Who Will Reply to/Retweet This Tweet? The Dynamics of Intimacy from Online Social Interactions

Nicholas Jing Yuan[†], Yuan Zhong[†‡], Fuzheng Zhang[†],
Xing Xie[†], Chin-Yew Lin[†], Yong Rui[†]
[†]Microsoft Research
[‡]Northeastern University, Boston, MA, USA
{nichy, fuzzhang, xingx, cyl, yongrui} at microsoft.com, zhong.yu at husky.neu.edu

ABSTRACT

Friendships are dynamic. Previous studies have converged to suggest that social interactions, in both online and offline social networks, are diagnostic reflections of friendship relations (also called social ties). However, most existing approaches consider a social tie as either a binary relation, or a fixed value (named *tie strength*). In this paper, we investigate the *dynamics* of dyadic friend relationships through online social interactions, in terms of a variety of aspects, such as *reciprocity*, *temporality*, and *contextuality*. In turn, we propose a model to predict *repliers* and *retweeters* given a particular tweet posted at a certain time in a microblog-based social network. More specifically, we have devised a learning-to-rank approach to train a ranker that considers elaborate user-level and tweet-level features (like sentiment, self-disclosure, and responsiveness) to address these dynamics. In the prediction phase, a tweet posted by a user is deemed a query and the predicted repliers/retweeters are retrieved using the learned ranker. We have collected a large dataset containing 73.3 million dyadic relationships with their interactions (replies and retweets). Extensive experimental results based on this dataset show that by incorporating the dynamics of friendship relations, our approach significantly outperforms state-of-the-art models in terms of multiple evaluation metrics, such as MAP, NDCG and Topmost Accuracy. In particular, the advantage of our model is even more promising in predicting the exact sequence of repliers/retweeters considering their orders. Furthermore, the proposed approach provides emerging implications for many high-value applications in online social networks.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Data mining; J.4 [**Social and Behavioral Science**]: Sociology

Keywords

dynamic tie strength, friendship relation, online social interaction

1. INTRODUCTION

Social ties have been extensively studied in domains as varied as sociology [11], psychology [30], and biology [6]. It is well recognized that *proximity* [1, 27] and *homophily* [18, 23] play key roles in the formation of friendships: People tend to be friends with others who are close by, and/or similar to themselves in terms of age, ethnicity, education, and even genotypes [6].

However, not all friendships are equal. The concept of tie strength was introduced by Granovetter [11] as early as the 1970's, where the frequency of interaction was used as a diagnosis of the tie strength. Until recently, however, the rapidly increasing digitalization of communications provided the availability to estimate tie strength using large-scale communication data such as emails [28, 12] and mobile phone logs [26, 5]. During the past few years, the emergence of online social networks has profoundly changed the way people communicate and socialize: People interact with much lower communication cost and greater variety of modes, (e.g., on Twitter, a user can *reply* to a friend's tweet, or *retweet* a tweet so as to broadcast it to his/her followers). Social ties and tie strength in online social networks were also brought to attention of researchers. Several studies have exploited rich interaction data in online social networks to investigate the strength between dyadic ties [35, 8, 17, 29].

Nevertheless, most existing studies consider tie strength as static: either a category (strong/weak tie) or a fixed value. However, the dynamics of tie strength, which have rarely been addressed, play a non-negligible role in understanding friend relationships. First, the strength of a tie may not be reciprocal [25]. For example, Alice considers Bob her best friend, but Bob may consider Carol his best friend. Second, tie strength can evolve over time, e.g., your best friend during high school may not be your current best friend. Third, given different contexts such as location, emotion, and environment, people may perceive different feelings of closeness, i.e., tie strength. For instance, when you travel to another city, you may frequently interact with some local friends (e.g., alumni) who you have not been in touch with for a long time. Given the above dynamic characteristics, a static measurement of tie strength is usually coarse-grained, and sometimes unrealistic.

In this paper, we examine dynamics of tie strength through a microblog-based social network and on this basis we propose a general model to predict the repliers and retweeters of a particular tweet considering friendship dynamics. Specifically, given *a tweet posted by a particular user at a certain time*, we aim to predict *which friends of this user will reply to/retweet this tweet.* Note that for replying and retweeting, the results may be quite different. To tackle this problem, we convert the prediction task into an information retrieval task: Given the tweet posted by u as a query, we retrieve a ranked list of friends of u, according to their inten-

tions of replying to/retweeting this tweet, where the ranking model is learned through a *learning-to-rank* framework.

Such prediction can enable a variety of promising applications in online social networks, e.g., people are more likely to adopt a product recommended by a close friend (after viewing his/her retweet talking about the product) than a casual acquaintance. Thus an accurate prediction of potential reactors can guide advertisers and marketers to design better incentive mechanisms. Our approach and findings can also help users create content to engage reactions and increase social impact, e.g., suggesting the best time of a day to post a tweet related to a certain topic that can trigger a discussion among the contacts in your social circle. Meanwhile, the proposed framework for analyzing friendship dynamics can also be applied to online communication platforms such as Skype and WhatsApp. For instance, creating and maintaining contact groups (or social circles) are usually tedious. However, given the precise estimation of dynamic tie strength, we can automatically provide a list of "local friends" a person would like to call when traveling to a given city.

To the best of our knowledge, this is the first work that predicts both repliers and retweeters at *tweet-level* (i.e., the prediction considers both the content of a tweet and the context of a user). Overall, this paper offers the following contributions:

• We exploit the dynamics of social interactions in terms of reciprocity, temporality, and contextuality, and craft a collection of features from online social networks to measure the interaction dynamics. In particular, we explicitly model the *responsiveness* of a user, which captures the availability, capacity, and tendency of a user's interaction behaviors.

• Based on a large dataset we have collected from Sina Weibo, we develop a learning-to-rank framework incorporating the above features, so as to predict both repliers and retweeters based on the content of a tweet and the context of a user. In addition, our model can predict the sequence of repliers and retweeters for a particular tweet, i.e., who replies/retweets the first, second, etc.

• We have conducted extensive experiments to validate the effectiveness of the proposed prediction model with multiple evaluation metrics. The results show that by considering the dynamics of tie strength, our approach significantly outperforms state-of-the-art models in terms of predictive power validated by multiple metrics such as MAP, NDCG, and Topmost Accuracy.

The rest of this paper is organized as follows: Sec. 2 explores the dynamics of social interactions in various aspects. Sec. 3 introduces the prediction model of online social interactions. Sec. 4 presents the experimental results based upon a large-scale dataset. Finally, related work is reviewed in Sec. 5 and the paper is briefly concluded in Sec. 6.

2. DYNAMICS OF SOCIAL INTERACTIONS

Given an online social network as a directed graph $G = (U, F)$, where U are users, and F are the following relations ($(u, v) \in F$ means a user $u \in U$ follows $v \in U$ on G), we define a *friend* and an *ego network* as follows.

