Hindawi Mathematical Problems in Engineering Volume 2019, Article ID 7348301, 15 pages https://doi.org/10.1155/2019/7348301



Research Article

Unsupervised Negative Link Prediction in Signed Social Networks

Pengfei Shen 📵, Shufen Liu, Ying Wang, and Lu Han 📵

College of Computer Science and Technology, Jilin University, Changchun 130012, China

Correspondence should be addressed to Lu Han; hanlu@jlu.edu.cn

Received 25 May 2018; Revised 15 March 2019; Accepted 14 April 2019; Published 30 April 2019

Academic Editor: Filippo Cacace

Copyright © 2019 Pengfei Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It has been proved in a number of applications that it is useful to predict unknown social links, and link prediction has played an important role in sociological study. Although there has been a surge of pertinent approaches to link prediction, most of them focus on positive link prediction while giving few attentions to the problem of inferring unknown negative links. The inherent characteristics of negative relations present great challenges to traditional link prediction: (1) there are very few negative interaction data; (2) negative links are much sparser than positive links; (3) social data is often noisy, incomplete, and fast-evolved. This paper intends to address this novel problem by solely leveraging structural information and further proposes the UN-PNMF framework based on the projective nonnegative matrix factorization, so as to incorporate network embedding and user's property embedding into negative link prediction. Empirical experiments on real-world datasets corroborate their effectiveness.

1. Introduction

With the increased popularity of smart devices and online applications, online behaviors can be regarded as an effective way to reflect people's lives. Social media has permeated every aspect of people's daily life, and people's behaviors and lifestyles are gradually transferred to the Internet nowadays. Huge volumes of opinion-rich data are usergenerated in social media at an unprecedented rate, easing the mining of valuable information [1]. Meanwhile, a largescale social network has been formed among online users. Social networks contribute to developing a vast variety of relations between users such as friendships in Facebook (https://www.facebook.com/), follower relations in Twitter (https://twitter.com/), trust relations and distrust relations in Epinions (http://www.epinions.com/), and Slashdot (http://slashdot.org/). Furthermore, these relations can be generally classified into two categories: positive relations and negative ones. It is significant to detect relationships among users when researches on social network are conducted, which can not only be helpful to some online applications, but also discover social problems through mining social relations. Besides, it is also helpful for sociological study.

Signed network analysis has attracted increasing attention in recent years. However, most of the current researches

on signed network concentrate on inferring positive links, for example, predicting trust relationships between users in social network [2, 3] or discovering the citation relationships between the citing reference and the cited reference in citation network [4]. Trust, which provides information about from whom users should accept information and with whom users should share information, plays an important role in helping online users collect reliable information [2]. Recent years have witnessed many trust-related online applications, such as trust-aware recommendation systems [5-7], high-quality user-generated content finding [8, 9], and viral marketing [2]. All in all, by studying the positive links in signed network, people can find potentially valuable knowledge. As the counterpart of positive links in signed network, negative links have not obtained the same degree of attention, which are usually ignored in many researches. In general, discovering distrust relationships can help users avoid fake information and Internet fraud. But essentially, inferring unknown negative links is worth far more than that, for example, a small number of negative links can significantly improve positive link prediction and they can also improve the performance of recommendation systems in social media [10, 11]. Furthermore, negative emotions are the latent factor of inducing contradictions. For instance, the crisis of trust between husband and wife could lead to marriage crisis; the credibility gap between employees and employers could influence the unity and harmony of the company, which would harm the company's interests, and the crisis of trust between citizens and governments could give rise to social unrest. Thus it can be seen that negative link prediction is significant because it can not only help many online applications to be more effective, but help to discover unsteadiness and solve contradictions ahead of time.

It is a nontrivial task to conduct researches on negative links in signed social network with many difficulties and challenges. Firstly, all links, positive or negative ones in signed social network, are very sparse. Negative links are even sparser and more are missing, thereby the signed social network has a lot of noise data. Secondly, there is little negative interactive data between users in social network. Regardless of it being users rating items or online interactions, users always tend to give favorable reviews or "Likes". Tang et al. [2] investigated two product review sites, Epinions and Ciao, both of which employ 5-star system to rate items, finding that the majority of ratings are scores of 4 and 5. The result is consistent with the study mentioned: users are likely to give positive ratings to items or others. Lastly, with the dramatic rise in the number of Internet users, social network data is growing explosively, and with the change in users' participation, the social network structure is fast-evolved, so negative link prediction becomes increasingly difficult. Based on the above difficulties, this paper proposes to predict negative links only with network topology and does not rely on any interactive data between users. To address the sparsity issue, this paper represents each node in a unified low-dimensional space through projective nonnegative matrix factorization, which can further alleviate the inconveniences caused by sparsity. Moreover, this study finds a latent factor of user pairs, which contributes to the formulation of social network links. Finally, it transforms the new negative link prediction problem into an unsupervised learning framework: UN-PNMF. The main contributions are summarized as follows:

- (i) The negative network structure is embedded into a low-dimensional vector space through projective nonnegative matrix factorization.
- (ii) Investigate the signed network only with positive links and exploit the activity degree and influence degree of users for negative link prediction via embedding these properties of user pairs.
- (iii) An unsupervised framework is proposed: UN-PNMF, which embeds network structure and a latent factor of user pairs into a low-dimensional vector space, simultaneously.
- (iv) The proposed UN-PNMF framework in real-world social media datasets intends to be evaluated so as to understand the effectiveness and mechanisms.

The rest of paper is organized as follows. Section 2 briefly reviews the related work. Section 3 describes the details of datasets. Section 4 defines the problem. The embedding of the negative network structure into a low-dimensional vector space through projective nonnegative matrix factorization is

mainly discussed in Section 5. Section 6 investigates a latent factor of user pairs, which is helpful to predict negative links in social network. Section 7 proposes the framework UN-PNMF and introduces the details of its algorithm. Section 8 presents experimental results and some observations. Section 9 is the conclusion part with some prospects mentioned in this field.

2. Related Work

This section briefly reviews work which is related to different variants of the link prediction in signed social network. The methods of most existing researches can be roughly divided into three categories: supervised, unsupervised, and semisupervised learning.

The supervised link prediction methods of unknown links in social network have two critical steps, first constructing features from available data sources and second training a binary classifier based on these features. Lichtenwalter et al. [12] show several advantages of supervised methods in link prediction such as superior performance adaptation to different domains and variance reduction [11]. Wang et al. [13] introduce a new objective function for signed network embedding guided by extended structural balance theory and propose a deep learning framework SiNE. The embedding learned by SiNE can significantly improve the link prediction performance. Leskovec et al. [14] first introduce status theory and balance theory to predict positive or negative link in signed networks. Zolfaghar et al. [15] develop a framework of several trust inducing factors then investigate C5.0 tree and neural network to predict trust and distrust relations. Tang et al. [16] apply user's interaction data to predict distrust. Wang et al. [17] explore the combination of Dempster-Shafer theory and neural network to predict trust and distrust. The Dempster-Shafer theory allows one to combine evidence from different sources and arrive at a belief function by taking into account all the available evidence. Because of this feature, the authors combine the Dempster-Shafer theory and neural network to predict trust and distrust.

The unsupervised learning is the learning task of inferring a function to describe hidden structure from unlabeled data. Unsupervised link prediction mostly uses matrix factorization methods, which can alleviate the sparsity of the links in social network. Tang et al. [2] investigate the homophily in trust prediction and formulate the trust prediction problem into an optimization problem integrated with homophily. Tang et al. [11] propose NeLp framework which can exploit only with positive links and content-centric interactions to predict negative links. Wang et al. [3] verify the status theory in trust relations and calculate the status of users by PageRank. They exploit status theory for trust prediction under the trust prediction framework based on low-rank matrix factorization. Oh et al. [18] propose a probability-based trust prediction model based on trust information transferring and the information includes explicit information and implicit information. Wu et al. [19] investigate the users' consumption behaviors and social behavior and propose an approach to predict unknown links in social network by jointly modeling

TABLE 1: Statistics of the datasets.

| | Epinions | Slashdot |
|-----------------------------|----------|----------|
| # of user | 7323 | 54052 |
| # of Positive Link | 118739 | 404153 |
| # of Negative Link | 19629 | 112307 |
| Positive Network Density | 0.0022 | 0.00014 |
| Negative Network Density | 0.00037 | 0.00004 |

users' consumption behaviors and social behaviors in social networking services.

