Exploiting Sentiment Homophily for Link Prediction

Guangchao Yuan Pradeep K. Murukannaiah Zhe Zhang Munindar P. Singh

North Carolina State University, Raleigh, NC, USA
{gyuan,pmuruka,zzhang13,singh}@ncsu.edu

ABSTRACT

Link prediction on social media is an important problem for recommendation systems. Understanding the interplay of users' sentiments and social relationships can be potentially valuable. Specifically, we study how to exploit sentiment homophily for link prediction. We evaluate our approach on a dataset gathered from Twitter that consists of tweets sent in one month during U.S. 2012 political campaign along with the "follows" relationship between users. Our first contribution is defining a set of sentiment-based features that help predict the likelihood of two users becoming "friends" (i.e., mutually mentioning or following each other) based on their sentiments toward topics of mutual interest. Our evaluation in a supervised learning framework demonstrates the benefits of sentiment-based features in link prediction. We find that Adamic-Adar and Euclidean distance measures are the best predictors. Our second contribution is proposing a factor graph model that incorporates a sentiment-based variant of cognitive balance theory. Our evaluation shows that, when tie strength is not too weak, our model is more effective in link prediction than traditional machine learning techniques.

Categories and Subject Descriptors

H.2.8 [**Information Systems**]: Database Applications—*Data mining*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology*

General Terms

Algorithms, Experimentation

Keywords

Social media; Social networks; Link prediction; Sentiment analysis; Topic affiliation

1. INTRODUCTION

Link prediction refers to inferring potential relationships from a snapshot of a social network. A common intuition behind link prediction approaches is the presence of *homophily* in networks—similarity breeds connections [25]. Existing works derive similarities among users from network topology [21] (structural similarity), and users' interests [30] and geography [29] (semantic similarity).

However, structurally or semantically similar users may express different sentiments toward a common characteristic. Thelwall [33] found some evidence of both positive and negative sentiment homophily among MySpace friends. We posit that there exists a *sentiment homophily* in networks—similarity in users' sentiments breeds connection. Sentiment can be an important trait for link prediction because of its application in different domains such as political election prediction [9], location recommendation [37], and event detection [34].

Romero et al. [28] found that topics of interest to users can predict social relationships. For example, two users interested in the topic "Obama for President" are likely to be friends. However, the two users may exhibit the same (both support or oppose Obama) or contradictory sentiments (one supports and the other opposes Obama) toward the topic. Motivated by sentiment homophily, we imagine that the two users are more likely to become friends in the former case than in the latter. Based on the above intuition, we ask two questions:

- How may we exploit sentiments for link prediction?
- Can sentiment homophily help in link prediction?

Two challenges arise when answering the first question. First, how can we design sentiment-based features between two users in order to quantify sentiment homophily? Unlike features such as number of common friends, age, number of common places, designing sentiment features is much more complex because interpreting the sentiment of a tweet depends upon its domain and topic.

Second, employing traditional machine learning techniques (e.g., logistic regression) for link prediction assumes independence among pairs of nodes in a network, i.e., whether A–B is connected is independent of other connected pairs. However, such a case seldom exists in the real world. Heider [16] proposed cognitive balance theory in social psychology suggesting that if strong ties A–B and A–C exist, the likelihood of B–C becoming a tie (whether weak or strong) increases because of the "psychological strain": C will want to maintain his or her own feelings to agree with A and A's friend, B. Granovetter claimed that the B–C tie is always

present in this case. The strength of a tie can be "a combination of the amount of time, the emotional intensity, the intimacy, and the reciprocal services" [11]. Therefore, we hypothesize that it is nontrivial to capture dependence between pairs of nodes (e.g., A–B and A–C ties predicting the B–C tie) via a machine learning technique. If sentiment homophily exists, can we leverage such homophily to quantify the strength of a tie? Will the sentiment-based cognitive balance theory help in link prediction? How to build such a model and how to define the strength of a tie through sentiments are our second challenge.

We employ a dataset of political tweets (and associated users) to address the second question. We extract users' sentiments toward different topics from their tweets, where sentiments are modeled as numeric scores and categorical values. Further, we design several sentiment-based features, and evaluate the effect of sentiment homophily in a supervised setting on two social networks: the mutual-follow graph and the graph formed by users referring to each other using "@" mentions (Section 4). We find that sentiment-based features improve the performance of link prediction in terms of the F_1 score on both networks. We also investigate each sentiment-based feature and find that sentiment features based on the Adamic-Adar and Euclidean distance measures are the best predictors (Section 6.1).

We further propose a factor graph model based on Dong et al. [10], incorporating Heider's cognitive balance theory, where the strength of ties is defined based on sentiment-based features (Section 5). Our model outperforms the other two well-known classifiers (logistic regression and random forest) in the mutual-follow graph and in mention graphs where the strength of ties is not too weak (number of mentions exceeds three) (Section 6.2).

Although our analyses focus on Twitter, we conjecture that our approach can extend to a broad setting involving online information sharing, e.g., for restaurant or movie recommendations.

Contributions

Sentiment-based features. We define features that quantify the likelihood of two users becoming friends based on their sentiments toward topics of mutual interest. We evaluate the potential benefits of each feature.