DEFINITION 1 (FRIEND). *A user $u \in U$ is referred to as a* friend *of user $v \in U$ (or u and v have a friendship relation), denoted by $u \sim v$, if $(u, v) \in F$ and $(v, u) \in F$.*

DEFINITION 2 (EGO NETWORK). *The* neighbours *of a user $u \in V$ are defined as*

$$\mathring{E}_u = \{v | u \sim v, v \in U\}, \qquad (1)$$

and the ego network *of u is $E_u = \mathring{E}_u \cup \{u\}$.*

Sutcliffe et al. [30] suggest that interaction frequency is one of the most reliable correlations of relationship strength in both humans and animals. In this section, exemplified with interaction data from Sina Weibo[1] (a microblog-based social network known as the Chinese Twitter), we introduce several key aspects of friendship dynamics. The dataset was crawled using open APIs (refer to Sec. 3.1 for the crawling procedure and Sec. 4.2.1 for the data description).

2.1 Reciprocity

From an economic perspective, social interaction can be considered an investment for maintaining tie strength and bringing benefit to the ego [30]. However, the investment will not necessarily equal the benefit.

To quantify the reciprocity between two users u and v, we first calculate the *percentile rank* of v in terms of outgoing interactions from u to v, which measures the percentage of users who have received more or as many interactions from u compared with v (the smaller PR is, the higher rank v has), denoted by $\text{PR}_u(v)$:

$$\text{PR}_u(v) = \frac{\|\{w | f_{u \to w} > f_{u \to v}\}\| + 0.5\|\{w | f_{u \to w} = f_{u \to v}\}\|}{\|\mathring{E}_u\|}, \qquad (2)$$

where $f_{u \to w}$ is the outgoing interaction frequency from u to w.

Then we define the Reciprocal Interaction Rank (RIR) from u to v as follows.

$$\text{RIR}_{u \to v} = (PR_u(v) - PR_v(u)) \times \mathfrak{D}. \qquad (3)$$

Here, we scale the original difference between $PR_u(v)$ and $PR_v(u)$ with $\mathfrak{D} = 150$, which is the so called Dunbar's number. Dunbar [4] believed that the number of stable social ties in an ego network is approximated to be 150 on average, which is correlated with the relative neocortex size. This number was recently validated using Facebook and Twitter data [10]. Note that RIR is an asymmetric measurement, which shows the degree of inequality in terms of perceived tie strength to a certain extent.

We plot the distribution of reciprocal interaction ranks of 1 million randomly sampled social ties on Sina Weibo as shown in Figure 1(a), where we only present the distribution of positive values since the negative ones are mirror symmetric. As is shown, for a large portion of social ties, the reciprocal interaction rank deviates from zero, which indicates the asymmetry of tie strength.

2.2 Temporality

The tie strength between a dyadic relation may change over a certain period of time.

For example, if we consider the friend who we reply to/retweet the most often as the "best friend", we can identify whether the best friend changes over time. We randomly sampled 1 million users and identified their best friends for each month. Then we count the number of distinct best friends for each user during about 51 months (the range of our dataset). Figure 1(b) and Figure 1(c) present the distributions of the number of distinct best friends in terms of replies and retweets respectively. Apparently, almost all users have changed their best friends, and most users (64% for reply and 74%) have 3 distinct best friends during this period. We note that although the best friends based on different interaction types (e.g., replies and retweets) may vary for an individual, the distributions look similar.

2.3 Contextuality

Given different contexts, people may perceive different degrees of intimacy. Such contexts include locations, availability (to inter-
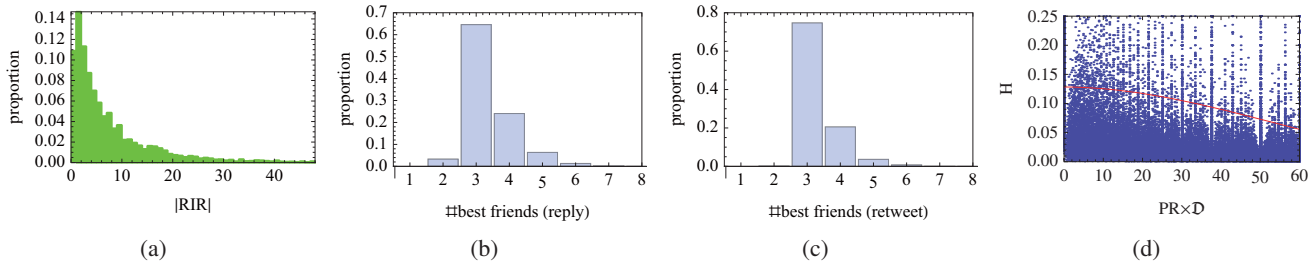
---

[1]http://weibo.com

Figure 1: Dynamics of tie strength w.r.t different aspects based on Sina Weibo data

act), emotional states and other factors (we will detail them later in Sec. 3.3).

For example, emoticons are popular in many online social networks, where some emoticons, such as a smiling face, represent delightful emotions while some others convey unhappiness, which imply the emotion of a user when posting a tweet/reply. We randomly sampled 1 million users and computed the ratio of replies that contain happy emoticons sent from u to each of u's friends v, denoted as $H_{u \rightarrow v}$, divided by all the replies sent from u to his/her friends (note that here we use the ratio for normalization over all users). Meanwhile, we calculated the percentile rank (PR) for each of the user pairs and scaled it with the Dunbar's Number introduced in Sec. 2.1. Then, we have a list: $\{(x_i, y_i)\}_{i=1}^{L}$, where x_i is the PR value scaled by $\mathfrak{D}$ and y_i is the ratio of happy-emoticon replies for the ith user pair. Figure 1(d) shows the scatter plot as well as a fitted curve of these pairs, which clearly indicates that when people reply to a friend's tweet, they are more likely to enclose an emoticon if the friend is ranked higher (in terms of the PR value).

3. LEARNING TO RANK FRIENDS

Given a tweet $\mathcal{T}$ posted on a microblog-based social network (such as Twitter or Sina Weibo) by user u, at time t, this section introduces a model to rank u's friends according to their intentions of replying and retweeting $\mathcal{T}$ respectively. We first provide an overview of the system in Sec. 3.1, and detail the key components later in Sec. 3.2, 3.3, and 3.4.

3.1 Overview

As illustrated in Figure 2 and Figure 3, our model works as a typical machine-learned search engine that includes three phases: the crawling phase, the offline training phase, and the online prediction phase, where the training and prediction phases are two core parts of our inference model.

● **Crawling.** In the crawling phase, we start from a seed user dataset obtained through the LifeSpec system [36]. By calling the APIs, we can retrieve all the friends of a user and recursively explore a large number of users, which is the *Expanding Step* of the crawler. Later on, in order to get all the tweets and replies of the users, we traverse their ego-networks and crawl all their tweets and their neighbours' tweets. This is the *Shrinking Step*. Note that Sina Weibo does not provide an API to crawl the replies posted by a user directly, so, we work around this by extracting a user's replies from each tweet posted by the user's friends (since there is an API to get the replies to a particular tweet). The Expanding Step and the Shrinking Step, as shown on the left side of Figure 2, are performed iteratively to achieve both the coverage (i.e., a large number of users) and the effectiveness (since we need all replies posted by these users to analyze their tie strength based on the online interaction histories). As a result, for each target user[2], we crawled pro-

---

[2] A target user is the user whose friends will be ranked.

files, tweets (including retweets), and replies of all the users in u's ego network E_u.