Embedding network data into a low-dimensional vector space has shown promising performance for link prediction in signed network. Most semisupervised link prediction methods have two key steps. The first step is to learn network representation and each node is represented by a low-dimensional vector. The second is to apply general machine learning techniques on these low-dimensional vectors. Wang et al. [20] propose a Structural Deep Network Embedding (SDNE) framework, to perform network embedding. Specifically, to capture the highly nonlinear network structure, they design a semisupervised deep model and the embedding learned by SDNE significantly improves the link prediction performance.

3. Data Analysis

This paper collects two publicly available datasets for study, i.e., Epinions and Slashdot, which include both positive and negative links to perform link prediction. Epinions and Slashdot are two product review sites, where users can decide to trust or distrust other users. Accordingly, the datasets are two signed graphs, so the trust relations can be seen as positive links and distrust relations can be seen as negative links in signed network. Similarly, trust network and distrust network can be regarded as positive network and negative network respectively. Firstly, these users with less than three (in-degree plus out-degree) in positive network are filtered. Then, this part deletes some users who only have negative links, aiming to achieve datasets with sufficient information of positive network. A number of key statistics of the datasets are shown in Table 1.

As seen from Table 1, links are very sparse in our datasets but negative links only take up a small proportion of total links. The negative network density of Epinions and Slashdot is 0.00037 and 0.00004, respectively. On average, users of Epinions have 16.21 positive links, while users of Epinions only have 2.68 negative links. Thus it can be seen that negative link prediction is more difficult and complicated than positive link prediction.

4. Problem Statement

This paper uses bold uppercase characters for matrices (e.g., **A**), bold lowercase characters for vectors (e.g., **a**), and normal lowercase characters for scalars (e.g., a). Also, this work

represents i-th row of matrix A as A (i,:), j-th column of matrix A as A(:,j), (i, j)-th entry as A_{i,j}, transpose as A^T, and trace as tr(A) if A is a square matrix.

Let $\mathcal{U} = \{\mu_1, \mu_2, \mu_3, \dots, \mu_n\}$ be the set of users where n is the number of users in the signed social network. A signed network can be decomposed into a positive network component $\mathcal{G}_{p}(\mathcal{U}, E_{p})$ and a negative network component $\mathcal{G}_n(\mathcal{U}, E_n)$ where E_p and E_n are the sets of positive and negative links, respectively. $\mathbf{G}_p \in \mathbb{R}^{n \times n}$ is the matrix representation of positive network where $\mathbf{G}_p(\mathfrak{i},\mathfrak{j})=1$ if there is a positive link from μ_i to μ_i and $\mathbf{G}_{\mathbf{p}}(\mathbf{i},\mathbf{j}) = 0$ otherwise. $\mathbf{G}_n \in \mathbb{R}^{n \times n}$ is the matrix representation of negative network where G_n (i,j) = 1 if there is a negative link from μ_i to $\mu_{\rm j}$ and $G_{\rm n}$ (i,j) = 0 otherwise. This paper also analyzes some properties of user pairs in positive network. $H \in$ $\mathbb{R}^{n\times n}$ denotes the property matrix where H (i, j) represents the property value of μ_i related to μ_i . Because the social network is fast-evolved, this study splits the whole dataset into 6 timestamps, i.e., $t = \{t_5, t_6, t_7, t_8, t_9, t_{10}\}$, where the subscript q of t_q represents q*10 percent of negative links in chronological order. $\mathcal{G}_{p}^{t}(\mathcal{U}, E_{p}^{t}), \mathcal{G}_{n}^{t}(\mathcal{U}, E_{n}^{t}), \text{ and } \mathbf{H}^{t}$ represent positive network, negative network, and property matrix at time t, respectively.

With the aforementioned notations and definitions, the problem of negative link prediction in social media is formally defined as follows.

Given the positive network $\mathcal{G}_p^t(\mathcal{U}, E_p^t)$, negative network $\mathcal{G}_n^t(\mathcal{U}, E_n^t)$, and property of pairwise matrix \mathbf{H}^t at time t, we aim to develop a predictor g to predict the negative network $\mathcal{G}_n^{t'}(\mathcal{U}, E_n^{t'})$ at time t', with $\mathcal{G}_p^t, \mathbf{H}^t$ as

$$g: \left\{ \mathcal{G}_{p}^{t}, \mathcal{G}_{n}^{t}, \mathbf{H}^{t} \right\} \longrightarrow \mathcal{G}_{n}^{t'}$$
 (1)

The negative link prediction problem model can be illustrated in Figure 1 and the relations among \mathcal{G}_{p}^{t} , \mathbf{H}^{t} , \mathcal{G}_{n}^{t} , and $\mathcal{G}_{n}^{t'}$ are demonstrated clearly.

5. Low-Rank Matrix Factorization for Negative Network

The sparsity of data and the huge volume of noise data are always the main difficulties for researches on signed social network. The quantity of users is huge and this number is rapidly increasing, but the links are very limited in social network. As excruciating as it is, negative links are less than positive links; thus the negative network is much sparser, and it is difficult to directly extract feature information. Owing to the advantages of network representation methods in sparse network, this paper intends to predict negative links in social network through network representation methods. Projective nonnegative matrix factorization [21] is employed to embed negative network, which reduces the dimension of representation vector of each node by matrix factorization and at the same time reduces the data sparseness. Our goal is to seek a low-rank representation $\mathbf{W} \in \mathbb{R}^{n \times k}$ with $k \ll n$ by a matrix factorization model. The low-dimensional vector can not only maximumly preserve effective information, but

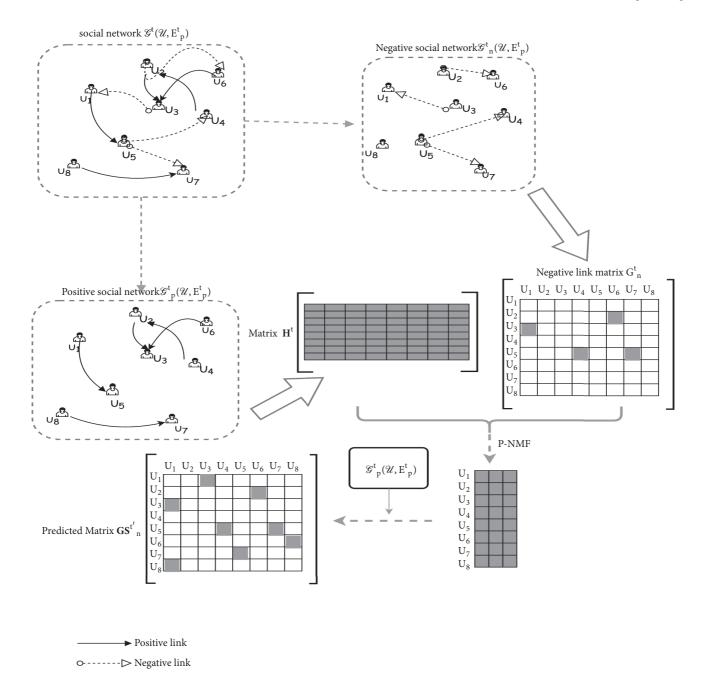


FIGURE 1: The illustration of unsupervised negative link prediction model.