Graphical model. We propose a model that incorporates the sentiment-based cognitive balance theory for link prediction. Our evaluation suggests that our model yields improved the performance (F_1 score) of link prediction when compared to traditional machine learning models.

2. DATA AND OBSERVATIONS

To obtain a dataset involving strong sentiments, we crawled Twitter during U.S. 2012 political campaign (from March 23 to April 23 in 2012) using the keywords "Obama" and "Romney." We preprocessed the dataset by first removing tweets that contain more than 10 hashtags. Because Twitter limits 140 characters in one tweet, a tweet containing too many hashtags is likely to be spam [18]. In addition, we treat users with less than five tweets as "inactive" and exclude them. The resulting dataset contains $3,970,974$ tweets from $123,073$ distinct user accounts.

Topics. A Twitter hashtag [35] is a string beginning with "#", which is viewed as a topic marker in the tweet. Typ-

ically, users adopt the same hashtag to discuss a particular topic. Thus, we use hashtags to represent different topics.

Graphs. We investigate two kinds of undirected graphs: the mention graph and the mutual-follow graph. The mention graph is based on "@" mentions: whether a retweet, a reply, or direct reference to a user. If two users mention each other more often than a certain threshold, we create an edge in their mention graph (we experiment with multiple thresholds). In the mutual-follow graph, we create an edge between two users if they follow each other.

2.1 Sentiment Extraction

We use an established sentiment lexicon, SentiWordNet [3], to obtain the sentiment scores of all tweets. SentiWordNet contains three real-valued scores for each word in its lexicon indicating *positivity*, *negativity*, and *objectivity*; the sum of the three scores is one. In addition, we extract emoticons from tweets and estimate the three sentiment scores of each emoticon. Agarwal et al. [2] provide a list of emoticons and categorize them into five categories: *extremely positive*, *positive*, *neutral*, *negative*, and *extremely negative*. We assign the sentiment scores to each category as triples of positivity, objectivity, and negativity scores, respectively, $\langle 1, 0, 0\rangle$, $\langle 3/4, 0, 0\rangle$, $\langle 0, 1, 0\rangle$, $\langle 0, 0, 3/4\rangle$, and $\langle 0, 0, 1\rangle$.

We adapt the major methods described by Bakliwal [5] to compute the sentiment score of a tweet. First, we choose only the adjectives in the lexicon and extract their stems [27] to build a pairwise stem-score mapping dictionary. We consider only adjectives because adjectives are strong indicators of sentiment [15] and can improve sentiment prediction accuracy [5]. Second, we use the Twitter part-of-speech tagger to extract the adjectives and emoticons in a tweet. Third, we handle the negation pattern through the Stanford parser [8], which contains a dependency schema (*neg*) to indicate negation. We reverse the sentiment polarity for each word marked in a *neg* schema. Finally, we obtain the positivity and negativity scores of a tweet by averaging the two scores for each adjective and emoticon; The objectivity score of the tweet is one minus the two polar scores.

2.2 Observations on the Data

As a sanity check on our intuition about sentiment homophily in the graphs we consider, we first determine the probability of two users sharing a same sentiment toward a topic of mutual interest, conditioned on whether they are connected.

We construct a mention graph choosing the threshold of mentions as three. That is, two users are connected if they mention each other at least three times. From the mutual-follow graph, we randomly choose a subgraph with 175 users and their friends. A pair of users is connected if there is an edge between them. For each graph, we construct pairs of unconnected users, whose number is identical to the number of connected pairs in the same graph. We choose the six most frequent topics (hashtags) in our dataset, and we want to compare the probability of two connected users sharing a sentiment with that of two unconnected users.

As Figure 1 shows, the probability of sharing a sentiment is 6% higher, on average, for connected users than unconnected users in the mention graph. In the mutual-follow graph (Figure 2), the mean difference in probability between connected and unconnected users is 4%, but the difference varies across topics. The probability difference is
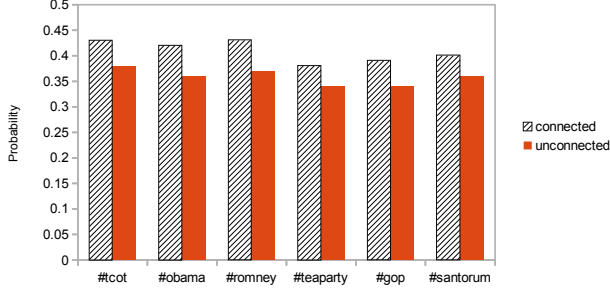
Figure 1: Probability of two users in the mention graph sharing a sentiment toward the six most frequent topics.
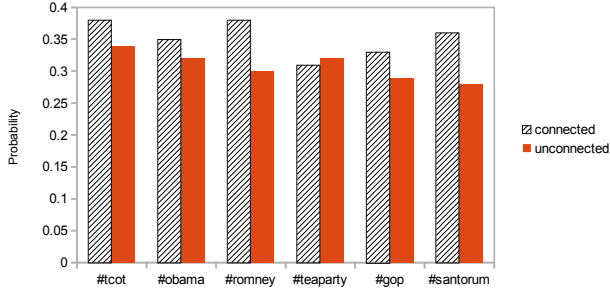

Figure 2: Probability of two users in the mutual-follow graph sharing a sentiment toward the six most frequent topics.

pronounced for "#romney" and "#santorum" (8%), whereas the difference is −1% for "#teaparty". These observations support our intuition that sentiments and connections are correlated.

