influence maximization, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 87–98.
[43] C. Long, R.C.-W. Wong, Viral marketing for dedicated customers, Inf. Syst. 46 (2014) 1–23.
[44] K. Feng, G. Cong, S.S. Bhowmick, S. Ma, In search of influential event organizers in online social networks, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 63–74.
[45] V. Chaoji, S. Ranu, R. Rastogi, R. Bhatt, Recommendations to boost content spread in social networks, in: WWW, 2012, pp. 529–538.
[46] J. Goldenberg, B. Libai, E. Muller, Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata, Acad. Mark. Sci. Rev. 9 (3) (2001) 1–18.
[47] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: Bringing order to the web. Tech. rep. Stanford InfoLab 1999.
[48] H. Yang, I. King, M.R. Lyu, Diffusionrank: a possible penicillin for web spamming, in: SIGIR, ACM, 2007, pp. 431–438.
[49] B. Golub, M.O. Jackson, Naive learning in social networks and the wisdom of crowds, Am. Econ. J. Microecon. (2010) 112–149.
[50] M.O. Jackson, Social and Economic Networks, Princeton University Press, 2010.