Unsupervised Feature Selection in Signed Social Networks

Kewei Cheng*
Arizona State University
kcheng18@asu.edu

Jundong Li*
Arizona State University
jundongl@asu.edu

Huan Liu
Arizona State University
huan.liu@asu.edu

ABSTRACT

The rapid growth of social media services brings a large amount of high-dimensional social media data at an unprecedented rate. Feature selection is powerful to prepare high-dimensional data by finding a subset of relevant features. A vast majority of existing feature selection algorithms for social media data exclusively focus on positive interactions among linked instances such as friendships and user following relations. However, in many real-world social networks, instances may also be negatively interconnected. Recent work shows that negative links have an added value over positive links in advancing many learning tasks. In this paper, we study a novel problem of unsupervised feature selection in signed social networks and propose a novel framework SignedFS. In particular, we provide a principled way to model positive and negative links for user latent representation learning. Then we embed the user latent representations into feature selection when label information is not available. Also, we revisit the principle of homophily and balance theory in signed social networks and incorporate the signed graph regularization into the feature selection framework to capture the first-order and the second-order proximity among users in signed social networks. Experiments on two real-world signed social networks demonstrate the effectiveness of our proposed framework. Further experiments are conducted to understand the impacts of different components of SignedFS.

KEYWORDS

Feature Selection; Signed Social Networks; Unsupervised Learning

ACM Reference format:
Kewei Cheng*, Jundong Li*, and Huan Liu. 2017. Unsupervised Feature Selection in Signed Social Networks. In *Proceedings of KDD '17, Halifax, NS, Canada, August 13-17, 2017,* 10 pages.
https://doi.org/10.1145/3097983.3098106

1 INTRODUCTION

With the rise of online social networks such as Facebook and Twitter, social network analysis has gained increasing attention in recent years. Huge volumes of data are user-generated at an unprecedented speed. For example, over 500 terabyte is are generated on Facebook every day and around 6000 tweets are tweeted on Twitter every second. The massive amount of high-dimensional social media data (e.g., posts, images, videos) presents challenges to traditional data mining and machine learning tasks due to the curse of dimensionality [8]. As a traditional way to reduce the dimensionality of data, feature selection shows its effectiveness by selecting a subset of relevant features for a compact and accurate representation [4, 19, 23, 27]. In addition to the rich source of feature information of high dimensionality, social media data is inherently linked, which contradicts the widely adopted data i.i.d. assumption of conventional feature selection algorithms. Also, in social networks, features and links are not independently presented but are strongly correlated, and the root cause of the correlations can be explained by some social science theories such as social influence and the principle of homophily [7, 30]. These theories are helpful in identifying relevant features in feature selection, especially in cases when label information is costly to obtain.

A vast majority of existing feature selection methods for social media data mainly leverage positive links among users to assess feature relevance. However, in addition to positive links, many real-world social networks may also contain negative links, such as the distrust relations in Epinions[1] and the foe links in Slashdot[2]. Recent work shows that negative links have additional value over positive interactions [32], which could be used to advance a variety of applications such as recommendation [31], sentiment analysis [5] and community detection [17, 25]. Recent advances of signed social network analysis motivate us to investigate if negative links can help us find relevant features for users in signed networks when label information is not available.

The existence of negative links in signed social networks not only brings potential opportunities but also presents great challenges for feature selection. First of all, without label information, existing feature selection methods mainly extract latent representations from positive links and then employ these latent representations to guide feature selection on the content space. However, negative links are distinct from positive links with unique properties. It is a difficult task to accurately model latent representations from negative links for feature selection. Secondly, most existing methods perform feature selection based on social theories for unsigned social networks. Yet, such theories may not be directly applicable to signed social networks. Therefore, the above mentioned challenges make feature selection in signed social networks a nontrivial task that needs further investigation.

In this paper, we study a novel problem of unsupervised feature selection in signed social networks, which has not been studied previously. In particular, we focus on answering the following two questions: (1) how to employ and adapt existing social science theories for feature selection in signed social networks? (2) how to

*Equal Contribution

[1] http://www.epinions.com/
[2] https://slashdot.org/

KDD '17, August 13-17, 2017, Halifax, NS, Canada

ACM ISBN 978-1-4503-4887-4/17/08…$15.00
https://doi.org/10.1145/3097983.3098106

mathematically model both positive and negative links for feature selection? To answer these two questions, we propose a novel unsupervised feature selection framework - SignedFS. The main contributions of this paper are summarized as follows:

- We verify the existence of user first-order proximity and second-order proximity in signed social networks;
- We propose an unsupervised feature selection framework SignedFS, which embeds the latent representation learning from positive and negative links into feature selection;
- We provide an effective alternating optimization algorithm for the proposed SignedFS framework;
- We evaluate the effectiveness of the proposed SignedFS framework on real-world datasets.

The rest of the paper is organized as follows. In Section 2, we formally define the problem of unsupervised feature selection in signed social networks. We introduce the datasets and conduct preliminary data analysis in Section 3. In Section 4, we present our proposed SignedFS framework for unsupervised feature selection in signed social networks. In Section 5, we provide an effective alternating optimization algorithm for the proposed SignedFS framework. Section 6 gives the experimental evaluation on two real-world signed social networks. Section 7 briefly reviews related work and Section 8 concludes the whole paper with future work.

2 PROBLEM STATEMENT

In this section, we first present the notations and then formally define the problem of unsupervised feature selection in signed social networks. We use bold uppercase characters for matrices (e.g., $\mathbf{A}$), bold lowercase characters for vectors (e.g., $\mathbf{a}$), normal lowercase characters for scalars (e.g., a). Also, We represent the i-th row of matrix $\mathbf{A}$ as $\mathbf{A}_{i*}$, the j-th column as $\mathbf{A}_{*j}$, the (i, j)-th entry as $\mathbf{A}_{ij}$, transpose as $\mathbf{A}'$, trace as $tr(\mathbf{A})$ if $\mathbf{A}$ is a square matrix. For any matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$, its Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{d} \mathbf{A}_{ij}^2}$, its $\ell_{2,1}$-norm is $\|\mathbf{A}\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{d} \mathbf{A}_{ij}^2}$. $\mathbf{I}_n$ denotes the identity matrix of size n-by-n.

Let $\mathcal{U} = \{u_1, u_2, ..., u_n\}$ be the set of n users in a signed social network $\mathcal{G}$. $\mathcal{G}$ can be decomposed into a positive component $\mathcal{G}_p$ and a negative component $\mathcal{G}_n$ in which $\mathbf{A}^p \in \mathbb{R}^{n \times n}$ is the corresponding adjacency matrix for the positive component $\mathcal{G}_p$ such that $\mathbf{A}_{ij}^p = 1$ if u_i has a positive link to u_j, and $\mathbf{A}_{ij}^p = 0$ otherwise. Similarly, $\mathbf{A}^n \in \mathbb{R}^{n \times n}$ denotes the adjacency matrix of $\mathcal{G}_n$ where $\mathbf{A}_{ij}^n = 1$ if u_i has a negative link to u_j, and $\mathbf{A}_{ij}^n = 0$ otherwise. Let $\mathcal{F} = \{f_1, f_2, ..., f_d\}$ be a set of d features and $\mathbf{X} \in \mathbb{R}^{n \times d}$ denotes the feature information of all n instances. With these notations, the problem of unsupervised feature selection in signed social networks can be formally stated as follows:

Given: feature set $\mathcal{F}$, feature matrix $\mathbf{X}$ and a signed social network $\mathcal{G}$ with positive links encoded in the adjacency matrix $\mathbf{A}^p$ and negative links encoded in the adjacency matrix $\mathbf{A}^n$;

Select: a subset of most relevant features $\mathcal{S} \subseteq \mathcal{F}$ by exploiting both feature matrix $\mathbf{X}$ and signed network structure encoded in the adjacency matrix $\mathbf{A}^p$ and the adjacency matrix $\mathbf{A}^n$.

Table 1: Statistics of the used datasets.

