Product Adoption Maximization Leveraging Social Influence and User Interest Mining

Ping Ji[1], Hui Huang[1], Xueliang Liu[2]*, and Xueyou Hu[1]
[1] Hefei University, Hefei 230601, China
[e-mail: {jiping,huihuang,xueyouhu}@hfuu.edu.cn]
[2] Hefei University of Technology, Hefei 230009, China
[e-mail:liuxueliang@hfut.edu.cn]
*Corresponding author: Xueliang Liu

Abstract

A Social Networking Service (SNS) platform provides digital footprints to discover users' interests and track the social diffusion of product adoptions. How to identify a small set of seed users in a SNS who is potential to adopt a new promoting product with high probability, is a key question in social networks. Existing works approached this as a social influence maximization problem. However, these approaches relied heavily on text information for topic modeling and neglected the impact of seed users' relation in the model. To this end, in this paper, we first develop a general product adoption function integrating both users' interest and social influence, where the user interest model relies on historical user behavior and the seed users' evaluations without any text information. Accordingly, we formulate a product adoption maximization problem and prove NP-hardness of this problem. We then design an efficient algorithm to solve this problem. We further devise a method to automatically learn the parameter in the proposed adoption function from users' past behaviors. Finally, experimental results show the soundness of our proposed adoption decision function and the effectiveness of the proposed seed selection method for product adoption maximization.

1. Introduction

With the prevalence of online social media, many SNSs have emerged in recent years[1], [2]. These websites provide online platforms for users to share adoption preferences over products within a social network. Millions of users' digital footprints on these platforms provide unparallel opportunities to discover users' interests (reflected in user-product interaction behavior) and track the social diffusion of products (reflected in user-user social network). Thus, the SNS platforms are becoming new sources for marketing[3]. Specifically, a key task in the marketing area is that: how to identify a small set of customers (i.e., seed users), if they are convinced to adopt a new product with relatively high preference evaluations, would finally trigger the largest number of future adoptions?[4]

Indeed, as the social networks embedded in these SNS platforms, play a fundamental role for users to spread product information to their friends, the social influence effect is well recognized as a main factor for product adoption [5]. To start with, Kemple et al. approached the above task as a social influence maximization problem, which asks that how to select a set of seed users (e.g., by providing them with free samples), such that they could influence the maximum expected number of users [6]. Due to the potential business values of this direction for viral marketing campaigns, lots of efforts have been devoted to it [7]–[10]. Later, the topic-aware influence maximization problem was discussed to consider user interests (which can be represented as topics) and item topics by extending classical influence propagation models [11]–[13]. Nevertheless, in real-world applications, sometimes it is hard to obtain reliable text representation (e.g., music, image) for learning product topics, thus the topic-aware influence models may fail under this circumstance. Meanwhile, in the product adoption maximization process, the selected seed users' different evaluations would have a large impact on the overall users' evaluations of the promoting product [14], [15], which was ignored by the topic-aware social influence models.

In this paper, we study the problem of new product adoption maximization with both user interest and social influence consideration. To overcome the limitations in previous works, we propose to exploit collaborative filtering techniques from seed users' ratings for the new product topic modeling to enable the general setting without any additional text information. Specifically, collaborative filtering provides a technique to discover user interests and product topics from the user-item interaction behavior, and has been successfully deployed in industry for user interest based recommender systems [16]–[18]. By leveraging seed users' evaluations in user interest modeling process, traditional topic-aware social influence models could hardly be applied directly to our problem due to the interdependency and mutual reinforcement relationship between users' interest and the social influence effect. Specifically, on one hand, as these collaborative filtering based interest models rely heavily on seed users' evaluations on the promoting product, different seeding strategies would alter the interest prediction scores of non-seed users. On the other hand, users' predicted interest scores would push them either towards or away from adopting the promoting product, thus change the future cascades of the promoting product with social influence effect.

To fully exploit the interplay between users' interest and social influence, in this paper, we first devise an adoption decision function that organically combines the social influence effect and users' interest score. Given this decision function, two main technical challenges arise in the product adoption maximization process: First, as there are exponential combinations of seed nodes from the user set, how to devise an efficient algorithm for seed user selection? Second, as different parameter setting in our proposed adoption decision function may lead to different predicted adoption results, how to find the right parameter setting and show the

0soundness of our proposed adoption function? To tackle the above challenges, we then prove the product adoption maximization problem is not only NP-hard, but also does not follow the same properties as the social influence maximization problem. Given the hardness, we provide an efficient algorithm to solve it. Furthermore, for the second challenge, we design a method to automatically learn the parameter in the adoption function given historical data. Finally, extensive experimental results on two datasets show the effectiveness of our proposed seed selection algorithm and the soundness of our proposed adoption function.

To summarize, the main contributions of this paper are listed as follows:
- We propose a solution to model the product adoption maximization problem by exploiting the social influence effect and users interest score jointly.
- We provide an efficient algorithm to solve the proposed adoption function and to automatically learn the parameter in the adoption function.

The rest of this paper is organized as follows. In Section 2, we review the related work of social inference and user behavior modeling. In Section 3, we propose our product adoption maximization solution. Section 4 validates the effectiveness of our approach on different datasets, and Section 5 concludes this paper.

2. Related Work

In a SNS platform with users U ($|U| = N$) and products V ($|V| = M$), let $R = [R_{ui}] \in \mathbb{R}^{N \times M}$ and $G = <U, \mathcal{E}, \mathbf{T}>$ denote users' consumption behavior and social link behavior respectively. Here, R_{ui} is the detailed rating preference value if user u consumes product i, otherwise it 0. $\mathcal{E}$ represents the edge connections among users and $\mathbf{T} = [t_{uv}] \in \mathbb{R}^{N \times N}$ denotes the influence transition matrix. If $(v, u) \in \mathcal{E}$ (i.e., v follows u), then $t_{uv} > 0$, otherwise $t_{uv} = 0$. Since learning the non-zero influence strength t_{uv} is beyond the scope of this paper, we follow the classical social influence research by assuming they are already known and usually $\sum_{(v,u) \in \mathcal{E}} t_{uv} \leq 1$ [6], [8], [19].