● **Training.** In the training phase shown on the right side of Figure 2, we construct a series of *user-level* and *tweet-level* features. Here, user-level features are comprised of reciprocal rank, profile affinity, topic affinity, and a responsiveness matrix, for each target user. The two categories of affinity capture the homophily between u and u's friends. For example, the similarity of educational background is one of the features in the profile affinity. For each affinity, we build a sorted list where each entry of the list is an affinity value calculated based on a social tie (u, v), where $v \in \mathring{E}_U$. Here, the topic affinity is computed based on the named-entities extracted from each tweet and reply posted by users in E_u. We note that for certain languages written without spaces (e.g., Chinese), text segmentation, which is a standard NLP task, should be pre-processed to segment the sentences into component words carrying meaningful units. All the affinity features are sorted in descending order of affinity value, which could also be considered the inverted indexes in a typical retrieval system. Through responsiveness analysis (detailed in Sec. 3.3), the responsiveness matrix is derived from u's historical interaction behaviors with $\mathring{E}_u$, which summarizes u's *tendency*, *availability*, and *capacity* to interact with his/her friends. In our model, all these indexes and the responsive matrix are periodically updated so as to keep track of users' temporal updates of their profile and topic affinity, as well as the responsiveness. Meanwhile, tweet-level features are extracted from each tweet and reply posted by users in E_u, which are derived from sentiment analysis and self-disclosure analysis. To learn a ranking model combining all the features, we sample a set of positive and negative instances to train the model. Given a tweet $\mathcal{T}$ posted by a user u at time t and the actual repliers/retweeters, we perform supervised learning to train a ranking model following the typical learning-to-rank framework.

● **Prediction.** In the prediction phase (detailed in Sec. 3.5), given a query, i.e., a tweet $\mathcal{T}$ posted by a user u at time t, we first perform query understanding to analyze the context (after named-entity recognition, segmentation, sentiment and self-disclosure analysis, as in the training phase), and feed the ranking model with all the extracted features to generate a score for each candidate, thus obtaining the final ranking list.

In summary, our model is designed to address the dynamics of tie strength and social interactions, where different components of models reflect various aspects of the dynamics as annotated in Figure 2.

Reciprocity. We model all interactions for each dyadic relationship as directed behaviors instead of undirected behaviors employed in many existing approaches [8], thus non-reciprocal patterns are explicitly measured, e.g., the responsiveness analysis reveals the relative tendency to interact with each friend in the ego network.
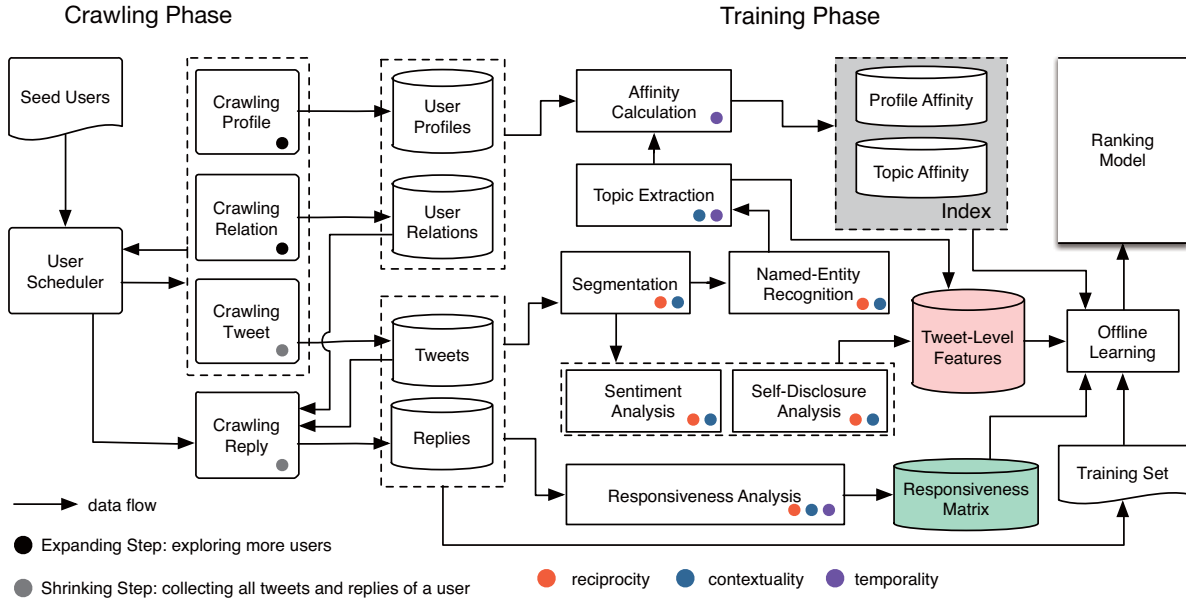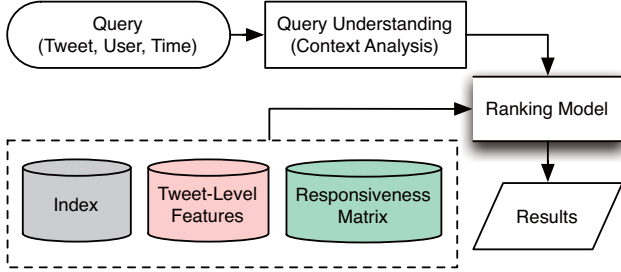
Figure 2: The crawling phase and training phase.

Figure 3: The prediction phase.

Temporality. Our prediction system works as a typical search engine, where the indexes and responsiveness matrices are updated periodically to ensure the freshness of the features. Meanwhile, the tweet-level features are streaming data that is updated in real-time to reflect the timely and subtle change in users' sentiment and degree of self-disclosure.

Contextuality. We extract rich contextual features from the tweets and replies, which diagnose a user's context when interacting with friends, such as mood, opinion, availability, and even physical environment.

3.2 Affinity Calculation

3.2.1 Profile Affinity

We use the following features to build indexes representing profile affinity: *Gender*, *Age*, *Education*, *City*, and *CommonFriends*. We build a sorted list for each feature to index u's friends $v_1, v_2, \ldots \in \mathring{E}_u$. Some features are binary, e.g., *Gender* and *City* respectively indicate whether a friend of u has the same gender and lives in the same city as u. Other features are real-valued features calculated using affinity functions. For *Education* and *CommonFriends*, we utilize Jaccard Similarity, e.g., given u's school list S_u (a user may publish several schools including high school, college, etc. on her profile page), and v's school list S_v, *Education* affinity is calculated

by

$$\text{Affinity}(\textit{Education}) = \frac{|S_u \cap S_v|}{|S_u \cup S_v|}. \quad (4)$$

Similarly, *CommonFriends* is obtained by

$$\text{Affinity}(\textit{CommonFriends}) = \frac{|\mathring{E}_u \cap \mathring{E}_v|}{|\mathring{E}_u \cup \mathring{E}_v|}. \quad (5)$$

For *Age*, we use the following affinity measurement:

$$\text{Affinity}(\textit{Age}) = \frac{\min(A_u, A_v)}{\max(A_u, A_v)}, \quad (6)$$

where A_u and A_v are the ages of u and v.

We note that the above features are provided by most users in our target Sina Weibo dataset (described later in Sec. 4.2.1), mainly because the system guides users to input these attributes once a user registers the website, and as long as such information has not been provided, the system will redirect to the page asking the user to input these attributes every time the user logs in to the system. Nevertheless, for the users without such attributes, we assign -1 for the corresponding features.