also eliminate a considerable quantity of noise data. What we will probably do is to predict new negative links by the existent social relationships (including both positive links and negative links) at the time of t. Therefore, for the negative network $\mathcal{G}_n^t(\mathcal{U}, E_n^t)$, G_n^t is the adjacency matrix of \mathcal{G}_n^t , this section embeds \mathcal{G}_n^t into a low-dimensional vector space by projective nonnegative matrix factorization. This method is based on the factorization of the matrix G_n^t , so this problem can be solved by the following optimization problem:

where $\| \bullet \|_F$ is the Frobenius norm of a matrix and $\mathbf{W} \in \mathbb{R}^{n \times k}$ is the low-dimensional representation of negative network. The k is the dimension of the low-dimensional representation while the value of k is a key element for quality of low-dimensional representation. How to determine the value of k is very important, which will be examined in detail in later sections. To avoid overfitting, smoothness regularization is added on \mathbf{W} , into (2),

$$\min_{\mathbf{W}} \left\| \mathbf{G}_n^t - \mathbf{W} \mathbf{W}^T \mathbf{G}_n^t \right\|_F^2 \tag{2} \qquad \min_{\mathbf{W}} \left\| \mathbf{G}_n^t - \mathbf{W} \mathbf{W}^T \mathbf{G}_n^t \right\|_F^2 + \alpha \left\| \mathbf{W} \right\|_F^2 \tag{3}$$

where α is nonnegative and is introduced to control the capacity of **W**. Nonnegative constraint is always applied to **W** in (3) as

$$\min_{\mathbf{W}} \quad \left\| \mathbf{G}_{n}^{t} - \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t} \right\|_{F}^{2} + \alpha \|\mathbf{W}\|_{F}^{2}$$

$$s.t. \quad W \ge 0$$
(4)

As for the optimization problem shown above, many optimization methods can be applied such as gradient decent. In the process of solving above model, which can not only reduce the dimensionality of input data through **W** but also effectively reduce Gaussian noise by setting suitable dimension of **W**, the main principle of this model is to obtain a proper projection operator that can project prior knowledge of user pairs to new feature space, and the projection operator guarantees minimum differences between new projection matrix and adjacency matrix of negative network. Accordingly, due to its flexibility, it allows us to include the prior knowledge such as some latent factors of user pairs, which will be introduced in next section.

6. Latent Factor Representation Learning

This section investigates the structural information of positive social network. A latent factor is found by conducting the research on positive social network topology. This latent factor contributes to the formulation of new negative links in the future.

In this study, given the positive network, the negative link prediction is to discover some user pairs which can establish negative links with high probability from a tremendous amount of user pairs without links. It needs to find out what kind of user pairs are easy to establish negative links in social network in the future, so our problem comes down to a clustering problem. In order to investigate which latent factor could contribute to the establishment of a new negative link between two users, the in-degree and out-degree of each user attract a great interest in positive network because in a trust network, if a user has a bigger in-degree, which shows that there are a lot of users who trust him and he has a great influence or authority in this network. Moreover, more and more users would be influenced by him. For example, the more influential a Twitter user is, the more followers he has. Also in some social media, the number of celebrities' fans is far higher than that of ordinary users' fans. This is also because celebrities have a greater influence than ordinary people. At the same time, the out-degree of a user could reflect his active degree, if a user has bigger out-degree in trust network, which shows that he communicates with a larger number of users in social network. From another point of view, this user likes interacting with other users. It can be considered in this way, if out-degree (μ_i) > out-degree (μ_i) , μ_i is more active than μ_i . In [22], researchers believe distrust is a low level of trust. In other words, the initial attitude of distrusted users is trust, and they convert from trust to distrust because of different notions or values during the later association or communication. The study of the negative link prediction aims to discover some user pairs that are more likely to interact with each other. Considering all user pairs in network, it is essential to combine the active degree of one user with the influential degree of another to investigate the social links between these two users. For the sake of convenience this latent factor is denoted by In_Out. Extending to signed social network, there is a question: if outdegree (μ_i) > out-degree (μ_m) and in-degree (μ_j) > in-degree (μ_n), is the tendency of establishing link between μ_i and μ_i higher than the tendency of establishing link between μ_m and μ_n ? To answer this question, this paper defines a latent factor matrix **In_Out**^t, where the superscript t refers to the time, defined as follows:

$$\mathbf{In_Out}^{t} = \begin{pmatrix} \zeta(1,1) & \zeta(1,2) & \cdots & \zeta(1,n) \\ \zeta(2,1) & \zeta(2,2) & \cdots & \zeta(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ \zeta(n,1) & \zeta(n,2) & \cdots & \zeta(n,n) \end{pmatrix}$$
(5)

where $\zeta(i, j)$ can be calculated as

$$\eta(i, j) = out - degree(i) * in - degree(j)$$
(6)

$$\zeta(i,j) = \frac{\eta(i,j)}{d_i} \tag{7}$$

$$d_i = \sum_{j=1}^n \eta(i, j) \tag{8}$$

The out-degree (i) and in-degree (j) are μ_i 's out-degree and μ_j 's in-degree in positive network $\mathcal{G}_p^t(\mathcal{U}, E_p^t)$, respectively. According to the correspondence between \mathbf{G}_n^t and \mathbf{In} - \mathbf{Out}^t , two vectors are conducted, \mathbf{s}_n and \mathbf{s}_r . For the arbitrary user μ_i and user μ_j , if the value of $\mathbf{G}_n^t(i,j)$ is equal to 1, then the value of \mathbf{In} - $\mathbf{Out}^t(i,j)$ is kept in \mathbf{s}_n ; otherwise, save it in \mathbf{s}_r .

In order to further prove the effectiveness of the latent factor In_Out contributing to the formulation of new negative links in social network, this part conducts a two-sample ttest on $\mathbf{s_n}$ and $\mathbf{s_r}$. The null hypothesis is $H_0\colon \mathbf{s_n}=\mathbf{s_r}$, and the alternative hypothesis is $H_1\colon \mathbf{s_n}>\mathbf{s_r}$. For two datasets, the null hypothesis is rejected at significance level $\alpha=0.01$ with p-value of 4.39e-16 and 6.68e-43, respectively. The evidence from two-sample t-test suggests a positive answer to the above question: in signed social network, the In_Out has a large difference between user pairs with negative link and the user pairs without any link. Moreover, user pairs with high value of In_Out may establish links with high probability in the future. Owing to the positive links given in our problem and compared with user pairs without links, the high value of In-Out makes it easier for two users to establish a new negative link.

Synthesizing the analysis and discussion above, it can be realized that the latent factor In_Out plays an important role in the formulation of negative links in signed social network. In social network, the number of users is great. However, in most time, the number of users who have established links with each user is really limited. Just as the conclusion obtained in social science, when one person communicates

with others in different periods of time, the people associated with him always change over time, but the number of people who communicate with him basically does not change or the number is constant in a range. Therefore, extracting features from all user pairs in social network, not only needs a large storage space but also causes much valuable information to be drowned by noise data; thus it is also a major difficulty in social media data mining. In short, the paper intends to embed the latent factors of user pairs into a low-dimensional vector space, which could eliminate the excessive noise data. In particular, there are a lot of favorable reviews without practical significance in social media, so the method for denoising is very necessary to be adopted. Therefore, our goal is to seek a low-rank representation $\boldsymbol{P} \in \, \mathbb{R}^{n \times h}$ with h≪ n, P being the effective representation of matrix In_Out. In this section, this part still adopts the method based on projective nonnegative matrix factorization, and the end projection operator is P. The low-dimensional matrix P only saves some valuable information of potential links among users in negative network. Each row of $\mathbf{P} \in \mathbb{R}^{n \times h}$ is the lowdimensional representation of a node in network, where h is the dimension of representative vector. Choosing an optimal number for h can not only effectively represent In_Out but also reduce the dimension and eliminate huge noise data. Therefore, embedding latent factor In_Out of user pairs into **P** can be solved by the following optimization problem:

$$\min_{\mathbf{p}} \left\| \mathbf{G}_{n}^{t} - \mathbf{P} \mathbf{P}^{T} \mathbf{In}_{-} \mathbf{Out}^{t} \right\|_{F}^{2}$$
 (9)

It is evident that this function is similar to (2) and the processing procedure is exactly the same as in the previous analysis. Due to the space limitations, the objective function of this submodule can be obtained directly, as shown below:

$$\min_{\mathbf{W}} \quad \left\| \left| \mathbf{G}_{n}^{t} - \mathbf{P} \mathbf{P}^{T} \mathbf{In}_{-} \mathbf{Out}^{t} \right| \right\|_{F}^{2} + \beta \|\mathbf{P}\|_{F}^{2}$$

$$s.t. \quad \mathbf{P} \ge 0$$
(10)

where β is nonnegative and is introduced to control the capacity of **P**. Nonnegative constraint is always applied to **P** in (10). In (10), $\mathbf{G}_{\mathbf{n}}^{t}$ and $\mathbf{In}_{-}\mathbf{Out}^{t}$ are nonnegative; therefore, function (10) of **P** is convex. In signed social network, nodes representation via embedding latent factors of user pairs provides a new approach for link prediction. It can get a flexible and unified model for embedding latent factors through (10), because researchers may discover more latent factors of user pairs in later study, and anyone could use this model to embed these new latent factors into low-dimensional space conveniently. This paper mines several latent factors which

contribute to the formulation of social links, but only In_Out has significant effect on formulating of negative links. Due to the space limitation, there is no need to enumerate other latent factors here. This section only demonstrates the analysis of representation learning of In_Out, but this does not limit the scalability and efficiency of this model, because this model could cope with different latent factors flexibly in further study.