2.1 Social Influence Maximization.

Several popular influence propagation models have been proposed, e.g., the IC model [20] and the LT model [21]. Under these models, the social influence maximization problem is shown to be NP-hard. Thus, many researchers focused on designing scalable algorithms to solve this problem [6]–[8]. A key finding is that, these influence models are monotone increasing and follow the sub-modularity property, thus a greedy algorithm to select seed nodes performs at least $1 - 1/e$ theoretical guarantee [22]. Meanwhile, to avoid the Monte-Carlo simulations of the classical influence models, recently a simple linear social influence model is proposed to effectively and efficiently calculate influence spread [9], [19]. Chen et al. extended the classical influence maximization problem and considered the scenario when users' negative opinions also propagate in the social diffusion process [15]. Instead of promoting one product for marketing, researchers also proposed to simultaneously promote the product set in a platform to achieve the largest adoptions of all products with product correlation modeling [23]. Recently, the topic-aware influence maximization problem was discussed to consider user interests (which can be represented as topics) and item topics by topic extensions of classical influence propagation models [11]–[13]. Researchers proposed a keyword-based targeted influence maximization to find a seed set that maximizes the expected influence over users that are relevant to a given advertisement, where the topic of users and items are measured by the keywords associated with them [12]. In [32], the authors proposed an event

recommendation system which leverage LDA topic model to compute the similarity of events based on their contents. In [33], the effect of online reviews on product sales is measured by a joint sentiment-topic model to empirically investigates the influence of numerical and textual reviews on product sales performance.

These topic-aware social influence models advance previous works on the topic modeling of products with text or keyword. However, in real-world applications, the promoting product either has few rating records from the social network or is hard to obtain text and keyword representations (e.g., music), making these topic-aware social influence models fail to obtain reliable topic distribution of the promoting item. In this paper, we propose to directly learn the product topics from the seed users' ratings with interest modeling techniques, thus the proposed technique is applicable to any kind of new product promoting problem.

2.2 User Interest Modeling

Given user-product interaction behavior, user interest modeling tries to uncover users' latent preferences from their past behaviors without any side information [16][34]. Among all user interest models, matrix factorization based collaborative filtering models have been widely deployed in industry due to their simplicity and relatively high performance [17], [24]. Specifically, the matrix factorization based models project both users and products into a low latent space by mining user-product rating matrix, thus a target user's interest on an unknown product is measured by the similarity between their latent vectors. With the emergence of social networks, researchers have devoted to embed the social influence theory for more accurate interest modeling and personalized recommendation [25], [26]. In addition, Lu et al. validated that users' future adoption preferences may be altered by the interest modeling algorithm in the system [27]. Goyal et al. studied the problem of using an existing interest model to launch a targeted marketing campaign without social influence consideration [14]. In [35], the authors proposed a user interest modeling approach in micro-blog system based on multi-tag semantic correlation via analyzing tag relationship and the limitations of the existing micro-blog user interest models. In [36], the user interest is modeled by users' activities on social media to exploit the change in one's behaviors over a period by following users' activities on social media over time.

In summary, all these user interest models focus on passively catering to users' interests and recommending products that suit users' previous interests. Nevertheless, the challenge of how to leverage users' interests for product manufactures to *actively* launch a marketing remains pretty much open.

3. The Proposed Model

In this section, we present our proposed model for adoption maximization problem. The overall flowchart of product adoption maximization is shown in **Fig. 1**. To evaluate the estimated product adoption effect with a selected seed set of users, we need to first define the product adoption function. This product adoption function considers both the user interest and social influence effect, and the parameter in this function could be learned by modeling users' historical behavior. After that, in order to predict each user's interest to the promoting product, the user interest part needs to learn the topics of the promoting product from seed users' ratings. We organize this section as follows. We would give an overview of our defined adoption decision function, followed by the adoption maximization problem definition. By proving the hardness of this problem, we provide an efficient algorithm to solve this problem. After that,

we propose a method to automatically learn the parameter in the adoption function.

3.1 Overview of the Adoption Function

Our model is based on the assumption that a user's adoption preference is determined by her/his interest to this product as well as how much she/he is influenced in the social network. Specifically, given a user u, her/his preference for a new promoting product ρ can be determined as a two-step stage: 1) First, she/he decides to choose one factor from her/his own interest and the social influence effect for future decision process. This stage could be represented as a binary hidden variable z, with $z = 1$ denotes the selection of this user's interest and $z = 0$ equals the social influence effect. 2) After the selection of value z, the user future decides whether to adopt the product based on the value on this effect. Thus, the expectation of each user's predicted preference $\hat{R}_{u\rho}$, could be mathematically formulated as:

$$\hat{R}_{u\rho} = \sum_z (p(z|u)\hat{R}_{(}\rho|u,z)) = \alpha \hat{R}_{u\rho}^P + (1 - \alpha)\hat{R}_{u\rho}^S \quad s.t. \quad 0 \leq \alpha \leq 1 \qquad (1)$$

where $\hat{R}_{u\rho}^P$ denotes the *Personal Interest* score and $\hat{R}_{u\rho}^S$ represents the *Social Influence* score. The selection of choosing one factor from two possible choices is characterised as a Bernoulli distribution with a mean value α. The larger the α, the more likely the user interest contributes to the adoption behavior with less social influence effect.
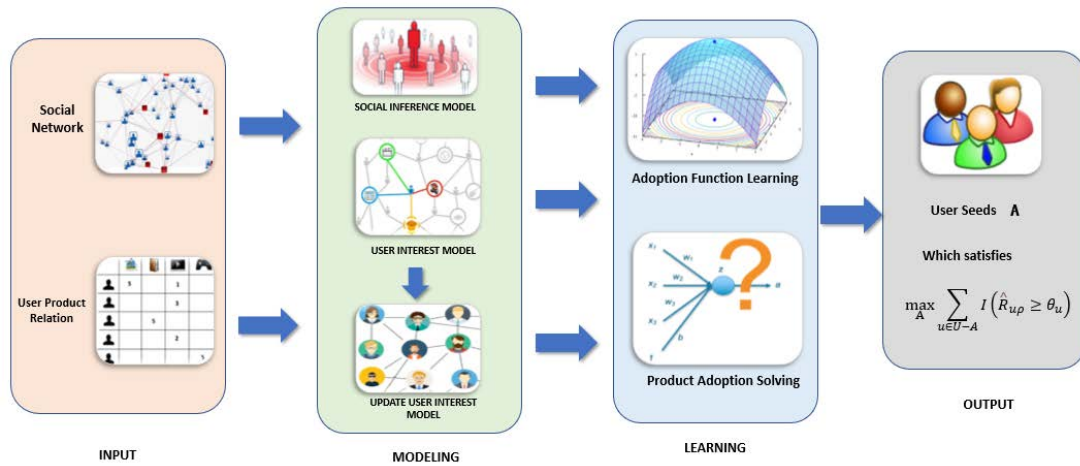

Fig. 1. Overall flowchart of the proposed adoption maximization solution.