Datasets	Epinions	Wiki-rfa
# of Users	7,140	7,096
# of Features	15,069	10,608
# of Classes (Ground Truth)	24	2
# of Positive Links	13,569	104,555
Density of Positive Links	2.7e(-4)	2.1e(-3)
# of Negative Links	3,010	23,516
Density of Negative Links	5.9e(-5)	4.7e(-4)

3 ANALYSIS OF SIGNED SOCIAL NETWORKS

In this section, we first introduce two real-world signed social networks used in this study and then investigate the first-order and the second-order proximity among users in signed social networks.

3.1 Datasets

We use two real-world datasets from Epinions[3] and Wiki-rfa[4] which include both positive and negative links.

Epinions: Epinions is a customer review website where users share their reviews about products. Users can both trust or distrust other users. Users can also write reviews for products from different categories. We collect positive and negative links between users to construct the signed social networks. For each user, features are formed by the unigram model based on all review comments posted by the user. To be specific, each feature denotes the frequency that a particular word appears, and the words that appear less than 10 times have already been removed. The major categories of reviews by users are taken as the ground truth of class labels.

Wiki-rfa: Wikipedia Requests for Adminship is a who-votes-for-whom network where a signed link indicates a positive or a negative vote by one user on the promotion of the other one. Each vote is typically accompanied by a short comment. Similarly, for each user, features are formed by the unigram model based on all comments posted by the user. The person voted by the user could be rejected or accepted, which is taken as the ground truth of labels.

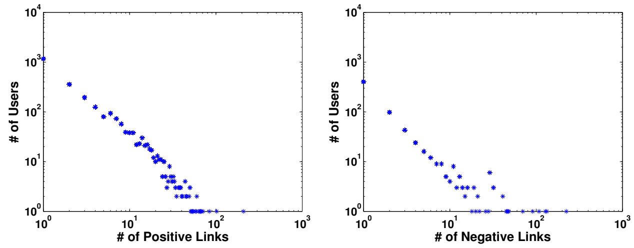
In order to better understand the distributions of positive and negative links, we further explore the degree distributions of these two types of links in Figure 1 and Figure 2. As can be observed from the figures, both positive and negative links present a power-law distribution, which is typical in most social networks [1].

Detailed statistics of these two datasets are presented in Table 1. With these two datasets, we now study the first-order user proximity and the second-order user proximity in signed social networks.

3.2 Analysis

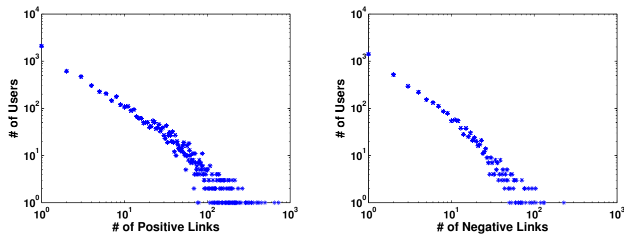
Social science theories such as the principle of homophily [28] and balance theory [14] suggest the correlations between user similarity and user positive/negative interactions. Hence, these theories are widely adopted in social network analysis. For example, the principle of homophily looks at all observed links, and implies that positively connected users are more likely to be similar to each

[3]http://jiliang.xyz/trust.html
[4]https://snap.stanford.edu/data/wiki-RfA.html



(a) (b)

Figure 1: Power-law distribution of the Epinions dataset: (a) degree distribution of positive links; (b) degree distribution of negative links.



(a) (b)

Figure 2: Power-law distribution of the Wiki-rfa dataset: (a) degree distribution of positive links; (b) degree distribution of negative links.

other than two randomly selected users. Second, balance theory looks at two-hop users' interactions and the observations are that "friends of friends are friends" and "enemies of enemies are friends". In this subsection, we revisit the principle of homophily and the balance theory in signed social networks and show how user similarity is correlated with the first-order and the second-order user proximity in signed social networks.

3.2.1 First-order Proximity. The principle of homophily indicates that users are similar to each other when they are interconnected. However, it is not appropriate to apply the principle directly on signed social networks [34] as instances may also be negatively connected. To explore the first-order user proximity in signed social networks, we revisit the principle of homophily by attempting to answering the following question: are users with positive relations tend to be more similar than users with negative relations?

To answer the question, we first define the similarity score between two users u_i and u_j as $sim_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|_2$, where $\mathbf{y}_i \in \mathbb{R}^{1\times c}$ and $\mathbf{y}_j \in \mathbb{R}^{1\times c}$ are the ground truth of labels for users u_i and u_j, respectively. c denotes the number of user labels.

With the definition of user similarity, we construct two vectors $\mathbf{p}_1$ and $\mathbf{n}_1$ of the same length to denote the user similarity between positively connected users and negatively connected users, respectively. To be specific, elements in $\mathbf{p}_1$ denote the similarity score between two users (u_i, u_j) with positive relations, and elements in $\mathbf{n}_1$ denote the similarity score between two users (u_i, u_j) with negative relations. We sample 500 pairs of users for both vectors $\mathbf{p}_1$ and $\mathbf{n}_1$ and conduct two sample t-test on these two vectors. The null hypothesis is rejected at the significance level of $\alpha = 0.01$ with a p-value shown in Table 2. Therefore, we verify the assumption that users with positive relations are more similar to each other than users with negative relations.

Table 2: The p-values of the t-test, where $\bar{\mathbf{p}}_1$, $\bar{\mathbf{n}}_1$, $\bar{\mathbf{p}}_2$, $\bar{\mathbf{n}}_2$ and $\bar{\mathbf{r}}$ denote the sample mean of the corresponding vectors.

Hypothesis	Epinions	Wiki-rfa
H_0: $\bar{\mathbf{p}}_1 >= \bar{\mathbf{n}}_1$; H_1: $\bar{\mathbf{p}}_1 < \bar{\mathbf{n}}_1$	$2.3974e(-7)$	$8.3255e(-4)$
H_0: $\bar{\mathbf{p}}_2 >= \bar{\mathbf{r}}$; H_1: $\bar{\mathbf{p}}_2 < \bar{\mathbf{r}}$	$1.3614e(-6)$	$9.8577e(-7)$
H_0: $\bar{\mathbf{n}}_2 >= \bar{\mathbf{r}}$; H_1: $\bar{\mathbf{n}}_2 < \bar{\mathbf{r}}$	$5.5854e(-5)$	$1.3126e(-12)$

3.2.2 Second-order Proximity. Balance theory in signed social networks suggests that "a friend of my friend is my friend" and "an enemy of my enemy is my friend". Based on the balance theory, we would like to investigate the correlation between user similarity and the second-order user proximity in signed social networks. Specifically, we aim to answer the following two questions: (1) are friends of my friends tend to be similar to me? (2) are enemies of my enemies tend to be similar to me?

With user similarity defined in section 3.2.1, we construct another three vectors $\mathbf{p}_2$, $\mathbf{n}_2$ and $\mathbf{r}$ to denote the user similarity between two users with a shared friend, two users with a shared enemy and two randomly selected users, respectively. For example, each element in $\mathbf{p}_2$ denotes the similarity score between two users u_i and u_k. Both u_i and u_k have a friend u_j. The element in $\mathbf{n}_2$ denotes the similarity score between two users u_i and u_k. Both u_i and u_k have an enemy u_j. And the element of $\mathbf{r}$ represents the similarity score between u_i and another randomly selected user u_r. We also sample 500 pairs of users for all these three vectors $\mathbf{p}_2$, $\mathbf{n}_2$ and $\mathbf{r}$ and then conduct two sample t-test on these three vectors. The null hypothesis is rejected at the significance level of $\alpha = 0.01$ with p-values shown in Table 2. From the table, we can conclude that both the friends of my friends and the enemies of my enemies are more similar to me than a randomly selected user.

4 THE PROPOSED FRAMEWORK

In this section, we illustrate our proposed framework SignedFS for unsupervised feature selection in signed social networks in details. As shown in Figure 3, it consists of three components: first, we show how to learn user latent representations from both positive and negative links (Section 4.1); second, we show how to embed the user latent representations into feature selection when we are lack of label information (Section 4.2); third, we show how to employ the first-order and the second-order user proximity in signed social networks to make the learned user latent representations consistent with the user proximity in signed social networks via a signed graph regularization (Section 4.3).

