Exploring and Inferring User–User Pseudo-Friendship for Sentiment Analysis with Heterogeneous Networks

Hongbo Deng[1]*, Jiawei Han[2], Hao Li[3], Heng Ji[3], Hongning Wang[2] and Yue Lu[4]

[1] *Yahoo Labs, Sunnyvale, CA, USA*

[2] *Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA*

[3] *City University of New York, New York, NY, USA*

[4] *Twitter Inc., San Francisco, CA, USA*

Received 11 September 2013; revised 6 January 2014; accepted 26 February 2014
DOI:10.1002/sam.11223
Published online in Wiley Online Library (wileyonlinelibrary.com).

Abstract: With the development of social media and social networks, user-generated content, such as forums, blogs and comments, are not only getting richer, but also ubiquitously interconnected with many other objects and entities, forming a heterogeneous information network between them. Sentiment analysis on such kinds of data can no longer ignore the information network, since it carries a lot of rich and valuable information, explicitly or implicitly, where some of them can be observed while others are not. However, most existing methods may heavily rely on the observed user–user friendship or similarity between objects, and can only handle a subgraph associated with a single topic. None of them takes into account the hidden and implicit dissimilarity, opposite opinions, and foe relationship. In this paper, we propose a novel information network-based framework which can infer hidden similarity and dissimilarity between users by exploring similar and opposite opinions, so as to improve post-level and user-level sentiment classification at the same time. More specifically, we develop a new *meta path*-based measure for inferring pseudo-friendship as well as dissimilarity between users, and propose a semi-supervised refining model by encoding similarity and dissimilarity from both user-level and post-level relations. We extensively evaluate the proposed approach and compare with several state-of-the-art techniques on two real-world forum datasets. Experimental results show that our proposed model with 10.5% labeled samples can achieve better performance than a traditional supervised model trained on 61.7% data samples. © 2014 Wiley Periodicals, Inc. Statistical Analysis and Data Mining 0: 000–000, 2014

Keywords: heterogeneous information network; semi-supervised refining model; sentiment analysis; dissimilarity

1. INTRODUCTION

Recently the rise of social media and social networks, such as blogs, forums, and Twitter, have fueled the online space with lots of reviews, ratings, and comments. For example, customers may give reviews for products, make ratings for movies, provide comments on services, present opinions on current events and politics, and so on. Over the years, sentiment has been a widely used measure of how customers view a company's products and services, and how people think about current events and politics. Sentiment analysis refers to the task of determining opinions, judgments, and other information related to the attitudes of a speaker or a writer with respect to some topics or the overall contextual polarity of a document. Based on such information, companies have the opportunity to examine what current and potential customers are saying about their products and services without costly and time-consuming surveys. Similarly, political organizations and candidates might be able to determine what issues the public is most interested in, as well as where they stand on those issues. Therefore, it is very important and highly desirable to conduct sentiment analysis automatically, which pertains to products, companies, and commercial and political entities.

Sentiment analysis has been studied with various approaches, such as lexicon-based methods 1,2 and learning-based methods 3, among which most of them are based on text content alone. But in social media, textual documents

Correspondence to: Hongbo Deng (hbdeng@yahoo-inc.com)

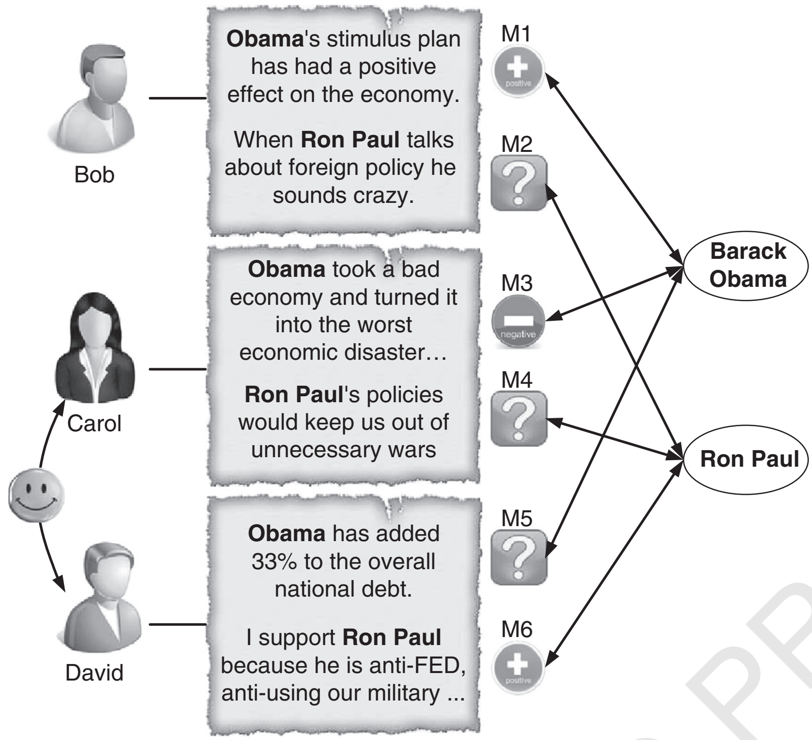

Fig. 1 An example data with User-Posts-Candidate heterogeneous networks and initial sentiment scores. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

are ubiquitously interconnected with many other entities, such as users and topics, in many ways, forming a heterogeneous network between documents and other entities. As shown in Figure 1, three users wrote six posts that were associated with 2 presidential candidates who were mentioned in these posts. In addition, we observe that users Caral and David are friends. Such a heterogeneous network carries rich semantic and valuable information, which should be utilized to improve sentiment analysis. The principles of homophily 4 and 'birds of a feature' 5 show the generation of a link depends on its context, and similar contexts could potentially lead to similar links, which suggests that users that are connected (become friends) may tend to hold similar opinions. For example, in a political forum, social network is formed between users due to homophily or influence, where friends are likely to share similar opinions to a certain topic. Therefore, it is essential to utilize the heterogeneous network, especially user–user relations, to enhance sentiment analysis.

Recently, although there are several studies 6, 7 proposed to improve sentiment analysis with user–user relationships 6 and other connections 7, the heterogeneous networks have not been fully exploited. These models rely on the explicit friendship between users or the similarity between objects, and can only handle a subgraph associated with a single topic/entity each time. In general, users may explicitly list their friends in the social media, but will not show their foes due to some hidden reason or lack of such function in the system. Actually, the foes usually hold opposite views on many issues, and there is significant difference or dissimilarity between them, which contain essential information as the friendship. However, none of the existing models takes into account the foe relationship, dissimilarity and opposite opinions.

To address the problem, we propose a novel information network-based framework by inferring pseudo-friendship between users and exploring post–post relations, so as to improve sentiment analysis. Given a number of labeled data with sentiment scores and observed friendship, we develop a semi-supervised refining model with user regularization (UserReg) to propagate the sentiment scores from labeled data to unlabeled data. It is worth noting that we not only consider the similarity but also the dissimilarity in the information network. With the sentiment scores, we develop a novel meta path-based measure for estimating the similarity and dissimilarity between entities, such as *pseudo-friends* and *pseudo-foes* between users. Consequently, the inferred pseudo-friendship along with the observed friendship may propagate the sentiment scores on the information network more effectively and consistently. Moreover, we also incorporate the post–post relations along with user–user relations to refine sentiment scores in a unified framework.

Color online, B&W in print

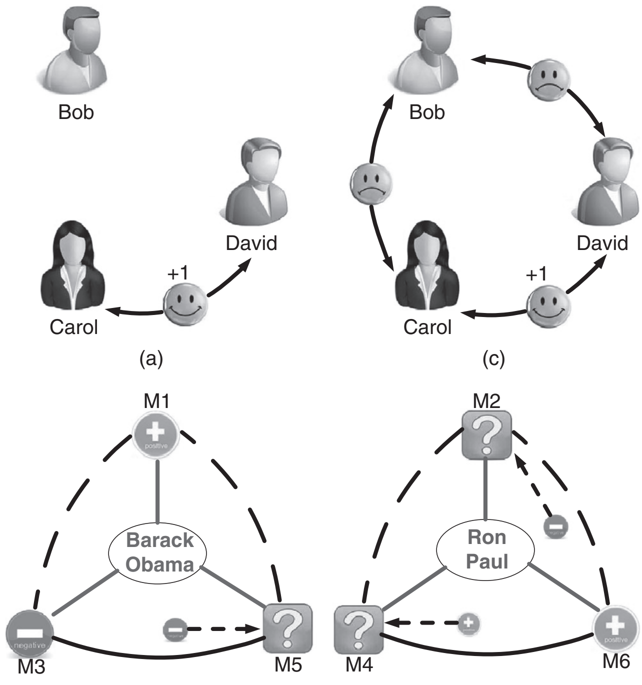

Fig. 2 The framework with (a) observed friendship, (b) refining sentiment scores based on the relationships between users and posts, and (c) the final inferred friendship. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The underlying intuition of our model is that better discovered friendship, including the observed and inferred friendship, can produce better sentiment scores, and better sentiment scores may help to discover more hidden friendship in turn, which will mutually enhance each other.