3.2 Problem Definition

Given a user's adoption function in Eq. 1, let ρ be a new product that needs to be promoted in the SNS platform, the problem we study in this paper is that: how to select a small seed set of users A with limited budget k ($|A| \leq k$), such that if they give relatively high evaluations of the promoting product, the expected adoption of this product is maximized?

Naturally, in order to calculate the expected adoptions of the promoting product, we need to figure out when each user u would adopt the promoting product. Here, as each user may has her/his own unique preference in adopting products, we give a user-specific threshold θ_u to depict u's threshold for adopting product. In fact, many classical social influence models are based on user-specific thresholds for adoption decisions, e.g., the well known LT model [6], [21] and the linear approach based social influence model [9]. If $\hat{R}_{u\rho} \geq \theta_u$, then u adopts the promoting product. Thus, the expected marketing gains, measured by the expected adoptions,

or the total active users, is:

$$f(A) = \sum_{u \in U-A} I(\hat{R}_{u\rho} \geq \theta_u), \tag{2}$$

where $I(x)$ is an indicator function that equals 1 if x is true and 0 otherwise. We now formally define the adoption maximization problem as:

Product Adoption Maximization Problem Definition Given a budget k, users' rating records R, and the social network G in a SNS platform, how to select a seed set A ($|A| \leq k$), such that the expected total adoptions, i.e., f(A) is maximized?

With the above problem definition, we next describe how to model the user preference score $\hat{R}_{u\rho}^{P}$ and the social influence propagation score $\hat{R}_{u\rho}^{S}$. Please note that, since both the user interest modeling techniques and the social diffusion models are rather mature, we focus on how to extend the classical models to combine the seed users' evaluations for our proposed problem.

User Interest Modeling. In this part, we select *P*robabilistic *M*atrix *F*actorization (PMF), a classical matrix factorization model for user interest modeling due to its effectiveness and efficiency [17]. Specifically, given the rating matrix R, PMF predicts R as a product of user latent matrix $P \subseteq \mathbb{R}^{K \times N}$ and a product latent matrix $Q \subseteq \mathbb{R}^{K \times M}$.

When the seed users' evaluations of the promoting product ρ come, an intuitive idea is to add this new product to the rating matrix and then update the model learning process of PMF. However, this naive method suffers from the huge storage and time cost as most matrix factorization models are batch-based methods, and we need to retrain the user and product latent matrix [17]. In fact, given the seed set of users' ratings to ρ, each user's latent factor would not change much since they already have lots of rating records. For the products, only the promoting product has gained lots of ratings from seed users, thus providing valuable data sources to model its profile. Hence, we borrow the ideas of online matrix factorization [28] and target at updating the promoting product ρ's latent factor (topic) Q_ρ with the loss function as: $\min_{Q_\rho} L = \sum_{u \in A} (R_{u\rho} - P'_u \times Q_\rho)^2$. As the above loss function is convex with regard to Q_ρ, a global minimum of it could be achieved by a simple gradient based method.

Social Influence Modeling. In the literature, a social influence model usually follows the two assumptions [6], [19]–[21], [29]: 1) A user in the seed set would influence herself with 100% probability due to the marketing efforts. 2) The influence score of a user who is not in the seed set depends on her/his neighbors' influence statuses. With this assumption, here we adopt a liner approach for modeling the social influence effect [9], [19]. The reason we choose this model is that, as compared to the classical descriptive influence models (e.g., LT and IC [6]) that are time-consuming to calculate social influence values, it computes the social influence of arbitrary node effectively and efficiently in linear time. Specifically, given the seed set A, the linear social influence model of any user is defined as:

$$\hat{R}_{u\rho}^{S} = \begin{cases} 1 & \text{if } u \in A \\ d \sum_{(u,v) \in \varepsilon} t_{vu} \hat{R}_{v\rho}^{S} & \text{if } u \notin A \end{cases} \tag{3}$$

where d is a damping factor that depicts influence decay in the diffusion process, which is usually set as 0.85 in practice [19]. For each user u that does not belong to the seed set, an iterative method could be used to calculate the influence propagation result $\hat{R}_{u\rho}^{S}$ with convergence guarantee [9], [19].

3.3 Hardness of the Proposed Problem

The proposed product adoption maximization problem is NP-hard.

Proof: The NP-hardness proof can be achieved by transforming a NP-hard problem, called the *Vertex Cover* problem to a special case of the product adoption maximization problem. Specifically, the vertex cover problem asks that: given an undirected graph $G = <V, \mathcal{E}>$ with n nodes ($|V| = n$), can we find a k element subset A ($|A| \subseteq V \wedge |A| = k$), such that $\forall (u, v) \in \mathcal{E}$, we have $u \in A \vee v \in A$ [30].

Given an arbitrary instance of the vertex cover problem involving a graph G, we define a corresponding instance of the product adoption maximization problem by directing all edges of G in both directions. Here, each node in this graph represents a user in the SNS platform. For each user u in SNS, we set a very small user-specified threshold value θ_u. If there is a vertex cover A with size k in G, we can deterministically set the seed set as A. Thus, for each user u that does not belong to the seed set A, i.e., $u \notin A$, according to the definition of the vertex cover problem, we have $\forall (u, v) \in \mathcal{E}, v \in A$. Thus $\hat{R}_{u\rho}^S = d$ based on Eq 3. The interest effect score $\hat{R}_{u\rho}^P$ is also non-negative as the preference value ranges from $[0,1]$ with any interest model. Finally, the predicted preference $\hat{R}_{u\rho}$ that linearly combines $\hat{R}_{u\rho}^P$ and $\hat{R}_{u\rho}^S$ keeps positive and exceeds the very small threshold value θ_u, leading u to become active. Hence, all users that do not belong to the seed set become active and adopt the promoting product. Conversely, this is the only way to get a set A with $f(A) = n - k$. Given the above analysis, the vertex cover problem is equivalent to deciding if there is a set A of k nodes in this graph such that $f(A) = n - k$.