4.1 Modeling Positive and Negative Links

In social media, a user establish relations with others due to a variety of hidden factors. These hidden factors can be the hobbies, geographical locations, religions, etc. It has been widely studied

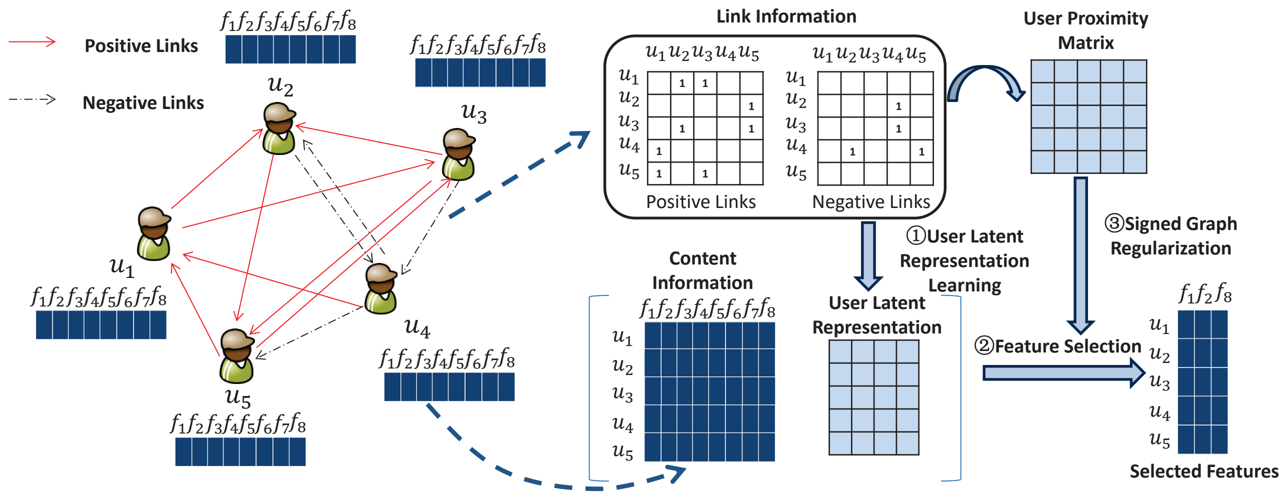


Figure 3: Illustration of the proposed SignedFS framework.

in previous research that both positive and negative links are relevant to these hidden factors [34]. According to the recent research, negative links possess some unique characteristics against positive links. For example, networks with negative links often exhibit distinct topological properties such as lower clustering coefficients compared with networks with only positive links [33]. Thus, we attempt to model positive and negative links independently to learn user latent representations (phase 1 in Figure 3). Let $\mathbf{U} \in \mathbb{R}^{n\times c}$ be the user latent representations where $\mathbf{U}_{i*} \in \mathbb{R}^{1\times c}$ denotes user latent representation of u_i and c is the number of user latent factors. It should be noticed that in real-world signed social networks, each user only has a limited number of interactions with others, which results in a sparse and low-rank network structure. Therefore, we employ low-rank matrix factorization to learn user latent representations. Specifically, to capture the distinct properties of positive and negative links, we collectively factorize $\mathbf{A}^p$ and $\mathbf{A}^n$ into a unified low-rank representation $\mathbf{U}$ via solving the following optimization problem:

$$\min_{\mathbf{U},\mathbf{V}^p,\mathbf{V}^n} \beta^+ \|\mathbf{O}^p \odot (\mathbf{A}^p - \mathbf{U}\mathbf{V}^p\mathbf{U}')\|_F^2 + \beta^- \|\mathbf{O}^n \odot (\mathbf{A}^n - \mathbf{U}\mathbf{V}^n\mathbf{U}')\|_F^2, \quad (1)$$

where β^+ and β^- balance the contribution of positive links and negative links in learning user latent representations, respectively. $\mathbf{O}^p$ and $\mathbf{O}^n$ are defined as follows:

$$\mathbf{O}^p_{ij} = \begin{cases} 1, & \text{if } \mathbf{A}^p_{ij} = 1 \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

$$\mathbf{O}^n_{ij} = \begin{cases} 1, & \text{if } \mathbf{A}^n_{ij} = 1 \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

In the above formulation, we approximate the positive link from u_i to u_j with $\mathbf{U}_i\mathbf{V}^p\mathbf{U}'_j$ where $\mathbf{V}^p \in \mathbb{R}^{c\times c}$ captures the correlations among user latent representations for positive links. $\odot$ is Hadamard product (element-wise product) where $(\mathbf{X} \odot \mathbf{Y})_{ij} = \mathbf{X}_{ij} \times \mathbf{Y}_{ij}$ for any two matrices $\mathbf{X}$ and $\mathbf{Y}$ of the same size. The Hadamard product operator is imposed since we only use observed positive links to learn user latent representations. Similarly, we approximate negative links with $\mathbf{U}\mathbf{V}^n\mathbf{U}'$. Since negative links are also related to user latent representations, we factorize $\mathbf{A}^n$ into the same low-rank space $\mathbf{U}$. The correlation matrix $\mathbf{V}^n$ is used to capture the unique properties of negative links.

4.2 Modeling Feature Information

After we model user latent representations, we now show how to make use of them to guide feature selection (phase 2 in Figure 3). Typically, in social media platforms, label information of users are costly and labor intensive to obtain. Without label information, it would be difficult to assess feature relevance. Fortunately, these latent representations encode the signed network structure which selected feature should preserve. Hence, we leverage the user latent representations $\mathbf{U}$ to guide feature selection via a multivariate linear regression model with a $\ell_{2,1}$-norm sparse regularization term [22]:

$$\min_{\mathbf{W}} \|\mathbf{X}\mathbf{W} - \mathbf{U}\|_F^2 + \alpha\|\mathbf{W}\|_{2,1}, \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{d\times c}$ is a feature weight matrix and each row of $\mathbf{W}$, i.e., $\mathbf{W}_{i*}$, measures the importance of the i-th feature. The $\ell_{2,1}$-norm regularization term is imposed on $\mathbf{W}$ to achieve a joint feature sparsity across c different dimensions of user latent representations. The parameter α controls the sparsity of the model.

4.3 Modeling User Proximity

In Section 3, we revisit the principle of homophily and balance theory. We verify the existence of the first-order and second-order user proximity in signed social networks. In this subsection, we introduce how to model the first-order and the second-order user proximity for unsupervised feature selection in signed social networks (phase 3 in Figure 3).

In particular, we have the following two important findings: (1) first-order proximity: two users with positive relations tend to be similar to each other than two users with negative relations; (2) second-order proximity: the friends of my friends and the enemies of my enemies are more similar to me than a randomly selected user. We first construct a user proximity matrix by employing both the first-order and the second-order user proximity. Given the adjacency matrix of a signed network $\mathbf{A}$ where $\mathbf{A}_{ij} = 1$, $\mathbf{A}_{ij} = -1$ and $\mathbf{A}_{ij} = 0$ denote positive, negative and missing links from u_i to u_j respectively, the first-order proximity matrix $\mathbf{P_1}$ is defined as $\mathbf{P_1} = \mathbf{A}$, where $\mathbf{P_1}_{ij} = 1$ indicates that u_j is a friend of u_i while $\mathbf{P_1}_{ij} = -1$ indicates that u_j is a foe of u_i. The second-order proximity matrix is defined as $\mathbf{P_2} = \mathbf{O} \odot \mathbf{A}^2$, where $\mathbf{O}$ is defined as follows:

$$\mathbf{O}_{ij} = \begin{cases} 0, & \text{if } \mathbf{P_1}_{ij} \neq 0 \text{ or } \mathbf{P_2}_{ij} < 0 \\ 1, & \text{otherwise} \end{cases}. \tag{5}$$

In the above formulation, we capture the second-order proximity from u_i to u_k by $(\mathbf{A}^2)_{ik} = \sum_{j=1}^{n} \mathbf{A}_{ij}\mathbf{A}_{jk}$. To show that $\mathbf{A}^2$ can capture the second-order user proximity, the proof is as follows: (1) to verify that $\mathbf{A}^2$ can capture the proximity between a friend of my friend and me, we should prove that if both u_i and u_k have a friend u_j, u_i and u_j should be similar to each other in the second-order proximity matrix. In other words, if $sgn(\mathbf{A}_{ij}) = 1$ and $sgn(\mathbf{A}_{jk}) = 1$, we should prove that $sgn(\mathbf{A}_{ik}) = 1$, which is obvious in the above formulation; (2) to verify that $\mathbf{A}^2$ can capture the proximity between an enemy of my enemy and me, we should prove that if both u_i and u_k have an enemy u_j, u_i and u_k should be similar to each other in the second-order proximity matrix. That is if $sgn(\mathbf{A}_{ij}) = -1$ and $sgn(\mathbf{A}_{jk}) = -1$, we should prove that $sgn(\mathbf{A}_{ik}) = 1$, which is also true in the above formulation. Though the second-order proximity (balance theory) may not always hold in signed networks [32], in an aggregate sense, the second-order proximity from network structure should be maximally preserved. Thus $(\mathbf{A}^2)_{ik} > 0$ can capture the second-order proximity from u_i to u_k. The Hadamard product operator is imposed to avoid the confliction between the first-order proximity and the second-order proximity. In this way, the user proximity matrix can be constructed by $\mathbf{P} = \mathbf{P_1} + \theta\mathbf{P_2}$. The parameter θ controls the weight of the first-order and the second-order proximity matrices in the model. In this paper, we empirically set the weight $\theta = 0.1$.