3.2.2 Topic Affinity

To summarize and compare users' topics of interest, we first identify named-entities from users' tweets and replies, following the method proposed in [20] (segmentation of Chinese words is preprocessed). Next, we remove the non-named entities from the corpus, and perform topic discovery and topical keyword extraction for each user. The reason for using a filtered corpus is that unlike formal text such as news reports, tweets are very short text strings, with more diverse topics and many noisy or meaningless words. For the same reason, traditional topic discovery methods designed for formal long text such as Latent Dirichlet Allocation (LDA) does not perform well on tweets [38]. To tackle this challenge, we employ the topical keyphrase extraction approach introduced by Zhao et al. [37]. In this method, a Twitter-LDA model is leveraged to discover topics, where each word in a tweet is either assigned as a topical word or a background word according to a Bernoulli test, then a

Sentiment	Emoticons
Happy	
Sad	

Figure 4: Classifications of emoticons on Sina Weibo

PageRank-based method is utilized to identify important words in a topic as topical key phrases.

As a result, we obtain a topic distribution $\boldsymbol{\theta}_u = (\boldsymbol{\theta}_u^{(1)}, \boldsymbol{\theta}_u^{(2)} \ldots \boldsymbol{\theta}_u^{(K)})$ for each user u, where $\boldsymbol{\theta}_u^{(i)}$ is a distribution over words, $i = 1, 2, \ldots, K$. And for each topic i, we identify a list of important words through cTPR (context-aware topical PageRank) [38], as the topical keywords of user u. In many social networks, users self-tag several keywords as their interests. We merge the identified keywords and the self-tags, as the final keywords of the users, which will be used in the online prediction phase. In addition, we compute the topic affinity between u and $v \in \mathring{E}_u$ using cosine similarity as follows:

$$\text{Affinity}(Topic) = \frac{\boldsymbol{\theta}_u \cdot \boldsymbol{\theta}_v}{\|\boldsymbol{\theta}_u\|\|\boldsymbol{\theta}_v\|}, \tag{7}$$

and index the affinity values with a sorted list, as in the profile affinity indexes.

3.3 Context Analysis

On microblog-based social networks, people may interact with both close friends and casual acquaintances. However, as discussed earlier, different contexts such as emotions, opinions, and level of self-disclosure capture the rich dynamics of tie strength. We perform sentiment analysis and self-disclosure analysis for each tweet (including retweets) and reply posted by users in $\mathring{E}_u$ to extract context features.

3.3.1 Sentiment

We perform sentiment analysis to mine both emotions (such as happy, sad) and polarity (positive/neutral/negative) from each tweet. On microblogs such as Twitter and Sina Weibo, people frequently use emoticons to express their emotions. These emoticons were also explicitly manually labeled to different classes as shown in Figure 4. To classify the tweets and replies according to polarity, we leverage the emoticons as noisy labels to train a SVM classifier using a linear kernel [9]. In addition to unigram and bigram features, we also adopt other contextual features such as punctuation, hashtags and several target-dependent features introduced by Jiang et al. [15] to train a classifier. For each tweet and reply, we obtain the polarity using the trained model and classify the emotion based on the emoticons (if no emoticons are used, we assign it a dummy class as the emotion).

3.3.2 Self-Disclosure Analysis

In social psychology, self-disclosure is found to be the key characteristic indicating the strength of social ties [24, 31]. However, canonical methods for measuring self-disclosure often require psychological assessments based on questionnaires. Nevertheless, some linguistic features reveal the degree of self-disclosure in online interactions. We measure self-disclosure for each tweet and reply with the following indicative linguistic features (only the first two features are applied to replies since the latter two are not available for replies in Sina Weibo):

• Length of tweets/replies. Although the length of tweets is limited (e.g., 140 words in Twitter and 140 Chinese characters in Sina

Weibo), when users tend to self-expose by sharing personal information, feelings, or thoughts with others, they often use more words [14].

• Occurrence of first person words such as "I," "me," "my," "mine," etc. Several studies have suggested that this feature is indicative of self-disclosure in both online and offline social interactions [3].

• Geo-Tweets. On many social networks such as Twitter and Sina Weibo, users can post geo tweets, indicating their geo-locations of themselves when posting that tweet, which is considered a typical behavior of self-disclosure [2]

• Client Types. We have the information of the clients (desktop client/mobile client) when people post a tweet, which exposes (or indicates) users' physical status, e.g., if users tweet using mobile clients, it is more likely that they are on the move, otherwise, i.e., using web client, they are probably sitting in front of their PCs at home or at work.

3.3.3 Responsiveness Analysis

Existing methods for predicting repliers/retweeters usually assume that all u's friends have read the tweets posted by u and are equally available to interact with u [29], which is often an unrealistic assumption. In responsiveness analysis, we use historical interaction data to explicitly measure 1) the availability that a user has to reply to/retweet a tweet; 2) the capacity of a user to read tweets and interact with others; and 3) the tendency of a user to interact with a particular friend. All these factors are time-dependent variables.

Actually, given a time bin $\Delta = [t_1, t_2]$, we can "reconstruct" the timeline of a user by retrieving all the tweets and replies posted by E_u during Δ. Let $\mathcal{T}_u^\Delta$ denote the tweets posted by u within Δ (including retweets), and let $\mathcal{R}_{v,u}^\Delta$ denote the replies sent from v to u within Δ. Then a user u's cost of reading the tweets and replies within Δ could be estimated by

$$\sum_{v \in E_u} \left(\|\mathcal{T}_v^\Delta\| + \|\mathcal{R}_{v,u}^\Delta\| \right). \tag{8}$$

In practice, we partition a day into discrete time bins and aggregate the above statistics over multiple days, to compute the average cost for each time bin of a day. Intuitively, the more cost u spends on reading these tweets and replies, the less available that u is to respond (reply to/retweet) a particular tweet posted within a given time period. Meanwhile, the capacity of u to reply to friends' tweets can be approximated by

$$\sum_{v \in \mathring{E}_u} \|\mathcal{R}_{u,v}^\Delta\|. \tag{9}$$

Thus we formally define the *Responsiveness* of v in terms of replying to u's tweets within Δ as

$$r_{v \to u}^\Delta|_{reply} = \frac{\|\mathcal{R}_{v,u}^\Delta\| - \frac{1}{\|\mathring{E}_v\|} \sum_{w \in \mathring{E}_v} \|\mathcal{R}_{v,w}^\Delta\|}{\frac{1}{\|\mathring{E}_v\|} \sum_{w \in \mathring{E}_v} \left(\|\mathcal{T}_w^\Delta\| + \|\mathcal{R}_{w,v}^\Delta\| \right)}, \tag{10}$$

where the numerator captures both the tendency of v to reply to u, as well as the capacity of v, and the inverse of the denominator captures the availability as discussed earlier. The responsiveness of v in terms of retweeting u can be similarly calculated by replacing $\mathcal{R}_{v,w}^\Delta$ in the numerator of Eq. (10) with $\mathcal{RT}_{v,w}^\Delta$, where $\mathcal{RT}$ represents retweets, i.e.,

$$r_{v \to u}^\Delta|_{retweet} = \frac{\|\mathcal{RT}_{v,u}^\Delta\| - \frac{1}{\|\mathring{E}_v\|} \sum_{w \in \mathring{E}_v} \|\mathcal{RT}_{v,w}^\Delta\|}{\frac{1}{\|\mathring{E}_v\|} \sum_{w \in \mathring{E}_v} \left(\|\mathcal{T}_w^\Delta\| + \|\mathcal{R}_{w,v}^\Delta\| \right)}. \tag{11}$$
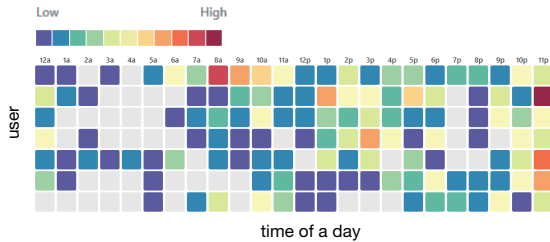
Figure 5: Responsiveness matrix of a user