3.4 Solution

Given the NP-hardness of the proposed problem, a natural idea is that since this problem resembles social influence maximization to some extent, can we borrow some nice properties of the solution to social influence maximization problem to inspire our algorithm design? In fact, researchers have proved that though the social influence maximization problem is NP-hard under most influence propagation models, its objective function is non-negative monotone and submodular [6], [7], thus a simple greedy algorithm that selects the seed user set one by one achieves at least $1 - 1/e$ theoretical performance guarantee [22].

Unfortunately, due to a careful analysis, our proposed objective function (Eq. 2), does not follow the monotone property. To explain this, let us consider a special case of the predicted adoption score in Eq. 1 with $\alpha = 1$. Under this case, each user's tendency to adopt a product is totally determined by the user interest model. However, due to the nature of the user interest model, u's predicted score $\hat{R}_{u\rho}^P$ does not positively increase as we increase the seed size k. Thus, our proposed objective function is also not monotone. Accordingly, a greedy algorithm could not provide any theoretical guarantee. In the meantime, the greedy algorithm also fails under the effectiveness concern. For example, even if we select the first seed user with a greedy algorithm, the user interest part still could not accurately capture users' interests for the new promoting product as we only have one rating record from the first seed user for model training, a phenomenon that is well recognized as the cold-start problem in user interest modeling [16]. Due to the above reasons, here we devise an efficient heuristic algorithm that only needs to scan the user set once no matter how large the seed set is. Specifically, when we revisit Eq. 1, we observe that each user's predicted adoption preference is balanced by his/her interest scores

and social neighbors' decisions. In other words, if a user has similar interests with many other users, and he/she also appears in many users' social neighbors (i.e., he/she influences many users), then he/she probably "contributes" a lot to others' decisions. We measure u's social contribution $CS(u)$ and interest contribution $CI(u)$ as:

$$CS(u) = \sum_{v:(v,u)\in\varepsilon} t_{uv}, \quad CI(u) = \sum_{v=1}^{N} w_{uv} = \sum_{v=1}^{N} P'_u \times P_v, \quad (4)$$

where $CS(u)$ measures the extent of the active user's total social influence strength on his/her social neighbors. The interest contribution counts the total interest similarity between this active user u and any other user v by comparing their latent interest vectors P_u and P_v from PMF. After that, each active user's contribution to others can be denoted as a combination of the interest contribution and social contribution:

$$C(u) = \alpha CI(u) + (1 - \alpha)CS(u). \quad (5)$$

Given the social graph and the rating matrix, we can calculate the weighted contribution score by Eq. 5 and find the top-k users with the largest score as the seed set A.

3.5 Parameter Learning

As discussed in Eq. 1, our proposed adoption decision function leverages the social influence and user interest with a Bernoulli distribution to model users' decisions for product adoption maximization. The parameter α in the Bernoulli distribution controls the likelihood of choosing from each user's own preference and the social influence for decision making. When $\alpha = 1$, the adoption decision function solely relies on user interest. With $\alpha = 0$, the problem degenerates to the social influence maximization. Hence, in order to select the seed users for product adoption maximization, it is necessary to "learn" the parameter α to better understand users' adoption decisions and estimate the production adoption maximization effect. However, this is a non-trivial task due to the unobserved selection process of choosing which factor for users' decision making process. In other words, when u adopts a product, this event may be caused by either the user's own interest or the social influence effect. Thus, the parameter α remains unobserved. To overcome this challenge, in the following, we propose a learning method that explores the historical adoption records of users in a SNS for model learning.

Before targeting a promoting product in a SNS, the system has gained lots of adoption records of various products. Each adoption record in this system could be represented as a four-element tuple as (u, i, t, R_{ui}), denoting user u adopts product i at time t with rating preference value R_{ui}. The time t starts from 0 and ends at T. Specifically, by selecting a small time range δ_t (e.g., $\delta_t = 10\% \times T$), all users that adopt product i before time δ_t could be summarized into a set $A_i = [u: R_{ui} > 0 \quad \& \quad t < \delta_t]$. Since these early adopters A_i rate product i in the earliest time, we could approximate these users as the seed users for product i in this SNS platform. Correspondingly, the final adoptions of i, denoted as $P_i = [u: R_{ui} > 0 \quad \& \quad t > \delta_t]$, are users that adopt the product after time δ_t. These users in P_i are either influenced by the earlier adopters A_i or adopt i based on their own interests. In summary, for each product $i \in V$, we have a triplet (i, A_i, P_i) to approximate the seed users and the final adopted users of this product.

With the above real-world adoption records, for each product i with seed set A_i, our goal is to correctly distinguish the adopted users and the non-adopters based on our proposed adoption function (Eq. 1). In other words, for each user u that is not in the seed set A_i, if he/she finally adopts i ($u \in P_i$), then the predicted rating value better exceeds θ_u. For users N_i that do not finally adopt i ($N_i = U - A_i - P_i$), the predicted values are better less than their corresponding adoption thresholds. Based on the above analysis, we now formalize the objective loss function as:

$$\min_{\alpha} L = \sum_{i=1}^{|V|}(\sum_{u \in P_i} h\,(\theta_u - \hat{R}_{ui}) + \sum_{v \in N_i} h\,(\hat{R}_{vi} - \theta_v)) \tag{6}$$

where $h(x)$ is a monotonically non-decreasing function. Empirically, it is better to set $h(x) \approx 1$ if $x > 0$ and $h(x) \approx 0$ if $x < 0$. To approximate the 0-1 loss, we empirically set $h(x)$ as:

$$h(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{1+e^{-cx}} & \text{if } x \geq 0. \end{cases} \tag{7}$$

where c is a large value (e.g., $c = 1000$).

The above loss function with the predicted preference is convex with regard to the model parameter α. Thus, a simple gradient descent method could guarantee the global optimal of the objective loss function. Specifically, the derivative of the parameter α is:

$$\begin{aligned} \frac{\partial L}{\partial \alpha} &= \sum_{i=1}^{|V|}[\sum_{u \in P_i} h\,'(\theta_u - \hat{R}_{ui})(\hat{R}_{ui}^S - \hat{R}_{ui}^P)] \\ &+ \sum_{i=1}^{|V|}[\sum_{v \in N_i} h\,'(\hat{R}_{vi} - \theta_v)(\hat{R}_{vi}^P - \hat{R}_{vi}^S)] \end{aligned} \tag{8}$$

where $h'(x) = h(x)(1 - h(x)) * c$ if $x > 0$ and 0 otherwise.

