Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Link Label Prediction in Signed Social Networks

Priyanka Agrawal* IBM Research Bangalore, India pagrawal@in.ibm.com

Vikas K. Garg* Toyota Technological Institute (TTI-C) Chicago, US vkg@ttic.edu, montsgarg@gmail.com

Ramasuri Narayanam IBM Research Bangalore, India ramasurn@in.ibm.com

Abstract

Online social networks continue to witness a tremendous growth both in terms of the number of registered users and their mutual interactions. In this paper, we focus on online signed social networks where positive interactions among the users signify friendship or approval, whereas negative interactions indicate antagonism or disapproval. We introduce a novel problem which we call the **link label prediction problem**: Given the information about signs of certain links in a social network, we want to learn the nature of relationships that exist among the users by predicting the sign, positive or negative, of the remaining links. We propose a matrix factorization based technique *MF-LiSP* that exhibits strong generalization guarantees. We also investigate the applicability of logistic regression [8] in this setting. Our experiments on Wiki-Vote, Epinions and Slashdot data sets strongly corroborate the efficacy of these approaches.

1 Introduction

The proliferation of user activity on online social communities, micro-blogging sites and other media such as Slashdot, Wikipedia, Facebook, and Twitter etc. offers a tremendous scope for mining interesting user behavior. Not only have these social networks been registering a steep growth in the number of new users, but also the interactions amongst the users. Typically, these interactions are either positive (indicating friendship or support) or negative (indicating antagonism or opposition) [6; 7]. How a user perceives his future interactions or considers his past experiences with other users involves underpinnings of several, possibly contradictory, factors. For instance, users with many common friends tend to develop an affinity for each other even in the absence of any direct interaction. The scenario becomes much more complicated in the case of a mixed bag of acquaintances: a user may have a negative impression of some contacts, and positive of some others, of a common friend. This effect becomes even more pronounced when the impact of several users is taken into account while determining the overall perception of each user toward others. We use the *sign* or *label* of a link to refer to the nature of this link. Given the information about signs of certain links in a social network, a central problem of interest concerns learning the nature of relationships that exist among the users by predicting the sign, positive or negative, of the remaining links. Hereafter we refer to this problem as the *link label prediction problem*.

Several challenges arise in the given context. First, the global behavior of the different users needs to be accounted for, albeit the local interactions do play a significant role too [8]. Second, in a realistic scenario, determining the most important or discriminative features may not be a straightforward task; more broadly, there may be correlations among the users and/or features that are not readily decipherable (or are hidden), or worse, the different social networks may not be amenable to a predetermined set of features. Third, determining the exact nature of relationships is often a tedious and costly process. In particular, due to manual effort involved in amassing the training data, information about only a subset of pairs of users may be available.

In this work, we address some of these key issues. We formulate link sign prediction as a matrix completion problem in a setting, where the data is represented as a partially observed (and typically asymmetric) matrix. We propose a technique, MF-LiSP (**M**atrix **F**actorization for **Li**nk **S**ign **P**rediction), which employs a trace norm regularizer with a particularly suited variation of the pair-wise hinge loss to approximate the given matrix. Our approach is strongly motivated by recent results [2; 14] that suggest the use of such a regularizer as an excellent alternative to other complexity measures such as the rank of the matrix. We provide rigorous generalization bounds for MF-LiSP via algorithmic stability [1]. Additionally, we investigate the suitability of the logistic regression model [8] in the multiple label prediction setting. Our experiments suggest that the maximum likelihood based logistic regression model works very well in the current setting too.

The rest of this paper is organized as follows. We first present the related work in §1.1, and introduce some definitions in §2. Next we formulate the problem and propose our main technique in §3. We then present the theoretical justification of this work by providing rigorous generalization bounds in §4. We provide a detailed analysis on the results of our experiments in §6 and a summary of our work in §7.

*P. Agrawal and V. K. Garg contributed equally to this work.

1.1 Related Work

Techniques based on matrix factorization have been recently used for analyzing social networks. For instance, Scripps *et al.* [13] and Menon and Elkan [12] proposed algorithms for sign prediction in graphs. Cui *et al.* [4] leveraged matrix factorization to rank users with the objective of maximizing the social influence for a given item, Ma *et al.* [11] looked into social network based recommendation systems, and Wang *et al.* [16] devised a method for community detection.

Recently, there has been much interest in the problem of predicting links and their signs. Liben-Nowell and Kleinberg [9] leveraged measures that captured proximity of nodes, whereas Chiang *et al.* [3] investigated the use of social imbalance measures and proposed a supervised learning approach. Kuter and Golbeck [10] proposed an algorithm for trust (positive sign) inference with a probabilistic interpretation. Yang *et al.* [15] designed algorithms to infer the signs of social ties completely based on decision making behavior of the users.

Leskovec *et al.*, in a seminal work [8], introduced the *edge sign prediction problem*: Given a social network with signs on all the edges, except that from node u to node v, how reliably can one infer this sign $s(u, v)$? In this paper, we generalize their setting to encompass several missing links. To the best of our knowledge, our work is the first to tackle simultaneous label prediction of multiple links.

2 Loss and Stability

Consider a learner that receives instances from a training set, $S = \{S^1, S^2, \ldots, S^m\}$, where $S^i = \{(x_1^i, y_1^i), \ldots, (x_n^i, y_n^i) | x_j^i \in \mathcal{X} \subseteq \mathbb{R}, y_j^i \in \mathcal{Y} \subseteq \mathbb{R}\}$ denotes the sample set of links[1] available with user i. Let $S_{i,i'}^{t,z}$ (and the abbreviated notation S') refer to the sample obtained by replacing (x_t^i, y_t^i) in S with $(x_z^{i'}, y_z^{i'})$ for some $1 \leq i, i' \leq m$ and $1 \leq t, z \leq n$. The goal is to learn a real-valued sign prediction function f using S.