As a result, for each target user u, we calculate the responsiveness with respect to each of u's friends, thus obtaining two responsive matrices (for replies and retweets), where each row of the matrix is associated with a user, and each column indicates a time bin as shown in Figure 5.

3.4 Learning the Ranking Model

Given a tweet q posted by user u as a query, and u's friends $\mathring{E}_u$ as candidate documents, a feature vector $\mathbf{x}(q, v)$ is generated for each query-document pair (q, v) where $v \in \mathring{E}_u$. Here $\mathbf{x}(q, v)$ consists of all the extracted user-level features associated with user u and v (e.g., the profile affinity between u and v), and tweet-level features associated with q. Suppose $Y = (1, 2, \ldots, m)$ is the label set for a query-document pair, where the label indicates the relevance, i.e., the intention to reply to/retweet a tweet. The training set is represented by $S = (\mathbf{x}_i, y_i)_{i=1}^N$, where $\mathbf{x}_i \in X$ is the feature vector of the ith query-document pair and $y_i \in Y$. We aim to train a ranking model $F(S) = f(\mathbf{x})$, which assigns a score for a feature vector $\mathbf{x}$, and can rank the documents for a given query.

The above setting follows the typical learning-to-rank framework for information retrieval. We note that by using the same learning framework, we can learn the ranking functions for different interaction types, such as replies and retweets. According to whether label set Y is binary, our prediction task has two sub-forms:

• **Binary Relevance Judgment** (|Y|=2). In this case, we neglect the orders of the repliers/retweeters given a particular tweet. Assume a tweet q posted by user u at time t. The users who reply to q are treated as relevant instances, and users who do not reply to q are regarded as irrelevant instances. Note that even though the label is binary, the ranking task is not equivalent to the typical classification/regression task since they have different optimization criteria (a perfect ranking does not necessarily mean a perfect classification/regression).

• **Ordinal Relevance Scale** (|Y|>2). In this case, we predict not only the repliers/retweeters of a tweet, but also their orders, i.e., who will reply/retweet first, second, etc. This is obviously harder than the first task, and requires finer-grained and more accurate understanding of the dynamic tie strength and users' interaction behavior patterns in the online social network.

Over the past decades, various algorithms have been proposed to learn the ranking functions, most of which convert the learning problem into an optimization problem by minimizing a defined loss function $L(F(S))$. These algorithms can be generally categorized to point-wise approaches, pair-wise approaches, and list-wise approaches, mainly in terms of the loss function [19]. We evaluated a number of well-adopted learning algorithms to learn the ranking functions separately and examined their performance in our framework (refer to Sec. 4.2.4 for more details). We note that the purpose of applying different learning algorithms is for demonstrating the applicability of our model, whereas a thorough comparison of these learning algorithms is beyond the focus of this paper.

Table 1: Percentage of users with profile information (%)

gender	city	education	age
100	92.9	35.03	31.89

3.5 Prediction

In the online prediction phase, given a query, i.e., tweet $\mathcal{T}$ posted by u at time t, we perform query understanding to generate contextual features from the query. We first extract topical keywords (refer to Sec. 3.2.2) after the pre-processing of segmentation and named-entity recognition. Next we perform sentiment analysis (Sec. 3.3.1) and self-disclosure analysis (Sec. 3.3.2) using the same approach as in the training phase. Then we feed the extracted tweet-level features as well as the user-level features from the candidates to the trained ranking models (different models for repliers and retweeters), and finally retrieve a ranking list by sorting the candidates according to the scores of the ranking function.

4. EVALUATION

4.1 Dataset

We collected a large dataset covering 1.116 billion tweets posted by 1.126 million users with all their replies and retweets to their friends, which range from Sep. 2009 to Dec. 2013. The covered users form 73.262 million dyadic friendship relations.

The percentage of users (among all the 1.1M users) who have different types of profile information on Sina Weibo is reported in Table 1. Such information is leveraged when we build the profile affinity index (Sec 3.2.1). For topic affinity calculation, we set the number of topics $K = 50$ and we retrieved the top 20 keywords for each topic. We chose these parameter settings because they generally gave coherent and meaningful topics for our dataset after our manual validation.

4.2 Settings

4.2.1 Preliminary

To evaluate the prediction models, we removed the tweets that had no replies/retweets) for replier/retweeter prediction, which resulted in two final datasets: 1) the tweet-reply dataset containing 215.165M tweets and 446.910M replies, 2) the tweet-retweet dataset containing 29.348M tweets and 38.152M retweets, which indicated that these users interacted with their friends more by replying than retweeting.

For both the replier prediction and retweeter prediction, we conducted a 10-fold cross validation for all the compared models. In each fold, we used 6 parts for training, 2 parts for validation, and the remaining 2 parts for testing. The validation set was used to tune the hyper-parameters of the learning algorithms, such as the combination coefficient in RankingSVM and the number of trees in LambdaMART, and then we fixed the parameters to train the models using the training set. The test set was used to evaluate the performance of the trained models. Note that all the reported results are the average performances over the 10 trials.

Furthermore, in order to evaluate the temporal effect on the performance, i.e., the temporality dynamics, we performed experiments using a batch training mode. We equally segmented the data according to the time that the tweet (i.e., query) was posted into 6 segments where each segment contains the data in a period of about 8.5 months. For each segment, we used the previous segment as the training (60%) and validation (40%) set, and the current segment as

the test set (note that the first segment was only used for burning up).

4.2.2 Baselines

To the best of our knowledge, there is no existing model that can predict both repliers and retweeters for a given tweet. Nonetheless, we compared our method against the following related state-of-the-art approaches.

• Social, Topical, and Activity features based model (**STA**) [29]. In this approach, a logistic regression model is trained to determine whether a user will or will not reply to another user, using a set of social features (such as the common friends), topical features (such as the topical similarity of tweets), and activity features (such as the number of previous replies). "List" related features in this model were not available for our dataset because Sina Weibo does not have the "List" feature that Twitter does (the comparable function in Sina Weibo is the social "groups," which are only visible to the users themselves).

• Homophily-based Graphical model (**HG**) [35]. In this graphical model, the tie strength between dyadic users is a latent variable solely dependent on the similarity of user profiles. The interactions between users (such as replies, retweets, and tags) are impacted by the latent tie strength.

Since both STA and HG are binary classifiers that can not directly be used for prediction, we obtained their prediction results by ranking the output probabilities for each classified instance.