Let us take Figure 1 and Figure 2 as a concrete example, suppose Bob, Caral, and David each post two messages about presidential candidates, i.e. 'Barack Obama' and 'Ron Paul', respectively. Suppose we observe Carol and David are friends as shown in Figure 2(a), and we already obtain some labeled data that M1 and M6 are positive, M3 are negative, and M2, M4, and M5 are unknown (i.e., positive or negative). According to the principle of homophily 4, 6, the sentiment of M4 tends to be similar with that of M6 on 'Ron Paul' since they are written by two friends, Carol and David, as shown in Figure 2(b). Similarly, we can predict M5 to be negative as M3 on 'Barack Obama'. Now let us look at M2, we do not observe any friends for Bob, the author of M2. There is a critical question when considering to predict the sentiment of posts which are written by a user without explicit friendship: *How can we predict the polarity of M2 based on the homophily?* One possible answer is to infer the *pseudo-friends* or *pseudo-foes* based on the information network as well as the user's other data. As we can see, Bob publishes a positive message M1 on 'Barack Obama', while Bob and Carol hold opposite opinions (M1 and M3) on 'Barack Obama', which implies they are pseudo-foes in some sense. In this example, we can infer that Caral and David are pseudo-foes of Bob, as shown in Figure 2(c). Consequently, M2 can be refined to have opposite opinions of M4 and M6 which are published by their pseudo-foes. Thus, all the unknown sentiments and unobserved users' relationship are resolved.

The basic idea of our model is that friends are more likely to hold similar opinions, while foes are more likely to have conflicting opinions with respect to a certain entity or topic. On the other hand, based on a user's sentiment scores on different topics, we can infer the similarity and dissimilarity (i.e., *pseudo-friend* and *pseudo-foe* relationship) between users. Furthermore, based on the inferred friendship, we may jointly improve post-level and user-level sentiment analysis by considering the global consistency on the heterogeneous networks. Intuitively, these two steps can mutual enhance each other. In our proposed framework, we combine them together in a unified way. In terms of general social network, it is better to identify a sub-network or community structure 8 where friendship links between users are generated by sentimental contexts, such that follows the principle of homophily about users' opinions. The details of identifying the correlated community or sub-network are beyond the scope of this paper, and we will focus on sentiment analysis with correlated heterogeneous networks in this paper.

To illustrate our methodology, we apply the proposed model UserReg to sentiment classification with two real-world datasets. For the sentiment classification task, we compare with several different state-of-the-art models, including lexicon-based models and supervised models. It is shown that our model with 100 (10.5%) labeled data can perform much better than SVM-based supervised model with 600 (61.7%) training data on the political forum dataset.

The rest of this paper is organized as follows. We first introduce the preliminaries in Section 2. In Section 3, we systematically present and develop the information network-based framework. In Section 4, we conduct extensive experiments on sentiment classification. Finally, we review some related work in Section 5, and present our conclusions and future work in Section 6.

2. PRELIMINARIES

More formally, a forum consists of a set of threads, along with 1) a set of users $U = \{u_1, u_2, ..., u_m\}$; 2) a set of documents $D = \{d_1, d_2, ..., d_n\}$ written by users, where each d_i is a post with textual content; and 3) a set of

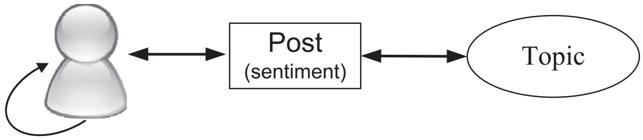

Fig. 3 A simple forum network schema. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

topics or issues T associated with them (we assume T is given which can be obtained from the forum). Let $F = \{f_1, f_2, \ldots, f_n\}$ denote the sentiment scores we want to identify for the set of posts D with respect to different topics. As shown in Figure 3, a heterogeneous information network is formed between these objects. For example, a user links with their written posts, and each post links with some specific topics if the post discusses that topic. Actually, each post in a forum is associated with many other information, such as the thread and the time. For simplicity and generality of this work, we will only focus on the simple network schema, as shown in Figure 3, and leave other information for future research.

We study sentiment analysis of forum discussions in this paper, and formulate the problem as a semi-supervised sentiment analysis: Given a forum information network with the relations, let the first $l \le n$ posts be labeled with sentiments $F_l = \{y_1, y_2, \ldots, y_l\} \in \{+1, -1\}$, the task is to predict the polarity of the remaining posts $F_u = \{f_{l+1}, \ldots, f_n\}$ which are unlabeled. For each unlabeled post, we can obtain an initial sentiment score y_i^0 $(l + 1 \le i \le n)$, whose value is the prediction of a separate method, such as lexicon-based method in our experiments. We aim to utilize the heterogeneous information network to better understand post-level sentiment as well as user-level opinions with respect to different topics.

3. INFORMATION NETWORK-ENHANCED FRAMEWORK

In this section, we propose a general framework to explore not only the post-level but also user-level relations on a heterogeneous information network.

3.1. Exploring User–User Relations

We discuss how to explore user–user relations so as to enhance post-level sentiment analysis.

3.1.1. Basic Principles

Our approach is based on some basic principles. First, similar users (e.g., friends) are prone to have similar and consistent opinions for a certain topic due to the principle of homophily 4. By contrast, two users tend to form a friendship if they share a lot of similar opinions. Inspired by the intuition, we can infer the pseudo-friendship based on their sentiment scores if the explicit friendship is not available.

Second, people tend to be foes if they hold conflicting or opposite opinions on many topics. On the other hand, the opposite opinions and dissimilarities between posts may indicate the *pseudo-foe* relationship between users. Such a pseudo-foe relationship, which is usually hidden, is as important as the friendship for improving sentiment analysis.

In order to validate these principles, we show two statistics on a political forum dataset (Details about the dataset are shown in Section 4.1). Figure 4(a) shows that the probability of two posts given by the same user or friends sharing the same sentiment on a topic is much higher than random. Figure 4(b) shows that it is more likely for users to be friends/connected if they share the same opinion on a certain topic than random (comparing blue and green bars); on the contrary, it is more unlikely for users to be friends (i.e., users tend to be foes) if they have opposite opinions on a certain topic than random (comparing red and green bars). As a whole, all the observations support our principles that the sentiment labels and social networks as well as heterogeneous networks are correlated, at least in the political domain.

3.1.2. Inferring user–user relations

Due to privacy and other issues, it is usually unavailable to obtain the explicit friendship between users. Therefore, it becomes very important to infer user–user relations, i.e. pseudo-friendship.

PathSim: We adopt a meta-path based similarity, i.e. *PathSim* 9, to estimate the similarity between objects or entities in a heterogeneous network, where two objects can be connected via different paths. For example, as shown in Figure 1, two users can be connected via 'user-post-candidate-post-user', or 'user-post-topic-post-user'. These paths are called *meta paths*, which consist of a sequence of relations defined between different object types (Please refer to 9 for more details).