7. Modeling Unsupervised Learning for Negative Link Prediction

7.1. The Proposed Framework: UN-PNMF. Section 5 introduces the representation learning of negative network structure, which adopts projective nonnegative matrix factorization. The matrix factorization can be applied to reduce the sparsity as well as to preserve the effective information of the first order proximity of negative networks [23–26]. Moreover, some valuable information of potential negative links is also kept. Sparsity is an inherent difficulty in social network and matrix factorization methods are widely used by many researchers, so the advantaged projective nonnegative matrix factorization is very suitable for the negative link prediction problem. Section 6 represents the latent factor matrix In_Out by a low-rank matrix. Obviously, the In_Out is equivalent to H in (1). Each element of **In_Out** represents the potential possibility of negative link formulation between corresponding users. However, as the number of users increases, the number of user pairs tends to grow exponentially. Hence, the scale of latent factor matrix In_Out will become infinitely large, which is not available by computer hardware. Besides, the matrix In_Out contains much noise data. The number of everyone establishing links (both positive links and negative ones) with others is always limited. Therefore, this study embeds In_Out into a low-dimensional matrix through the projective nonnegative matrix factorization. With this solution, the achieved low-dimensional matrix not only filters huge noise data but also decreases the space complexity of algorithm in a large degree.

In this paper, the negative network structure and the latent factor of user pairs are not treated independently. For these two problems, we intend to make them be complementary with each other, because the structure of negative network with the latent factor of each user pair can more precisely orientate the user pairs who would establish negative links in the future. With the combination of the above two analyses, this part proposes the framework, UN-PNMF, based on the projective nonnegative matrix factorization. UN-PNMF is to solve the following optimization problem:

$$\min_{\mathbf{W},\mathbf{P}} F = \left\| \mathbf{G}_{n}^{t} - \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t} \right\|_{F}^{2} + \lambda \left\| \mathbf{G}_{n}^{t} - \mathbf{P} \mathbf{P}^{T} \mathbf{I} \mathbf{n}_{-} \mathbf{O} \mathbf{u} \mathbf{t}^{t} \right\|_{F}^{2} + \alpha \|\mathbf{W}\|_{F}^{2} + \beta \|\mathbf{P}\|_{F}^{2}$$

$$s.t. \quad \mathbf{P} \ge 0, \qquad (11)$$

Quite evidently, (11) is the simple summation of (4) and (10), where λ is a parameter that controls the degree of latent factor In_Out. In this way it will obtain two low-dimensional vectors which represent the network structural information and the latent information of users respectively,

but this way is opposite to the original intention. This paper intends to embed network structural information and **In_Out** into a low-dimensional projection operator simultaneously. Therefore, (11) can be rewritten as shown below:

$$\min_{\mathbf{W}} F = \left\| \mathbf{G}_{n}^{t} - \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t} \right\|_{F}^{2} + \lambda \left\| \mathbf{G}_{n}^{t} - \mathbf{W} \mathbf{W}^{T} \mathbf{I} \mathbf{n}_{-} \mathbf{O} \mathbf{u} \mathbf{t}^{t} \right\|_{F}^{2} + \alpha \|\mathbf{W}\|_{F}^{2}$$

$$s.t. \quad \mathbf{W} \ge 0$$
(12)

Through the above adjustment, the UN-PNMF can help to save storage space and improve the computation efficiency as well. There are two iteration vectors **P** and **W** in the process of solving (11), but only one iteration vector **W** in the process of solving (12). Moreover, in the later experiments, it is also proved that (12) is more efficient than (11) in negative link prediction, so (12) can be the objective function of this problem. By removing constants in the objective function, (12) can be rewritten as

$$F = tr \left(-2\left(\mathbf{G}_{n}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t} + \left(\mathbf{G}_{n}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t}\right)$$

$$+ \lambda tr \left(-2\left(\mathbf{G}_{n}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \left(\mathbf{In_Out}^{t}\right)$$

$$+ \left(\mathbf{In_Out}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{W} \mathbf{W}^{T} \left(\mathbf{In_Out}^{t}\right)\right)$$

$$+ \alpha tr \left(\mathbf{W} \mathbf{W}^{T}\right)$$

$$(13)$$

It can be clearly seen from (13) that this objective function only has one unknown variable **W**. However, the traditional methods of matrix factorization have two unknown variables, which is difficult to find optimal solutions for two variants simultaneously. For (12), the **W** can be updated according to the following updating rule:

$$\mathbf{W}(i,j) \longleftarrow \mathbf{W}(i,j) \sqrt{\frac{\mathbf{A}(i,j)}{\mathbf{B}(i,j)}}$$
 (14)

where A and B are defined as

$$\mathbf{A} = 2\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\mathbf{W}$$

$$+ \lambda\left(\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{In_{-}Out}^{t}\right)^{T} + \left(\mathbf{In_{-}Out}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\right)\mathbf{W}$$

$$\mathbf{B} = \left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\mathbf{W}\mathbf{W}^{T}\mathbf{W} + \mathbf{W}\mathbf{W}^{T}\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\mathbf{W}$$

$$+ \alpha\mathbf{W} + \lambda\mathbf{W}\mathbf{W}^{T}\left(\mathbf{In_{-}Out}^{t}\right)\left(\mathbf{In_{-}Out}^{t}\right)^{T}\mathbf{W}$$

$$+ \lambda\left(\mathbf{In_{-}Out}^{t}\right)\left(\mathbf{In_{-}Out}^{t}\right)^{T}\mathbf{W}\mathbf{W}^{T}\mathbf{W}$$

$$(15)$$

To ensure the final W is the optimum solution to the objective function, next the correctness of the updating rules in (14) shall be proved by showing that the final W would

satisfy the KKT condition. The Lagrangian function of (12) can be written as follows:

$$L_{F} = tr\left(-2\left(\mathbf{G}_{n}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t}\right)$$

$$+\left(\mathbf{G}_{n}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{G}_{n}^{t}\right)$$

$$+\lambda tr\left(-2\left(\mathbf{G}_{n}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T}\left(\mathbf{In_Out}^{t}\right)\right)$$

$$+\left(\mathbf{In_Out}^{t}\right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{W} \mathbf{W}^{T}\left(\mathbf{In_Out}^{t}\right)\right)$$

$$+\alpha tr\left(\mathbf{W} \mathbf{W}^{T}\right) - tr\left(\Lambda \mathbf{W}\right)$$
(16)

where the Λ is the Lagrangian multiplier for nonnegative **W**. Take the derivative of (16) with respect to **W**; then we have

$$\frac{\partial L_F}{\partial \mathbf{W}} = 2\left(-2\left(\mathbf{G}_n^t\right)\left(\mathbf{G}_n^t\right)^T \mathbf{W} + \left(\mathbf{G}_n^t\right)\left(\mathbf{G}_n^t\right)^T \mathbf{W} \mathbf{W}^T \mathbf{W} + \mathbf{W} \mathbf{W}^T \left(\mathbf{G}_n^t\right)\left(\mathbf{G}_n^t\right)^T \mathbf{W} - \left(\lambda \left(\mathbf{G}_n^t\right)\left(\mathbf{In_Out}^t\right)^T + \lambda \left(\mathbf{In_Out}^t\right)\left(\mathbf{G}_n^t\right)^T\right) \mathbf{W} \\
+ \lambda \mathbf{W} \mathbf{W}^T \left(\mathbf{In_Out}^t\right) \left(\mathbf{In_Out}^t\right)^T \mathbf{W} \\
+ \lambda \left(\mathbf{In_Out}^t\right) \left(\mathbf{In_Out}^t\right)^T \mathbf{W} \mathbf{W}^T \mathbf{W} + \alpha \mathbf{W}\right) - \Lambda^T$$