From the above learning process, we can calculate the complex cost of Eq.6 as $O(S(|U|^2 + |U| * |V|))$, where S is the loop number of gradient descent, which can be estimated by

$$S = \frac{\|a^0 - a^*\|^2}{\varepsilon \eta} \tag{7}$$

where the a^0, a^*, ε and η are the initial and optimized value, optimized approximation error and learning step respectively.

4. Experiments

We perform experimental results on two real-world datasets. Specifically, the goal of our experiments is two-fold as follows. First, we use real-world adoption records to show the soundness of our proposed adoption function with the learned parameter (Section 4.1). After that, with our proposed adoption function, we show the effectiveness and efficiency of different seed selection models for product adoption maximization performance (Section 4.2). **Datasets.** We use two real-world datasets: the undirected movie sharing SNS platform *Flixster* [25] and the directed location based social network dataset *Gowalla* [31]. We filter out users and items that have less than 5 ratings, then we normalize each users' rating preference into the range [0,1]. After pruning, the detailed statistics are shown in **Table 1**.

Table 1. The statistics of the two datasets

Dataset	Users	Items	Ratings	Links	Type
Flixster	49,979	15,672	7,817,180	651,799	undirected
Gowalla	21,755	71,139	278,154	251,296	directed

4.1 Effectiveness of the Adoption Function

Before targeting a promoting product for product adoption maximization, a key element is to evaluate when each user would adopt a product with the predefined product adoption function. In this part, we would show the effectiveness of our proposed product adoption function (Eq. 1) with the parameter learning method in Section. 3.5. Our goal is to predict the accuracy performance of the potential adopters based on the given adoption function.

Algorithms for Comparison. Specifically, there are special cases of our proposed adoption function that are widely researched in parallel. On one hand, when the balance parameter $\alpha = 0$ in Eq. 1, the Social Influence (SI) is the only factor that determines users' adoption decision, thus our proposed product adoption maximization problem degenerates to the classical social influence maximization problem [6]. On the other hand, if $\alpha = 0$ in Eq. 1, each user solely relies on User Interest (UI) for the product adoption decision, which is well researched in CF [17]. Thus, we select the SI and UI as baselines for comparison. Please note that, there are various user interest modeling and social influence modeling techniques. As the focus in this paper is to show the effectiveness of our proposed adoption function with the two parts, the SI and UI baselines are selected as depicted in Section. 3.2. We would explore different algorithms for user interest and social influence modeling in the future.

As shown in Section 3.5, from users' historical adoption records, for each product $i \in V$, there is a triple (i, A_i, P_i) of adoption record that shows the seed users A_i and the adopted users P_i. We select 20% of the triples as the training data to learn parameter α, and then use the learned parameter α to test the effectiveness. In other words, for each product j that is not in the training data, given the seed set A_j, we could simulate the adoption decision process with the learned parameter α and predict each user u's final adoption status with the predicted value $\hat{R}_{uj}$. We summarize the predicted adopters of each test product j in a set $\hat{P}_j$. Specifically, to consider the accuracy of the predicted adoptions, for each item j, we select the widely used ranking-based metrics of Precision, Recall and F1-measure as:

$$Pre(j) = \frac{\#hit}{|\hat{P}_j|}, \quad Rec(j) = \frac{\#hit}{|P_j|}$$

$$F1(j) = \frac{2 * Pre(j) * Rec(j)}{Pre(j) + Rec(j)},$$

where $\#hit$ denotes the number of users in the interaction between P_j and $\hat{R}_j$. Then the three measures are averaged among all products in the test data. The above metrics measure the prediction accuracy of the potential adopters from a micro perspective. Besides, as our goal is to predict the number of potential adopters of each product j, we choose *MAE* and *RMSE* to measure the difference between the number of real adopters and the number of the predicted adopters as:

$$MAE = \sum_j \frac{|(|\hat{P}_j| - |P_j|)|)|}{L}, \quad RMSE = \sqrt{\frac{\sum_j (|\hat{P}_j| - |P_j|)^2}{L}},$$

where L denotes the number of products in the test data. For the MAE and RMSE metrics, the smaller the value, the better the performance. For the remaining three metrics, the larger the value, the better the performance.

Table 2. The measurements under different adoption functions.

	Flixster					Gowalla				
	MAE	RMSE	Pre	Rec	F1	MAE	RMSE	Pre	Rec	F1
The Proposed Method	**121.97**	**13.95**	**0.293**	0.0187	**0.0195**	**10.24**	**4.13**	**0.1869**	0.0912	**0.1134**

| UI | 3885.68 | 65.65 | 0.0054 | **0.1672** | 0.0091 | 752.26 | 30.16 | 0.0019 | **0.2802** | 0.0037 |
| SI | 149.9 | 15.61 | 0.0266 | 0.0172 | 0.0172 | 16.91 | 5.19 | 0.1298 | 0.0869 | 0.1035 |

As can be seen from **Table 2**, our proposed adoption function has the smallest MAE and RMSE values, showing the number of predicted adoptions is closer to the number of real adoptions in the real world. For the UI method, on one hand, it has the largest MAE and RMSE values. On the other hand, UI has the lowest Pre while the Rec value is quite high. Combing both observations, we conclude that UI usually generates a much larger predicted adoption set than the real world adoptions, since UI only models users' latent interest to the product without considering whether the user has known the product or not. Our proposed adoption function produces the similar adoption results as real-world adoptions, and the F1 measure is also the highest among all methods, showing the soundness of our proposed adoption function.