4.2.3 Criteria

Given a test set with m tweets $q_1, q_2, \ldots, q_m$ (as queries), the predicted result for q_i is represented by a ranked list π_i for $i = 1, 2, \ldots, m$. We evaluated the results for binary judgment, using the following criteria for all the compared methods:

• Topmost Accuracy (Acc). This measures the ratio of successfully predicted repliers/retweeters to the number of tweets, i.e.,

$$\text{Acc} = \frac{\left\| \sum_{i=1}^{m} \delta_{\{l_{\pi_i^{-1}(1)}=1\}} \right\|}{m}, \qquad (12)$$

where $\delta_{\{\cdot\}}$ is the indicator function and $l.$ is the label of a user (replier/retweeter is labeled 1 and non-replier/non-retweeter is labeled 0).

• Mean Average Precision (MAP). The Average Precision for tweet q_i, denoted as AP_i, is calculated by

$$\text{AP}_i = \frac{\sum_{k=1}^{m} \text{P@}k \cdot \delta_{\{l_{\pi_i^{-1}(k)}=1\}}}{R_i}, \quad \text{P@}k = \frac{\sum_{t \leq k} \delta_{\{l_{\pi_i^{-1}(t)}=1\}}}{k}$$

where R_i is the number of actual repliers/retweeters of the tweet q_i. MAP is the average over all AP_i for $i = 1, 2, \ldots, m$.

• Normalized Discounted Cumulative Gain (NDCG). This is a widely-adopted measurement for evaluating ranking quality. The Discounted Cumulative Gain (DCG) at position k for a query with ranked result π is given by

$$\text{DCG@}k = \frac{\sum_{j=1}^{k} G(l_{\pi^{-1}(j)})}{\log(1+j)}, \qquad (13)$$

where $G(\cdot)$ is the gain function (we set $G(x) = 2^x - 1$ in our experiments). Then NDCG@k is calculated by normalizing DCG@k

with its maximum value (i.e., replacing π with the ideal ranking in Eq. (13)).

For evaluating the sequence of repliers/retweeters (ordinal relevance scale), we also employed the NDCG metric, where we labeled the first replier/retweeter with 3, the second replier/retweeter with 2, the third and latter repliers/retweeters with 1, and non-repliers/non-retweeters were still labeled 0. Furthermore, we replace $l_{\pi_i^{-1}(1)} = 1$ in Eq. (12) with $l_{\pi_i^{-1}(1)} = 3$ (which is a stricter metric) to evaluate whether a model was able to identify the actual first replier/retweeter in Sec. 4.3.2.

4.2.4 Ranking Functions

We evaluated the following models for training the ranker (Sec. 3.4).

LogisticRegression is a widely used model for binary classification that can also be considered a pointwise approach in the learning-to-rank framework [19]. The ranking function is defined as follows:

$$f(\mathbf{x}) = \Pr(y = 1 | \mathbf{x}) = \frac{1}{1 + e^{-c - \langle \mathbf{w}, \mathbf{x} \rangle}}, \qquad (14)$$

where $\mathbf{w}$ is the weight vector for the feature vector $\mathbf{x}$, which is estimated using maximum likelihood estimation.

RankingSVM [16] is a pairwise approach that applies the SVM on pairwise training instances, so as to learn a linear ranking function $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$, where the weight feature $\mathbf{w}$ is trained by minimizing the regularized hinge loss function, given by

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{(i,j) \in \mathcal{P}} [\langle \mathbf{w}, \mathbf{x}_i - \mathbf{x}_j \rangle]_+, \qquad (15)$$

where $[x]_+ = \max(0, 1 - x)$, and $(i, j) \in \mathcal{P}$ means candidate i is preferred over candidate j.

LambdaMART [34] is a listwise approach, which learns a combination of regression trees as the ranking function, denoted as

$$f(\mathbf{x}) = \sum_{i=1}^{T} \alpha_i f_i(\mathbf{x}), \qquad (16)$$
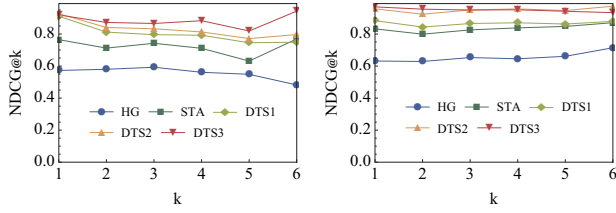
where each $f_i(\mathbf{x})$ is a regression tree over the features. LambdaMart is trained by empirically optimizing IR measures such as NDCG directly, which has shown excellent performance on several real-world ranking problems including the 2010 Yahoo! Learning-to-Rank Challenge.

In the evaluation, we label the results of our model by considering the dynamic tie strength using the above ranking functions as **DTS1**, **DTS2**, and **DTS3**, respectively.
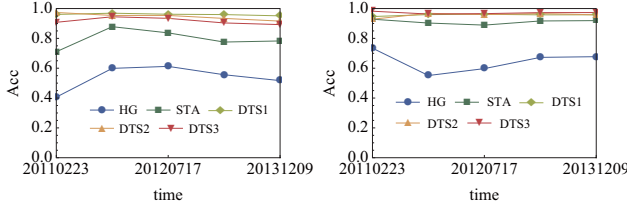
4.3 Results

4.3.1 Predicting the Replier/Retweeter

Figure 6 presents the NDCG over k for replier/retweeter predictions in terms of binary judgment, i.e., actual repliers and retweeters were all considered relevant documents regardless of their order. According to the evaluation results, our method significantly outperforms the baseline methods, with the STA method performing better than the HG method. For example, as depicted in Figure 6(a), even the worst ranking function of our method DTS1 improves the NDCG@1 of STA with 0.15. Note that both DTS1 and STA employ Logistic Regression. However, DTS1 considers more dynamic features such as reciprocal rank, sentiment, self-disclosure, and responsiveness. Furthermore, we found that the DTS3 model, which
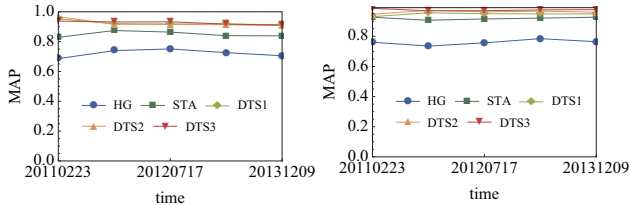
(a) replier prediction   (b) retweeter prediction

Figure 6: NDCG@k in terms of binary judgment

(a) replier prediction   (b) retweeter prediction

Figure 7: Acc in terms of binary judgment

(a) replier prediciton   (b) retweeter prediction

Figure 8: MAP in terms of binary judgment

(a) replier prediction   (b) retweeter prediction

Figure 9: NDCG@k in terms of ordinal relevance scale

(a) replier prediction   (b) retweeter prediction

Figure 10: Acc in terms of ordinal relevance scale

is a listwise learning algorithm achieving the highest performance in terms of NDCG@k for $k = 1, 2, \ldots, 6$.
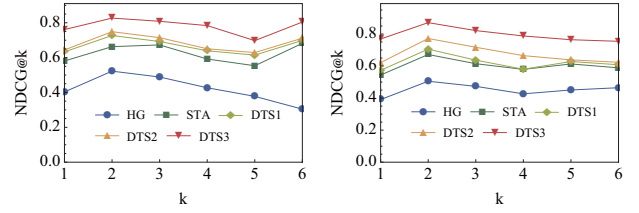
Figure 6(b) similarly presents the corresponding results for retweeter prediction. Overall, we observe that all algorithms show better performance compared with replier prediction, while the dynamic features demonstrate the advantage again. As mentioned earlier in Sec. 4.2.1, users retweet their friends less often than they reply. This finding is also in coherence with previous studies, which suggest that retweeting is more driven by the content while replying is more related to social relationships [21, 29].