The KKT complementary condition is

Then, (19) into (18),

$$\mathbf{W}(i,j) \Lambda(i,j) = 0, \quad \forall i \in [1,n], \ j \in [1,k]$$

$$\operatorname{Set} \partial L_{F} / \partial \mathbf{W} = 0,$$

$$2 \left(-2 \left(\mathbf{G}_{n}^{t} \right) \left(\mathbf{G}_{n}^{t} \right)^{T} \mathbf{W} + \left(\mathbf{G}_{n}^{t} \right) \left(\mathbf{G}_{n}^{t} \right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{W} \right)$$

$$+ \mathbf{W} \mathbf{W}^{T} \left(\mathbf{G}_{n}^{t} \right) \left(\mathbf{G}_{n}^{t} \right)^{T} \mathbf{W}$$

$$- \left(\lambda \left(\mathbf{G}_{n}^{t} \right) \left(\mathbf{In_{-}Out}^{t} \right)^{T} + \lambda \left(\mathbf{In_{-}Out}^{t} \right) \left(\mathbf{G}_{n}^{t} \right)^{T} \right) \mathbf{W}$$

$$+ \lambda \mathbf{W} \mathbf{W}^{T} \left(\mathbf{In_{-}Out}^{t} \right) \left(\mathbf{In_{-}Out}^{t} \right)^{T} \mathbf{W}$$

$$+ \lambda \left(\mathbf{In_{-}Out}^{t} \right) \left(\mathbf{In_{-}Out}^{t} \right)^{T} \mathbf{W} \mathbf{W}^{T} \mathbf{W} + \alpha \mathbf{W} \right) = \Lambda$$

$$2\left(-2\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\mathbf{W}+\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\mathbf{W}\mathbf{W}^{T}\mathbf{W}\right)$$

$$+\mathbf{W}\mathbf{W}^{T}\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\mathbf{W}$$

$$-\left(\lambda\left(\mathbf{G}_{n}^{t}\right)\left(\mathbf{In_{-}Out}^{t}\right)^{T}+\lambda\left(\mathbf{In_{-}Out}^{t}\right)\left(\mathbf{G}_{n}^{t}\right)^{T}\right)\mathbf{W}$$

$$+\lambda\mathbf{W}\mathbf{W}^{T}\left(\mathbf{In_{-}Out}^{t}\right)\left(\mathbf{In_{-}Out}^{t}\right)^{T}\mathbf{W}$$

$$+\lambda\left(\mathbf{In_{-}Out}^{t}\right)\left(\mathbf{In_{-}Out}^{t}\right)^{T}\mathbf{W}\mathbf{W}^{T}\mathbf{W}+\alpha\mathbf{W}\right)\odot\Lambda$$

$$=0$$

where \odot is the Hadamard product, e.g., $(\mathbf{A} \odot \mathbf{B})_{i,j} = \mathbf{A}_{i,j} \times \mathbf{B}_{i,j}$. According to gradient decent, it is evident that the updating rule equation (14) satisfies the above KKT condition. Furthermore, since $\mathbf{G}_{\mathbf{n}}^{t}$ and $\mathbf{In}_{\cdot}\mathbf{Out}^{t}$ are nonnegative, \mathbf{W} is negative during the updating process. Because the objective function only has an independent variable \mathbf{W} , it is easy to verify that the updating rule equation (14) is guaranteed to converge.

7.2. UN-PNMF Algorithm. The detailed algorithm for the proposed framework, UN-PNMF, is shown in Algorithm 1. The input of the framework is the adjacency matrix of negative network at time t and hyperparameters α and λ . It constructs the latent factor matrix of user pairs, In_Out^t, in line (1). From line (3) to line (8), the algorithm alternatingly updates W until it achieves convergence. Note that, in practice, Algorithm 1 will stop when reaching predefined maximal iterations or there is little change for the objective function value. After obtaining the optimal W, the final predicted matrix P_G^{t'} can be calculated following line (12). The likelihood of μ_i and μ_j to establish negative link is indicated by P_G^{t'}(i, j). The parameter σ could coordinate between network topology and In_Out of user pairs in predicting negative links. The σ is set to 0.7 and 0.2 in Epinions and Slashdot, respectively.

For this algorithm, each iteration for updating W occupies most of the running time. The updating rules for W may limit the application of this proposed framework, so it is essential to analyze the time complexity. First we consider the time complexity of $\mathbf{A} = 2(\mathbf{G}_n^t)(\mathbf{G}_n^t)^T \mathbf{W} + \lambda((\mathbf{G}_n^t)(\mathbf{G}_n^t)^T \mathbf{W}$ + $(\mathbf{In_Out}^t)^T (\mathbf{G}_n^t)^T \mathbf{W}$). The matrix \mathbf{G}_n^t is very sparse; thus the time complexity of $(\mathbf{G}_n^t)(\mathbf{G}_n^t)^T \mathbf{W}$ and $(\mathbf{In_Out}^t)^T (\mathbf{G}_n^t) \mathbf{W}$ is O(nk) and O(n²k), respectively. The $G_n^t(In_Out^t)^T$ can be computed inO(n²); therefore, the time complexity of A is $O(n^2k)$. For **B**, we can calculate $(G_n^t)(G_n^t)^TWW^TW$ by either $((\mathbf{G}_n^t)(\mathbf{G}_n^t)^T)(\mathbf{W}\mathbf{W}^T)\mathbf{W}$ or $(\mathbf{G}_n^t)(\mathbf{G}_n^t)^T\mathbf{W}\mathbf{W}^T\mathbf{W}$. The former costs $O(n^2k^2)$, while the latter takes $O(n^3k^2)$, so the former is more efficient. Similarly, $(In_Out^t)(In_Out^t)^T(WW^T)W$ can be computed inO(n^2k^2); therefore, the **B** can be computed in $O(n^2k^2)$. Owing to $k \ll n$, the overall time complexity of Algorithm 1 is #iterations $*O(n^2)$. We also counted the time of each iteration for updating W, which costs about 1.699s and 1.689s on Epinions and Slashdot, respectively.

8. Experiments

This section conducts extensive experiments on real-world dynamic networks to evaluate the effectiveness of the framework for negative link prediction. This part also compares different prediction methods and analyzes the impact of parameters. The data and code used in the paper can be available. Anyone could access the data and code by emailing the author.

8.1. Experiment Setting. The experiment setting of the datasets is demonstrated in Figure 2. The whole datasets have been divided into 6 timestamps, i.e., $\mathcal{E} = \{t_5, t_6, t_7, t_8, t_9, t_{10}\}$. \mathbf{G}_n^t has been chosen as old negative network and $\mathbf{G}_n^{t'}$ as the new or missing links negative network which needs to be predicted. The t is varied as $\{t_5, t_6, t_7, t_8, t_9\}$.