PathSim is a similarity measure which can capture the semantics of peer similarity without bias to either highly visible objects or highly concentrated objects. Given a meta path $\mathcal{P}$, the corresponding similarity between two objects of the same type x and y is defined as:

$$\mathrm{Sim}(x, y) = \frac{2 \times |\{p_{x \rightsquigarrow y} : p_{x \rightsquigarrow y} \in \mathcal{P}\}|}{|\{p_{x \rightsquigarrow x} : p_{x \rightsquigarrow x} \in \mathcal{P}\}| + |\{p_{y \rightsquigarrow y} : p_{y \rightsquigarrow y} \in \mathcal{P}\}|},$$

where $p_{x \rightsquigarrow y}$ is a path instance between x and y that follows the defined meta path $\mathcal{P}$, $p_{x \rightsquigarrow x}$ and $p_{y \rightsquigarrow y}$ denote a path

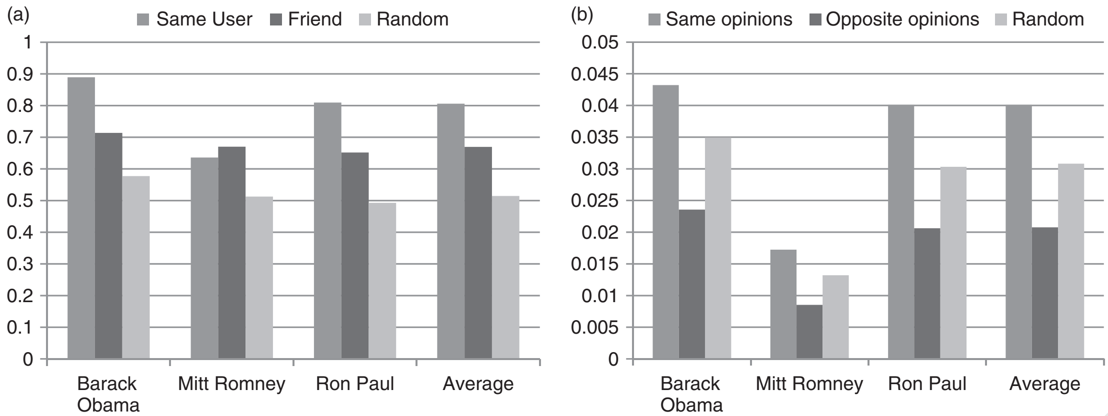

Fig. 4 Two statistics on a political forum dataset: (a) Probability of two posts having the same sentiment label with respect to different candidates/topics, conditioned on three types of relationships: two posts given by the same user, two friends, or two random users, respectively; (b) Probability that two users are friends/connected, conditioned on any two posts from those two users have the same opinions, opposite opinions, or random opinions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

starting from the object x or y and ending with the same object. This formulation is defined in terms of two parts: (1) the *connectivity* which is captured by the number of paths between two objects (i.e., $|\{p_{x\rightsquigarrow y} : p_{x\rightsquigarrow y} \in \mathcal{P}\}|$), and the *visibility* which is defined by the number of path instances between themselves. The intuition behind this formulation is that two objects should be similar if they are strongly connected and share comparable visibility.

Based on the meta path 'user-post-topic-post-user' (i.e., U-P-T-P-U), we can calculate the similarity between users, for example, $\mathrm{sim}(\mathrm{Bob}, \mathrm{Carol}) = 1$ indicates that there are two such path instances between Bob and Carol which are equal to the number of path instances between themselves. However, when looking into the sentiment of user's messages on 'Barack Obama', Bob and Carol hold the opposite opinions rather than similar opinions. Therefore, it is more reasonable to take the sentimental context into account, such that dissimilarity is introduced in the following subsection.

User–User Similarity and Dissimilarity: According to previous studies 9, two users are similar if they are strongly connected. However, when looking into the sentiment of user's posts, highly connected users may be very dissimilar in terms of sentiment context if they hold opposite opinions rather than similar opinions. Therefore, it is more reasonable to take the sentimental context into account, as well as the heterogeneous networks between users and posts. Intuitively, two users are likely to be similar if they are connected and hold consistent opinions on different topics. By contrast, two users are likely to be dissimilar if they are connected but hold opposite opinions.

To infer the pseudo-friendship between users, we introduce a similarity function $\mathrm{Sim}(u_i, u_j)$ with a real value ranging from -1 to $+1$. A score $\mathrm{Sim}(u_i, u_j)$ close to 1 indicates that these two users are *pseudo-friends* who have similar and consistent opinions, while a score $\mathrm{Sim}(u_i, u_j)$ close to -1 indicates that these two users are *pseudo-foes* who hold conflicting and opposite opinions on many topics. Moreover, a score close to 0 indicates that two users do not associate with each other, or they share both consistent and conflicting opinions.

Before we introduce the definition of the dissimilarity, let us define the *signed connectivity* of two posts, which will be used to estimate the similarity and dissimilarity between users. Two posts d_i and d_j on the same topic may be consistent or conflict with each other. A function $\mathrm{Sc}(d_i, d_j)$ is defined to indicate the degree of consistency or confliction between them ($-1 \le \mathrm{Sc}(d_i, d_j) \le 1$), $\mathrm{Sc}(d_i, d_j) = f(d_i) \cdot f(d_j)$, where $f(d_i)$ and $f(d_j)$ are current sentiment scores of posts. In the semi-supervised setting, we set $f(d_i) = y_i$ for labeled data ($1 \le i \le l$), and $f(d_i) = y_i^0$ for unlabeled data ($l + 1 \le i \le n$). For reliable inference, only labeled data and highly confident data are used, while setting some untrustworthy data $f(d_i) = 0$ if $-0.1 < y_i^0 < 0.1$. Thus we obtain that $\mathrm{Sc}(d_i, d_j) > 0$ if d_i and d_j are both positive or negative, and $\mathrm{Sc}(d_i, d_j) < 0$ if d_i and d_j hold conflicting or opposite opinions.

In order to capture both sentimental context and heterogeneous networks, we define a new *meta path*-based similarity between two users. For example, two users u_i and u_j can be connected via the path 'user-post-topic-post-user' (i.e., U-P-T-P-U) if both u_i and u_j have written at least one post about a same topic T. Given a meta path (e.g., $\mathcal{P}$ = U-P-T-P-U), we can infer the similarity and dissimilarity between users with the following definition:

$$\mathrm{Sim}(u_i, u_j) = \frac{2\sum_{d_k \in D_{u_i}, d_l \in D_{u_j}} \left(1_{(p_{u_i \rightsquigarrow u_j} \in \mathcal{P})} \cdot \mathrm{Sc}(d_k, d_l)\right)}{|\{p_{u_i \rightsquigarrow u_i} \in \mathcal{P}\}| + |\{p_{u_j \rightsquigarrow u_j} \in \mathcal{P}\}|},$$

where $p_{u_i \rightsquigarrow u_j}$ is a path instance between u_i and u_j that follows the defined meta path $\mathcal{P}$, and D_{u_i} and D_{u_j} are the set of posts written by user u_i and u_j, respectively. This formulation is defined in terms of two parts: (1) (numerator) - the *signed connectivity* which is captured by the number of paths connecting two users along with their consistent/conflicting opinions on the topics, and (2) (denominator) - the *visibility* which is defined by the number of path instances between themselves.

The underlying intuition of this formulation is that two users are more likely to be similar, i.e. $\mathrm{Sim}(u_i, u_j)$ close to 1, if they share similar opinions (both positive or negative) with respect to different topics, and we denote them as *pseudo-friends*. On the contrary, two users are more likely to be dissimilar, i.e. $\mathrm{Sim}(u_i, u_j)$ close to -1, if they have conflicting or opposite opinions on many topics, and we denote them as *pseudo-foes*. It is obvious that the similarity is symmetric, i.e. $\mathrm{Sim}(u_i, u_j) = \mathrm{Sim}(u_j, u_i)$, and we set $\mathrm{Sim}(u_i, u_i) = 1$ for all users.