Definition 1. ***(Loss function)*** *A loss function $l : \mathbb{R}^{\mathcal{X}} \times (\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y}) \rightarrow \mathbb{R}^+ \cup \{0\}$ assigns to each $f : \mathcal{X} \rightarrow \mathbb{R}$ and $(x_j^i, y_j^i), (x_k^i, y_k^i) \in (\mathcal{X} \times \mathcal{Y})$, a non-negative real number $l(f, (x_j^i, y_j^i), (x_k^i, y_k^i))$, which is interpreted as the penalty or loss incurred by f due to its relative predictions of x_j^i and x_k^i given the corresponding labels y_j^i and y_k^i. We shall require that the loss function l be symmetric with respect to (x_j^i, y_j^i) and (x_k^i, y_k^i), i.e., for all f, (x_j^i, y_j^i), and (x_k^i, y_k^i),*

$$l(f, (x_j^i, y_j^i), (x_k^i, y_k^i)) = l(f, (x_k^i, y_k^i), (x_j^i, y_j^i)).$$

Definition 2. ***(Expected** l**-error)*** *Let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a sign prediction function on $\mathcal{X}$, and $l : \mathbb{R}^{\mathcal{X}} \times (\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y}) \rightarrow \mathbb{R}^+ \cup \{0\}$ be a loss function. Define the expected l-error of f: $R_l(f) = E_{((X_j^i, Y_j^i), (X_k^i, Y_k^i)) \in \mathcal{D} \times \mathcal{D}}[l(f, (X_j^i, Y_j^i), (X_k^i, Y_k^i))]$, where both the training and the unseen examples are assumed to be drawn randomly and independently according to some (unknown) distribution D.*

[1] We use links and examples interchangeably in the sequel.

Definition 3. ***(Empirical** l**-error)*** *Let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a function on $\mathcal{X}$ and $l : \mathbb{R}^{\mathcal{X}} \times (\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y}) \rightarrow \mathbb{R}^+ \cup \{0\}$ be a loss function. Define the empirical l-error of f, denoted by $\hat{R}_l(f; S)$, as*

$$\hat{R}_l(f; S) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|Q_i^S|} \sum_{(j,k) \in Q_i^S} l(f, (x_j^i, y_j^i), (x_k^i, y_k^i)),$$

where Q_i^S is the set of link pairs that are available with the user i in S.

Definition 4. ***(Uniform loss stability)*** *Let $\mathcal{A}$ be an algorithm whose output on a training sample S we denote by f_S, and let l be a loss function. Let $\beta : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$. We say that $\mathcal{A}$ has uniform loss stability β with respect to l if $\forall\ m, n \in \mathbb{N}$, $S \in (\mathcal{X} \times \mathcal{Y})^{m \times n}, 1 \leq i, i' \leq m$ and $1 \leq t, z \leq n$, we have $\forall (x_j^i, y_j^i), (x_k^i, y_k^i) \in (\mathcal{X} \times \mathcal{Y})$,*

$$|l(f_S, (x_j^i, y_j^i), (x_k^i, y_k^i)) - l(f_{S'}, (x_j^i, y_j^i), (x_k^i, y_k^i))| \leq \beta(m, n).$$

Definition 5. ***(Uniform score stability)*** *Let $\mathcal{A}$ be a sign prediction algorithm whose output on a training sample S is denoted by f_S. Let $\nu : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}$. We say that $\mathcal{A}$ has uniform score stability ν if $\forall\ m, n \in \mathbb{N}$, $S \in (\mathcal{X} \times \mathcal{Y})^{m \times n}, 1 \leq i, i' \leq m$ and $1 \leq t, z \leq n$, we have $\forall x \in \mathcal{X}$,*

$$|f_S(x) - f_{S'}(x)| \leq \nu(m, n).$$

Definition 6. *($\sigma-$admissibility) Let $\mathcal{F}$ be a class of real-valued functions on $\mathcal{X}$. Let l be a sign prediction loss function and let $\sigma > 0$. We say that l is σ-admissible (or Lipschitz continuous) with respect to $\mathcal{F}$ if for all $f_1, f_2 \in \mathcal{F}$ and all $(x_j^i, y_j^i), (x_k^i, y_k^i) \in (\mathcal{X} \times \mathcal{Y})$, we have*

$$\begin{aligned} &|l(f_1, (x_j^i, y_j^i), (x_k^i, y_k^i)) - l(f_2, (x_j^i, y_j^i), (x_k^i, y_k^i))| \\ &\leq \sigma(|f_1(x_j^i) - f_2(x_j^i)| + |f_1(x_k^i) - f_2(x_k^i)|) \end{aligned}$$