This paper follows a common metric for unsupervised link prediction in [2] to evaluate the effectiveness of negative link prediction. In detail, let $N = f(\boldsymbol{G}_n^{t'} - \boldsymbol{G}_n^t)$, where $f(\boldsymbol{M})$ is a function for calculating the number of nonzero elements in matrix \boldsymbol{M} . Each negative link predictor ranks pairs in descending order of confidence and takes the first N pairs as the set of predicted negative links. These pairs' corresponding values are set to 1 in matrix $\boldsymbol{P}_{-}\boldsymbol{G}^{t'}$; the rest of elements of $\boldsymbol{P}_{-}\boldsymbol{G}^{t'}$ are set to 0. Then the prediction accuracy (PA) can be calculated as follows. The $g(\boldsymbol{M})$ refers to the function of the number of the elements which are equal to 2 in matrix \boldsymbol{M} .

$$C = g\left(\mathbf{P}_{-}\mathbf{G}^{t'} + \left(\mathbf{G}^{t'}_{n} - \mathbf{G}^{t}_{n}\right)\right) \tag{21}$$

$$PA = \frac{C}{N} \tag{22}$$

8.2. Comparison of Different Predictors. In this section, to evaluate the effectiveness of the proposed framework UN-PNMF, this paper compares UN-PNMF with several baseline methods, and the detailed descriptions are listed as follows:

- (i) Random: it randomly establishes negative links between two users in signed social network.
- (ii) MF: it is the representative method of traditional matrix factorization, which conducts a matrix factorization on the matrix representation of negative links [27].
- (iii) hTrust: it is an unsupervised framework, which exploits the homophily effect for positive link prediction in social media[2].
- (iv) PMF: it performs a low-rank representation based on the projective matrix factorization as shown in (4).
- (v) UN-PNMF_1: it is a variant of the proposed method, and it embeds network structure representation and latent factor In_Out, respectively, as shown in (11).
- (vi) triNMF: it predicts social links based on nonnegative matrix factorization[2].

For all baseline methods, this paper uses the implementation released by the original authors. Note that this

```
Input: G_n^t, G_p^t, \alpha, \lambda
Output: Rank list of pairs of users
 (1) Construct the latent factor matrix In_Out<sup>t</sup>
 (2) Initialize W randomly
 (3) while Not convergent do
                                      Set \mathbf{A} = 2(\mathbf{G}_n^t)(\mathbf{G}_n^t)^T \mathbf{W} + \lambda((\mathbf{G}_n^t)(\mathbf{In\_Out}^t)^T + (\mathbf{In\_Out}^t)(\mathbf{G}_n^t)^T)\mathbf{W}
                                                  Set \mathbf{B} = (\mathbf{G}_{n}^{t})(\mathbf{G}_{n}^{t})^{T}\mathbf{W}\mathbf{W}^{T}\mathbf{W} + \mathbf{W}\mathbf{W}^{T}(\mathbf{G}_{n}^{t})(\mathbf{G}_{n}^{t})^{T}\mathbf{W} + \alpha \mathbf{W} + \lambda \mathbf{W}\mathbf{W}^{T}(\mathbf{In\_Out}^{t})(\mathbf{In\_Out}^{t})^{T}\mathbf{W} + \alpha \mathbf{W} + \alpha \mathbf{
                                                                                       \lambda(\mathbf{In}_{-}\mathbf{Out}^{t})(\mathbf{In}_{-}\mathbf{Out}^{t})^{T}\mathbf{W}\mathbf{W}^{T}\mathbf{W}
                                                          for i = 1:n do
(6)
(7)
                                                                         for j = 1:k do
                                                                                                   Update W(i, j) \leftarrow W(i, j) \sqrt{A(i, j)/B(i, j)}
(8)
(9)
 (10) end for
(11) end while
(12) Set \mathbf{P}_{-}\mathbf{G}^{t'}(\mathbf{i}, \mathbf{j}) = ((\mathbf{W}\mathbf{W}\mathbf{G}_{n}^{t}) + \sigma(\mathbf{W}\mathbf{W}\mathbf{I}\mathbf{n}_{-}\mathbf{O}\mathbf{u}\mathbf{t}))(\mathbf{i}, \mathbf{j})
(13) Ranking pairs of users according to P_G<sup>t'</sup> in a descending order
```

ALGORITHM 1: The framework UN-PNMF for negative link prediction.

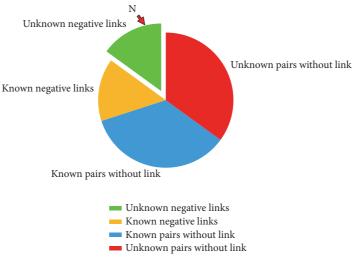


FIGURE 2: Separation of the dataset.

study does not compare the proposed framework with the methods proposed in [13, 17, 20]. Firstly, these methods are either supervised or semisupervised methods. Secondly, these methods use extra sources such as users' interaction data. Although, the hTrust framework is designed for trust prediction, Rotter et al. [22] suggest distrust is a low level of trust, and the essence of our problem is to discover some user pairs which would establish links from huge user pairs without any link; therefore this comparison is significant. This paper calculates the neighbors' similarity of each user pair to replace the homophily of each user pair. This paper empirically sets these parameters as $\{\alpha = 0.5, \lambda = 20, k = 5\}$ and $\{\alpha = 0.5, \lambda = 20, k = 6\}$ in Epionions and Slashdot, respectively. The effectiveness of these parameters will be discussed later. This part uses a random sample of 2000 users as the experiment data. In this section, eight groups of experiments are designed to evaluate the efficiency of different methods. The old negative link matrix G_n^t can be G_n^5 , G_n^6 , G_n^7 , and G_n^8 ,

respectively; the predicting matrix $\mathbf{G}_n^{t'}$ can be \mathbf{G}_n^6 , \mathbf{G}_n^7 , \mathbf{G}_n^8 , \mathbf{G}_n^9 , and \mathbf{G}_n^{10} , respectively, e.g., if $\mathbf{G}_n^t = \mathbf{G}_n^5$, then $\mathbf{G}_n^{t'} = \{\mathbf{G}_n^6, \mathbf{G}_n^7, \mathbf{G}_n^8, \mathbf{G}_n^9, \mathbf{G}_n^{10}\}$, if $\mathbf{G}_n^t = \mathbf{G}_n^6$, then $\mathbf{G}_n^{t'} = \{\mathbf{G}_n^7, \mathbf{G}_n^8, \mathbf{G}_n^9, \mathbf{G}_n^{10}\}$, and so on. To ensure the accuracy and reliability of experiment results, the experiments have been repeated 5 times and the average performance is reported. The comparison results of various unsupervised link prediction algorithms on Slashdot and Epinions are shown in Figure 3.

From Figure 3, we make the following observations:

(i) The proposed UN-PNMF framework almost outperforms all baseline methods. As $\mathbf{G}_n^t = \mathbf{G}_n^5$, the prediction accuracy of UN-PNMF predicting \mathbf{G}_n^6 achieves the maximum value of 16.71% in Slashdot, but the prediction accuracy only reaches 14.06% in Epinions. The average prediction accuracy of UN-PNMF is approximately 3.7% higher than average prediction accuracy of hTrust. Thus it can be seen the similarity



FIGURE 3: Performance comparison for different negative link predictors.

between two users has only a limited effect on the negative link prediction in social network. There are many user pairs with a high similarity, but they are not linked. Moreover, we cannot consider that user pairs without links do not have similarity between them in social network.

- (ii) UN-PNMF_1 is a variant of UN-PNMF. Besides UN-PNMF, UN-PNMF_1 has a better performance compared with other predictors. The average prediction accuracy of UN-PNMF is approximately 1.36% higher than UN-PNMF_1, which shows that jointly embedding network structure and latent factor In_Out are more efficient than embedding them separately. At the same time, it proves that the negative network structure and latent factor of user pairs interacting with each other could be helpful to predict negative links in social media.
- (iii) UN-PNMF always performs far better in predicting negative links than PMF, so it is helpful that embedding the latent factor In_Out can explore some node pairs with negative links. The average prediction accuracy of UN-PNMF is approximately 3.06% higher than PMF, which shows that embedding In_Out can improve the performance of predicting negative links.
- (iv) The performance of MF, UN-PNMF_1, PMF, hTrust, and UN-PNMF is much better than that of random, which supports that modeling negative link properties can improve the performance significantly. The average prediction accuracy of UN-PNMF is much higher than random. However, we find that the performance of predicting long-term negative links is not very good. As $\mathbf{G}_n^t = \mathbf{G}_n^5$, the prediction accuracy of UN-PNMF predicting \mathbf{G}_n^{10} just reaches 9.15% and 9.41% in Slashdot and Epinions, respectively. Therefore, UN-PNMF cannot capture the characteristics of node pairs in dynamic networks very well and is not effective in predicting long-term negative links.