3.1.3. User regularization

Since different posts may be associated with different topics, we denote two posts as *on-topic* posts if they discuss the same topic, and *off-topic* if they are about different topics. Based on the basic principles, two *on-topic* posts written by *pseudo-friends* are more likely to be consistent. To capture this assumption, we define to minimize the following user regularization term so as to constrain sentiment scores between posts,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \delta^T_{d_i,d_j} U_{d_i,d_j} (f_i - f_j)^2, \text{ if } U_{d_i,d_j} > 0,$$

where U is a matrix with entry $U_{d_i,d_j} = \mathrm{Sim}(u_{d_i}, u_{d_j})$, u_{d_i} is the user that wrote d_i, and $\delta^T_{d_i,d_j}$ is the delta function that equals 1 if d_i and d_j are *on-topic* posts, and 0 otherwise. On the other hand, two *on-topic* posts written by two *pseudo-foes* are more likely to be dissimilar. Correspondingly, we define to minimize the following user regularization term

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \delta^T_{d_i,d_j} (-U_{d_i,d_j}) (f_i + f_j)^2, \text{ if } U_{d_i,d_j} < 0,$$

which indicates that f_i and f_j should have opposite sentiment scores or both close to zero. By combining the previous two terms together, we would like to minimize a new penalty term

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \delta^T_{d_i,d_j} |U_{d_i,d_j}| (f_i - s^U_{ij} f_j)^2,$$

where $s^U_{ij} = 1$ if $U_{d_i,d_j} \geq 0$, and $s^U_{ij} = -1$ if $U_{d_i,d_j} < 0$. If the explicit friendship is available, we can combine the inferred pseudo-friendship U^{Infer} with the explicit friendship U^{Explicit} by simply aggregating them together $U = U^{\text{Infer}} + U^{\text{Explicit}}$.

3.2. Incorporating Post–Post Relations

Now we discuss how to explore relations between posts based on multiple evidences. To explore the most useful relations between posts, we will only explore relations between on-topic posts including cross-thread posts.

3.2.1. Similarity relation

Each post can be represented as a feature vector, e.g. bag of words, and a similarity ($\mathrm{Sim}(d_i, d_j) \geq 0$) between any two posts d_i and d_j can be calculated given a similarity measure. Generally, a large similarity implies that the two posts tend to express the same sentiment 10, 11. Although counter examples are easy to construct, it has been found that exploiting the similarity between posts can boost the performance of sentiment classification 10, 11.

Given a similarity measure, we can construct a kNN graph, where each post is connected to its k nearest on-topic posts. Let M^{Sim} be a similarity matrix with

$$M^{\text{Sim}}_{ij} = \begin{cases} \mathrm{Sim}(d_i, d_j), & j \in kNN(i) \\ 0, & \text{Otherwise} \end{cases}$$

We experiment with word-vector cosine similarity. To encode the assumption that two posts with a large similarity share similar sentiment, we define to minimize the following loss function

$$\sum_{i=1}^{n}\sum_{j\in kNN(i)} M^{\text{Sim}}_{ij} \left(f_i - f_j\right)^2.$$

3.2.2. Reply to relation

A forum discussion on a topic typically consists of many seed posts, and a large number of posts that are responses to a seed post or responses to responses. Typically there are explicit quotations from earlier posts for responses. Based on such quotations, we can extract 'Reply-to' relations between posts, and create a link between these two posts. An interesting characteristic of many forum discussions, especially political discussion board, is that users tend to quote posts by users with different views 12, 13. For example, users often debate a controversial topic, quoting and disputing each others' previous claims. To treat the 'Reply-to' relation between two posts, we can simply assume these two posts have opposite opinions with respect to a topic. Recently, some advanced methods 14, 15 were proposed to discover supporting and opposite 'Reply-to' relations between posts, which can be easily adapted into our framework.

Let R be a matrix to denote 'Reply-to' graph with entries $R_{ij} = 1$ if d_i replies to d_j with supporting opinions, $R_{ij} = -1$ if d_i replies to d_j with opposite opinions, and $R_{ij} = 0$ otherwise. In order to handle both positive and negative relations, we choose a suitable penalty term as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |R_{ij}| \left(f_i - s_{ij}^{R} f_j\right)^2,$$

where $s_{ij}^{R} = 1$ if $R_{ij} \geq 0$, and $s_{ij}^{R} = -1$ if $R_{ij} < 0$. In order to minimize the term, f_i and f_j should have similar sentiment scores when $R_{ij} > 0$ (i.e., d_i and d_j are mutual supportive). When $R_{ij} < 0$ (i.e., d_i and d_j are mutual exclusive/unsupportive), d_i and d_j should have opposite scores or both close to zero.

3.2.3. *User consistency*

Generally users may publish many posts on a topic in different threads. Different on-topic posts from the same user tend to express consistent opinion. More specifically, suppose we know one post from u_i shows strong positive opinion for a given topic, then all the other posts written by u_i on that topic would be likely to share a similar opinion. Following this assumption, we can encode it as a matrix A with $A_{ij} = 1$ if on-topic posts d_i and d_j are written by the same user (i.e., $\mathrm{user}(d_i) = \mathrm{user}(d_j)$). To force consistent sentiment scores among on-topic posts from the same user, we define to minimize the following cost function

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} \left(f_i - f_j\right)^2.$$

3.3. Regularization Framework

To put all the information together, we have the following objective function:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \left\{ \mu \sum_{i=l+1}^{n} (f_i - y_i^0)^2 + \frac{\beta}{2} \sum_{i=l+1}^{n}\sum_{j=1}^{n} |U'_{d_i,d_j}|(f_i - s_{ij}^{U'} f_j)^2 + \frac{(1-\beta)}{2} \sum_{i=l+1}^{n}\sum_{j=1}^{n} |P_{ij}|(f_i - s_{ij}^{P} f_j)^2 \right\}$$

subject to

$$\begin{aligned} f_i &= y_i,\ 1 \leq i \leq l, \\ U'_{d_i,d_j} &= \delta_{d_i,d_j}^{T} U_{d_i,d_j} \\ P_{ij} &= M_{ij}^{Sim} + R_{ij} + A_{ij} \end{aligned}$$

where $s_{ij}^{U'} = 1$ if $U'_{ij} \geq 0$, and $s_{ij}^{U'} = -1$ if $U'_{ij} < 0$, $s_{ij}^{P} = 1$ if $P_{ij} \geq 0$ and $s_{ij}^{P} = -1$ if $P_{ij} < 0$. μ and β are the weights to trade off three different components: The first component enforces the refined sentiment scores to fit the initial sentiment scores; the second component enables the user-level constraints based on inferred pseudo-friendship; and the third component incorporates the post-level supporting and opposite relations. The key idea is to refine sentiment scores by encoding similarity and dissimilarity from both user-level and post-level relations.

In the above semi-supervised setting, we are given a labeled set $F_l = [y_1, y_2, \ldots, y_l]^{\mathrm{T}}$, an unlabeled set $F_u = [f_{l+1}, \ldots, f_n]^{\mathrm{T}}$, and an initial sentiment score vector $Y_u^0 = [y_{l+1}^0, \ldots, y_n^0]^{\mathrm{T}}$ for each unlabeled post whose value is the prediction of a separate method (e.g., lexicon-based method in our experiments). Basically, we can split the weight matrices U' and P into four blocks as $U' = \begin{bmatrix} U'_{ll} & U'_{lu} \\ U'_{ul} & U'_{uu} \end{bmatrix}$ and $P = \begin{bmatrix} P_{ll} & P_{lu} \\ P_{ul} & P_{uu} \end{bmatrix}$, where P_{xy} is an $|x| \times |y|$ matrix. In the semi-supervised setting, the labeled set F_l is fixed, and our goal is to estimate the final sentiments F_u for unlabeled posts. Thus the original objective function can be rewritten as follows:

$$\begin{aligned} L \;=\; & \mu(F_u - Y_u^0)^T (F_u - Y_u^0) \\ & + \beta F_u^T (D_{|U'_{uu}|} - U'_{uu}) F_u - \beta F_u^T U'_{ul} F_l \\ & + (1-\beta) F_u^T (D_{|P_{uu}|} - P_{uu}) F_u - (1-\beta) F_u^T P_{ul} F_l, \end{aligned}$$

where $D_{|U'_{uu}|}$ is a diagonal matrix with $d_{ii} = \sum_{j=1}^{n} |U'_{ij}|$, and $D_{|P_{uu}|}$ is a diagonal matrix with $d_{ii} = \sum_{j=1}^{n} |P_{ij}|$ (Please note we ignore $F_l^T D_{|U'_{lu}|} F_l$ and $F_l^T D_{|P_{lu}|} F_l$ as they are constant given fixed F_l).