3 MF-LiSP

Formally, we consider a sparse matrix $(A)_{m \times m} \in \mathbb{R}^{m \times m}$ corresponding to a social network with m nodes. Each known entry A_{ij} represents the sign (1 for positive links and 0 for negative links) of the edge from node i to node j. Let A_i denote the i^{th} row of A, and S the set of known entries in A. Assuming the entries in A are sampled from a uniform distribution, our objective is to employ matrix factorization based on the unweighted trace-norm to fill the remaining entries in A. We accomplish this by approximating the matrix A with a matrix $(X)_{m \times m} = (U)_{m \times d}(V)'_{m \times d}$, for some suitable d, such that the *discrepancy* between A and X is minimized. For our sign prediction problem, instead of using the point-wise error, we introduce the following pair-wise empirical error:

$$\hat{R}_l(X; S) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|Q_i^S|} \sum_{(j,k) \in Q_i^S} ((A_{ij} - A_{ik}) - (X_{ij} - X_{ik}))_+,$$

where Q_i^S is the set of link pairs that are available with the user i in S. This empirical error loss function is reminiscent of the hinge loss convex surrogate for 0/1 loss in classification. We use this particular variation since it elegantly captures the correlations amongst the users and thereby makes the technique more robust to fluctuations in individual behaviors. Moreover, albeit we are concerned with binary prediction in this work, we emphasize that this loss function can be readily translated to the more generic multi-label settings, especially owing to its invariance to scale. For the current setting, our goal is to determine appropriate U and V to minimize the empirical prediction error $\hat{R}_l(X;S)$, or equivalently

$$\min_{U,V} \ \hat{R}_l(UV';S)$$

To avoid over-fitting, we incorporate a convex regularization term:

$$\min_{U,V'} \ \hat{R}_l(UV';S) + \frac{\lambda}{2}(||U||_F^2||V||_F^2),$$

and obtain the following primal formulation:

$$\min_{U,V,\xi} \frac{1}{2}(||U||^2||V||^2) + \frac{C}{m}\sum_{i=1}^{m}\frac{1}{|Q_i^S|}\sum_{(j,k)\in Q_i^S}\xi_{jk}^i$$

subject to

$$\xi_{jk}^i \geq 0 \quad \forall i,j,k, \text{ and}$$
$$\xi_{jk}^i \geq (A_{ij} - A_{ik}) - U_i^T(V_j - V_k) \quad \forall i,j,k$$

The corresponding dual is obtained by introducing Lagrange multipliers:

$$\min_{\boldsymbol{\alpha},\boldsymbol{V}} \frac{\mathbf{1}}{\mathbf{2}||\boldsymbol{V}||^{\mathbf{2}}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}}\sum_{\substack{(\boldsymbol{j},\boldsymbol{k})\in \boldsymbol{Q}_{\boldsymbol{i}}^{\boldsymbol{S}}\\(\boldsymbol{j'},\boldsymbol{k'})\in \boldsymbol{Q}_{\boldsymbol{i}}^{\boldsymbol{S}}}}\boldsymbol{\alpha}_{\boldsymbol{jk}}^{\boldsymbol{i}}\boldsymbol{\alpha}_{\boldsymbol{j'k'}}^{\boldsymbol{i}}\boldsymbol{K}_{\boldsymbol{jk,j'k'}}^{\boldsymbol{i}}$$
$$-\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}}\sum_{(\boldsymbol{j},\boldsymbol{k})\in \boldsymbol{Q}_{\boldsymbol{i}}^{\boldsymbol{S}}}\boldsymbol{\alpha}_{\boldsymbol{jk}}^{\boldsymbol{i}}(\boldsymbol{A_{ij}} - \boldsymbol{A_{ik}})$$

subject to

$$\mathbf{0} \leq \boldsymbol{\alpha}_{\boldsymbol{jk}}^{\boldsymbol{i}} \leq \frac{\boldsymbol{C}}{\boldsymbol{m}|\boldsymbol{Q_i}|} \quad \forall \boldsymbol{i,j,k}$$

where

$$\boldsymbol{K}_{\boldsymbol{jk,j'k'}}^{\boldsymbol{i}} = (\boldsymbol{V_j} - \boldsymbol{V_k})\cdot(\boldsymbol{V_{j'}} - \boldsymbol{V_{k'}})$$

Our algorithm, MF-LiSP, solves this convex quadratic program (QP) by following a stochastic gradient projection approach [5]. See Algorithm 1 for more details.

Algorithm 1 MF-LiSP

Inputs:
Training sample $\boldsymbol{\mathcal{Q}_i} = (\boldsymbol{\mathcal{Q}_{i+}}, \boldsymbol{\mathcal{Q}_{i-}}) \in \boldsymbol{\mathcal{A}^{n_+}} \times \boldsymbol{\mathcal{A}^{n_-}}$
$\forall 1 \leq i \leq m$
Kernel function $\boldsymbol{K} : (\mathcal{A}\times\mathcal{A})\times(\mathcal{A}\times\mathcal{A}) \to \mathbb{R}$
Parameters C, η_0, p_{max}

Initialize:
$\alpha_{jk}^{i(1)} \leftarrow \frac{C}{1000m|\boldsymbol{\mathcal{Q}_i}|}, \forall 1 \leq i \leq m, (j,k) \in \boldsymbol{\mathcal{Q}_i}$
$\boldsymbol{V}^{(1)}$ to a random matrix of size $m \times d$.

For $p = 1$ to p_{max} **do**
Randomly select a row i.