In order to explore the impact of different input data on UN-PNMF, eight groups of experiments are designed to observe the results. Other parameters are fixed, and \mathbf{G}_n^t can be $\{\mathbf{G}_n^5, \mathbf{G}_n^6, \mathbf{G}_n^7, \mathbf{G}_n^8\}$. This part integrates the different experimental results of different \mathbf{G}_n^t , and the comparison results are shown in Figure 4 and Table 2.

The first observation is that, with the increase of t, the performance of the proposed UN-PNMF framework reduces. And it also can be found that UN-PNMF achieves best performance when the input data \mathbf{G}_n^t is equal to \mathbf{G}_n^5 and \mathbf{G}_n^6 in Slashdot and Epinions, respectively. In general, with more old negative links, more effective data information can be learned to predict new negative links, and the predictor should also obtain better performance. However, the experiment results are quite contrary to this situation. By analyzing the reasons, as the negative links add in input matrix, the negative links which need to be predicted become less and less and the sparsity gets more and more serious, so inferring new negative links becomes more and more difficult. It also finds that the prediction accuracy of short-term is much better

than long-term. For example, when the input matrix is \mathbf{G}_n^5 , the accuracy of UN-PNMF predicting \mathbf{G}_n^6 is 16.71%, but the accuracy of UN-PNMF predicting \mathbf{G}_n^{10} is 9.15% in Slashdot. The difference between them is 7.56%; hence the results of each group of experiments present descending trend in Figure 4. With the result of the social network being fast-evolved, the interactive data among users vary from hour to hour, which leads to the decrease of the reference value of current data. Therefore, the prediction accuracy of long-term negative links gets low.

8.3. Parameters Setting. This section investigates the impact of parameters with different values on UN-PNMF framework. Because the values of parameters play an important role in machine learning algorithms, appropriate parameters could improve the performance of algorithms. The proposed framework UN-PNMF includes three parameters which are the dimension of representative vector k, regularization coefficient α , and parameter λ . These parameters are important but are not to tune. The range of regularization coefficient α generally is from 0 to 1, and α empirically is set to 0.5. This part intends to explore the impact of different k on negative link prediction. Because the too large value of k not only cannot reduce the sparsity but also could preserve some noise data; however, the too small value of k must lose some effective information. Finding an appropriate k has great significance for the framework UN-PNMF. Eight groups of experiments are designed to compare the performance of UN_PNMF with different k. The k is varied as {3, 5, 6, 8, 10, 20, 50, 100} and the input data \mathbf{G}_{n}^{t} is fixed to \mathbf{G}_{n}^{5} . The results are shown in Figure 5. Since the selection process of k is similar in two datasets and the space is limited, we take Slashdot dataset as the example.

In general, with the increase of k, the performance of predictors shows similar patterns: first increasing, reaching its peak value and then degrading. These patterns can be used to determine the optimal value of k for UN-PNMF in practice. In Figure 5, it can be observed that

- (i) When k increases from 3 to 6, the performance improves a lot, which shows that, with the increase of k, the low-dimensional vector **W** contains more and more effective information data, and it helps UN-PNMF improve the ability of negative link prediction.
- (ii) UN-PNMF achieves its best performance when k = 6, which shows that the **W** contains relative optimal data information.
- (iii) From k = 6 to k = 100, the performance decreases rapidly. This can be explained by the fact that, with the increase of k, more and more noise data are contained, which is harmful to the effectiveness of UN-PNMF.

Thus it can be seen that, in social network, the negative link prediction is faced with two difficulties, respectively, the data sparseness of links and much noise data of latent factors. Only by finding an appropriate dimension k, can UN-PNMF achieve the best performance.

Parameter λ controls the degree of latent factor In_Out in the formulation of negative links. UN-PNMF can control the influence of latent factor In_Out on predicting new negative

| Slashdot | | | | Epinions | | | | | | |
|----------|--------|--------|--------|----------|--------|--------|--------|--------|--------|--------|
| | t'=6 | t'=7 | t'=8 | t'=9 | t'=10 | t'=6 | t'=7 | t'=8 | t'=9 | t=10 |
| t=5 | 0.1671 | 0.1408 | 0.1194 | 0.1022 | 0.0915 | 0.1406 | 0.1165 | 0.1032 | 0.0993 | 0.0941 |
| t=6 | | 0.1558 | 0.1211 | 0.1080 | 0.0958 | | 0.1429 | 0.1199 | 0.1073 | 0.1025 |
| t=7 | | | 0.1549 | 0.1127 | 0.0901 | | | 0.1357 | 0.1126 | 0.0992 |
| t=8 | | | | 0.1127 | 0.0873 | | | | 0.1336 | 0.0921 |

TABLE 2: Performance comparison for different input G_n^t .

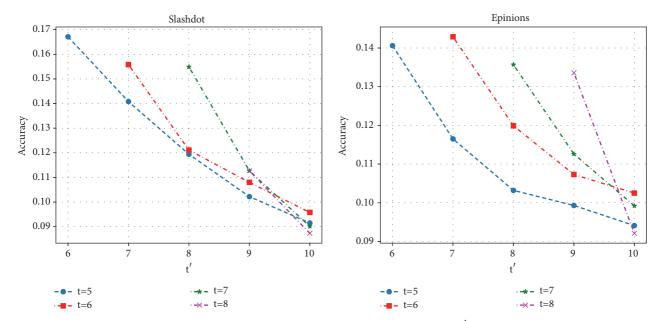


Figure 4: Performance comparison for different input G_n^t .

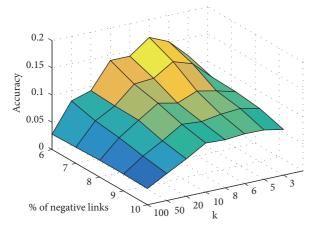


FIGURE 5: Effect of different dimensions k.

links by setting an appropriate value of λ . In the development of social network, the formulation of negative links is a result that might be influenced by multiple latent factors and social network structure. Due to the limited research level about this problem, this paper only introduces the latent factor In_Out. However, with the in-depth research, more efficient latent factors could be found, so controlling parameters setting

is crucial to UN_PNMF framework. If the value of λ is set to be too large, UN-PNMF must exaggerate the effect of latent factor in formulating negative links and mislead the negative link prediction. For example, some user pairs without links have a long distance in network topology, but they would be wrongly predicted to establish negative links only because the framework exaggerates the effect of In_Out. If the value of λ is set to be too small, UN-PNMF would neglect the effect of In_Out in establishing negative links between two users. In order to seek an appropriate value of λ , eight groups of experiments are designed to compare the performance of UN-PNMF with different λ , and the λ is varied as {0.1, 0.5, 1, 5, 10, 20, 50, 100}. The results are shown in Figure 6. Similarly, this part also takes Slashdot dataset as the example. It can be observed, when λ increases from 0.1 to 20, the performance improves a lot, which shows that, with the increase of dimension λ , the effect of In_Out can help to predict new or missing negative links. UN-PNMF achieves its best performance when $\lambda = 20$, which shows that the effect of In_Out is controlled to the relative optimal state. From λ = 20 to λ = 100, the performance decreases rapidly. This can be explained by the fact that, with the increase of λ , the effect of In_Out could be greatly exaggerated. Accordingly, this paper chooses $\lambda = 20$ as the most suitable value in the UN-PNMF framework.

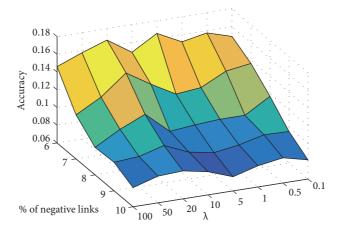


FIGURE 6: Performance of UN-PNMF by varying the λ .

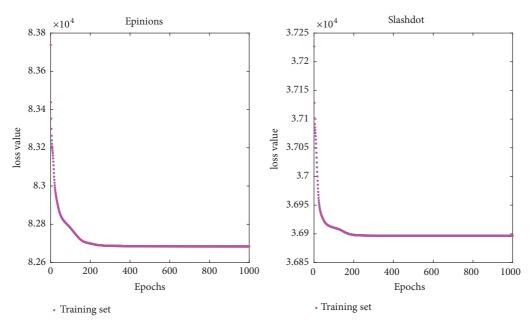


FIGURE 7: Convergence of the UN-PNMF.