The above optimization problem can be solved directly as the objective function is convex 13, and a closed-form solution can be derived. To minimize L we only need to find F_u^* such that

$$\begin{aligned} \left.\frac{\partial L}{\partial F_u}\right|_{F_u = F_u^*} = & (\beta D_{|U'_{uu}|} + (1-\beta) D_{|P_{uu}|} + \mu I) F_u \\ & - (\beta U'_{uu} + (1-\beta) P_{uu}) F_u \\ & - (\beta U'_{ul} + (1-\beta) P_{ul}) F_l - \mu Y_u^0 = 0, \end{aligned}$$

where I is an identity matrix. Therefore, a closed-form solution can be derived as

$$F_u^* = \Big(\beta D_{|U'_{uu}|} + (1-\beta) D_{|P_{uu}|} + \mu I - \beta U'_{uu} - (1-\beta) P_{uu}\Big)^{-1} \left(\beta U'_{ul} F_l + (1-\beta) P_{ul} F_l + \mu Y_u^0\right). \quad (1)$$

Basically, the computational complexity of the closed-form solution is determined by the inverse of the matrices. Fortunately, both U'_{uu} and P_{uu} are sparse matrices, thus the cost of computation is determined by the number of nonzero entries, rather than the size of the matrix.

For a large-scale network, an iterative algorithm would be more effective and preferable to solve the optimization problem. Suppose F_u^t is the sentiment score vector of unlabeled set after t iterations. We repeat the following step until F_u converges,

$$F_u^{t+1} = \Big(\beta D_{|U'_{uu}|} + (1-\beta) D_{|P_{uu}|} + \mu I\Big)^{-1} \Big(\beta U'_{uu} F_u^t + (1-\beta) P_{uu} F_u^t + (\beta U'_{ul} + (1-\beta) P_{ul}) F_l + \mu Y_u^0\Big). (2)$$

Note that $(\beta D_{|U'_{uu}|} + (1-\beta) D_{|P_{uu}|} + \mu I)$ is a diagonal matrix, therefore its inverse is very easy to compute. Based on our empirical experiments, the iterative algorithm tends to converge after 10–20 loops. Given the initial sentiment scores Y_u^0 and the matrices U' and P, we can estimate the final sentiment scores F_u^* effectively.

3.4. The Overall Algorithm

By unifying the above regularization framework, we summarize the proposed algorithm in Algorithm 1. In this algorithm, note that we first perform preprocessing in a collection to get the information network and the observed friendship. The algorithm infers the pseudo-friendship, incorporates the post–post relations, and then iteratively updates F_u^{t+1} until it converges. For a small network, we can also replace steps 3–8 with Eq. (1) to estimate F_u^* directly.

Algorithm 1 Regularization Framework for Sentiment Analysis with Heterogeneous Networks

Input: An information network with the relations $G = (U \cup D \cup T, E)$, a set of posts with labeled sentiments $F_l = \{f_1, ..., f_l\}$, and the parameters μ and β.
Output: Sentiment scores for unlabeled set $F_u = \{f_{l+1}, ..., f_n\}$.

```
1: Infer pseudo-friendship Sim(u_{d_i}, u_{d_j}) and combine
   with explicit friendship if available by simply aggregating
   them together U = U^{Infer} + U^{Explicit};
2: Incorporate post-post relations P = M^{Sim} + R + A
3: t ← 0, diff ← 10e6, F_u^0 ← Y_u^0;
4: while t < MaxIteration and diff > MinThreshold do
5:     Update sentiment scores F_u^{t+1} as in Eq. (3.2);
6:     diff ← ∑(|F_u^{t+1} − F_u^t|);
7:     t ← t + 1
8: end while
9: Return F_u^*
```

In our experiments, we compared our models with several state-of-the-art methods, including SentiWordNet 1, SSL+WV 11, SSL+Dissim 13, and LP 15. Our regularization framework has strong connections with these methods, and can cover these methods by setting different parameters. We utilize SentiWordNet 1 to calculate the initial sentiment scores for unlabeled data. The method SSL+WV only considered the similarity graph on both labeled and unlabeled posts, which is equivalent to our model by setting $\beta = 0$ and $P = M^{Sim}$. Furthermore, SSL+Dissim introduced the dissimilarity on labeled and unlabeled posts, which is similar to our model by setting $\beta = 0$ and $P = M^{Sim} + R$ (where R consists of both positive and negative relations). Finally, LP is proposed to use three post–post relations $P = M^{Sim} + R + A$, but does not consider user–user friendship which is similar to our model by setting $\beta = 0$. The experimental results show that our proposed method achieves the best performance by considering both user–user and post–post relations.

4. EXPERIMENTAL EVALUATION

In the following experiments, we compare our proposed models with other methods on the sentiment analysis task through an empirical evaluation. In the rest of this section, we introduce the data collection, experimental setup, and report the experimental results.

4.1. Data Collection

We created our datasets from two online forums. The first dataset is crawled from the 'Election & Campaigns' board of a political forum[1], which is actively discussing elections and campaigns as well as political candidates. We collected the most recent posts from March 2011 to April 2012, containing 608 threads and 31 991 posts. In order to make it easier for the human judges to annotate, we further narrowed down to three popular US presidential candidates, and applied information extraction method to extract relevant posts. The basic statistics are shown in Table 1. There are totally 1901 labeled posts written by 232 unique users. Moreover, we manually labeled 419 positive and 553 negative posts with respect to the associated candidates, the rest of the posts are either neutral or not sure about their polarity. *For these 232 users, we crawled their profiles and observed a total of 782 friendship edges, which indicates that each user has 3.37 friends on average.*

The second dataset is crawled from a military forum[2], containing 43 483 threads and 1 343 427 posts. In order to make it easier for the human judges to annotate[3], we further narrowed down to five popular and controversial topics, and applied information retrieval method to retrieve the top five most relevant threads for each topics. The basic statics are shown

[1] http://www.politicalforum.com/elections-campaigns/
[2] http://forums.military.com/
[3] Please refer to ref. 15 for more details about human annotation.

Table 1. Basic statistics of datasets

Candidates	#Post	#Reply	#User	#Pos/#Neg
PF1901: Political Forum (with explicit user–user friendship)				
Barack Obama	681	188	167	93/220
Mitt Romney	335	118	92	50/75
Ron Paul	885	258	153	276/258
MF1560: Military Forum (without explicit user–user friendship)				
Topics	#Post	#Reply	#User	#Pos/#Neg
Abortion	297	41	93	110/70
Healthcare reform	323	62	84	90/121
Illegal immigrants	307	49	105	54/194
Iraq war	324	51	98	124/136
President Obama	309	42	96	59/97

in Table 1. There are totally 1560 labeled posts written by 320 unique users. We manually labeled 437 positive and 618 negative posts with respect to the topics, the rest of the posts are either neutral or not sure about their polarity. Unlike the above political forum, there is no explicit friendship between users listed in this forum.

4.2. Experimental Setup

The proposed algorithm can be applied to sentiment analysis. We evaluate our proposed model and compare with several state-of-the-art methods as follows:

SentiWordNet We tag each post by taking the average SentiWordNet 1 score of words. It represents an unsupervised sentiment analysis method which only relies on the text. Note the output is used as initial sentiment scores for test (unlabeled) data in the semi-supervised learning methods.

SSL+WV This method is proposed for semi-supervised sentiment classification in 11. They create a similarity graph on both labeled and unlabeled posts, and the graph is formed based on the word- or sentence-level similarity. Here we choose the same similarity measure as our proposed method.

SSL+Dissim The authors 13 further improved the above method by considering both similarity and dissimilarity on labeled and unlabeled posts. The dissimilarity edge is created between two posts if they have exhibited the 'Reply-to' relation.

LP LP is proposed in 15 for analyzing agree/disgree relations between posts. Then the post–post relations are used in a linear programming framework for sentiment analysis (while they did not consider user–user relations). The method is defined in the unsupervised setting, and cannot be directly applied in semi-supervised setting. We develop LP+ for comparison in the semi-supervised setting. Note that this method uses all the three post–post relations, but does not use user–user friendship.

UserReg This is our proposed method by exploring and utilizing both post–post relations and user–user relations. It is the first method to explore the inferred user–user relations[4] for enhancing sentiment classification.

In the semi-supervised setting, we randomly choose a small number of labeled data with equal size of positive and negative points. In order to randomize the experiments and make the comparison fair, we conduct the evaluation with the size of labeled data ranging from 50 (5.3%) to 250 (26.3%), choose the same random initializations for different models, and use the same similarity measure and 'Reply-to' relation. For each method, 10 test runs were conducted, and the final performance score represents the average result across 10 trials.