Update α^i:
- **[Gradient step]**
  $\boldsymbol{\alpha}^{i(p+1/2)} \leftarrow \boldsymbol{\alpha}^{i(p)} - \frac{\eta_0}{\sqrt{p}}\nabla_\alpha O(\boldsymbol{\alpha}^{i(p)}, \boldsymbol{V}^{(p)})$
- **[Projection step]**
  For all $(j,k) \in \boldsymbol{\mathcal{Q}_i}$
  If $\left(\alpha_{jk}^{i(p+1/2)} < 0\right)$ **Then** $\alpha_{jk}^{i(p+1)} = 0$
  Else If $\left(\alpha_{jk}^{i(p+1/2)} > \frac{C}{m|\boldsymbol{\mathcal{Q}_i}|}\right)$ **Then**
  $\alpha_{jk}^{i(p+1)} = \frac{C}{m|\boldsymbol{\mathcal{Q}_i}|}$
  Else $\alpha_{jk}^{i(p+1)} = \alpha_{jk}^{i(p+1/2)}$

Update V:
$\boldsymbol{V}^{(p+1)} \leftarrow \boldsymbol{V}^{(p)} - \frac{\eta_0}{\sqrt{p}}\nabla_V O(\boldsymbol{\alpha}^{i(p+1)}, \mathbf{V}^{(p)})$

Output:
$\boldsymbol{X} = (\boldsymbol{U}^{(p^*)})(\boldsymbol{V}^{(p^*)})^T$,
where $p^* = \arg\min_{1\leq p\leq(p_{max}+1)} O(\boldsymbol{\alpha}^{(p)}, \mathbf{V}^{(p)})$
and $\boldsymbol{U}^{\boldsymbol{i(p^*)}} = \sum_{(j,k)\in\boldsymbol{\mathcal{Q}_i}} \alpha_{jk}^{i(p^*)}(\boldsymbol{V}_j^{(p^*)} - \boldsymbol{V}_k^{(p^*)})$

4 Generalization Bounds

We now provide generalization bounds for MF-LiSP. All our theoretical results hold very generally about labels on links, covering link strength as well as link sign. We present guarantees that hold for real (multi) valued matrices and subsume the current 0/1 setting. Therefore, our algorithm can be readily adapted to more generic settings: for instance, when in addition to the (weighted) positive and negative links, there may be no links or unsigned links. But the main application in this paper is link sign prediction. To the best of our knowledge, MF-LiSP is the first technique that has provable guarantees for such general link labels. Due to space constraints, we only outline the main steps of the proofs and omit the details.

Theorem 1. *Let $\mathcal{A}$ be an algorithm whose output, for a user $i \in [m]$, on a training sample S we denote by f_{iS}, and let l be a bounded loss function such that $0 \leq l(f_i,(x_j^i,y_j^i),(x_k^i,y_k^i)) \leq B$ for all $f_i : \mathcal{X} \to \mathbb{R}$ and $(x_j^i,y_j^i),(x_k^i,y_k^i) \in (\mathcal{X}\times\mathcal{Y})$. Let $\beta : \mathbb{N} \to \mathbb{R}$ be such that $\mathcal{A}$ has uniform loss stability β with respect to l. Then for any $0 < \delta \leq 1$, with probability at least $1-\delta$ over the draw of S,*

$$R_l(f_{iS}) - \hat{R}_l(f_{iS};S) \leq 2\beta(|S_i|) + (|S_i|\beta(|S_i|) + B)\sqrt{\frac{2}{|S_i|}\ln(\frac{1}{\delta})}.$$

Proof. Define $\phi_i(S) = R_l(f_{iS}) - \hat{R}_l(f_{iS};S)$. We first show that ϕ_i satisfies the bounded difference requirement of the McDiarmid's inequality. We have $\left|\phi_i(S) - \phi_i(S_{i',i''}^{t,z})\right|$

$$= \left|\left(R_l(f_{iS}) - \hat{R}_l(f_{iS};S)\right) - \left(R_l(f_{iS_{i',i''}^{t,z}}) - \hat{R}_l(f_{iS_{i',i''}^{t,z}};S_{i',i''}^{t,z})\right)\right|$$

$$\leq \left|R_l(f_{iS}) - R_l(f_{iS_{i',i''}^{t,z}})\right| + \left|\hat{R}_l(f_{iS};S) - \hat{R}_l(f_{iS_{i',i''}^{t,z}};S_{i',i''}^{t,z})\right|.$$

It is easy to show that

$$\begin{aligned}\left|R_l(f_{iS}) - R_l(f_{iS^{t,z}_{i',i''}})\right| &\leq \mathbf{E}_{((X^i_j,Y^i_j),(X^i_k,Y^i_k))}\left[\beta(|S_i|)\right] \\ &= \beta(|S_i|).\end{aligned}$$

Now consider $\left|\hat{R}_l(f_{iS};S) - \hat{R}_l(f_{iS^{t,z}_{i',i''}};S^{t,z}_{i',i''})\right|$. Considering the four different cases separately (whether $i' = i$ and/or $i'' = i$ is true), we can show that for all $1 \leq i', i'' \leq n$

$$\left|\phi_i(S) - \phi_i(S^{t,z}_{i',i''})\right| \leq 2\left(\beta(|S_i|) + \frac{B}{|S_i|}\right).$$

Then, from McDiarmid's inequality, we have for any $\epsilon > 0$,

$$\mathbf{P}\left(\left(R_l(f_{iS}) - \hat{R}_l(f_{iS};S)\right) - \mathbf{E}\left[R_l(f_{iS}) - \hat{R}_l(f_{iS};S)\right] \geq \epsilon\right) \leq \exp^{-\frac{\epsilon^2|S_i|}{2|S_i|(|S_i|\beta(|S_i|)+B)^2}}$$

It can be shown that $\mathbf{E}\left[R_l(f_{iS}) - \hat{R}_l(f_{iS};S)\right] \leq 2\beta(|S_i|)$, which immediately gives the result on setting $\delta = \exp^{-\frac{\epsilon^2|S_i|^2}{2(|S_i|\beta(|S_i|)+B)^2}}$. ☐