This section also investigates the convergence of the UN-PNMF framework. For illustrative purpose, the change of the value of objective function can be drawn for two datasets, in Figure 7. As shown in the graph, the value of the objective function continuously decreases and then stabilizes. The results show that the Algorithm 1 usually converges to a stable value.

9. Conclusion

This paper studies the problem of the negative link prediction based on network embedding, which only focuses on the topology of social network and does not rely on any interaction data. Firstly, the paper seeks to learn the low-dimensional vector representation of social negative network through the projective nonnegative matrix factorization. Secondly, the latent factor In_Out is discovered, which contributes to

the formulation of negative links in social media, and the embedding of latent factor matrix **In_Out** is conducted by a low-dimensional projection operator. Lastly, the network structure and latent factor matrix are embedded into the same low-dimensional space corporately, and an unsupervised framework UN-PNMF is proposed. Extensive experiments are conducted on two datasets from real-world product review sites to evaluate the proposed framework, and the experimental results demonstrate that UN-PNMF consistently outperforms other negative link prediction methods.

The negative links are very sparse, making the negative link prediction very difficult. However, studying the negative links has great significance for the development of social network and the discovery of social problems, so the negative link prediction is worth having further studies. There are several interesting directions that need to be investigated in future. With in-depth analysis in social network, more and

more valuable latent factors of user pairs could be found. The future study can also combine matrix factorization with supervised methods to propose a semisupervised method to predict negative links. A flexible and unified framework for a specific study on social network can also be proposed in the future, which can not only do link prediction but also do other assignments, such as node classification, community detection, and recommendation system.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [Grant no. 61472160], the National Key Technology Research and Development Program of China [Grant no. 2014BAH29F03], the National Natural Science Foundation of China [Grant no. 61872161], the National Natural Science Youth Fund [Grant no. 61602057], the Jilin Province Science and Technology Development Plan Project [Grant no. 2018101328JC], and the Project of Excellent Young Talents Fund of Jilin Provincial Science and Technology Department [Grant no. 20170520059JH].

References

- [1] K. Cheng, J. Li, J. Tang, and H. Liu, "Unsupervised sentiment analysis with signed social networks," in *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pp. 3429–3435, San Francisco, Calif, USA, 2017.
- [2] J. Tang, H. Gao, X. Hu, and H. Liu, "Exploiting homophily effect for trust prediction," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, pp. 53–62, Rome, Italy, 2013.
- [3] Y. Wang, X. Wang, and J. Tang, "Modeling status theory in trust prediction," in *Proceedings of the Twenty-Ninth AAAI* Conference on Artificial Intelligence, pp. 1875–1881, Austin, Tex, USA, 2015.
- [4] A. Garciaduran and M. Niepert, "Learning graph representations with embedding propagation," in *Proceedings of the 31st Conference on Neural Information Processing Systems*, pp. 1–12, Long Beach, Calif, USA, 2017.
- [5] Y. H. Koura, Y. Zhang, and H. Liu, "Competitive interaction model for online social networks' users' data forwarding at a subnet," *Mathematical Problems in Engineering*, vol. 2017, Article ID 3971803, 9 pages, 2017.
- [6] H. Ma, D. Zhou, C. Liu, and I. King, "What your images reveal: exploiting visual contents for point-of-interest recommendation," in *Proceedings of the 26th International World Wide Web Conference Committee*, pp. 391–400, Perth, Australia, 2017.
- [7] L. Wu and H. Liu, "Tracing fake-news footprints: characterizing social media messages by how they propagate," in *Proceedings of*

- the 11th ACM International Conference on Web Search and Data Mining, pp. 637–645, Los Angeles, Calif, USA, 2018.
- [8] X. Wang, Y. Liu, and Y. Nan, "A stable-matching-based user linking method with user preference order," *Mathematical Problems in Engineering*, vol. 2017, Article ID 3247627, 8 pages, 2017.
- [9] Y. Hu, A. John, F. Wang, and S. Kambhampati, "ET-LDA: joint topic modeling for aligning events and their twitter feedback," in *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pp. 59–65, Toronto, Ontario, Canada, 2012.
- [10] C. Huang, M. Liu, H. Gong, and F. Xu, "Season-aware attraction recommendation method with dual-trust enhancement," *Jour*nal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, vol. 33, no. 4, pp. 2437–2449, 2017.
- [11] J. Tang, S. Chang, C. Aggarwal, and H. Liu, "Negative link prediction in social media," in *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, pp. 87–96, Shanghai, China, 2015.
- [12] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New perspectives and methods in link prediction," in *Proceedings of* the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 243–252, Washington, Wash, USA, 2010.
- [13] S. Wang, J. Tang, C. Aggarwal, Y. Chang, and H. Liu, "Signed network embedding in social media," in *Proceedings of the 17th SIAM International Conference on Data Mining*, pp. 327–335, Houston, Tex, USA, 2017.
- [14] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in Proceedings of the 19th International Conference on World Wide Web, pp. 641–650, Raleigh, NC, USA, 2010.
- [15] K. Zolfaghar and A. Aghaie, "A syntactical approach for interpersonal trust prediction in social web applications: combining contextual and structural data," *Knowledge-Based Systems*, vol. 26, pp. 93–102, 2012.
- [16] J. Tang, X. Hu, Y. Chang, and H. Liu, "Predictability of distrust with interaction data," in *Proceedings of the 23rd ACM Interna*tional Conference on Conference on Information and Knowledge Management, pp. 181–190, Shanghai, China, 2014.
- [17] X. Wang, Y. Wang, and H. Sun, "Exploring the combination of Dempster-Shafer theory and neural network for predicting trust and distrust," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 5403105, 12 pages, 2016.
- [18] H. Oh, J. Kim, S. Kim, and K. Lee, "A probability-based trust prediction model using trust message passing," in *Proceedings of the 22nd International Conference on World Wide Web Companion*, pp. 161-162, Janeiro, Brazil, 2013.
- [19] 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 *Proceedings of the 30th AAAI* Conference on Artificial Intelligence, pp. 279–286, Phoenix, Ariz, USA, 2016.
- [20] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1225–1334, San Francisco, Calif, USA, 2016.
- [21] Z. Yang and E. Oja, "Projective nonnegative matrix factorization with α-divergence," in *Proceedings of the 19th International Conference on Artificial Neural Networks: Part I*, pp. 20–29, 2009.
- [22] J. B. Rotter, "Interpersonal trust, trustworthiness, and gullibility," *American Psychologist*, vol. 35, no. 1, pp. 1–7, 1980.

- [23] X. Chen, L. Cao, C. Li, Z. Xu, and J. Lai, "Ensemble network architecture for deep reinforcement learning," *Mathematical Problems in Engineering*, vol. 2018, Article ID 2129393, 6 pages, 2018.
- [24] J. Tang, C. Aggarwal, and H. Liu, "Node classification in signed social networks," in *Proceedings of the Siam International Conference on Data Mining*, pp. 54–62, Fla, USA, 2016.
- [25] B. M. Brentan, E. Campbell, G. L. Meirelles, E. Luvizotto, and J. Izquierdo, "Social network community detection for dma creation: criteria analysis through multilevel optimization," *Mathematical Problems in Engineering*, vol. 2017, Article ID 905323, 12 pages, 2017.
- [26] L. Guo, W. Zuo, T. Peng, and L. Yue, "Text matching and categorization: mining implicit semantic knowledge from treeshape structures," *Mathematical Problems in Engineering*, vol. 2015, Article ID 723469, 9 pages, 2015.
- [27] S. Zhu, K. Yu, Y. Chi, and Y. Gong, "Combining content and link for classification using matrix factorization," in *Proceedings* of the SIGIR, pp. 487–494, Amsterdam, The Netherlands, 2007.

















Submit your manuscripts at www.hindawi.com