To quantitatively compare with these methods, we use several popular metrics to evaluate the sentiment classification, including accuracy, precision, and recall[5]. The accuracy (AC) is defined as

$$\mathrm{AC} = \frac{\sum_{i=1}^{n} \delta(f(m_i) * t(m_i) > 0)}{n},$$

where n denotes the total number of test/unlabeled data, $t(m_i)$ is the true polarity of message m_i, and $\delta(f(m_i) * t(m_i) > 0)$ is the delta function that equals one if $f(m_i)$ and $t(m_i)$ are both positive or negative, and equals zero otherwise.

4.3. Evaluation of Post-Level Classification

We first evaluate the accuracy of the post-level sentiment classification. The ground truth is derived from the positive and negative posts based on post level judgment, and ignore the neutral and 'Not Sure' cases.

4.3.1. Comparison with different models

In Table 2, we compare with different models on the dataset *PF1901*. Top part of the table shows the performance of a Lexicon-based unsupervised method SentiWordNet 1 and a SVM-based supervised method 16. The accuracy of SentiWordNet 1 without using any supervision is 44.4%, indicating the poor performance of unsupervised method. Support vector machines (SVMs) 16 have been shown to be highly effective at traditional

Table 2. Accuracy of post-level classification *-PF1901*

Method	Accuracy	Precision	Recall
SentiWordNet	0.4444	0.4573	0.6134
SVM	0.6124	0.556	0.6548
Semi-supervised setting with 100 (10.5%) labels			
SSL+WV	0.5325	0.4619	0.6144
SSL+Dissim	0.5431	0.4711	0.6084
LP+	0.6607	0.5879	0.6707
UserReg	**0.6903**	**0.6125**	**0.7344**
UserReg*	**0.7275**	**0.6578**	**0.7493**

UserReg* uses both the inferred user–user relation and the explicit friendship, while UserReg only use the inferred user–user relation.

[4] Previous studies only consider explicit friendship.
[5] http://en.wikipedia.org/wiki/Precision_and_recall

text categorization as well as sentiment analysis. We used LIBSVM 17 for training and testing. The model was trained on 300 positive and 300 negative samples, and tested on the remaining 372 samples. We tried different features including N-gram feature, lexicon feature, POS feature, punctuation feature, and so no. The best classification accuracy we obtained is 61.2%. From the results, although supervised model achieves better performance than Lexicon-based unsupervised model, it is still much lower than the results reported in previous studies on movie reviews 16[6]. The reason is that the sentiment analysis task on forum data is fairly difficult and complicated. For example, 'NONE of the GOP candidates have a significant advantage on national polls against Obama'. This post is supporting 'Obama' by arguing with GOP supporters. Within the context, 'significant' and 'advantage' are positive words while 'against' and 'NONE' are negative words, so it is not a trivial work to identify the correct sentiment label of the sentence by using lexicon words. If the user of this post wrote some other posts to support 'Obama', it becomes easier to classify it correctly by combining such information with previous posts. Our proposed model can connect complicated posts with straightforward ones through heterogeneous networks, which can enhance the performance significantly.

In the lower part of Table 2, we compare three baselines with our UserReg method in the semi-supervised setting. For each method, we use 100 (10.5%) posts as labeled data and test on the rest of the posts. The matrix M^{Sim} is constructed by building a kNN graph ($k = 20$) between posts based on word-vector (TF-IDF) cosine similarity. For 'Reply-to' relation, we follow 13 by simply setting $R_{ij} = -1$ if d_i replies to d_j. As shown in the table, it is obvious that (1) SSL+WV and SSL+Dissim perform slightly better than SentiWordNet, indicating that only considering similarity and reply-to relations among posts is not effective enough; (2) LP+ provides better accuracy than the previous two baselines, suggesting that user consistency is a reasonable relation; (3) our proposed methods, both UserReg and UserReg*, perform much better than all the other methods with a statistically significant improvement ($p < 0.05$) in all measures. This is because we further explore and infer pseudo-friends and pseudo-foes among users, and our model can benefit from the inferred pseudo-friendship as well as the explicit observed friendship.

In Table 3, we compare with different methods on another dataset *MF1560*. In order to fairly compare with the unsupervised

Table 3. Accuracy of• post-level classification -*MF1560*

Method	Accuracy	Precision	Recall
SentiWordNet	0.4664	0.4275	0.5469
LP	0.4844	0.4271	0.5629
Semi-supervised setting with 100 (9.5%) labels			
SSL+WV	0.5086	0.4272	0.6191
SSL+Dissim	0.5044	0.4217	0.5956
LP+	0.5353	0.4466	0.6121
UserReg	**0.556**	**0.4659**	**0.6496**

[6] Note that the accuracy of SentiWordNet is 72% for movie reviews, but it is only 44.4% for this forum data

method LP proposed in 15, we use the same topic-based similarity measure and post–post relations discovered in 15 (i.e., set M^{Sim} and R to be T^{agr} and $(R^{agr} - R^{dis})$, respectively). The top part of this table shows the performance of two unsupervised methods SentiWordNet and LP 15, and LP performs better than SentiWordNet because it considers three relations between posts. The lower part of the table shows the comparison of different methods in the semi-supervised setting with 100 (9.5%) labels. Similarly, SSL+WV and SSL+Dissim perform slightly better than SentiWordNet, and LP+ achieves better results by considering user consistency. As expected, our proposed method UserReg achieves the best performance. The overall results on the dataset *MF1560* are worse than the results on *PF1901*, and one major reason is the heterogeneous networks between users, posts and topics on *MF1560* are more sparse which somehow limits the advantage of our method.

Varying Labeled Set: In previous experiments, we fixed $|L| = 100$. Now we systematically vary labeled set size $|L| \in \{50, 100, 150, 200, 250\}$ to investigate the effect of semi-supervised learning. For each $|L|$, we run 10 trials where we randomly split the corpus into labeled and test (unlabeled) sets. We ensure that all two classes are represented in each labeled set. The same random splits are used for all methods, allowing paired t-tests for statistical significance. All reported results are average test set accuracy. Figure 5 shows the results on datasets *PF1901* and *MF1560*. Generally, the models perform better with the increase of the size of labeled data, especially for LP+ and our model UserReg. In the dataset *PF1901*, SSL+WV and SSL+Dissim approaches perform slightly better with larger labeled set sizes. However, in the dataset *MF1560*, these two approaches do not perform better with large labeled set sizes. Apparently, the reason is that topic-based similarity measure is too rough and not accurate enough to capture the semantics. Final, we can observe UserReg and UserReg* achieve much higher performance than other methods in all the settings.

4.3.2. *Parameter analysis*

There are two parameters, decay factor β and regularization parameter μ, in our method. Previous experimental results were obtained by empirically setting $\beta = 0.5$ and $\mu = 0.1$. The optimal parameters can be obtained by cross validation. In this subsection, we study and evaluate the effect of parameters β and μ based on the dataset *PF1901* using 100 labeled data (The results on *MF1560* illustrate similar results).

Figure 6(a) shows the performance of UserReg and UserReg* by varying the decay factor β from 0.1 to 0.9 (and fixing $\mu = 0.1$). As mentioned before, β is used to control the balance between user–user relations and post–post relations. We can see that the performance is relatively stable, and it is better to set β between 0.5 and 0.75. The performance of UserReg* is higher than that of UserReg.