4.2 Adoption Maximization Performance

In this subsection, we would compare the performance of product adoption maximization performance with our proposed adoption function. Specifically, we compare the expected adoptions (Eq. 2) under different seed selection methods.
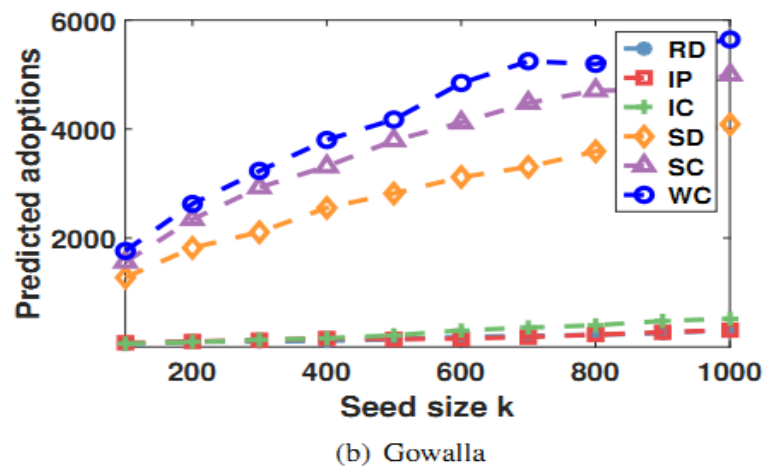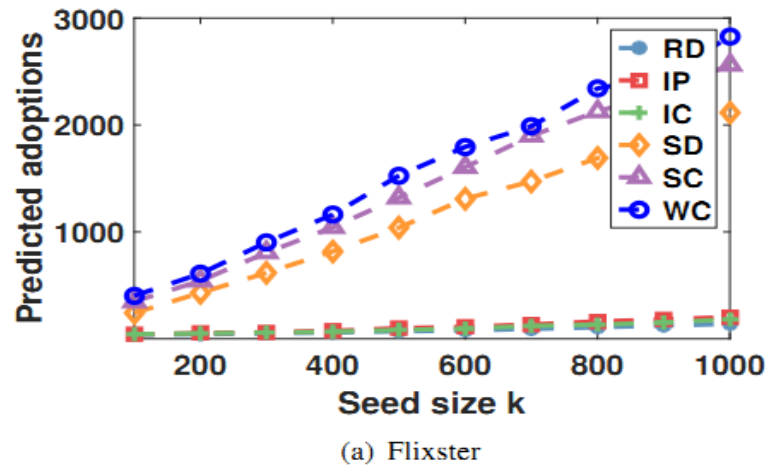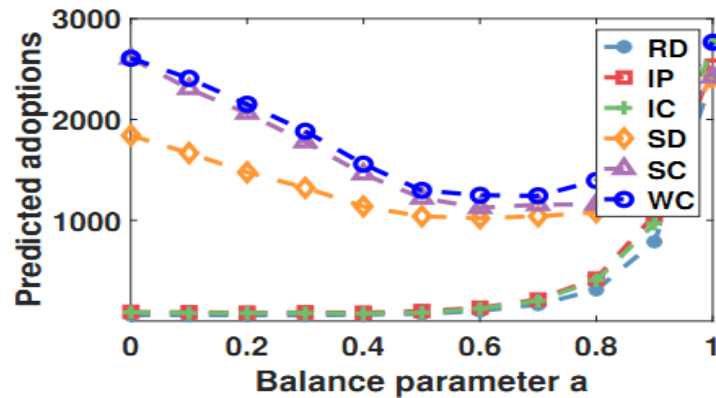

(a) Flixster

(b) Gowalla

Fig. 2. The predicted adoptions with different k.

Algorithms for Comparison. Given the hardness of the product adoption maximization problem with our proposed adoption function, we design a WC algorithm to select the seed sets. Here, we also explore several heuristics for seed selection, and then compare their performance on product adoption maximization with our proposed WC algorithm. Please note that, though some topic-aware social influence models also consider user interest and social influence for product adoption maximization [11], [13], these methods could not be directly used for our problem as they rely heavily on the text information of the new promoting product for topic modeling, which is not available in our datasets. Specifically, these baselines include:

- *RD*: This method denotes using *RanD*om selected users as seed users for comparison.
- *IP*: *I*nterest *P*ositive baseline selects users that have the largest average positive ratings in the history. The intuition is that these users' high ratings would possibly influence others to generate high ratings.
- *IC*: This *I*nterest *C*ontribution baseline selects users that have the largest interest contribution of $CI(u)$, which can be computed in Eq. 4.
- *SD*: This *S*ocial *D*egree baseline selects users that have the highest degrees, with the intuition that these users would influence a large number of social neighbors.
- *SC*: It is a special case of our model with $\alpha = 0$. Under this case, i.e., the seed users are selected based on their *S*ocial *C*ontribution scores.
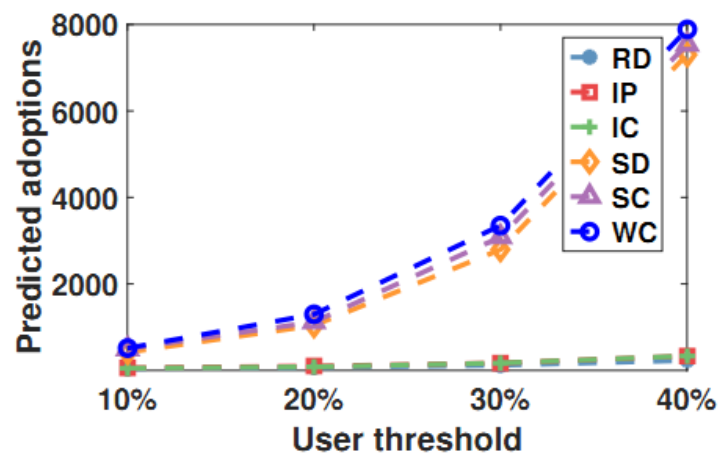

(a) Flixster
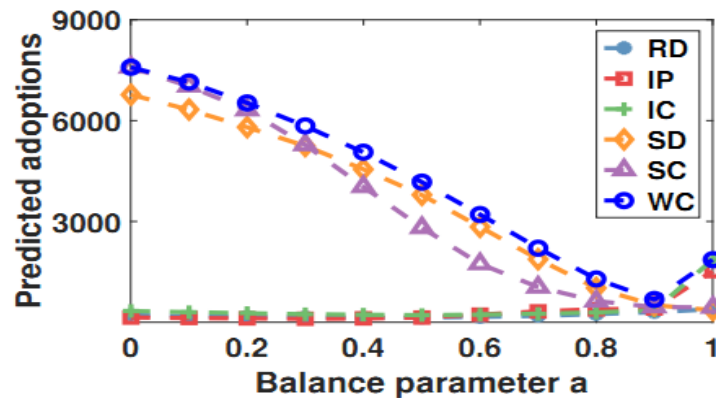
(b) Gowalla

Fig. 3. The predicted adoptions with different (k = 500).