Figure 7 and Figure 8 show the Topmost Accuracy and MAP in the batch training mode. The results show that our methods that consider the dynamic tie strength between dyadic relations reveal higher predictive power for both replier and retweeter. Furthermore, the results of our method (for all the ranking functions) are more stable than competing approaches. For ACC and MAP, the three ranking functions achieve similar performances.

We further conducted greedy forward selection to investigate the contribution of different features in our model using the best performer, DTS3. As presented in Table 2 (NDCG@1 and NDCG@3 are presented as examples while the results for $k = 1, 2, \ldots, 6$ show similar patters), every type of feature improves the NDCG measurement. The responsiveness shows critical improvement for replier prediction and topic affinity shows the most important improvement for retweeter prediction. As more and more dynamic features are incorporated, the NDCG is monotonically increased.
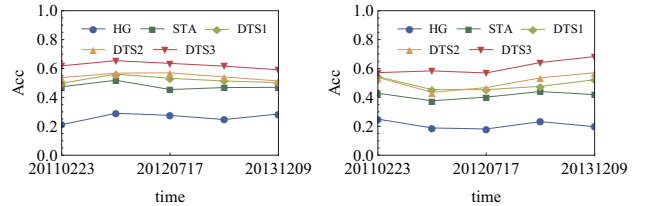
4.3.2 Predicting the Replier/Retweeter Sequence

Figure 9 shows the evaluation results of NDCG and Acc respectively in terms of ordinal relevance scale. It is clear that the NDCG

results of all the compared models are lower than the previous results, which indicates that predicting the order of repliers and retweeters is much more challenging. For example, NDCG@1 of both STA and DTS3 are about 0.15 lower than the binary judgment case. However, the dynamic models still perform better than the baselines, suggesting that the subtle dynamics of tie strength are faithfully reflected in users' online interaction behaviors. Besides this, we found that the advantage of DTS3 over the other two ranking functions is more significant when the order of replying/retweeting is considered. On average, DTS3 improves the NDCG with 0.16 gain for replier prediction and 0.19 gain for retweeter prediction over STA.

Figure 10 presents the Topmost Accuracy for replier and retweeter prediction. Note that, the Topmost Accuracy here is stricter than the previous Acc measurement, since the prediction for a query is considered correct only if a model output the actual first replier or retweeter, while the Acc in the binary judgment case considers all repliers as correct. According to the results, DTS3 model achieves the best performance with an Acc higher than 0.6, and the gap between DTS3 with other methods (including other ranking functions) are much more prominent than in the previous binary judgment setting. The contributions of different features were also studied using DTS3 and reported in Table 3. Similar patterns are shown as the previous for replier prediction, which suggest the difference between users' replying and retweeting behaviors and reveal the predictive power of tie strength dynamics on predicting repliers and retweeters.

5. RELATED WORK

5.1 Estimation of Tie Strength

The strength of social ties has been studied for decades in social science. In 1973, Granovetter [11] explicitly defined the tie strength as the "combination of amount of time, the emotional intensity, the intimacy, and the reciprocal services which characterize the tie," and suggested using the frequency of interaction to estimate tie strength, which was later adopted by many subsequent studies [30, 22].

The proliferation of digital communications brought the opportunity to estimate tie strength with large-scale communication data. For example, using mobile phone logs, Onnela et al. [26] studied

Table 2: NDCG w.r.t features in terms of binary relevance
P: profile affinity, T: topic affinity, R: reciprocity
r: responsiveness, S: sentiment, D: self-disclosure

	replier			retweeter	
Method	NDCG@1	NDCG@3	Method	NDCG@1	NDCG@3
P	0.786	0.735	P	0.858	0.808
Pr	0.832	0.781	PT	0.908	0.881
PrR	0.854	0.819	PTR	0.934	0.921
PrRT	0.886	0.842	PTRr	0.948	0.933
PrRTD	0.905	0.855	PTRrD	0.960	0.945
PrRTDS	0.919	0.866	PTRrDS	0.969	0.950

Table 3: NDCG w.r.t features in terms of relevance scale
P: profile affinity, T: topic affinity, R: reciprocity
r: responsiveness, S: sentiment, D: self-disclosure

	replier			retweeter	
Method	NDCG@1	NDCG@3	Method	NDCG@1	NDCG@3
P	0.623	0.693	P	0.626	0.677
Pr	0.664	0.740	PT	0.679	0.730
PrR	0.703	0.764	PTR	0.710	0.772
PrRT	0.736	0.785	PTRr	0.739	0.795
PrRTD	0.750	0.799	PTRrD	0.758	0.809
PrRTDS	0.761	0.808	PTRrDS	0.769	0.821

the local and global structure of a society-wide communication network and identified that social networks are robust to the removal of strong ties but fall apart after a phase transition if the weak ties are removed. Roth et al. [28] proposed a measurement termed "Interaction Rank" to model the group-level tie strength, where the interactions were weighted by recency to measure the tie strength, and outflowing interactions were considered to be more indicative than inflowing interactions.

Recently, online social networks has rapidly developed. There has been increasing interest in measuring the tie strength in online social networks [8, 35, 13, 17]. Gilbert and Karahalios [8] proposed a set of variables to classify strong ties and weak ties using Facebook data. Later, Gilbert [7] applied similar approaches to Twitter data and found that the Facebook tie strength model can generally be applied to Twitter. With Linkedin data, Xiang et al. [35] proposed a graphical model to learn the tie strength of a dyadic relationship. In this model, the tie strength is a latent variable directly determined by profile similarity and the interaction behaviors are conditionally independent given the tie strength. Wiese et al. [33] studied the sharing behaviors of online social networks and found that frequency of communication can predict both tie strength and willingness to share. Jones et al. [17] identified strong ties using Facebook interaction data, and compared with the ground truth obtained from users' self-reported data, where they found that interaction frequency is more indicative than profile attributes for estimating tie strength, regardless of whether the interactions are public or private.

Our work differs from existing approaches in the following ways: 1) We focus on mining the dynamics of tie strength from social interaction data in terms of various aspects such as temporality, contextuality, and reciprocity; 2) We have developed a learning-to-rank approach to predict interaction behaviors in online social networks by considering the dynamics of tie strength.

5.2 Prediction of Online Social Interactions

Wang and Huberman [32] addressed the predictability of online social interactions by measuring conditional entropy and mutual information with data from Epinions (a consumer review site) and Whrrl (a location-based social network), and suggested that strong deterministic components were embedded in social interactions.