Figure 6(b) shows the performance of our models by varying the regularization parameter μ from 0.02 to 50 (and fixing $\beta = 0.5$). As mentioned before, the parameter μ is used to control the trade-off between the consistency of sentiment scores on the information network and the initial scores. When μ is set

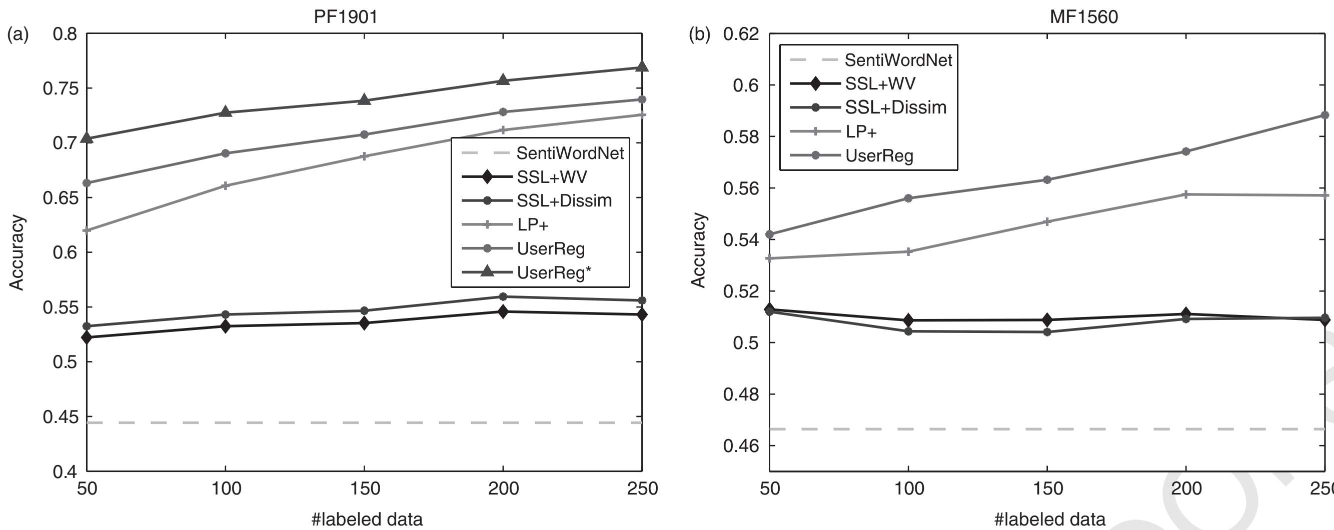

Fig. 5 Varying the number of labeled data. (a) PF1901 and (b) MF1560 [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

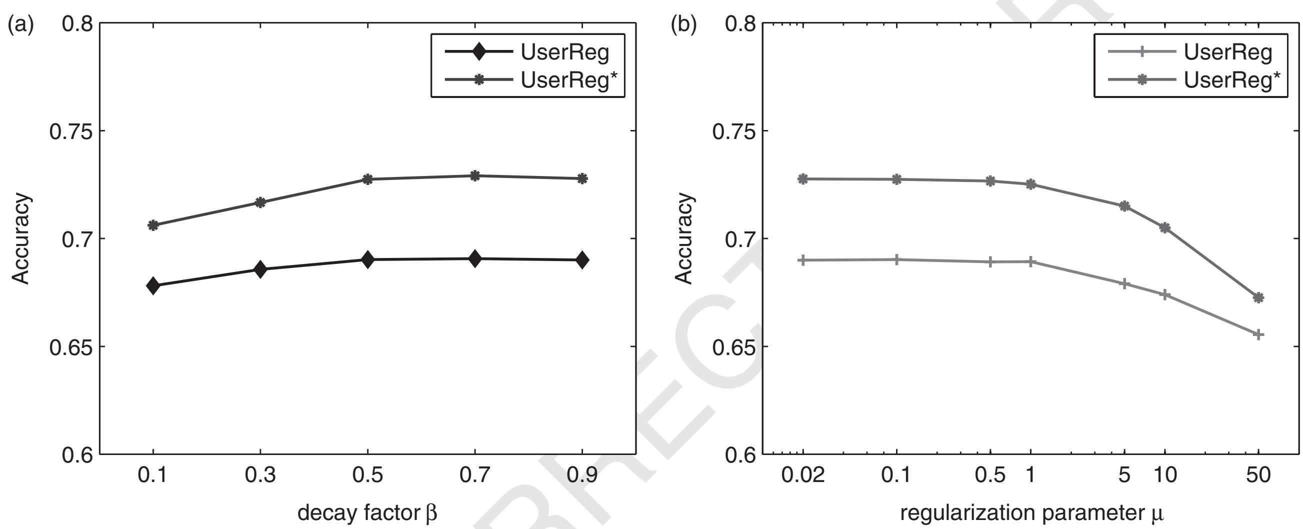

Fig. 6 The effect of varying parameters β and μ with 100 labels. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

to 1 or less than 1, the model tends to trust more on the information network (e.g., explicit or inferred friendship and post–post relations). With the increase of μ, the model tends to trust more on the initial scores. As we can see, when μ is set to 5 and 10, the final results start to decrease. It is clear that the performance of our model is very stable by setting μ to be 1 or less than 1, which confirms the effectiveness of our method. We empirically set the parameter $\mu = 1$ in other experiments.

4.4. Evaluation of User-Level Sentiment

Beyond evaluating the quality of post-level sentiment polarity, it would be more meaningful to assess the model's capability of identifying user-level sentiment polarity, which can be further used as higher level summarization of user's opinion toward specific topics.

Following the setting used in ref. 15, we only evaluate on a subset of uses, who possess strong opinions. In particular, we selected the users in the MF1560 dataset with at least two posts and their aggregated ground-truth opinion score $s > 0.5$. This results in 57 users with strong positive opinions and 78 users with strong negative opinions for the selected five topics. To compare the reported results in 15, we also evaluated prediction accuracy for supporting users (i.e., 'for'), against users (i.e., 'against') and both of them. The experiment results are listed in Table 4

Comparing to the unsupervised methods list on the top half of Table 4, all the semi-supervised models achieved promising

Table 4. Accuracy of user opinion prediction

Method	Accuracy (For)	Accuracy (Against)	Accuracy (For+Against)
SentiWordNet	0.6250	0.5250	0.5896
LP	0.6429	0.5513	0.5896
Semi-supervised models with 100 (9.5%) labels			
SSL+WV	0.7143	0.4746	0.5746
SSL+Dissim	0.6786	0.4615	0.5522
LP+	0.7143	0.5385	0.6119
UserReg	**0.7857**	**0.6410**	**0.7015**

accuracy improvement in the '*for*' category, especially for our proposed UserReg method (accuracy improvement over 22.2% over the unsupervised LP method). And for the 'against' category, our method also achieved the largest improvement over other semi-supervised methods (accuracy improved over 19.0% over the runner-up LP+ method). These improvements confirm the usefulness of prorogating information through the heterogenous network for better classification performance.

5. RELATED WORK

Traditional sentiment analysis has mostly focused on formal genres such as newswire, many levels of granularity such as document level 3, sentence level 18, and phrase level 19. During recent years informal texts in forum discussions and microblogging have become one of the major forms of online communication, enabling the sharing of real-time updates by both individuals and organizations. Informal genres have exhibited advantages over traditional news agencies in the success of reporting news more timely and local information. For example, some work has successfully used sentiment analysis for presidential elections 20,21, incorporating profile information of posters and president candidates. Some recent studies have moved to informal genres such as tweets 22, 20, 23 and forum discussions 15. However, most of them were based on text content alone. The debating nature and informality of forum posts have brought great additional challenges to sentiment analysis. It is thus impractical to use standard supervised machine learning techniques alone which are dependent on annotated training examples.

Several studies 6, 7 have been proposed to improve sentiment analysis with user–user relationships 6 and other connections 7. In 7, the authors exploited knowledge about word types encoded in a lexicon, in combination with the Twitter follower graph for label propagation to improve sentiment analysis for tweets. However, this model treats different types of objects in a similar way, which does not fully explore the heterogeneous information network. In 24, the authors exploited the retweet networks to classify the users into coarse-grained sentiments (left and right), without analyzing the content of messages. Tan et al. 6 proposed another approach to predict user-level sentiment with extracted social networks from followers and '@' mentions. However, this model cannot handle the sentiment in the post level, and it heavily relies on the explicit observed friendship. Our proposed model is different, which can not only predict both post-level and user-level sentiment in a unified way, but also infer the hidden friendship as well as dissimilarity between users. Previous research on sentiment analysis tried to associate aspects with sentiments, but they mainly focused on lexical level 25, 26 or event level 27. By contrast, we extracted explicit posts discussing certain topics and target entities automatically, and analyzed sentiments that are linked to topics and entities.

