Link Prediction for New Users in Social Networks

Xiao Han*, Leye Wang*, Son N. Han*, Chao Chen†, Noël Crespi* and Reza Farahbakhsh*
* Institut-Mines Télécom, Télécom SudParis
† Chongqing University

***Abstract*—Link prediction for new users who have not created any link is a fundamental problem in Online Social Networks (OSNs). It can be used to recommend friends for new users to start building their social networks. The existing studies use cross-platform approaches to predict a new user's links on a certain OSN by porting his existing links from other OSNs. However, it cannot work when OSNs are not willing to share their data or users do not want to connect different OSN accounts. In this paper, we use a single-platform approach to carry out the link prediction. We explore the users' profile attributes (e.g., workplace, high school and hometown) which can be easily obtained during the new users' sign up procedure. Based on the limited available information from the new user, along with the attributes and links from existing users, we extract three types of social features: basic feature, derived feature and latent relation feature. We propose a link prediction model using these social features based on Support Vector Machines. Eventually, we rely on a large Facebook data set consisting of $479,000$ users to evaluate our proposed model. The result reveals that our model outperforms the baselines by achieving the AUC value of 0.83; it also demonstrates that each of the proposed social features contribute significantly to the prediction model.**

***Index Terms*—Social Network, New-User Link Prediction.**

I. Introduction

In Online Social Networks (OSNs), social links (e.g., *friends* in Facebook, *follower-followee* in Twitter) play an important role in users' experience as well as in the success of the OSN. If a user's links are well-established, he can use the social network more frequently [1][2][3]. Therefore, a high-quality link prediction is required to allow OSNs to recommend useful links to users. Especially, when it comes to a new user who has not created any link, the link prediction (*new-user link prediction*) becomes even more crucial, because it can be used to recommend friends for new users to start building their social networks. A poor prediction in the first place may discourage new users from using the platform.

Many approaches have been proposed to predict users' potential links depending on the existing ones that they have already established [2][3][4]. However, these approaches cannot be adopted to the new-user link prediction since the new users have not created any link. Recently, by using *cross-platform* approaches, a few studies have begun to tackle this new-user link prediction problem. These studies predict a new user's links in a certain OSN platform by porting that user's well-established links from other OSNs [5][6]. Nevertheless, the application of these cross-platform methods in real-life scenarios may face some problems. First, two OSNs may not agree to share users' links as users' information is generally private, confidential and valuable to them [7]. Second, users may not give their consent to be tracked back to their information in other OSNs or users intend to use different OSNs for different purposes (e.g., LinkedIn for professional and Facebook for personal). Due to these problems, in this work, we study the new-user link prediction problem in a *single-platform* instead of cross-platform.

Our single-platform approach is to leverage the attributes (e.g., workplace, high school and hometown) provided by new users when they register their accounts, as well as information from existing users. It is inspired by the previous studies showing that the similarity between users' attributes reflects their relationship to some extent. For instance, people who share more common interests are more likely to be friends [8][9][10][11]. As it is practical for OSNs to request user attributes during the registration, our approach is applicable in real-life scenarios.

Without the loss of generality, new-user link prediction problem can be considered to predict whether a new user s will link to a given existing user u or not. In particular, given s with attributes and u with both attributes and friends, we attempt to extract some social features that can indicate the probability of s linking to u. By using Support Vector Machines (SVM) [12], we train a link prediction model to determine whether s will link to u based on the combination of the extracted social features.

Exploring appropriate social features is crucial and challenging since it directly affects the capacity of the prediction model and the available information is very limited. We propose to fully use the obtainable information and extract the following three types of social features:

- *Basic features:* There are two types of basic features: *binary similarity* and *number of common attributes*. The former is calculated by comparing two users s and u by each attribute (e.g., current city). The latter is the total number of the same attributes between them.
- *Derived features:* We further describe the relation between two users' attributes by various ways, e.g., the geographic distance between their current cities or their interest similarity.
- *Latent relation:* We use a latent relation score to estimate how much s and friends of u share the same attributes. We show that two users probably obtain a higher score if they are friends.

In summary, this paper has the following contributions: (1) We explore multiple social features to predict links for new users who have not created any link. To the authors' best knowledge, this is the first work to address the new-user link prediction problem by leveraging the information from a single-platform. (2) To evaluate our approach, we use a Facebook data set including 479,000 users [13]. For each user, we record his demographics, interests and links (friends). Results show that all the features we proposed in this paper can significantly improve the performance of the new-user link prediction.