Theorem 2. *Let $\mathcal{A}$ be a symmetric sign prediction algorithm whose output, for a user $i \in [m]$, on a training sample S we denote by f_{iS}, and let l be a bounded sign prediction loss function such that $0 \leq l(f_i, (x^i_j, y^i_j), (x^i_k, y^i_k)) \leq B$ for all $f_i : \mathcal{X} \to \mathbb{R}$ and $(x^i_j, y^i_j), (x^i_k, y^i_k) \in (\mathcal{X} \times \mathcal{Y})$. Let $\mathcal{A}$ have uniform loss stability $\beta : \mathbb{N} \to \mathbb{R}$ with respect to l and uniform score stability $\nu : \mathbb{N} \to \mathbb{R}$. Then, we must have $\beta(|S_i|) = 2\nu(|S_i|)$. Furthermore, for any $0 \leq \delta \leq 1$, the following holds with probability at least $1 - \delta$ over the draw of S*

$$R_l(f_{iS}) - \hat{R}_l(f_{iS};S) < 4\nu(|S_i|) + (2|S_i|\nu(|S_i|)+B)\sqrt{\frac{2}{|S_i|}\ln(\frac{1}{\delta})}.$$

Proof. Without introducing any ambiguity, for the purpose of this proof, we succinctly denote $S^{t,z}_{i',i''}$ by S'. Now, to prove $\beta(|S_i|) = 2\nu(|S_i|)$, we need to show for $(x^i_j, y^i_j), (x^i_k, y^i_k) \in (\mathcal{X} \times \mathcal{Y})$,

$$|l(f_S, (x^i_j, y^i_j), (x^i_k, y^i_k)) - l(f_{S'}, (x^i_j, y^i_j), (x^i_k, y^i_k))| \leq 2\nu(|S_i|).$$

In order to avoid triviality of the result, let us assume without loss of generality that

$$l(f_S, (x^i_j, y^i_j), (x^i_k, y^i_k)) > l(f_{S'}, (x^i_j, y^i_j), (x^i_k, y^i_k)).$$

It is easy to argue that we must have $\max\left(f_S(x^j_i) - f_S(x^k_i), f_{S'}(x^j_i) - f_{S'}(x^k_i)\right) < y^i_j - y^i_k$, whence

$$\begin{aligned}&|l(f_S, (x^i_j, y^i_j), (x^i_k, y^i_k)) - l(f_{S'}, (x^i_j, y^i_j), (x^i_k, y^i_k))| \\ =\ & ((y^i_j - y^i_k) - (f_S(x^j_i) - f_S(x^k_i))) \\ -\ & ((y^i_j - y^i_k) - (f_{S'}(x^j_i) - f_{S'}(x^k_i))) \\ \leq\ & |f_S(x^j_i) - f_{S'}(x^j_i)| + |f_S(x^k_i) - f_{S'}(x^k_i)| \\ =\ & 2\nu(|S_i|).\end{aligned}$$

The theorem follows immediately by substituting this result in Theorem 1. ☐

4.1 Stability Results

A stable algorithm is understood to be one whose output does not change significantly with a small change in the input. In this section, we derive the stability bounds for our algorithm. Consider the following regularized empirical l-error of a sign prediction function $f_i \in \mathcal{F}_i$ with respect to a sample S and a regularization parameter $\lambda_i > 0$, where $\mathcal{F}_i \subseteq \mathcal{F}$ is a class of real valued functions and $N : \mathcal{F} \to \mathbb{R}^+ \cup \{0\}$ is a regularized functional: $\hat{R}^\lambda_l(f_i;S) = \hat{R}_l(f_i;S) + \lambda_i N(f_i)$.

Let $f_{iS} \in \mathcal{F}$ denote the minimizer of this regularized empirical error on S, i.e., $f_{iS} = \arg\min_{f_i \in \mathcal{F}_i} \hat{R}^\lambda_l(f_i;S)$.

Lemma 1. *Let $\mathcal{F}_i \subseteq \mathcal{F}$ be a class of real-valued functions on $\mathcal{X}$. Let l be a sign prediction loss function that is convex in f_i, and let $\sigma > 0$ be such that l is σ-admissible with respect to $\mathcal{F}_i$. Let $\lambda_i > 0$ and let $N : \mathcal{F} \to \mathbb{R}^+ \cup \{0\}$ be a functional defined on $\mathcal{F}$ such that for any sample S, the regularized empirical l-error $\hat{R}^\lambda_l(f_i;S)$ has a minimizer (not necessarily unique) f_{iS} in $\mathcal{F}$. Then for any $p \in [0,1]$,*

$$\begin{aligned}&N(f_{iS}) - N(f_{iS} + p\Delta f_{iS}) + N(f_{iS^{t,z}_{i',i''}}) \\ &\qquad - N(f_{iS^{t,z}_{i',i''}} - p\Delta f_{iS}) \\ &\leq \frac{p\sigma}{\lambda_i \binom{|S_i|}{2}} \sum_{j \neq t}(|\Delta f_{iS}(x^t_i)| + 2|\Delta f_{iS}(x^j_i)| + |\Delta f_{iS}(x^z_{i''})|),\end{aligned}$$

where $\Delta f_{iS} = f_{iS^{t,z}_{i',i''}} - f_{iS}$.