In this work, a new PathSim is proposed to estimate the similarity and dissimilarity between users in terms of similar or opposite opinions. This work is very relevant to the link prediction and social tie inferring, which has been studied in 28, 29, 30. The difference is that these methods focus on the prediction of positive and negative links on social network, which do not have strong correlation with similar or opposite opinions. In addition, we propose a novel regularization framework to utilize the discovered similarity and dissimilarity to improve sentiment analysis. This work is also related to graph-based semi-supervised learning 31, 32, 33, 34, 11, 13, 35, which usually assumes label smoothness over the graph. These types of graph regularization methods have been successfully applied in many data mining tasks 36, 37, 38. In 37, the authors developed a graph-based re-ranking model by regularizing the smoothness of relevance scores over the latent graph. NetPLSA 36 and TMBP 38 explored graph-based regularizers with topic modeling. Glodberg and Zhu 11 applied a graph-based semi-supervised learning algorithm to address sentiment analysis, by constructing similarity graphs to ensure that similar reviews receive similar labels. Our work is different from theirs, as we explore not only the relations between posts, but also the relations between users, including explicitly observed and implicitly inferred friendship.

6. CONCLUSIONS AND FUTURE WORK

We have presented a novel information network-based framework for enhancing sentiment analysis. Consequently, we have investigated a new meta path-based measure which can estimate not only the similarity but also the dissimilarity between users. Furthermore, a semi-supervised refining model with user regularization was developed to estimate the sentiment scores by exploring both user-level and post-level relations. The key to refining the sentiment scores is the global consistency over the heterogeneous information network, which leverages both the explicit information such as observed user–user friendship and the implicit information such as inferred friendship and dissimilarity between objects. Experimental results on the sentiment classification task show the effectiveness and correctness of our proposed approach, and the improvement of our approach is promising.

For future work, there are a few research directions to improve the current framework: (1) We have a strong assumption in this paper that friends hold similar options with respect to all topics/objects. One research direction is to relax the assumption and extend our framework by exploring the correlation between different topics/objects, such that friends who hold similar options in one topic will hold similar options in another correlated topic. (2) It is very promising to improve the current framework by adapting advanced link prediction methods 28, 29, 30 to estimate the similarity and dissimilarity between users in terms of sentiment analysis. (3) It would be interesting to take more information such as time and more complicated heterogeneous networks into consideration.

ACKNOWLEDGMENTS

The work was supported in part by the U.S. National Science Foundation grants IIS-0905215, by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053 (NS-CTA), and U.S. Air Force Office of Scientific Research MURI award FA9550-08-1-0265. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] S. Baccianella, A. Esuli, and F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, In *Proceedings of LREC*, xxx, 2010.

[2] B. OConnor, R. Balasubramanyan, B. Routledge, and N. Smith, From tweets to polls: Linking text sentiment to public opinion time series, In Proceedings of the ICWSM, xxx, (2010), 122–129.

[3] B. Pang and L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, In *Proceedings of ACL*, xxx, (2004), 271–278.

[4] P. Lazarsfeld and R. Merton, Friendship as a social process: a substantive and methodological analysis, In Freedom and *Control in Modern Society*, Vol. 18, xxx, (1954), 18–66.

[5] M. McPherson, L. Smith-Lovin, and J. Cook, Birds of a feather: homophily in social networks, Annu Rev Sociol 27 ((2001)), 415–444.

[6] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, User-level sentiment analysis incorporating social networks, In Proceedings of KDD, xxx, (2011), 1397–1405.

[7] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, Twitter polarity classification with label propagation over lexical links and the follower graph, In Proceedings of EMNLP, xxx, (2011), 53–63.

[8] M. Girvan and M. Newman, Community structure in social and biological networks, PNAS 99(12) ((2002)), 7821–7826.

[9] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, Pathsim: meta path-based top-k similarity search in heterogeneous information networks, PVLDB, 4(11) ((2011)), 992–1003.

[10] B. Pang and L. Lee, Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales, In *Proceedings of• ACL,* xxx, (2005), 115–124. AQ3

[11] A. Goldberg and X. Zhu, Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization, In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, xxx, (2006), 45–52.

[12] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, Mining newsgroups using networks arising from• social behavior, In Proceedings of WWW, xxx, ACM, (2003), 529–535. AQ4

[13] A. B. Goldberg, X. Zhu, and S. Wright, Dissimilarity in graph-based semi-supervised classification, In Proceedings of AISTATS, xxx, (2007).

[14] A. Murakami and R. Raymond, Support or oppose?: classifying positions in online debates from reply activities and opinion expressions, In Proceedings of COLING, xxx, (2010), 869–875.

[15] Y. Lu, H. Wang, C. Zhai, , and D. Roth, Unsupervised discovery of opposing opinion networks from forum discussions, In Proceedings of CIKM, xxx, (2012).

[16] P. D. Turney, Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews, In Proceedings of ACL, xxx, (2002), 417–424.

[17] C. Chang and C. Lin, Libsvm: a library for support vector machines, TIST, 2(3) ((2011)), 27–.

[18] S.-M. Kim and E. Hovy, Determining the sentiment of opinions, In Proceedings of COLING, xxx, (2004).

[19] A. Agarwal, F. Biadsy, and K. Mckeown, Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams, In Proceedings of EACL, xxx, (2009).

[20] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, Predicting elections with twitter: What 140 characters reveal about political sentiment, In Proceedings of ICWSM, xxx, (2010).

[21] M. Choy, M. L. Cheong, M. N. Laik, and K. P. Shung, A sentiment analysis of singapore presidential election 2011 using twitter data with census correction, In Proceedings of CoRR, Vol. abs/1108. 5520, xxx, (2011).

[22] A. Bermingham and A. Smeaton, Classifying sentiment in microblogs: Is brevity an advantage? In Proceedings of CIKM, xxx, (2010).

[23] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, Sentiment analysis of twitter data, In Proceedings of ACL/HLT Workshop on Language and Social Media, xxx, (2011).

[24] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, A. Flammini, and F. Menczer, Political polarization on twitter, In Proceedings of ICWSM, xxx, (2011).

[25] X. Zhao, J. Jiang, H. Yan, and X. Li, Jointly modeling aspects and opinions with a maxent-lda hybrid, In Proceedings of EMNLP, xxx, (2010).

[26] S. Brody and N. Elhadad, An unsupervised aspect-sentiment model for online reviews, In Proceedings of NAACL, xxx, (2010).

[27] A. Kolya, D. Das, A. Ekbal, and S. Bandyopadhyay, Identifying event-sentiment association using lexical equivalence and co-reference approaches, In ACL Workshop on Relational Models of Semantics, xxx, (2011).

[28] D. Liben-Nowell and J. M. Kleinberg, The link-prediction problem for social networks, JASIST, 58(7) ((2007)), 1019–1031.

[29] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, Predicting positive and negative links in online social networks, In Proceedings of WWW, xxx, (2010), 641–650.

[30] J. Tang, T. Lou, and J. M. Kleinberg, Inferring social ties across heterogenous networks, In Proceedings of WSDM, xxx, (2012), 743–752.

[31] X. Zhu, Z. Ghahramani, and J. D. Lafferty, Semi-supervised learning using gaussian fields and harmonic functions, In Proceedings of ICML, xxx, (2003), 912–919.
[32] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, Learning with local and global consistency, In Proceedings of NIPS, xxx, (2003).
[33] A. Smola and R. Kondor, Kernels and regularization on graphs. In Proceedings of COLT, xxx, (2003).
[34] D. Zhou, B. Schölkopf, and T. Hofmann, Semi-supervised learning on directed graphs, In Proceedings of NIPS, xxx, (2004).
[35] M. Li, X.-B. Xue, and Z.-H. Zhou, Exploiting multi-modal interactions: A unified framework, In Proceedings of IJCAI, xxx, (2009), 1120–1125.
[36] Q. Mei, D. Cai, D. Zhang, and C. Zhai, Topic modeling with network regularization, In Proceedings of WWW, xxx, (2008), 101–110.
[37] H. Deng, M. R. Lyu, and I. King, Effective latent space graph-based re-ranking model with global consistency, In Proceedings of WSDM, xxx, (2009), 212–221.
[38] H. Deng, J. Han, B. Zhao, Y. Yu, and C. X. Lin, Probabilistic topic models with biased propagation on heterogeneous information networks, In Proceedings of KDD, xxx, (2011), 1271–1279.

QUERIES TO BE ANSWERED BY AUTHOR

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. DO NOT mark your corrections on this query sheet.

Queries from the Copyeditor:

AQ1. Please provide significance for bold values in Tables 2, 3 and 4.

AQ2. Please provide significance for bold values in Tables 2, 3 and 4.

AQ3. Please provide place and date details for all proceedings references.

AQ4. Please provide place of publication for ACM.