II. Related work

The links among users are the foundation of a social network platform. Link prediction thus is an interesting and challenging research topic where a number of researchers devote themselves to it. Hasan *et al.* rely on the co-authorship graph to predict the likelihood of the future co-author relation between users. They investigate various supervised learning algorithms and discover that SVM outperforms the others [2]. In addition, Menon *et al.* improve the prediction by using a supervised matrix factorization method to learn several latent features from the social network graph [3]. Based on supervised random walks, Backstrom *et al.* combine information from users and links to guide the random walk on the existing network graph to improve the link prediction [4]. Most of these approaches aim at *existing users* who already exhibited many links; they predict new links relying primarily on the existing users' established links (i.e., existing friends). Whereas, this paper concentrates on the link prediction for new users who have not linked to anyone.

Recently, a few studies started to deal with the problem of link prediction for new users using cross-platform approaches. A common idea is to explore the new users' established links in other OSNs to carry out the prediction in the target OSN. Yan *et al.* devise a random walk based algorithm drawing on the cross-platform information [6]; Zhang *et al.* solve the new-user link prediction problem by using information transferred from both existing active users in the target network and other networks [5]. The main concern on these cross-platform approaches is its feasibility on real-life scenarios. For instance, a Facebook user is about to open a new account in Foursquare. If Foursquare tends to use cross-platform approaches to recommend friends to the user, it ought to retrieve data from Facebook. However, Facebook may refuse to share user data to Foursquare. Besides, there is another work addressing the cold start link prediction problem without knowing any user-to-user links in a platform by using some information outside of the platform (e.g., shopping history) [7]. In this paper, we focus on single-platform instead of cross-platform to predict the links for new users.

Some existing work illustrates the potential that users' demographic information or interests may enhance the performance of link prediction. For instance, interest similarity is leveraged to improve friend prediction for existing users [10][11][14]. Previous work also verifies that the probability of two users becoming friends relates to their geographic closeness [8][9]. Referring to these results, we fully use the available information and extract appropriate social features to predict links for new users.

III. Link prediction

The goal of this work is, given an undirected social network graph and a new user s who has not created any link yet, to determine which users in the given graph the new user s will connect to. In this section, we formulate the link prediction problem and describe our solution.

A. Problem statement

Considering a given undirected social network graph $\mathcal{G} = (\mathcal{U}, E)$, where U is a set of *existing users* in the social network graph; E is a set of undirected links $e\langle u, v\rangle$ between users u and v where $u,\ v \in U$. Apart from the links, users on social networks usually expose other personal attributes such as age, hometown, college and work. Therefore, for each *existing user* u, we generate an *attribute vector*, denoted as $\mathcal{A} = \langle a_1(u), a_2(u), \cdots a_m(u)\rangle$. We also gather all of u's links into a *friend set*, denoted as $\mathcal{F}(u) = \{f | f \in U \wedge e\langle u, f\rangle \in E\}$. Then, we can use a tuple to represent an *existing user* as $u : \langle a_1(u), a_2(u), \cdots a_m(u), \mathcal{F}(u)\rangle$. Note that, as the user u may not complete all the attributes or not expose his friend set, some elements in the tuple can be *null*.

For a new user who has not constructed any link, OSNs usually request him to provide some personal information when he is signing up. For this reason, without existing links (i.e., no friend set), a new user s can be represented by a tuple merely with attributes as: $s : \langle a_1(s), a_2(s), \cdots a_m(s)\rangle$.

According to the goal of this work — to distinguish which of the *existing users* $u \in U$ are preferred to construct a link by s and which are not, we classify a candidate set of *existing users* (i.e., $\mathcal{C}$) into two categories: *linked-users* (i.e., $\mathcal{L}$) and *de-linked-users* (i.e., $\mathcal{D}$). Note that $\mathcal{C} = \mathcal{L} \bigcup \mathcal{D}$, where $\mathcal{C} \in U$. We assume that the users in $\mathcal{L}$ are more likely to get linked by s than the users in $\mathcal{D}$.