There are several parameters of these models. For interest model of PMF, we randomly select 90% of the rating data for training and tune all the parameters in the test data to ensure the best performance. Since learning the social influence strength is not the focus of this paper, we set the social influence strength as $t_{uv} = \frac{weight(u,v)}{Indegree(v)}$, which is widely used in the literature [6], [19], [29]. Specifically, we need to set the initial rating values for the seed users. In fact, we do not require each seed user to give a highest rating score to the promoting product. As these seed users are provided with marketing efforts (e.g., free samples), we set each seed user u's rating as sampled from a uniform distribution that ranges from the user's average rating score to the highest rating score.

(a) Flixster
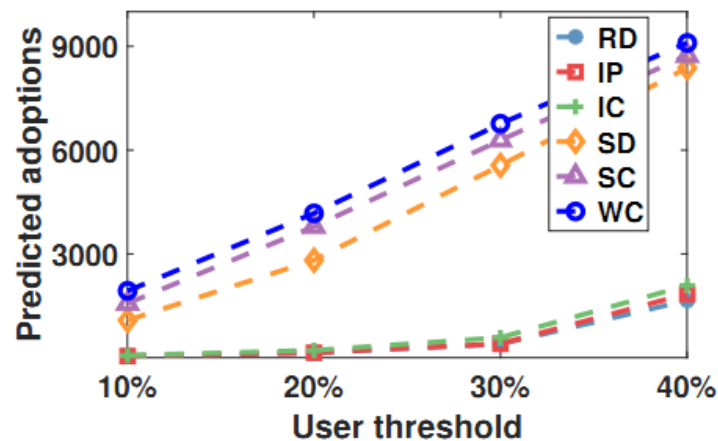
(b) Gowalla

Fig. 4. The predicted adoptions with different threshold

Effectiveness. We first compare the performance of different seed selection methods on the product adoption maximization problem under different parameter settings. The performance is measured by the expected adoptions (Eq. 2). Thus, the larger the predicted value, the better the performance. Specifically, at each time, we only change the parameter we study and keep

the remaining parameters fixed. The default setting is that $\alpha = 0.5, k = 500$ and each user's threshold is set as the top 20% ratings of the user.

Fig. 2 and **Fig. 3** depict the results under different seed size k and parameter α. As can be seen from **Fig. 2**, as the seed size k increases, the performances of all methods improve. Our proposed algorithm WC always performs the best, followed by SC and SD. Particularly, since the baseline of RD, IP and IC do not leverage the social structure into consideration, their results are far from satisfactory. We show the performance comparison under different α in **Fig. 3**. When $\alpha = 0$, our proposed problem degenerates to the social influence maximization, so both WC and SC have the best results. As α increases, the weight of user interest increases in the adoption decision, leading the IC baseline improves over SC when $\alpha = 1$. We observe that SC (IC) always performs better than SD (IP) under any α, which empirically validates the effectiveness of our proposed social contribution and interest contribution scores (Eq. 4). **Fig. 4** investigates the setting of different user thresholds, e.g., 10% means for each user u, θ_u is set as the top 10% ratings of this user. The overall trend is that all methods' performances increase as we decrease the user threshold (Eq. 2) from the top 10% rating to 40%. In summary, it is obvious that our proposed WC method always has the best performance under any parameter setting.

Efficiency. We conduct all the experiments on an Ubuntu server with two 3.10GHz Four-Core Intel Core i5-2400 and 16G memory. **Table 3** shows the runtime of all methods, which include seed selection, interest and social influence estimation, and the adoption prediction process. As can be seen from this table, all the algorithms run very fast. The interest calculation and liner social influence estimation time is the same for all methods. Specifically, the time complexity of interest calculation step is linearly with the seed size. The linear social influence estimation method used in this paper has a complexity that linearly increases with the user size. In seed selection step, all these methods only need to scan the SNS platform once to select the seed set. Our proposed method costs a little more time than other baselines since it needs to balance the interest contribution and social contribution. Given the above analysis, it is obvious that our proposed WC method is very time-efficient, thus can be applied to real-world SNSs with billions of users.

Table 3. Runtime of different methods (seconds).

Methods	RD	IP	IC	SD	SC	WC
Flixster	30.53	32.24	35.35	34.78	36.28	38.24
Gowalla	21.57	23.72	24.32	22.54	23.87	25.73

5. Conclusion

In this paper, we provided the problem of product adoption maximization in a SNS platform. Specifically, we first introduced a general adoption function that combines both user interest and social influence, where the user interest part relies on historical user behavior and seed users' evaluations without any external data. Then, we formally defined the product adoption maximization problem and proved the hardness of it. Furthermore, we designed an efficient algorithm to solve this problem. Experimental results showed the superiority of our proposed adoption function and seed selection method for product adoption maximization. In the future, as different algorithms for user interest and social influence may lead to different predicted adoption results, we plan to explore the impact of different models for the product adoption maximization problem.

References

[1] M. A. Al-Garadi et al., "Analysis of online social network connections for identification of influential users: Survey and open research issues," *ACM Comput. Surv.*, vol. 51, no. 1, Jan. 2018.

[2] H. Wang, Z. Le, and X. Gong, "Recommendation system based on heterogeneous feature: A survey," *IEEE Access*, vol. 8, pp. 170779–170793, 2020. Article(CrossRef Link)

[3] F. García-Sánchez, R. Colomo-Palacios, and R. Valencia-García, "A social-semantic recommender system for advertisements," *Information Processing and Management*, vol. 57, no. 2, p. 102153, 2020. Article(CrossRef Link)

[4] A. Fazeli and A. Jadbabaie, "Game theoretic analysis of a strategic model of competitive contagion and product adoption in social networks," in *Proc. of 2012 ieee 51st ieee conference on decision and control (cdc)*, pp. 74–79, 2012. Article(CrossRef Link).