To integrate user proximity in feature selection, the basic idea is to make latent representations $\mathbf{U}_{i*}$ and $\mathbf{U}_{j*}$ of two users as close as possible if u_i and u_j are similar while as far as possible if u_i and u_j are dissimilar. It could be mathematically formulated by the signed graph regularization [17]:

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} |\mathbf{P}_{ij}| \times \|\mathbf{U}_{i*} - sgn(\mathbf{P}_{ij})\mathbf{U}_{j*}\|_2^2 = tr(\mathbf{U}'\mathbf{L}\mathbf{U}), \tag{6}$$

where $sgn(\mathbf{P}_{ij})$ denotes the sign of $\mathbf{P}_{ij}$. $\mathbf{L} = \mathbf{D} - \mathbf{P}$ is a signed Laplacian matrix [17] constructed from $\mathbf{P}$ and the signed degree matrix $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with $\mathbf{D}_{ii} = \sum_{j=1}^{n} |\mathbf{P}_{ij}|$.

With the modeling of user proximity by signed graph regularization, the final objective function of the proposed SignedFS framework is formulated as follows:

$$\begin{aligned} \min_{\mathbf{W},\mathbf{U},\mathbf{V}^p,\mathbf{V}^n} & \|\mathbf{X}\mathbf{W} - \mathbf{U}\|_F^2 + \alpha\|\mathbf{W}\|_{2,1} + \frac{\gamma}{2}tr(\mathbf{U}'\mathbf{L}\mathbf{U}) \\ & + \frac{\beta^+}{2}\|\mathbf{O}^p \odot (\mathbf{A}^p - \mathbf{U}\mathbf{V}^p\mathbf{U}')\|_F^2 \\ & + \frac{\beta^-}{2}\|\mathbf{O}^n \odot (\mathbf{A}^n - \mathbf{U}\mathbf{V}^n\mathbf{U}')\|_F^2, \end{aligned} \tag{7}$$

where γ is a regularization parameter for the modeling of user proximity in signed social networks.

5 OPTIMIZATION

In this section, we introduce an effective alternating optimization algorithm for solving the optimization problem of the proposed SignedFS framework with a convergence analysis.

5.1 Alternating Optimization Algorithm

In Eq.(7), the coupling between $\mathbf{U}$, $\mathbf{V}^p$, $\mathbf{V}^n$ and $\mathbf{W}$ makes it difficult to find the global optimal solutions for all four variables simultaneously. Therefore, we propose to employ an alternating optimization scheme to solve it which has been widely adopted for a variety of real-world problems. First, we fix $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ and update $\mathbf{W}$. Specifically, when $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ are fixed, the objective function is convex w.r.t. the feature weight matrix $\mathbf{W}$. We take the partial derivative of objective function w.r.t. $\mathbf{W}$ and set it to be zero:

$$2\mathbf{X}'(\mathbf{X}\mathbf{W} - \mathbf{U}) + 2\alpha\mathbf{H}\mathbf{W} = 0, \tag{8}$$

where $\mathbf{H} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with its i-th diagonal element as:

$$\mathbf{H}_{ii} = \frac{1}{2\|\mathbf{W}_{i*}\|_2}{}^5. \tag{9}$$

It can be noticed that $\mathbf{X}'\mathbf{X}$ is a positive semidefinite matrix and $\alpha\mathbf{H}$ is a diagonal matrix with positive entries which is positive semidefinite as well. Therefore, their summation should also be positive semidefinite. Hence, $\mathbf{W}$ has a closed-form solution:

$$\mathbf{W} = (\mathbf{X}'\mathbf{X} + \alpha\mathbf{H})^{-1}\mathbf{X}'\mathbf{U}. \tag{10}$$

By substituting the above solution of $\mathbf{W}$ into Eq.(7), we have:

$$\begin{aligned} & \min_{\mathbf{U},\mathbf{V}^p,\mathbf{V}^n} \mathcal{J}(\mathbf{U},\mathbf{V}^p,\mathbf{V}^n) \\ = & \, tr(\mathbf{U}'\mathbf{U}) - tr(\mathbf{U}'\mathbf{X}\mathbf{M}^{-1}\mathbf{X}'\mathbf{U}) + \frac{\beta^+}{2}\|\mathbf{O}^p \odot (\mathbf{A}^p - \mathbf{U}\mathbf{V}^p\mathbf{U}')\|_F^2 \\ & + \frac{\beta^-}{2}\|\mathbf{O}^n \odot (\mathbf{A}^n - \mathbf{U}\mathbf{V}^n\mathbf{U}')\|_F^2 + \frac{\gamma}{2}tr(\mathbf{U}'\mathbf{L}\mathbf{U}) \\ = & \, tr(\mathbf{U}'(\mathbf{I}_n - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')\mathbf{U}) + \frac{\beta^+}{2}\|\mathbf{O}^p \odot (\mathbf{A}^p - \mathbf{U}\mathbf{V}^p\mathbf{U}'))\|_F^2 \\ & + \frac{\beta^-}{2}\|\mathbf{O}^n \odot (\mathbf{A}^n - \mathbf{U}\mathbf{V}^n\mathbf{U}'))\|_F^2 + \frac{\gamma}{2}tr(\mathbf{U}'\mathbf{L}\mathbf{U}), \end{aligned} \tag{11}$$

where $\mathbf{M} = \mathbf{X}'\mathbf{X} + \alpha\mathbf{H}$.

Similarly, we fix other variables to update $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ iteratively. Since their closed-form solutions are hard to obtain, we employ the gradient descent method to update them. In particular, the partial derivative of the objective function w.r.t. $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ can be calculated as follows:

[5] In practice, $\|\mathbf{W}_{i*}\|_2$ could be close to zero. Thus, we make $\mathbf{H}_{ii} = \frac{1}{2\|\mathbf{W}_{i*}\|_2 + \epsilon}$, where ϵ is a very small constant.

$$\frac{\partial \mathcal{J}}{\partial \mathbf{U}} = (\mathbf{I}_n - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')\mathbf{U} + (\mathbf{I}_n - \mathbf{X}\mathbf{M}^{-1}\mathbf{X}')'\mathbf{U}$$
$$+\beta^{+}(-(\mathbf{O}^p \odot \mathbf{O}^p \odot \mathbf{A}^p)\mathbf{U}\mathbf{V}^{p\prime} - (\mathbf{O}^p \odot \mathbf{O}^p \odot \mathbf{A}^p)'\mathbf{U}\mathbf{V}^p$$
$$+(\mathbf{O}^p \odot \mathbf{O}^p \odot \mathbf{U}\mathbf{V}^p\mathbf{U}')\mathbf{U}\mathbf{V}^{p\prime}$$
$$+(\mathbf{O}^p \odot \mathbf{O}^p \odot \mathbf{U}\mathbf{V}^p\mathbf{U}')'\mathbf{U}\mathbf{V}^p)$$
$$+\beta^{-}(-(\mathbf{O}^n \odot \mathbf{O}^n \odot \mathbf{A}^n)\mathbf{U}\mathbf{V}^{n\prime} - (\mathbf{O}^n \odot \mathbf{O}^n \odot \mathbf{A}^n)'\mathbf{U}\mathbf{V}^n$$
$$+(\mathbf{O}^n \odot \mathbf{O}^n \odot \mathbf{U}\mathbf{V}^n\mathbf{U}')\mathbf{U}\mathbf{V}^{n\prime}$$
$$+(\mathbf{O}^n \odot \mathbf{O}^n \odot \mathbf{U}\mathbf{V}^n\mathbf{U}')'\mathbf{U}\mathbf{V}^n) + \gamma\mathbf{L}\mathbf{U}, \quad (12)$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{V}^p} = \beta^{+}(\mathbf{U}'(\mathbf{O}^p \odot \mathbf{O}^p \odot \mathbf{U}\mathbf{V}^p\mathbf{U}')\mathbf{U}$$
$$-\mathbf{U}'(\mathbf{O}^p \odot \mathbf{O}^p \odot \mathbf{A}^p)\mathbf{U}), \quad (13)$$