Proof. We have, for any $p \in [0,1]$, using the convexity of $\hat{R}_l(f_{iS};S)$ in f (since l is convex in f):

$$\begin{aligned}&\hat{R}_l(f_{iS} + p\Delta f_{iS};S) - \hat{R}_l(f_{iS};S) \\ &\qquad \leq p\left(\hat{R}_l(f_{iS^{t,z}_{i',i''}};S) - \hat{R}_l(f_{iS};S)\right).\end{aligned} \tag{1}$$

Likewise,

$$\begin{aligned}&\hat{R}_l(f_{iS^{t,z}_{i',i''}} - p\Delta f_{iS};S) - \hat{R}_l(f_{iS^{t,z}_{i',i''}};S) \\ &\qquad \leq p\left(\hat{R}_l(f_{iS};S) - \hat{R}_l(f_{iS^{t,z}_{i',i''}};S)\right).\end{aligned} \tag{2}$$

Since $\mathcal{F}$ is convex, $(f_{iS} + p\Delta f_{iS};S), (f_{iS^{t,z}_{i',i''}} - p\Delta f_{iS};S) \in \mathcal{F}$. Further, since f_{iS} and $f_{iS^{t,z}_{i',i''}}$ are, respectively, the minimizers of $\hat{R}^\lambda_l(f_i;S)$ and $\hat{R}^\lambda_l(f_i;S^{t,z}_{i',i''})$, we must have

$$\hat{R}^\lambda_l(f_{iS};S) - \hat{R}^\lambda_l(f_{iS} + p\Delta f_{iS};S) \leq 0, \text{ and} \tag{3}$$

$$\hat{R}^\lambda_l(f_{iS^{t,z}_{i',i''}};S^{t,z}_{i',i''}) - \hat{R}^\lambda_l(f_{iS^{t,z}_{i',i''}} - p\Delta f_{iS};S^{t,z}_{i',i''}) \leq 0. \tag{4}$$

The result follows by combining (1), (2), (3) and (4). ☐

4.2 Regularization in the Hilbert Space

We now prove that MF-LiSP can be used for efficient sign prediction in a Reproducible Kernel Hilbert Space (RKHS).

Theorem 3. *Let $\mathcal{F}$ be an RKHS with kernel K such that for all $x \in \mathcal{X}, K(x,x) \leq \kappa^2 < \infty$. Let l be a sign prediction loss function that is convex in $f_i \in \mathcal{F}_i \subseteq \mathcal{F}$, and let $\sigma > 0$ be such that l is σ-admissible with respect to $\mathcal{F}$. Let $\lambda_i > 0$, and let $\mathcal{A}$ be a sign prediction algorithm that, given a training* *sample S, outputs a sign prediction function $f_{iS} \in \mathcal{F}_i$ that satisfies $f_{iS} = \arg\min_{f_i \in \mathcal{F}_i} \{\hat{R}_l(f_i; S) + \lambda_i ||f_i||_K^2\}$. Then $\mathcal{A}$ has uniform score stability ν given by $\nu(|S_i|) = \frac{8\sigma\kappa^2}{\lambda_i |S_i|}$. Furthermore, for any $0 < \delta \leq 1$, the following holds with probability at least $1 - \delta$ over the draw of S:*

$$R_l(f_{iS}) < \hat{R}_l(f_{iS}; S) + \frac{32\sigma\kappa^2}{\lambda_i |S_i|} + (\frac{16\sigma\kappa^2}{\lambda_i} + B)\sqrt{\frac{2}{|S_i|} \ln(\frac{1}{\delta})}$$

Proof. Let $N(f) = ||f||_K^2$. Then, defining $\Delta f_{iS} = f_{iS^{t,z}_{i',i''}} - f_{iS}$ and applying Lemma 1 with $p = 0.5$, we get

$$||f_{iS}||_K^2 - ||f_{iS} + 0.5\Delta f_{iS}||_K^2 + ||f_{iS^{t,z}_{i',i''}}||_K^2 - ||f_{iS^{t,z}_{i',i''}} - 0.5\Delta f_{iS}||_K^2$$

$$\leq \frac{\sigma}{\lambda_i |S_i|(|S_i| - 1)} \sum_{j \neq t} (|\Delta f_{iS}(x_i^t)| + 2|\Delta f_{iS}(x_i^j)| + |\Delta f_{iS}(x_{i''}^z)|) \tag{5}$$

Also,
$$\begin{aligned} & ||f_{iS}||_K^2 - ||f_{iS} + 0.5\Delta f_{iS}||_K^2 + ||f_{iS^{t,z}_{i',i''}}||_K^2 \\ - & ||f_{iS^{t,z}_{i',i''}} - 0.5\Delta f_{iS}||_K^2 \\ = & ||f_{iS}||_K^2 + ||f_{iS^{t,z}_{i',i''}}||_K^2 - \frac{1}{2}||f_{iS} + f_{iS^{t,z}_{i',i''}}||_K^2 \\ - & \langle f_{iS}, f_{iS^{t,z}_{i',i''}} \rangle_K \\ = & \frac{1}{2}||\Delta f_{iS}||_K^2. \end{aligned}$$

Using (5), we get $\frac{1}{2}||\Delta f_{iS}||_K^2$

$$\begin{aligned} \leq & \frac{\sigma}{\lambda_i |S_i|(|S_i| - 1)} \sum_{j \neq t} (|\Delta f_{iS}(x_i^t)| \\ + & 2|\Delta f_{iS}(x_i^j)| + |\Delta f_{iS}(x_{i''}^z)|). \end{aligned}$$