Recently, a number of models have been proposed to predict various types of social interactions in online networks. Xiang et al. [35] proposed a generative model to estimate the occurrence of certain interactions such as posting on Facebook Walls. Macskassy and Michelson [21] investigated the motifs of users' retweeting behaviors on Twitter and proposed different models: general model, content model, homophily model, and recency model, where the content-based propagation was found to be better at explaining retweeting behavior. Replying behavior was also studied in several literatures [13, 29]. In contrast to retweeting behavior, replying is found to be more driven by tie strength and homophily, e.g., Schantl et al. [29] proposed a regression model combining a set of topical and social factors. Using this model, they found that social features are more indicative of user replying behavior. However, this model is only capable of classifying whether a user will reply to another user without considering the effect of current tweet, i.e., the prediction results are identical for different tweets posted by a user.

In this work, we aim to predict multiple types of interactions (reply/retweet) with a general framework, and we focus on the predictions at a finer granularity. First, a key difference between our model and previous ones (such as STA and HG) is that unlike existing ones which provide predictions at *user-level*, our model is at *tweet-level*, i.e., previous models will generate the same prediction for different tweets posted by a user while our model incorporates dynamic features such as the content of a tweet and the context of a user, which are crucial because users' replying and retweeting behaviors are typically affected by these factors.

Second, almost all existing models [35, 21, 29] assume that the user reads all tweets posted by her friends (shown on the user's own timeline) and is equally available to interact with others, which is unrealistic. Instead, by reconstructing users' timelines and analyzing their historical interaction behavior, we propose the responsiveness measure to explicitly consider a user's availability, tendencies and capacity to consume a tweet and respond to their friends. Finally, in contrast to previous approaches that consider only binary situations (replier/non-replier, retweeter/non-retweeter), our model is able to predict the order of repliers/retweeters by employing a learning-to-rank framework.

6. CONCLUSION

In this paper, we have addressed the dynamics of tie strength and social interactions from multiple aspects, and developed a comprehensive framework to predict repliers and retweeters given a tweet posted by a user at a certain time. Extensive evaluations were performed to validate the effectiveness of our method. As a result, the proposed model considering a variety of dynamic features significantly outperforms state-of-the-art methods. While the dynamics of tie strength show compelling predictive power for online social interactions, we have also been concerned with some potential limitations in this work, e.g., the sampling bias of the dataset. However, we believe that the friendship dynamics in online social networks will bring more and more attention in both industry and academia, and have a profound impact on many high-value applications, such as online communication, social recommendation, and viral marketing.

References

[1] R. Athanasiou and G. A. Yoshioka. The spatial character of friendship formation. *Environment and Behavior*, 1973.

[2] J. Y. Bak, S. Kim, and A. Oh. Self-disclosure and

relationship strength in twitter conversations. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 60–64. Association for Computational Linguistics, 2012.

[3] A. Barak and O. Gluck-Ofri. Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior*, 10(3):407–417, 2007.

[4] R. I. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6):469–493, 1992.

[5] N. Eagle, A. S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

[6] J. H. Fowler, J. E. Settle, and N. A. Christakis. Correlated genotypes in friendship networks. *Proceedings of the National Academy of Sciences*, 108(5):1993–1997, 2011.

[7] E. Gilbert. Predicting tie strength in a new medium. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pages 1047–1056. ACM, 2012.

[8] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211–220. ACM, 2009.

[9] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 2009.

[10] B. Gonçalves, N. Perra, and A. Vespignani. Modeling users' activity on twitter networks: Validation of dunbar's number. *PloS one*, 6(8):e22656, 2011.

[11] M. S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[12] M. Gupte and T. Eliassi-Rad. Measuring tie strength in implicit social networks. In *Proceedings of the 3rd Annual ACM Web Science Conference*, pages 109–118. ACM, 2012.

[13] M. Gupte and T. Eliassi-Rad. Measuring tie strength in implicit social networks. In *Proceedings of the 3rd Annual ACM Web Science Conference*, pages 109–118. ACM, 2012.

[14] D. J. Houghton and A. N. Joinson. Linguistic markers of secrets and sensitive self-disclosure in twitter. In *Proceedings of the 2012 45th Hawaii International Conference on System Sciences*, pages 3480–3489. IEEE Computer Society, 2012.

[15] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-dependent twitter sentiment classification. In *Proc. ACL*, pages 151–160, 2011.

[16] T. Joachims. Optimizing search engines using clickthrough data. In *Proc. SIGKDD*, pages 133–142. ACM, 2002.

[17] J. J. Jones, J. E. Settle, R. M. Bond, C. J. Fariss, C. Marlow, and J. H. Fowler. Inferring tie strength from online directed behavior. *PloS one*, 8(1):e52168, 2013.

[18] P. F. Lazarsfeld, R. K. Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

[19] T.-Y. Liu. *Learning to rank for information retrieval*. springer, 2011.

[20] X. Liu, S. Zhang, F. Wei, and M. Zhou. Recognizing named entities in tweets. In *Proc. ACL*, pages 359–367, 2011.

[21] S. A. Macskassy and M. Michelson. Why do people retweet? anti-homophily wins the day! In *Proc. ICWSM*, 2011.

[22] P. V. Marsden and K. E. Campbell. Measuring tie strength. *Social forces*, 63(2):482–501, 1984.

[23] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[24] D. G. Myers. *Social psychology*. McGraw-Hill New York, 2010.

[25] P. M. Olk and D. E. Gibbons. Dynamics of friendship reciprocity among professional adults. *Journal of Applied Social Psychology*, 40(5):1146–1171, 2010.

[26] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[27] P. Preciado, T. A. Snijders, W. J. Burk, H. Stattin, and M. Kerr. Does proximity matter? distance dependence of adolescent friendships. *Social networks*, 34(1):18–31, 2012.

[28] M. Roth, A. Ben-David, D. Deutscher, G. Flysher, I. Horn, A. Leichtberg, N. Leiser, Y. Matias, and R. Merom. Suggesting friends using the implicit social graph. In *Proc. SIGKDD*, pages 233–242. ACM, 2010.

[29] J. Schantl, C. Wagner, R. Kaiser, and M. Strohmaier. The utility of social and topical factors in anticipating repliers in twitter conversations. In *Proceedings of ACM Web Science*, 2013.

[30] A. Sutcliffe, R. Dunbar, J. Binder, and H. Arrow. Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British journal of psychology*, 103(2):149–168, 2012.

[31] A. L. Vangelisti and D. Perlman. *The Cambridge handbook of personal relationships*. Cambridge University Press, 2006.

[32] C. Wang and B. A. Huberman. How random are online social interactions? *Scientific reports*, 2, 2012.

[33] J. Wiese, P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. In *UbiComp*, pages 197–206, 2011.

[34] Q. Wu, C. J. Burges, K. M. Svore, and J. Gao. Adapting boosting for information retrieval measures. *Information Retrieval*, 13(3):254–270, 2010.

[35] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.

[36] N. J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu, and X. Xie. We know how you live: exploring the spectrum of urban lifestyles. In *Proceedings of the first ACM conference on Online social networks*, pages 3–14. ACM, 2013.

[37] W. X. Zhao, J. Jiang, J. He, Y. Song, P. Achananuparp, E.-P. Lim, and X. Li. Topical keyphrase extraction from twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 379–388. Association for Computational Linguistics, 2011.

[38] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In *Advances in Information Retrieval*, pages 338–349. Springer, 2011.