$$\frac{\partial \mathcal{J}}{\partial \mathbf{V}^n} = \beta^{-}(\mathbf{U}'(\mathbf{O}^n \odot \mathbf{O}^n \odot \mathbf{U}\mathbf{V}^n\mathbf{U}')\mathbf{U}$$
$$-\mathbf{U}'(\mathbf{O}^n \odot \mathbf{O}^n \odot \mathbf{A}^n)\mathbf{U}). \quad (14)$$

Algorithm 1: The proposed SignedFS framework.

Input : $\{\mathbf{X}, \mathbf{A}^p, \mathbf{A}^n, c, \alpha, \beta^{+}, \beta^{-}, \gamma, \theta\}$
Output : ranking of features in a descending order

1 Initialize $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ randomly;
2 Initialize $\mathbf{H}$ as an identity matrix;
3 $\mathbf{A} = \mathbf{A}^p - \mathbf{A}^n$, $\mathbf{P}_1 = \mathbf{A}$, $\mathbf{P}_2 = \mathbf{O} \odot \mathbf{A}^2$, $\mathbf{P} = \mathbf{P}_1 + \theta\mathbf{P}_2$;
4 $\mathbf{L} = \mathbf{D} - \mathbf{P}$;
5 **while** *not converge* **do**
6   Set $\mathbf{M} = \mathbf{X}'\mathbf{X} + \alpha\mathbf{H}$;
7   Calculate $\frac{\partial \mathcal{J}}{\partial \mathbf{U}}$, $\frac{\partial \mathcal{J}}{\partial \mathbf{V}^p}$ and $\frac{\partial \mathcal{J}}{\partial \mathbf{V}^n}$;
8   Update $\mathbf{U} \leftarrow \mathbf{U} - \lambda_u \frac{\partial \mathcal{J}}{\partial \mathbf{U}}$;
9   Update $\mathbf{V}^p \leftarrow \mathbf{V}^p - \lambda_p \frac{\partial \mathcal{J}}{\partial \mathbf{V}^p}$;
10   Update $\mathbf{V}^n \leftarrow \mathbf{V}^n - \lambda_n \frac{\partial \mathcal{J}}{\partial \mathbf{V}^n}$;
11   Update $\mathbf{W} \leftarrow \mathbf{M}^{-1}\mathbf{X}'\mathbf{U}$;
12   Update $\mathbf{H}$ through Eq.(9);
13 **end**
14 Rank features according to the values of $\|\mathbf{W}_{i*}\|_2$ in a descending order;

With these equations, the detailed algorithm of the proposed SignedFS framework is illustrated in Algorithm 1. At first, we initialize $\mathbf{U}$, $\mathbf{V}^p$, $\mathbf{V}^n$, $\mathbf{H}$ and calculate user proximity matrix and signed Laplacian matrix. From line 5 to 13, we update $\mathbf{U}$, $\mathbf{V}^p$, $\mathbf{V}^n$ and $\mathbf{W}$ alternatively until achieving convergence. In each iteration, we first calculate $\mathbf{M}$, the computation cost of $\mathbf{M}$ is $O(nd^2)$. After obtaining $\mathbf{M}$, we fix $\mathbf{W}$ and update $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ with the gradient descent method. $\lambda_u, \lambda_p, \lambda_n$ are the step size for the update $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$, respectively. These step sizes can be determined by line search according to the Armijo rule. The computation cost of updating $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ are $O(nd^2) + O(n^2d) + O(n^2c) + O(nc^2)$, $O(nc^2) + O(n^2c) + O(n^3)$ and $O(nc^2) + O(n^2c) + O(n^3)$, respectively. Then we employ Eq.(10) to update $\mathbf{W}$, the computational cost of updating $\mathbf{W}$ is $O(nd^2) + O(dn^2) + O(d^3) + O(d^2c) + O(ncd)$. After we obtain the local optimal solution of $\mathbf{W}$, we rank the features in a descending order according to the values of $\|\mathbf{W}_{i*}\|_2$.

5.2 Convergence Analysis

In this subsection, we show the alternating optimization algorithm monotonically decreases the value of the objective function and it is guaranteed to converge. We start the convergence analysis with the following lemma.

Lemma 5.1. *The following inequality holds if $(\mathbf{W}_{i*})^k$ and $(\mathbf{W}_{i*})^{k+1}$ are non-zero vectors ($i = 1, 2, ..., d$), where $(\mathbf{W}_{i*})^k$ and $(\mathbf{W}_{i*})^{k+1}$ denote the update of $\mathbf{W}_{i*}$ in the k-th and $(k+1)$-th iteration, respectively [29]:*

$$\begin{aligned} &\|(\mathbf{W})^{k+1}\|_{2,1} - \sum_i \frac{\|(\mathbf{W}_{i*})^{k+1}\|_2^2}{2\|(\mathbf{W}_{i*})^k\|_2} \\ \leq &\|(\mathbf{W})^k\|_{2,1} - \sum_i \frac{\|(\mathbf{W}_{i*})^k\|_2^2}{2\|(\mathbf{W}_{i*})^k\|_2}. \end{aligned} \quad (15)$$

Theorem 5.2. *The value of the objective function in Algorithm 1 monotonically decreases in each iteration.*

Proof. As described by Algorithm 1, we update $\mathbf{U}$, $\mathbf{V}^p$ and $\mathbf{V}^n$ by the gradient descent method. Thus, the following inequality condition holds:

$$\begin{aligned} &\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^{k+1}, (\mathbf{V}^n)^{k+1}, (\mathbf{W})^k) \\ \leq &\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^{k+1}, (\mathbf{V}^n)^k, (\mathbf{W})^k) \\ \leq &\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^k, (\mathbf{V}^n)^k, (\mathbf{W})^k) \\ \leq &\mathcal{J}((\mathbf{U})^k, (\mathbf{V}^p)^k, (\mathbf{V}^n)^k, (\mathbf{W})^k), \end{aligned} \quad (16)$$

where $(\mathbf{U})^k$, $(\mathbf{V}^p)^k$, $(\mathbf{V}^n)^k$, $(\mathbf{W})^k$ and $(\mathbf{U})^{k+1}$, $(\mathbf{V}^p)^{k+1}$, $(\mathbf{V}^n)^{k+1}$, $(\mathbf{W})^{k+1}$ denote the update of $\mathbf{U}$, $\mathbf{V}^p$, $\mathbf{V}^n$, $\mathbf{W}$ in the k-th and $(k+1)$-th iteration, respectively.

After that, we fix $(\mathbf{U})^{k+1}$, $(\mathbf{V}^p)^{k+1}$ and $(\mathbf{V}^n)^{k+1}$ to update $\mathbf{W}$. As can be observed in Eq.(7), $(\mathbf{W})^{k+1}$ is the optimal solution of the following objective function:

$$\min_{\mathbf{W}} \mathcal{J}(\mathbf{W}) = \|\mathbf{X}\mathbf{W} - (\mathbf{U})^{k+1}\|_F^2 + \alpha\, tr(\mathbf{W}'(\mathbf{H})^k\mathbf{W}). \quad (17)$$

Thus, we obtain the following inequality:

$$\begin{aligned} &\|\mathbf{X}(\mathbf{W})^{k+1} - (\mathbf{U})^{k+1}\|_F^2 + \alpha\, tr((\mathbf{W})^{k+1\prime}(\mathbf{H})^k(\mathbf{W})^{k+1}) \\ \leq &\|\mathbf{X}(\mathbf{W})^k - (\mathbf{U})^{k+1}\|_F^2 + \alpha\, tr((\mathbf{W})^{k\prime}(\mathbf{H})^k(\mathbf{W})^k) \end{aligned} \quad (18)$$
.