3. PROBLEM DEFINITION

Let $G(V, E)$ be a social network, where V is the set of users and E is the social relationship between the users. For a given node v_s and a candidate set $C = \{v_1, v_2, \ldots, v_{|C|}\}$, our goal is to predict whether there is a link between v_s and v_i ($v_i \in C$). Specifically, the task is to find a predictive function for v_s such that: $Y = f(G, v_s, C)$, where $Y = \{y_{si}|v_i \in C\}$ is a vector of inferred results; i.e., $y_{si} = p(1|G, v_s, C)$ represents the probability that v_s will create a link with v_i.

To do so, we take two steps. First we generate a candidate set for a source node v_s as described next. Second, we learn a predictive function by defining prediction features (Section 4) and applying a factor graph model incorporating cognitive balance theory (Section 5).

3.1 Candidate Generation

For each source node v_s, we choose its two-hop neighborhood as its candidate set: friends and friends of friends. We choose the two-hop neighborhood as the candidate set because (1) the number of candidates increases exponentially with the number of hops [22]; (2) the number formed friendships decays exponentially with the number of hops [20].

We model our problem as a classification problem, where friends are positive instances and friends of friends are neg-

ative instances. We assign half of the source nodes into a training set and half into a test set. We measure F_1 scores to validate our recommendations with respect to the existing friends of v_s in the test set.

We can derive the potential friends of v_s from its friends of friends and then make recommendations based on ranked probabilities of links between v_s and its friends of friends. However, measuring the validity of such recommendations requires that we train a model from candidates at one time instance and test for candidates at a future time instance. Since our dataset does not have information about when the links were formed, such an evaluation is out of our scope.

4. PREDICTION FEATURES

With the development of online information sharing, the coevolution of social and affiliation networks is gaining attention, e.g., [38]. We consider a user as *affiliating* with a topic if the user evinces interest in it, and we call the affiliation *topical affiliation*. On Twitter, topical affiliation happens when a user includes a hashtag in his or her tweet.

Combining sentiment analysis of users' messages and topical affiliation, we call such an affiliation the *topic-sentiment affiliation*. That is, a user affiliates with a set of topics, and associates a sentiment with each topic. We now describe how a user's topic-sentiment affiliation can help link prediction by describing our prediction features.

We consider three kinds of features: sentiment, structural, and topical. Sentiment features are extracted from topic-sentiment affiliation; structural features are based on the graph-based similarity between two users; topical features are based on the topical affiliation of two users, measuring the similarity in their usage of topics. The sentiment features are our contribution whereas the other two categories serve as baseline predictors.

4.1 Topic-Sentiment Affiliation Construction

We compute each tweet's positivity, negativity, and objectivity scores using the methods of Section 2.1. If a user mentions a hashtag in one of his or her tweets, we affiliate him with the topic-sentiment pair; if a user mentions the same hashtag in several tweets, we take the mean of the three scores of these tweets. We further adopt the *sentiment-volume-objectivity* (SVO) function [13] to measure the aggregate effect of a user's level of interest and his or her sentiment scores toward a topic. The SVO score is a real value between 0 and 1 that incorporates three elements: polar sentiment (positivity or negativity), number of times a user mentions a topic, and objectivity. Therefore, the sentiment in the affiliation system consists of four numeric scores: positivity, negativity, objectivity, and the SVO score. In this way, we obtain a topic-sentiment affiliation for each user from his or her tweets.

4.2 Sentiment Features

We use the difference between the positivity and negativity scores to decide a user's categorical sentiment toward a hashtag. The opinion is positive (negative) if the difference is greater (less) than zero; otherwise, the opinion is objective. In addition, the *size* of a hashtag is the number of users who have adopted it. We adopt the following notation:

- Let $v_1, v_2, \ldots, v_N$ be N users.
- Let $h_1, h_2, \ldots, h_M$ be M hashtags.

- Let P_i, N_i, O_i be user v_i's adopted hashtags set with positive, negative, objective sentiments, respectively.
- Let $\overline{P_j}$, $\overline{N_j}$, $\overline{O_j}$ be the set of users who expressed positive, negative, objective sentiments in a hashtag h_j, respectively.
- Let $H_i = P_i \cup N_i \cup O_i$
- Let $U(h_j) = \overline{P_j} \cup \overline{N_j} \cup \overline{O_j}$
- Let $s_i(h_j)$ be user v_i's SVO score toward hashtag h_j.