[5] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proc. of WWW*, pp. 519–528, 2012. Article(CrossRef Link)

[6] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. of SIGKDD*, pp. 137–146, 2003. Article(CrossRef Link)

[7] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. of SIGKDD*, pp. 420–429, 2007. Article(CrossRef Link)

[8] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. of SIGKDD*, pp. 1029–1038, 2010. Article(CrossRef Link)

[9] Q. Liu, B. Xiang, E. Chen, H. Xiong, F. Tang, and J. X. Yu, "Influence maximization over large-scale social networks: A bounded linear approach," in *Proc. of CIKM*, pp. 171–180, 2014. Article(CrossRef Link)

[10] Y. Yang and J. Pei, "Influence analysis in evolving networks: A survey," *TKDE*, vol. 33, no. 3, pp. 1045-1063, 2019. Article(CrossRef Link)

[11] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," *Knowledge and Information Systems*, vol. 37, pp. 555–584, 2013. Article(CrossRef Link)

[12] Y. Li, D. Zhang, and K.-L. Tan, "Real-time targeted influence maximization for online advertisements," *Proceedings of the VLDB Endowment*, vol. 8, no. 10, pp. 1070–1081, 2015. Article(CrossRef Link)

[13] S. Chen, J. Fan, G. Li, J. Feng, K.-l. Tan, and J. Tang, "Online topic-aware influence maximization," *Proceedings of the VLDB Endowment*, vol. 8, no. 6, pp. 666–677, 2015. Article(CrossRef Link)

[14] A. Goyal and L. V. Lakshmanan, "Recmax: Exploiting recommender systems for fun and profit," in *Proc. of SIGKDD*, pp. 1294–1302, 2012. Article(CrossRef Link)

[15] W. Chen *et al.*, "Influence maximization in social networks when negative opinions may emerge and propagate," in *Proc. of SDM*, vol. 11, pp. 379–390, 2011. Article(CrossRef Link)

[16] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005. Article(CrossRef Link)

[17] A. Mnih and R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. of NIPS*, pp. 1257–1264, 2007. Article(CrossRef Link)

[18] L. Chen, L. Wu, R. Hong, K. Zhang, and M. Wang, "Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach," in *Proc. of AAAI*, vol. 34, pp. 27–34, 2020. Article(CrossRef Link)

[19] B. Xiang, Q. Liu, E. Chen, H. Xiong, Y. Zheng, and Y. Yang, "PageRank with priors: An influence propagation perspective," in *Proc. of the Twenty-Third international joint conference on Artificial Intelligence*, pp. 2740-2746, 2013. Article(CrossRef Link)

[20] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing letters*, vol. 12, no. 3, pp. 211–223, 2001. Article(CrossRef Link)

[21] M. Granovetter, "Threshold models of collective behavior," *American journal of sociology*, vol. 83, no. 6, pp. 1420–1443, 1978. Article(CrossRef Link)

[22] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978. Article(CrossRef Link)

[23] H.-J. Hung et al., "When social influence meets item inference," in *Proc. of SIGKDD*, pp. 915–924, 2016. Article(CrossRef Link)

[24] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Computer*, vol. 42, no. 8, pp. 30–37, 2009. Article(CrossRef Link)

[25] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. of RecSys*, pp. 135–142, 2010. Article(CrossRef Link)

[26] L. Wu, Y. Ge, Q. Liu, E. Chen, B. Long, and Z. Huang, "Modeling users' preferences and social links in social networking services: A joint-evolving perspective," in *Proc. of AAAI*, pp. 279–286, 2016. Article(CrossRef Link)

[27] W. Lu, S. Ioannidis, S. Bhagat, and L. V. Lakshmanan, "Optimal recommendations under attraction, aversion, and social influence," in *Proc. of SIGKDD*, pp. 811–820, 2014. Article(CrossRef Link)

[28] S. Rendle and L. Schmidt-Thieme, "Online-updating regularized kernel matrix factorization models for large-scale recommender systems," in *Proc. of RecSys*, pp. 251–258, 2008. Article(CrossRef Link)

[29] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. of ICDM*, pp. 88–97, 2010. Article(CrossRef Link)

[30] R. M. Karp, "Reducibility among combinatorial problems," *50 Years of Integer Programming 1958-2008*, pp 219-241, 2009. Article(CrossRef Link)

[31] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting place features in link prediction on location-based social networks," in *Proc. of SIGKDD*, pp. 1046–1054, 2011. Article(CrossRef Link)

[32] T. Trinh, N. Nguyen, D. Wu, J. Z. Huang and T. Z. Emara, "A New Location-Based Topic Model for Event Attendees Recommendation," in *Proc. of 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, Danang, Vietnam, pp. 1-6, 2019. Article(CrossRef Link)

[33] Xiaolin Li, Chaojiang Wu, Feng Mai, "The effect of online reviews on product sales: A joint sentiment-topic analysis," *Information and Management*, vol. 56, no. 2, pp. 172-184, 2019. Article(CrossRef Link).

[34] Zarrinkalam, Fattane, Hossein Fani, and Ebrahim Bagheri. "Social user interest mining: methods and applications," in *Proc. of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3235-3236, 2019. Article(CrossRef Link).

[35] Wang, Yanru, Shifei Ding, Xiao Xu, and Weikuan Jia, "The multi-tag semantic correlation used for micro-blog user interest modeling," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 765-772, 2019. Article(CrossRef Link).

[36] Shah, Bansari, Ananda Prakash Verma, and Shailendra Tiwari, "User interest modeling from social media network graph, enriched with semantic web," in *Proc. of International Conference on Computational Intelligence and Data Engineering*, Springer, Singapore, pp. 55-64, 2018. Article(CrossRef Link).

Ping JI is Associate Professor at School of Advanced Manufacture Engineering, Hefei University. She received her Master's degree in photonics from Goettingen College of science and art (Germany) in 2010. Her research interests include Media technology and the signal processing and pattern recognition.

Hui HUANG is Associate Professor at School of Advanced Manufacture Engineering, Hefei University. She received her Master's degree in surface science from Tsukuba University (Japan) in 2010. Her research interests include superconductor electronics, two-dimensional semiconductors, and integrated electronic circuits.
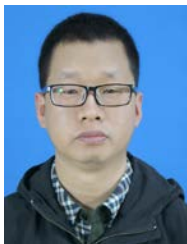
Xueliang Liu is a professor in school of computer science and information engineering in Hefei University of Technology, China. He received his Ph.D and MSc from EURECOM France and USTC China in 2013 and 2008 respectively. His research interests include multimedia retrieval, and social media analysis. He has authored over 40 journal and conference papers in these areas.

Xueyou HU is currently a professor at School of Advanced Manufacturing Engineering, Hefei University. He received the Master's degree in signal and information processing from Hefei University of Technology in 2003 and the Ph.D degree in computer application technology from Anhui University in 2011. His research interests include the signal processing and pattern recognition.