Applying the Cauchy-Schwarz inequality in conjunction with reproducing property of the RKHS, we have for all $x \in \mathcal{X}$ and all $f \in \mathcal{F}$,

$$|f(x)| \leq ||f||_K ||K_x||_K = ||f||_K \sqrt{K(x, x)}. \tag{6}$$

Since $\mathcal{F}$ is an RKHS and $\Delta f_{iS} \in \mathcal{F}$, therefore $\frac{1}{2}||\Delta f_{iS}||_K^2$

$$\begin{aligned} \leq & \frac{\sigma}{\lambda |S_i|(|S_i| - 1)} ||\Delta f_{iS}||_K \sum_{j \neq t} \Bigg(\sqrt{K(x_i^t, x_i^t)} \\ + & 2\sqrt{K(x_i^j, x_i^j)} + \sqrt{K(x_{i''}^z, x_{i''}^z)} \Bigg) \\ \leq & \frac{4\sigma\kappa}{\lambda_i |S_i|} ||\Delta f_{iS}||_K, \end{aligned}$$

which yields $||\Delta f_{iS}||_K \leq \frac{8\sigma\kappa}{\lambda_i |S_i|}$. Using (6), the following holds for all $x \in \mathcal{X}$,

$$|f_{iS}(x) - f_{iS^{t,z}_{i',i''}}(x)| = |\Delta f_{iS}(x)| \leq \frac{8\sigma\kappa^2}{\lambda_i |S_i|}.$$

Thus, $\nu(|S_i|) = \frac{8\sigma\kappa^2}{\lambda_i |S_i|}$. The error bound follows immediately from Theorem 2. □

5 Logistic Regression based Approach

The Logistic Regression model [8] provides a sound approach to address the link label prediction problem. As suggested in [8], we represent each signed edge as a 23-dimensional vector of following features: (a) 7 features based on (signed) degree on nodes that include the *embeddedness* of an edge, which corresponds to the number of neighbors its vertices have in common; and (b) 16 features corresponding to 16 triads involving that edge [8]. We use the following model for predicting the sign of an edge x (in the test set) represented as a vector of k features $(x_1, \ldots, x_k)$

$$P(+|x) = \frac{1}{1 + e^{-(w_0 + \sum_{i=1}^k w_i x_i)}}$$

where the parameters $w_0, \ldots, w_k$ are estimated using the training data. Note that in [8], the logistic regression model is trained on the entire set of the edges except for a test edge (Leave-one-out (LOO) analysis is conducted using each edge as a test edge exactly once). The same model can be readily extended to applications that generate limited training data, and learn the model on the training data all at once.

6 Empirical Evaluation

We evaluated the empirical performance of MF-LiSP and logistic regression (LR) using both the synthetic and real world data sets. We observed the performance of LR based approach for three different sets of features: 7-dimensional vector of degree features, 16-dimensional vector of triad features and 23-dimensional vector of all features. We first describe the network data sets that we used for our experiments.

Real World Network Data Set: We carried out experiments on the voting network of Wikipedia, henceforth referred to as the *Wiki-Vote* data set. The network comprises 7118 users and 107080 links among these users. Each of these 107080 links is signed: the sign of each link indicates a positive or negative vote by a user about the promotion in status to admin for some other user. Out of these 107080 links, 78.41% are positive and the rest negative. We employed three-fold cross validation, i.e., two-thirds of the data was used for training and the rest for testing.

Synthetic Data Sets: We generated two datasets each comprising 10000 users drawn randomly from real-world datasets Epinions and Slashdot, so that the network structure and distribution of links in these synthetic data sets closely follows the real-world social networks.

For all the data sets, we executed our experiments over 10 runs and then averaged the results to account for statistical significance. We compared the link sign prediction accuracy of methods based on MF-LiSP and LR with two baseline heuristics, namely *random* and *weighted random*. In the context of *random* sign prediction method, we randomly predict the sign of any given link to be either positive or negative. Note that the distribution of the positive and negative links could be very skewed (e.g. in the Wiki-Vote data set about 78.41% of the links are positive). To take into account this skewness, for any given link, we also predicted the sign of this link to be positive or negative in proportion to the respective share, and we refer to this as the *weighted* method.

We followed the methodology of Leskovec et al. [8] to evaluate the performance of these methods for different threshold embeddedness, Em. Specifically in the case of logistic regression, as described in Section 5, we represented each signed edge as a 23-dimensional vector (of which 16 features (triads) become relevant only when the edge vertices have common neighbors). Therefore, we expect LR methods to perform better with increasing embeddedness. In our experiments, we used three different levels of minimum embeddedness, $Em = 0$ (no filtering), $Em = 10$ and $Em = 25$.

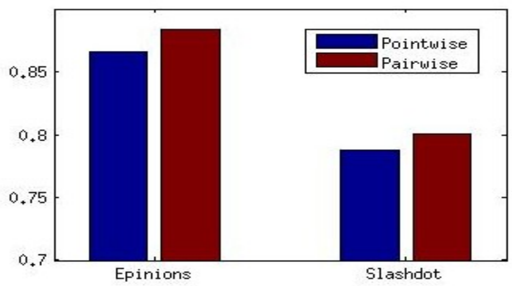
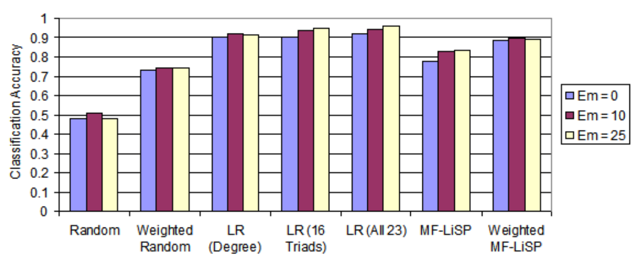
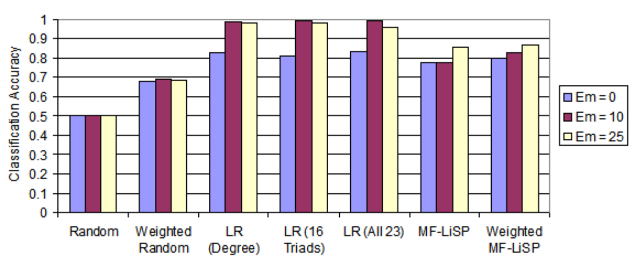


Figure 1: Comparison between pointwise and pairwise loss function on synthetic datasets.