Given two users v_s and v_t, we divide sentiment features into the following seven categories:

1. **The number of hashtags for which they have the same sentiments**.
   - SENTIMENT-AGREEMENT: $|P_s \cap P_t| + |N_s \cap N_t| + |O_s \cap O_t|$

2. **Sentiment alignment coefficient**: among the common hashtags, the number that involve the same or opposite polar sentiments.
   - SENTIMENT-ALIGNED: $(|P_s \cap P_t| + |N_s \cap N_t|)/|H_s \cap H_t|$
   - SENTIMENT-MISALIGNED: $(|P_s \cap N_t| + |N_s \cap P_t|)/|H_s \cap H_t|$

3. **Size of the rarest common hashtags**: among the hashtags for which the users share polar sentiments, the one that is the least adopted by all of them.
   - SENTIMENT-RAREST: $\min(\min_{h_j \in P_s \cap P_t} |\overline{P_j}|, \min_{h_j \in N_s \cap N_t} |\overline{N_j}|)$

4. **Adamic-Adar**: sum of the Adamic-Adar distances for each hashtag set affiliated with the three opinions.
   - SENTIMENT-AA: $\sum_{h_j \in P_s \cap P_t} 1/\log |\overline{P_j}| + \sum_{h_j \in N_s \cap N_t} 1/\log |\overline{N_j}| + \sum_{h_j \in O_s \cap O_t} 1/\log |\overline{O_j}|$

5. **Sum of inverse size**.
   - SENTIMENT-INVERSE: $\sum_{h_j \in P_s \cap P_t} 1/|\overline{P_j}| + \sum_{h_j \in N_s \cap N_t} 1/|\overline{N_j}| + \sum_{h_j \in O_s \cap O_t} 1/|\overline{O_j}|$

6. **Mean size of common hashtags for which they share the same sentiment**.
   - SENTIMENT-MEAN: $\frac{1}{\text{SENTIMENT-AGREEMENT}} \times (\sum_{h_j \in P_s \cap P_t} |\overline{P_j}| + \sum_{h_j \in N_s \cap N_t} |\overline{N_j}| + \sum_{h_j \in O_s \cap O_t} |\overline{O_j}|)$

7. **Topic-SVO distance**.
   - EUCLIDEAN: $\sqrt{\sum_{h_j \in H_s \cap H_t} (s_s(h_j) - s_t(h_j))^2}$
   - COSINE: $d_s d_t / \|d_s\| \|d_t\|$, where d_s and d_t are the SVO score vectors for common hashtags

4.3 Structural Features

These features are based on graph structure without considering semantic information. We choose four predictors introduced by Liben-Nowell et al. [21].

1. Number of common neighbors (CN) between two users.
2. Jaccard's coefficient (JC): CN divided by the total number of neighbors.
3. Adamic-Adar (STRUCTURAL-AA) [1]: weighting the importance of a common neighbor by the degree of the neighbor.
4. Preferential attachment (PA): the product of two users' degrees.

4.4 Topical Features

These features, introduced by Romero et al. [28], are baseline predictors in our study.

1. The number of common hashtags (COMMON): $|H_s \cap H_t|$
2. Size of the smallest common hashtag (SMALLEST): $\min_{h_j \in H_s \cap H_t} |U(h_j)|$
3. Adamic-Adar distance (TOPICAL-AA): $\sum_{h_j \in H_s \cap H_t} 1/\log |U(h_j)|$
4. Sum of inverse sizes (INVERSE): $\sum_{h_j \in H_s \cap H_t} 1/|U(h_j)|$
5. Mean size of common hashtags (MEAN): $\frac{1}{|H_s \cap H_t|} \sum_{h_j \in H_s \cap H_t} |U(h_j)|$

5. PROPOSED MODEL: TSAM

To investigate how to better exploit sentiments for link prediction, we propose a topic-sentiment affiliation based graphical model (TSAM). The motivation underlying TSAM arises from cognitive balance theory: if A–B and A–C are strong ties, then these two links are not independent because B–C is likely to be present. We seek (1) a way of building such relationships where the strength of a tie incorporates sentiment and (2) to study whether this sentiment-based cognitive balance theory could improve link prediction.

A TSAM model is an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of variables and $\mathcal{E}$ is the set of edges in the graphical model. Below, *variables* and *edges* refer to entities in TSAM, and *nodes* and *links* refer to entities in the social network.

We now describe how we build a TSAM model $\mathcal{G}(\mathcal{V}, \mathcal{E})$ by representing links and their relationships in a social network $G(V, E)$. Given the social network $G(V, E)$, a source node v_s and its candidate set $C = \{v_1, v_2, \ldots, v_{|C|}\}$, link prediction seeks to infer the probability y_{si} that v_s will create a link with v_i. Thus, we treat y_{si} as a variable (hidden) and the relationships between such variables as edges $\mathcal{E}$ in the TSAM model.

There is an edge between any two hidden variables if they contain the same source node v_s. Thus, for each source node, there is a clique in the TSAM model. For each hidden variable, an observed variable is connected with it, representing a vector of features associated with the hidden variable.

For example, suppose v_s has five candidates: $\{v_1, v_2\}$, its one-hop friends, and $\{v_3, v_4, v_5\}$, its friends of friends. The resulting TSAM model of Figure 3 has five hidden variables $\{y_{s1}, \cdots, y_{s5}\}$ and five observed variables $\{x_{s1}, \cdots, x_{s5}\}$.