It is equivalent to:

$$\begin{aligned} &\|\mathbf{X}(\mathbf{W})^{k+1} - (\mathbf{U})^{k+1}\|_F^2 + \alpha\|(\mathbf{W})^{k+1}\|_{2,1} \\ &-\alpha(\|(\mathbf{W})^{k+1}\|_{2,1} - \sum_i \frac{\|(\mathbf{W}_{i*})^{k+1}\|_2^2}{2\|(\mathbf{W}_{i*})^k\|_2}) \\ \leq &\|\mathbf{X}(\mathbf{W})^k - (\mathbf{U})^{k+1}\|_F^2 + \alpha\|(\mathbf{W})^k\|_{2,1} \\ &-\alpha(\|(\mathbf{W})^k\|_{2,1} - \sum_i \frac{\|(\mathbf{W}_{i*})^k\|_2^2}{2\|(\mathbf{W}_{i*})^k\|_2}). \end{aligned} \quad (19)$$

According to LEMMA 5.1, we obtain the following inequality condition:

$$\begin{aligned} &\|\mathbf{X}(\mathbf{W})^{k+1} - (\mathbf{U})^{k+1}\|_F^2 + \alpha\|(\mathbf{W})^{k+1}\|_{2,1} \\ \leq &\|\mathbf{X}(\mathbf{W})^k - (\mathbf{U})^{k+1}\|_F^2 + \alpha\|(\mathbf{W})^k\|_{2,1}. \end{aligned} \quad (20)$$

Therefore, integrating the inequality condition in Eq.(16), we have the following:

$$
\begin{aligned}
&\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^{k+1}, (\mathbf{V}^n)^{k+1}, (\mathbf{W})^{k+1}) \\
\leq &\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^{k+1}, (\mathbf{V}^n)^{k+1}, (\mathbf{W})^{k}) \\
\leq &\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^{k+1}, (\mathbf{V}^n)^{k}, (\mathbf{W})^{k}) \\
\leq &\mathcal{J}((\mathbf{U})^{k+1}, (\mathbf{V}^p)^{k}, (\mathbf{V}^n)^{k}, (\mathbf{W})^{k}) \\
\leq &\mathcal{J}((\mathbf{U})^{k}, (\mathbf{V}^p)^{k}, (\mathbf{V}^n)^{k}, (\mathbf{W})^{k}),
\end{aligned}
\tag{21}
$$

which completes the proof. □

6 EXPERIMENTS

In this section, we perform experiments on real-world signed social networks to validate the effectiveness of the proposed SignedFS framework. We begin by introducing the experimental settings. After that, we present the comparison results between SignedFS and the state-of-the-art unsupervised feature selection methods. Finally, we discuss the impact of various components of SignedFS and the effects of model parameters.

6.1 Experimental Setting

Following a commonly adopted way to assess unsupervised feature selection, we evaluate the proposed SignedFS in terms of clustering performance. To be specific, after we obtain the selected features, we employ K-means clustering based on the selected features. Since K-means may converge in local minimal, we repeat it 20 times and report the average clustering results. Two clustering evaluation metrics, clustering accuracy (ACC) and normalized mutual information (NMI) are used. The higher the ACC and NMI values, the better the selected features are.

SignedFS is compared with the following state-of-the-art unsupervised feature selection algorithms.

- Laplacian Score [13] selects features based on their ability to preserve data manifold structure.
- SPEC [38] evaluates features by spectral regression.
- NDFS [26] selects features by a joint nonnegative spectral analysis and $\ell_{2,1}$-norm regularization.
- LUFS [36] utilizes social dimension extracted from links to guide feature selection.
- NetFS [22] embeds latent representations extracted from links into feature selection.

Among these baseline methods, Laplacian Score, SPEC and NDFS are traditional unsupervised feature selection methods which only use feature information $\mathbf{X}$. LUFS and NetFS are unsupervised feature selection algorithms for unsigned networks with only positive links. To fairly compare different methods, we set the parameters for all methods by a grid search strategy from the range of {0.001,0.01,...,100,1000}. Afterwards, we compare the best clustering results of different unsupervised feature selection methods.

6.2 Quality of Selected Features by SignedFS

In this subsection, we compare the quality of features selected by SignedFS and aforementioned baseline algorithms. The number of selected features are varied among {400,800,...,1800,2000}. In SigendFS, we have four regularization parameters α, β^+, β^- and γ. We empirically set these parameters as $\{\alpha = 1, \beta^+ = 10, \beta^- = 1000, \gamma = 1000\}$ in Epinions and $\{\alpha = 1, \beta^+ = 1, \beta^- = 100, \gamma = 1000\}$ in Wiki-rfa. The comparison results of various feature selection algorithms on Epinions and Wiki-rfa datasets are shown in Table 3 and Table 4, respectively. We make the following observations from these two tables:

- SignedFS outperforms traditional feature selection algorithms LapScore, SPEC and NDFS on both datasets with significant clustering performance gain in most cases. We also perform pairwise Wilcoxon signed-rank test between SignedFS and these three traditional unsupervised feature selection methods, it shows SignedFS is significantly better (p-value=0.05). The superiority of SignedFS can be attributed to the utilization of additional link information while traditional methods are mainly based on the data i.i.d. assumption.
- SignedFS also obtains better clustering performance than the other two feature selection methods LUFS and NetFS on linked data. The major reason is that LUFS and NetFS only exploit positive links while SignedFS incorporates both positive links and negative links into a coherent model to obtain better features. It indicates the potential of using negative links for feature selection.
- We can see that when we gradually increase the number of selected features from 400 to 2000, the clustering performance in terms of clustering accuracy and NMI does not vary a lot. In particular, when a small number of features are selected, SignedFS already gives us very good performance. A small number of selected features is very appealing in practice as it significantly reduces the memory storage costs and computational costs for further learning tasks.

6.3 Further Analysis

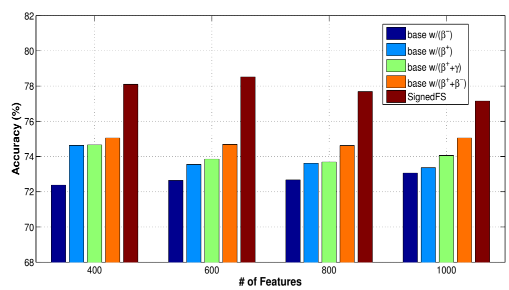


Figure 4: The impact of different components of SignedFS on Wiki-rfa.

In this subsection, we conduct experiments to further analyze the impact of each component in SignedFS for feature selection. As can be shown in the objective function of SignedFS in Eq.(7), the base model of our framework is $\|\mathbf{XW} - \mathbf{U}\|_F^2 + \alpha\|\mathbf{W}\|_{2,1}$. Upon the base model, we derive four different variants as follows:

- base w/β^+: We use the positive links for user latent representations learning ($\beta^- = 0, \gamma = 0$).
- base w/ β^-: We use the negative links for user latent representations learning ($\beta^+ = 0, \gamma = 0$).

Table 3: Clustering performance of different feature selection algorithms on Epinions.

Accuracy (%)									
	400	600	800	1000	1200	1400	1600	1800	2000
LapScore	11.48	11.34	10.95	11.79	12.54	11.61	11.29	11.19	12.79
SPEC	21.10	16.93	17.73	17.96	17.91	18.73	18.75	18.57	17.38
NDFS	12.18	11.29	11.92	12.16	12.32	12.14	11.92	13.19	11.78
LUFS	16.23	17.02	18.47	17.44	17.54	19.10	19.29	17.63	18.54
NetFS	18.59	19.62	19.21	18.80	18.43	18.77	17.82	19.76	19.98
SignedFS	**23.24**	**21.76**	**21.69**	**22.11**	**21.27**	**21.88**	**20.04**	**20.64**	**21.20**
NMI (%)									
	400	600	800	1000	1200	1400	1600	1800	2000
LapScore	2.74	2.72	2.68	3.68	2.68	2.73	2.75	2.63	2.67
SPEC	1.66	1.75	1.74	2.41	2.50	2.53	2.64	2.62	2.60
NDFS	1.49	1.47	1.46	1.46	1.46	1.46	1.46	1.46	1.46
LUFS	1.61	1.60	1.76	1.82	1.86	1.89	1.91	1.72	1.99
NetFS	1.80	1.90	2.28	1.75	1.79	1.11	1.56	1.47	2.08
SignedFS	**3.82**	**3.68**	**3.84**	**3.87**	**3.72**	**3.84**	**4.00**	**3.86**	**3.79**

Table 4: Clustering performance of different feature selection algorithms on Wiki-rfa.