(a) Epinions Dataset



(b) Slashdot Dataset

Figure 2: Synthetic Datasets: Classification accuracy of MF-LiSP, Logistic Regression, and Random.

To bring forth the importance of pairwise loss function, we present the results in Fig. 1 comparing the pointwise and pairwise loss functions for weighted MF-LiSP with $Em = 0$. Clearly, our pairwise loss function helps in reducing the classification error relative to the pointwise loss function. We also report the prediction accuracy of the MF-LiSP, LR and random methods with uniform and weighted priors on the synthetic (Figure 2.a and 2.b) and Wiki-Vote (Figure 3) data sets, for varying threshold on embeddedness, Em. Note that we do not include the results for weighted LR methods, as performance gain for LR is achieved using increasing level of minimum embeddedness, Em and we observe that due to greater dependency of LR on features, a weighted prior does not make significant difference.

From the results with $Em = 0$, i.e. considering all the edges irrespective of their embeddedness, we can clearly infer that MF-LiSP and LR methods outperform the random and the weighted random heuristics. Furthermore, although MF-LiSP has a comparable (slightly inferior) performance to LR, it is crucial to note that determining the important set of features, as required by LR, might not be feasible in all scenarios (for instance, due to privacy issues). MF-LiSP overcomes the need of using features for the learning task, and hence is very generic in its applicability.

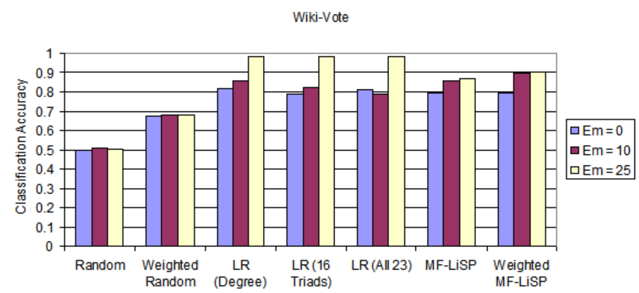


Figure 3: Wiki-Vote: Classification accuracy of MF-LiSP, Logistic Regression, and Random.

The experiments for $Em = \{10, 25\}$ show that the LR model is most effective for edges with higher embeddedness, while MF-LiSP is robust to varying embeddedness. Naturally, the embeddedness of the edges has no effect on the random methods. Therefore, LR is more useful in specific scenarios, e.g., when the nodes in the network are well connected to each other. However, since in general the social networks exhibit strong connectivity, we expect LR to perform well across a wide spectrum of network topologies.

7 Conclusions

We introduced the *link label prediction problem*: simultaneously predict the label of several links in social networks, and proposed a technique, *MF-LiSP*, for this problem. We provided strong generalization guarantees for MF-LiSP thereby theoretically establishing its efficacy for link label prediction. Since the bounds are generic, MF-LiSP can be readily adapted to more generic social network settings (e.g. when no links or only unsigned links are considered), or other domains (e.g. recommender systems). To the best of our knowledge, MF-LiSP is the first technique that has provable guarantees for the general link label prediction problem. We also investigated the applicability of the logistic regression based method [8] in the multiple label prediction setting. Both the techniques outperform the random and weighted-random heuristics on several synthetic data sets and real world benchmarks.

References

[1] O. Bousquet and A. Elisseeff. Stability and Generalization. *JMLR*, 2:499-526, 2002.

[2] E.J. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56:2053-2080, 2009.

[3] K.-Y. Chiang, N. Natarajan, A. Tewari, I.S. Dhillon. *Exploiting longer cycles for link prediction in signed networks.* In Proceeding of ACM CIKM, pages 1157-1162, 2011.

[4] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun. Who should share what?: item-level social influence prediction for users and posts sign prediction. In *SIGIR*, 2011.

[5] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient Projections onto the l_1-Ball for Learning in High Dimensions. In *ICML*, 2008.

[6] J. Kunegis, A. Lommatzsch, and C. Bauckhage. The Slashdot zoo: Mining a social network with negative edges. In *WWW*, pages 740-750, 2009.

[7] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *ACM SIGCHI*, pages 1361-1370, 2010.

[8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting Positive and Negative Links in Online Social Networks. In *WWW*, 2010.

[9] D. Liben-Nowell and J.M. Kleinberg. The link prediction problem for social networks. Proceedings of the 12th ACM International Conference on Information and Knowledge Management (CIKM), pages 556-559, 2003.

[10] U. Kuter and J. Golbeck. *SUNNY: A New Algorithm for Trust Inference in Social Networks Using Probabilistic Confidence Models*. In the Proceedings of AAAI, pages 1377-1382, 2007.

[11] H. Ma, D. Zhou, C. Liu, M.R. Lyu, and I. King. Recommender Systems with Social Regularization. In *WSDM*, 2011.

[12] A.K. Menon and C. Elkan. sign prediction via matrix factorization. In *ECML PKDD*, 2011.

[13] J. Scripps, P.-N. Tan, F. Chen, and A.-H. Esfahanian. A matrix alignment approach for sign prediction. In ICPR, pages 1-4, 2008.

[14] N. Srebro and A. Shraibman. Rank, trace-norm and max-norm. In *COLT*, pages 545560, 2005.

[15] S.-H. Yang, A.J. Smola, B. Long, H. Zha, and Y. Chang. *Friend or frenemy?: predicting signed ties in social networks*. In Proceedings of SIGIR, pages 555-564, 2012.

[16] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493-521, 2011.