Even though the figure only shows the factor graph for one source node, our model captures a scenario with multiple source nodes. Correspondingly, a TSAM model with multiple source nodes is composed of multiple disconnected components, where all the hidden variables in one component form a clique.
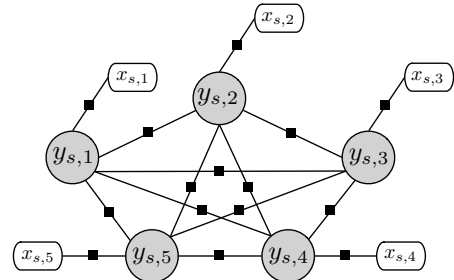

Figure 3: Graphical representation of TSAM.

Let T represent the set of indices for any link between a source node and one of its candidates. Then $Y = \{y_t | t \in T\}$ and $X = \{x_t | t \in T\}$ denote the set of hidden and observed variables in the TSAM model, respectively.

The graph can be modeled as a conditional random field [19] that defines a distribution over the graph:

$$P(Y|X) = \frac{1}{Z} \prod_{t \in T} \varphi(y_t, x_t) \prod_{(y_t, y_{t'}) \in \mathcal{E}} \psi(y_t, y_{t'})$$

where Z is a constant that ensures $\sum_Y P(Y|X) = 1$.

The model incorporates two factor functions, which we instantiate by the Hammersley-Clifford theorem [14]. We follow the presentation of Dong et al. [10]: the **attribute factor** models the influence of different features on the hidden variable (link).

$$\varphi(y_t, x_t) = \exp\{\sum_{m=1}^{d} \alpha_m f_m(y_t, x_{t_m})\} \quad (1)$$

where α_m is a weight constant and d is the number of features associated with y_t. We include all the features in Section 4 to build the attribute factor. Second, the **edge factor** encodes the relationships between connected hidden variables.

$$\psi(y_t, y_{t'}) = \exp\{\sum_{n=1}^{k} \beta_n g_n(y_t, y_{t'})\} \quad (2)$$

where β_n is a weight constant and k is the number of features associated with the edge $(y_t, y_{t'})$ in the TSAM model.

We define $g_n(y_t, y_{t'})$ as a binary function. For any two hidden variables y_t and $y_{t'}$, a triad is potentially involved because y_t and $y_{t'}$ contain the same source node. The triad would be cognitively balanced if both y_t and $y_{t'}$ are strong ties. We can use any sentiment feature that satisfies the axioms of Gupte and Eliassi-Rad [12] to define tie strength. Let's take the feature SENTIMENT-AA as an example. Following Hopcroft et al.'s [17] definition of the importance of an user, we select the top 1% edges in the social network $G(V, E)$ in terms of SENTIMENT-AA features as strong ties. Therefore, $g_n(y_t, y_{t'})$ is one when both y_t and $y_{t'}$ are strong ties; otherwise, it is zero.

Accordingly, the log-likelihood objective function is

$$\mathcal{O}(\theta, Y|X) = \log P(Y|X) = \sum_{i \in T} \sum_{m=1}^{d} \alpha_m f_m(y_i, x_{i_m}) +$$

$$\sum_{(i,j) \in \mathcal{E}} \sum_{n=1}^{k} \beta_n g_n(y_i, y_j) - \log Z$$

Here $\theta = (\alpha, \beta)$ is the model (parameter configuration) that we seek to learn to maximize the log-likelihood objective function: $\theta^\star = \arg\max_\theta \mathcal{O}(\theta, Y|X)$.

We adapt the methods in [32] to learn the model except that we conduct experiments in a supervised setting; we use gradient descent to optimize the objective function, where the gradient is approximated by loopy belief propagation [26]. With the estimated model $\theta^\star$, the goal of link prediction is to determine the probabilities of hidden variables that maximize the joint probability $P(Y|X, \theta^\star)$:

$$Y^\star = \arg\max_Y P(Y|X, \theta^\star) \quad (3)$$

6. EXPERIMENTAL EVALUATION

We seek to answer two questions:
- Do sentiment features help in link prediction?

- Does our proposed graphical model incorporating the sentiment-based cognitive balance theory benefit link prediction?

6.1 Sentiment Features Evaluation

Evaluation Strategy. We conduct our experiments using the mention and mutual-follow graphs. For the mention graphs, the number of "@" references between users can be viewed as the strength of a tie. We therefore define several mention graphs by setting different strengths of ties. Table 1 shows the statistics of mention graphs with different thresholds and of the mutual-follow graph.

Table 1: Graph statistics.

Graph	Nodes	Edges	Mean degree
@ ≥ 1	23,915	53,009	4.43
@ ≥ 2	8,936	12,978	2.90
@ ≥ 3	4,770	5,933	2.49
@ ≥ 5	2,106	2,134	2.03
@ ≥ 7	1,201	1,100	1.83
@ ≥ 9	762	639	1.68
Follow	11,239,979	48,572,793	8.64

For each mention graph, we choose all users whose degree is less than 50 as v_s. For the mutual-follow graph, we select users who adopt at least 50 hashtags; additionally, we only select "active" users, with the degree falling within range $[50, 100]$. Then we generate the two-hop candidate set for each source node. A pair is constructed between each source node and any one of its candidates. Doing so leads to the class imbalance being extremely high, a common problem in link prediction [22]. Because we first want to investigate the effect of sentiment features in link prediction, we undersample the negative instances to obtain a balanced dataset. Table 2 shows the number of instances of all the learning datasets after preprocessing.

Table 2: Learning dataset statistics.