Accuracy (%)									
	400	600	800	1000	1200	1400	1600	1800	2000
LapScore	70.92	70.94	70.93	70.31	70.52	70.89	70.92	71.13	71.37
SPEC	71.76	72.11	72.02	71.76	71.76	71.90	71.76	71.83	71.56
NDFS	72.94	72.73	72.94	72.75	72.78	72.94	72.94	72.94	72.94
LUFS	75.55	75.55	73.79	74.11	74.14	73.24	73.21	73.28	73.89
NetFS	72.81	72.91	72.94	72.73	72.68	72.70	72.97	72.97	72.97
SignedFS	**78.10**	**78.52**	**77.59**	**77.15**	**77.18**	**77.63**	**78.27**	**78.94**	**78.63**
NMI (%)									
	400	600	800	1000	1200	1400	1600	1800	2000
LapScore	0.24	0.22	0.23	0.26	0.26	0.26	0.26	0.26	0.26
SPEC	0.70	1.05	0.83	0.45	0.29	0.30	0.26	0.26	0.26
NDFS	0.26	0.30	0.28	0.25	0.26	0.26	0.27	0.26	0.26
LUFS	0.14	0.14	0.07	0.04	0.05	0.07	0.07	0.08	0.11
NetFS	0.44	0.38	0.36	0.36	0.38	0.38	0.37	0.37	0.37
SignedFS	**1.54**	**1.57**	**1.49**	**1.47**	**1.40**	**1.56**	**1.81**	**3.37**	**3.34**

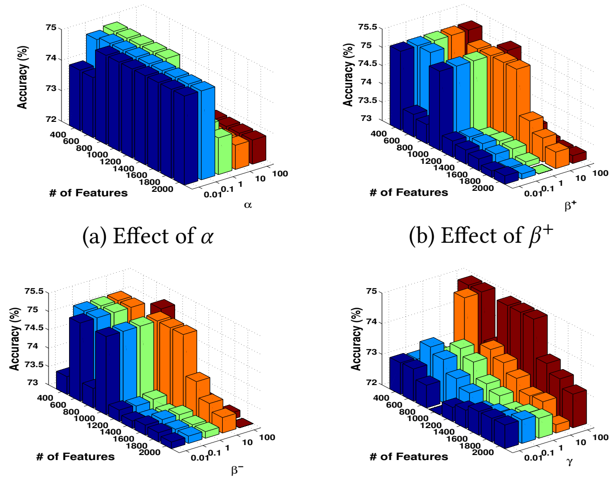
- base w/$(\beta^+ + \gamma)$: We use the positive links for user latent representations learning, and consider user proximity in terms of both positive and negative links ($\beta^+ = 0$).
- base w/$(\beta^+ + \beta^-)$: We use both positive and negative links for user latent representations learning ($\gamma = 0$).

We compare these four variants with the original SignedFS framework. The comparison results are shown in Figure 4. Due to space limit, we only show the results on the Wiki-rfa dataset as we have the similar observations on the Epinions dataset. We have several interesting observations from the figure:

- The variant base w/β^+ outperforms base w/β^- consistently. It indicates that positive links are more useful than negative links for feature selection. Also, negative links are not as useful as positive links in learning informative latent representations for feature selection.
- The variant base w/$(\beta^+ + \beta^-)$ obtains better clustering performance than base w/(β^+). It indicates that negative links have some added values over positive links, and can help finding more relevant features.
- The variant base w/$(\beta^+ + \gamma)$ outperforms base w/(β^+), and SignedFS outperforms base w/$(\beta^+ + \beta^-)$. It implies that user proximity modeling term is very helpful and could contribute to obtain better clustering performance.
- The proposed SignedFS framework achieves the best clustering results, showing the necessities of both user latent representation modeling and user proximity modeling.

6.4 Parameter Analysis

The proposed SignedFS has four important parameters. Among them, α controls the sparsity of the model; β^+ and β^- balances the contribution of positive and negative links in learning user latent representations; γ controls the modeling of the first-order and the second-order user proximity in signed social networks for feature selection. We study the effect of each parameter by fixing the others to investigate how it affects the performance of unsupervised feature selection. Since we make similar observations on both datasets, we only report the experimental results w.r.t. clustering accuracy on Wiki-rfa dataset to save space. First, we fix



(a) Effect of α

(b) Effect of β^+

(c) Effect of β^-

(d) Effect of γ

Figure 5: Parameter analysis of SignedFS on Wiki-rfa.

$\{\beta^+ = 10, \beta^- = 10, \gamma = 100\}$ and vary α among $\{0.01, 0.1, 1, 10, 100\}$. As shown in Figure 5 (a), the clustering performance first increases and then reaches the peak values when $\alpha = 0.1$. If we continuously increase the value of α, the clustering performance decreases. Therefore, we could empirically set α among the range of 0.01 to 0.1. Second, to investigate how β^+ affects the clustering performance, we vary β^+ among $\{0.01, 0.1, 1, 10, 100\}$ by fixing $\{\alpha = 0.1, \beta^- = 10, \gamma = 100\}$. The result is presented in Figure 5 (b). Similarly, the clustering performance first increases, reaches its maximal value when $\beta^+ = 10$ and then degrades. Next, to study the impact of β^-, we set $\{\alpha = 0.1, \beta^+ = 10, \gamma = 100\}$, and vary β^- among $\{0.01, 0.1, 1, 10, 100\}$. The result is presented in Figure 5 (c). The performance variation w.r.t. β^- has a similar trend as the variation of β^+, which suggests that negative links are also very important in finding relevant features in signed social networks. Finally, we fix $\{\alpha = 0.1, \beta^+ = 10, \beta^- = 10\}$ and vary γ among $\{0.01, 0.1, 1, 10, 100\}$ to investigate the effect of γ. As depicted in Figure 5 (d), with the increase of γ, the clustering performance gradually increases and then keeps stable. The clustering performance is relatively more sensitive to the number of selected features than these regularization parameters, which is still an open problem in unsupervised feature selection.

7 RELATED WORK

In this section, we briefly review related work from three aspects: (1) traditional feature selection; (2) feature selection in networks; and (3) signed social network analysis.

7.1 Traditional Feature Selection

Feature selection algorithms can be either supervised or unsupervised according to the availability of labels [19]. Supervised feature selection algorithms take advantage of the class labels to evaluate feature relevance by its ability to distinguish instances from different classes, which can be broadly divided into three categories: wrapper methods [16] which evaluate feature by its predictive accuracy of a predetermined learning algorithm, filter methods [12] which select features according to the general characteristics of the training data and embedded methods [2, 29] which embed feature selection into the learning algorithms. Since most real-world data is usually unlabeled, unsupervised feature selection received increasingly attention in recent years. Due to the lack of label information, these methods exploit different criteria such as data similarity [2, 13, 38], local discriminative information [26, 37] and data reconstruction error [9, 24] to define feature relevance.

7.2 Feature Selection in Networks

Feature selection methods for networked data are distinct from traditional feature selection methods as traditional methods assume that data is independent and identically distributed. In [10], a supervised feature selection algorithm FSNet was proposed for networked data. FSNet captures the correlation between content information and class labels by a linear classifier and it incorporates link information via graph regularization. Distinct from traditional networked data, social media data present its unique characteristics with the existence of complex linkage structure such as CoPost, CoFollowing, CoFollowed and Following. Motivated by these observations, Tang and Liu [35] made the first attempt to perform feature selection for social media data. Since networked data is usually costly to label, an unsupervised feature selection framework LUFS was proposed in [36]. In particular, LUFS extracts social dimensions from link information to help select relevant features. However, link information may contain a lot of noise and itself may be incomplete [3]. In order to alleviate the negative impacts from noisy and incomplete links, Li et al. [22] proposed a robust unsupervised feature selection framework for networked data. The authors further studied how to perform feature selection on networks with streaming features [21] and dynamic network structure [20].