Graph	Overall	Positive	Negative
@ ≥ 1	183,359	78,301	105,058
@ ≥ 2	46,357	22,087	24,270
@ ≥ 3	21,816	10,763	11,053
@ ≥ 5	7,429	3,981	3,448
@ ≥ 7	3,653	2,094	1,559
@ ≥ 9	2,064	1,232	832
Follow	11,153	5,193	5,960

We normalize all the features to $[0, 1]$. We apply logistic regression and random forest models from the WEKA framework, and we conduct a 10-fold cross-validation with default parameters. The F_1 score and the Area under the Receiver-Operating-Characteristic Curve (AUC) are two widely used metrics in evaluating performance of classifiers [7]. In our setting, we are more interested in the *true positive* than the *true negative* metric for two reasons. First, positive instances (links between source nodes and their existing friends) are ground truth, and ensuring high *recall* of positive instances is nontrivial. Second, *false positive* with zero, indicating no new friends to recommend, is not necessarily

desirable. Because AUC incorporates both positive and negative instances equally whereas F_1 ignores the true negative metric, we adopt F_1 as our performance metric.

Results. Table 3 shows F_1 scores on different combinations of features for logistic regression and random forest classifiers. In general, sentiment features yield better performance in terms of F_1 scores, no matter whether they are combined with structural features to build the model. To investigate whether sentiment features indeed help improve the F_1 score, we conduct a paired t-test: each value in sample one is the F_1 score with sentiment or sentiment plus structural features; each value in sample two is the F_1 score with topical or topical plus structural features. The p-value is 0.0012, indicating the difference is statistically significant. In addition, structural features perform much better than both sentiment and topical features, but adding sentiment features can generally improve performance. Thus, sentiment features can indeed help in link prediction, but as adjuncts to the structural features.

Individual Feature Evaluation. We show the performance of each sentiment feature for random forest since it outperforms logistic regression in Table 3. Table 4 shows the F_1 score for individual sentiment feature. We highlight the top three ranked features in each graph. We find that in the mention graphs, the features Euclidean, sentiment-mean, and sentiment-AA perform best. We find that in the mutual-follow graph, the feature sentiment-agreement performs best, but Euclidean and sentiment-AA are good indicators. And, sentiment-rarest is not as informative as smallest (topical features). In addition, sentiment-aligned outperforms sentiment-misaligned in general in all graphs, indicating the existence of sentiment homophily. Therefore, our evaluation suggests that sentiment homophily exists and can benefit link prediction.

6.2 TSAM Model Evaluation

Following the preprocessing strategy proposed by Backstrom and Leskovec, we choose "active" source nodes in all graphs. That is, active nodes are those whose degree is within the range [lower, upper]. After constructing pairs consisting of each source node and each of its candidates, we remove those whose number of common friends is less than a threshold, because users with only a few common friends are unlikely to form friendships. Table 5 shows the criteria we used. The criteria differ for graphs with differing statistics. As the ties become stronger, the mention graph becomes smaller. Thus an overfitting problem may arise if the dataset is too small. We therefore limit our evaluation to graphs with more than 100 source nodes left after preprocessing.

Table 5: Preprocessing parameters for evaluating TSAM.

Graph	Lower	Upper	Threshold
@ $\geq$ 1	10	50	4
@ $\geq$ 2	10	50	4
@ $\geq$ 3	5	50	4
@ $\geq$ 5	3	50	2
@ $\geq$ 7	3	50	2
@ $\geq$ 9	3	50	2
Follow	50	100	4

Because the feature sentiment-AA ranks in the top three features for each graph in Table 4, we choose it to define the strength of a tie in the edge-factor function. For each graph, we assign half of the source nodes into a training and half into a test set. In the training phase, we set the learning rate $\lambda = 0.001$ and the number of iterations as 500. We split each graph five times and report the mean precision, recall, and F_1 scores of the TSAM model as well as for logistic regression and random forest, and identify the best performing model in terms of F_1. TSAM outperforms the other two models in the mutual-follow graph. In the mention graphs, TSAM performs best when the ties are strong (@ $\geq$ 3 and @ $\geq$ 5), but not for weak ties (@ $\geq$ 1 and @ $\geq$ 2). This suggests that when the ties are weak, sentiment-based cognitive balance theory does not help because the links between users are somewhat random; hence, a balanced cognitive structure does not necessarily mean any relationship between two pairs.

Overall, our results indicate that performance of link prediction could be improved when we incorporate sentiment-based cognitive balance theory, especially on graphs where the strength of relationship is not too weak (mutual-follow graph or mention graphs where the number of mentions exceeds three).

7. RELATED WORK

Link Prediction. Existing work on link prediction can be classified into two categories. For the unsupervised methods, Liben-Nowell and Kleinberg evaluated different "proximity" features extracted from network topology. They found that the Adamic-Adar metric performs best in predicting links. Most recent works are based on supervised methods. Lichtenwalter et al. [22] provided a detailed analysis of challenges, such as class imbalance, of using supervised methods in link prediction. Dong et al. [10] proposed a probabilistic graphical model to predict links. Backstrom and Leskovec [4] developed a supervised random walk algorithm for friend recommendation on Facebook. Scellato et al. [29] exploited place features in predicting links on location-based social networks. Whereas we adopt a supervised approach, we additionally consider sentiment features and investigate how they improve link prediction.

Collaborative Tagging Systems. These are based on a tripartite structure: users, tags, and resources, enabling users to share their tags for particular resources. Combined with social structure, collaborative tagging systems provide new modalities of link prediction. Marlow et al. [24] found that users tend to have a larger similarity of tag vocabularies with their friends compared with random users. Markines et al. [23] built a foundation for the folksonomy-based similarity measures, such as matching, overlap, Jaccard, and cosine similarity. Romero et al. [28] studied the relationship between topical affiliations and social network on Twitter. They found that the adoption of hashtags can predict users' social relationships. We design sentiment features based on Romero et al.'s findings. Further, we propose a graphical model based on sentiment features.

Sentiment Analysis. With the popularity of social media, sentiment analysis brings us deeper understanding of social network analysis. Twitter enables researchers to access huge amounts of data to discover collective sentiments [6, 34], predict political elections [9, 36], and so on. In other applications, Yang et al. [37] extracted sentiments from users'

Table 3: F_1 scores on the positive instances for different classifiers on different combination of features.

Classifier	Set of features	@ ≥ 1	@ ≥ 2	@ ≥ 3	@ ≥ 5	@ ≥ 7	@ ≥ 9	Follow
Logistic Regression	Topical	0.295	0.489	0.530	0.650	0.720	0.735	0.479
	Sentiment	0.312	0.500	0.550	0.650	0.712	0.739	0.499
	Structural	0.736	0.740	0.759	0.792	0.812	0.846	0.666
	Topical+Structural	0.689	0.777	0.792	0.818	0.846	0.880	0.678
	Sentiment+Structural	0.717	0.789	0.800	0.834	0.861	0.877	0.679
	All	0.705	0.778	0.795	0.838	0.858	0.880	0.682
Random Forest	Topical	0.628	0.765	0.804	0.839	0.853	0.880	0.516
	Sentiment	0.723	0.818	0.845	0.859	0.876	0.883	0.517
	Structural	0.875	0.908	0.913	0.939	0.946	0.957	0.667
	Topical+Structural	0.914	0.946	0.956	0.971	0.972	0.972	0.680
	Sentiment+Structural	0.916	0.949	0.959	0.972	0.975	0.979	0.685
	All	0.916	0.949	0.958	0.972	0.970	0.979	0.684

Table 4: F_1 scores on individual sentiment feature with Random Forest classifier.

Feature	@ ≥ 1	@ ≥ 2	@ ≥ 3	@ ≥ 5	@ ≥ 7	@ ≥ 9	Follow
SENTIMENT-AGREEMENT	0.166	0.430	0.498	0.636	0.696	0.711	**0.488**
SENTIMENT-ALIGNED	0.192	0.291	0.555	0.650	0.700	0.735	0.452
SENTIMENT-MISALIGNED	0.122	0.509	0.526	0.640	0.689	0.710	0.390
SENTIMENT-RAREST	0.122	0.307	0.360	0.691	0.721	0.742	0.295
SENTIMENT-AA	**0.430**	**0.589**	**0.706**	**0.753**	**0.789**	**0.822**	**0.466**
SENTIMENT-INVERSE	0.349	0.501	0.648	0.741	0.785	0.812	0.452
SENTIMENT-MEAN	**0.455**	**0.593**	**0.706**	**0.757**	**0.791**	**0.825**	0.454
EUCLIDEAN	**0.589**	**0.721**	**0.751**	**0.788**	**0.825**	**0.832**	**0.467**
COSINE	0.247	0.480	0.525	0.625	0.692	0.748	0.456

Table 6: Evaluation results for the TSAM.

Feature	Logistic Regression			Random Forest			TSAM		
	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
@ ≥ 1	0.350	0.443	**0.391**	0.445	0.196	0.272	0.193	0.890	0.317
@ ≥ 2	0.456	0.780	**0.576**	0.616	0.535	0.573	0.602	0.539	0.569
@ ≥ 3	0.727	0.596	0.655	0.667	0.628	0.647	0.712	0.620	**0.663**
@ ≥ 5	0.752	0.701	0.726	0.863	0.806	0.834	0.859	0.863	**0.861**
Follow	0.794	0.636	0.705	0.759	0.664	0.708	0.677	0.801	**0.734**

reviews on venues to improve location recommendation services; Tan et al. [31] used the social relationship to improve user-level sentiment prediction. We conduct our work with a different purpose: using sentiment homophily to predict links.

8. CONCLUSIONS

We study how to exploit sentiments for link prediction, and evaluate the extent to which sentiment homophily can help improve link prediction. By extracting users' sentiments from their tweets on different topics, we describe a set of sentiment features to quantify the likelihood of two users becoming friends. The evaluation results suggest that sentiment features improve the performance of link prediction in terms of F_1 in both mutual-follow and mention graphs. Upon investigating the predictive power of each sentiment feature, we find that Adamic-Adar and Euclidean distance based measures perform best. We pro-

pose a factor graph model considering the sentiment-based cognitive balance theory. The results show that our model outperforms the other two well-known classifiers (logistic regression and random forest) in the mutual-follow graph and mention graphs where the strength of ties is not too weak (@ ≥ 3). In future work, we plan to evaluate our work in a friend recommendation framework by exploiting temporal information regarding how links form.