7.3 Signed Networks Analysis

Even though mining signed graph is still in its early stage, some problems in signed networks have already been well studied, such as link prediction, community detection and genetic association analysis. Existing link prediction methods on signed social network include both supervised methods, which usually leverage local topology features [18] and features derived from long cycles [6] for link prediction, and unsupervised methods, which predict signs of links according to the topological properties of signed networks [11]. Community detection is another fundamental problem in signed social network analysis. In [25], the authors extended modularity maximization to signed networks. A spectral clustering algorithm for signed social networks was proposed in [17]. It is the first attempt to define a signed laplacian matrix which can separate users with negative links far away and force users with positive links to be close. In addition, signed networks are also used to represent quantitative trait network for unraveling the causal genetic variations [15]. However, since the signed network structure is over the output variables instead of input instances, it is inherently different from our work.

8 CONCLUSIONS

Feature selection has shown its effectiveness in preparing high-dimensional social media data for many learning tasks. A vast majority of existing efforts only consider positive interactions among connected instances while negative links are also prevailing in real-world social networks. In this paper, we attempt to perform unsupervised feature selection in signed social networks by leveraging both positive and negative interactions among linked instances. Methodologically, we propose a principled framework SignedFS. It first models both positive and negative links for a unified user latent representation. Then it embeds the user latent representation learning into feature selection. In addition, we revisit the principle of homophily and balance theory to model user proximity by a signed graph regularization. Also, we conduct experiments on two real-world datasets, the results show that SignedFS significantly improves the clustering performance and further experiments show the impacts of various components of SignedFS.

Future work can be focused on two aspects. First, in addition to social media data, other networks such as gene networks also have implicit negative interactions. We would like to investigate how to adapt SignedFS to different kinds of signed networks based on their unique characteristics. Second, as shown in [32], negative links can be predicted from explicit positive user interactions. Therefore, we would like to apply the SignedFS framework to social networks when negative links are not explicitly available.

ACKNOWLEDGMENTS

The authors would like to thank Jiliang Tang for helpful discussions. This material is, in part, supported by National Science Foundation (NSF) under grant number 1614576.

REFERENCES

[1] Lada A Adamic and Bernardo A Huberman. 2000. Power-law distribution of the world wide web. *science* 287, 5461 (2000), 2115–2115.

[2] Deng Cai, Chiyuan Zhang, and Xiaofei He. 2010. Unsupervised feature selection for multi-cluster data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 333–342.

[3] Chen Chen, Hanghang Tong, Lei Xie, Lei Ying, and Qing He. 2016. FASCINATE: fast cross-layer dependency inference on multi-layered networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 765–774.

[4] Kewei Cheng, Jundong Li, and Huan Liu. 2016. FeatureMiner: a tool for interactive feature selection. In *Proceedings of the 25th ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2445–2448.

[5] Kewei Cheng, Jundong Li, Jiliang Tang, and Huan Liu. 2017. Unsupervised sentiment analysis with signed social networks. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. AAAI, 3429–3435.

[6] Kai-Yang Chiang, Nagarajan Natarajan, Ambuj Tewari, and Inderjit S Dhillon. 2011. Exploiting longer cycles for link prediction in signed networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM, 1157–1162.

[7] Patrick Doreian. 1989. Network autocorrelation models: Problems and prospects. *Spatial statistics: Past, present, future* (1989), 369–89.

[8] Richard O Duda, Peter E Hart, and David G Stork. 2012. *Pattern classification*. John Wiley & Sons.

[9] Ahmed K Farahat, Ali Ghodsi, and Mohamed S Kamel. 2011. An efficient greedy method for unsupervised feature selection. In *Proceedings of the 2011 IEEE International Conference on Data Mining*. IEEE, 161–170.

[10] Quanquan Gu and Jiawei Han. 2011. Towards feature selection in network. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM, 1175–1184.

[11] Ramanthan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2004. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*. ACM, 403–412.

[12] Mark A Hall. 1999. *Correlation-based feature selection for machine learning*. Ph.D. Dissertation. The University of Waikato.

[13] Xiaofei He, Deng Cai, and Partha Niyogi. 2005. Laplacian score for feature selection. In *Advances in Neural Information Processing Systems*. 507–514.

[14] Fritz Heider. 1946. Attitudes and cognitive organization. *The Journal of psychology* 21, 1 (1946), 107–112.

[15] Seyoung Kim and Eric P Xing. 2009. Statistical Estimation of Correlated Genome Associations to a Quantitative Trait Network. *PLoS Genetics* 5, 8 (2009).

[16] Ron Kohavi and George H John. 1997. Wrappers for feature subset selection. *Artificial intelligence* 97, 1 (1997), 273–324.

[17] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto William De Luca, and Sahin Albayrak. 2010. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM, 559–570.

[18] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, 641–650.

[19] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Trevino Robert, Jiliang Tang, and Huan Liu. 2016. Feature selection: a data perspective. *arXiv:1601.07996* (2016).

[20] Jundong Li, Xia Hu, Ling Jian, and Huan Liu. 2016. Toward time-evolving feature selection on dynamic networks. In *Proceedings of the 2016 IEEE International Conference on Data Mining*. IEEE, 1003–1008.

[21] Jundong Li, Xia Hu, Jiliang Tang, and Huan Liu. 2015. Unsupervised streaming feature selection in social media. In *Proceedings of the 24th ACM International Conference on Conference on Information and Knowledge Management*. ACM, 1041–1050.

[22] Jundong Li, Xia Hu, Liang Wu, and Huan Liu. 2016. Robust unsupervised feature selection on networked data. In *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 387–395.

[23] Jundong Li and Huan Liu. 2017. Challenges of feature selection for big data analytics. *IEEE Intelligent Systems* 32, 2 (2017), 9–15.

[24] Jundong Li, Jiliang Tang, and Huan Liu. 2017. Reconstruction-based unsupervised feature selection: an embedded approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. IJCAI/AAAI.

[25] Yadong Li, Jing Liu, and Chenlong Liu. 2014. A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks. *Soft Computing* 18, 2 (2014), 329–348.

[26] Zechao Li, Yi Yang, Jing Liu, Xiaofang Zhou, and Hanqing Lu. 2012. Unsupervised feature selection using nonnegative spectral analysis. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. AAAI Press, 1026–1032.

[27] Huan Liu and Hiroshi Motoda. 2007. *Computational methods of feature selection*. CRC Press.

[28] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001), 415–444.

[29] Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding. 2010. Efficient and robust feature selection via joint $\ell 2$, 1-norms minimization. In *Advances in Neural Information Processing Systems*. 1813–1821.

[30] Cosma Rohilla Shalizi and Andrew C Thomas. 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research* 40, 2 (2011), 211–239.

[31] Jiliang Tang, Charu Aggarwal, and Huan Liu. 2016. Recommendations in signed social networks. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 31–40.

[32] Jiliang Tang, Shiyu Chang, Charu Aggarwal, and Huan Liu. 2015. Negative link prediction in social media. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 87–96.

[33] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. 2015. A survey of signed network mining in social media. *arXiv preprint arXiv:1511.07569* (2015).

[34] Jiliang Tang, Xia Hu, and Huan Liu. 2014. Is distrust the negation of trust?: the value of distrust in social media. In *Proceedings of the 25th ACM conference on Hypertext and Social Media*. ACM, 148–157.

[35] Jiliang Tang and Huan Liu. 2012. Feature selection with linked data in social media. In *SDM*. SIAM, 118–128.

[36] Jiliang Tang and Huan Liu. 2012. Unsupervised feature selection for linked social media data. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 904–912.

[37] Bo Yang, William K Cheung, and Jiming Liu. 2007. Community mining from signed social networks. *IEEE Transactions on Knowledge and Data Engineering* 19, 10 (2007), 1333–1348.

[38] Zheng Zhao and Huan Liu. 2007. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th International Conference on Machine Learning*. ACM, 1151–1157.