9. REFERENCES

[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, 25:211–230, 2001.

[2] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau. Sentiment analysis of Twitter data. In *Proc. LSM*, pages 30–38. 2011.

[3] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. LREC*. May 2010.

[4] L. Backstrom and J. Leskovec. Supervised random walks: Predicting and recommending links in social networks. In *Proc. WSDM*, pages 635–644. 2011.

[5] A. Bakliwal, J. Foster, J. van der Puil, R. O'Brien, L. Tounsi, and M. Hughes. Sentiment analysis of political tweets: Towards an accurate classifier. In *Proc. LASM*, pages 49–58. June 2013.

[6] J. Bollen, H. Mao, and A. Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proc. ICWSM*, 2011.

[7] S. Daskalaki, I. Kopanas, and N. Avouris. Evaluation of classifiers for an uneven class distribution problem. *Applied Artificial Intelligence*, 20:381–417, Sept. 2006.

[8] M.-C. de Marneffe, B. MacCartney, and C. D. Manning. Generating typed dependency parses from phrase structure parses. In *Proc. LREC*, pages 449–454, 2006.

[9] N. A. Diakopoulos and D. A. Shamma. Characterizing debate performance via aggregated Twitter sentiment. In *Proc. CHI*, pages 1195–1198. 2010.

[10] Y. Dong, J. Tang, S. Wu, J. Tian, N. V. Chawla, J. Rao, and H. Cao. Link prediction and recommendation across heterogeneous social networks. In *Proc. ICDM*, pages 181–190. 2012.

[11] M. S. Granovetter. The strength of weak ties. *American J. Sociology*, 78(6):1360–1380, May 1973.

[12] M. Gupte and T. Eliassi-Rad. Measuring tie strength in implicit social networks. In *Proc. WebSci*, pages 109–118. 2012.

[13] D. F. Gurini, F. Gasparetti, A. Micarelli, and G. Sansonetti. A sentiment-based approach to Twitter user recommendation. In *Proc. RsWeb*. 2013.

[14] J. M. Hammersley and P. E. Clifford. Markov random fields on finite graphs and lattices. Unpublished manuscript, 1971.

[15] V. Hatzivassiloglou and J. M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proc. COLING*, pages 299–305. 2000.

[16] F. Heider. *The Psychology of Interpersonal Relations*. Wiley, 1958.

[17] J. Hopcroft, T. Lou, and J. Tang. Who will follow you back?: Reciprocal relationship prediction. In *Proc. CIKM*, pages 1137–1146. 2011.

[18] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proc. WWW*, pages 591–600. 2010.

[19] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. ICML*, pages 282–289. 2001.

[20] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. In *Proc. KDD*, pages 462–470. 2008.

[21] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proc. CIKM*, pages 556–559. 2003.

[22] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla. New perspectives and methods in link prediction. In *Proc. KDD*, pages 243–252. 2010.

[23] B. Markines, C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. Evaluating similarity measures for emergent semantics of social tagging. In *Proc. WWW*, pages 641–650. 2009.

[24] C. Marlow, M. Naaman, D. Boyd, and M. Davis. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proc. HYPERTEXT*, pages 31–40. 2006.

[25] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, Aug. 2001.

[26] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: An empirical study. In *Proc. UAI*, pages 467–475. 1999.

[27] M. F. Porter. An algorithm for suffix stripping. *Program: electronic library and information systems*, 14(3):130–137, July 1980.

[28] D. M. Romero, C. Tan, and J. Ugander. On the interplay between social and topical structure. In *Proc. ICWSM*, 2013.

[29] S. Scellato, A. Noulas, and C. Mascolo. Exploiting place features in link prediction on location-based social networks. In *Proc. KDD*, pages 1046–1054. 2011.

[30] P. Singla and M. Richardson. Yes, there is a correlation: From social networks to personal behavior on the Web. In *Proc. WWW*, pages 655–664. 2008.

[31] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li. User-level sentiment analysis incorporating social networks. In *Proc. KDD*, pages 1397–1405. 2011.

[32] W. Tang, H. Zhuang, and J. Tang. Learning to infer social ties in large networks. In *Proc. ECML PKDD*, pages 381–397. 2011.

[33] M. Thelwall. Emotion homophily in social network site messages. *First Monday*, 15(4), 2010.

[34] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment in Twitter events. *J. Amer. Soc. Info. Sci. Tech.*, 62(2):406–418, Feb. 2011.

[35] O. Tsur and A. Rappoport. What's in a hashtag?: Content based prediction of the spread of ideas in microblogging communities. In *Proc. WSDM*, pages 643–652. 2012.

[36] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proc. ICWSM*, pages 178–185, 2010.

[37] D. Yang, D. Zhang, Z. Yu, and Z. Wang. A sentiment-enhanced personalized location recommendation system. In *Proc. HYPERTEXT*, pages 119–128. 2013.

[38] E. Zheleva, H. Sharara, and L. Getoor. Co-evolution of social and affiliation networks. In *Proc. KDD*, pages 1007–1016. 2009.