On the basis of the above establishments, the problem of ***new-user link prediction*** can be formally stated as: *Given a social network graph $\mathcal{G} = (U, E)$ where each $u \in U$ contains an attribute vector and a friend set, $u : \langle a_1(u), a_2(u), \cdots a_m(u), \mathcal{F}(u)\rangle$, a set of existing user candidates $\mathcal{C} \subseteq U$, a given new user s who is represented by an attribute vector (i.e., $s : \langle a_1(s), a_2(s), \cdots a_m(s)\rangle$), predict which of the users in $\mathcal{C}$ that s may create links to, labeled as $\mathcal{L}$(linked-users), and which of the users that s may not, labeled as $\mathcal{D}$ (de-linked-users).*

B. Workflow of the link prediction

Given a new user s who reveals some profile attributes and a set $\mathcal{C}$ of existing users who exhibit both attributes and friends, the basic idea is exploiting all the obtainable information to figure out existing users that are similar to s from $\mathcal{C}$ as the *linked-users* ($\mathcal{L}$), for much existing work has proved that people are likely to connect to another if they are similar to each other [10][11][14][8][9].

Based on the mentioned idea, we model the friend probability, which measures the probability that s will create a link to $u \in \mathcal{C}$, by computing their similarity based on their obtainable information (i.e., s and u's *attribute vector*, u's *friend set* and

u's friends' *attribute vector*). Specifically, we leverage SVM [12] to train a link prediction model, which describes the friend probability by a combination of multiple social features (i.e., ψ_{us}). To train the model, we generate a training data set which gathers information of a number of user pairs. Each user pair corresponds to a label z_i and multiple social features $\boldsymbol{x}_i$. Note that z_i equals 1 if two users are friends; otherwise, z_i equals 0. With the data set, we aim at training a set of parameters (w) and making the social features' parameterized combination describe the pattern of connectivity between users. In other words, with taking the social features that are parameterized by the trained w, we can compute the friend probability between u and s, and then determine whether u belongs to $\mathcal{L}$ or $\mathcal{D}$ for the new user s.

Thus, constructing the SVM-based link prediction model is addressing the optimization problem as follows:

$$\min F(w) = \tfrac{1}{2}\|w\|^2 + \lambda \textstyle\sum_{i=1}^{q} \xi_i$$
$$\text{subject to: } \begin{cases} \xi_i \geqslant 0 \\ z_i \langle w, \boldsymbol{x}_i \rangle \geqslant 1 - \xi_i \end{cases} \tag{1}$$

where q stands for the total number of the user pairs and i denote the ith pair; λ is a constant and $\xi_i (i = 1, ..., q)$ are slack variables for optimization.

IV. Exploring social features

Capturing good social features that are exploited in the learning algorithm is critical and challenging [4]. For training the model with enough features, we take various ways to extract plenty of features with limited social attributes. Particularly, we conduct this study based on a real social data set which we have crawled from Facebook [15]. We first briefly introduce the data set and then illustrate multiple captured social features. We also reveal some relations between friend probability and social features. Note that, although the social features seem tightly depending on the social attributes in Facebook, our work is easy to be extended to other social network platforms.

A. Data description

We have crawled users' public available information on Facebook from March to June 2012 and collected profile data from $479,048$ users. For each user in our data set, we have three types of information including user's demographics, interests and friendships. In this work, we think of the user's demographics and interests as social attributes. Specially, we consider ten social attributes: current city, hometown, high school, college, work, age, gender, user's favorite music, movies and TV shows.

B. Basic features

With a new user's social attributes, the most straight-forward way to predict his link is to look for some users who exhibit some common attributes with the new user. For instance, if a new user s states that he is working at TELECOM SUDPARIS, he might know others working at TELECOM SUDPARIS. Therefore, in this OSN, if there is an existing user u stating that he is working for TELECOM SUDPARIS, it is more probable that s will link to u than others. We define two types of basic features: *binary similarity* and *number of common attributes*.

Binary similarity estimates whether two users are same or not in one certain attribute. For instance, binary similarity on work of s and u equals 1 if they work in the same company or organization; otherwise, it is 0. Moreover, we sum up the binary similarity on all the attributes to obtain the *number of common attributes* as another basic feature, since two users are more likely to be friends if they share more attributes. With the Facebook data set, we study the relations between *friend probability* and the two basic features.

Figure 1(a) displays the friend probability of two users if they have same value on a certain attribute. We observe that users from the same high school and workplace might connect with each other with the highest probabilities — around 15% and 7.4% respectively. Figure 1(b) reveals the increase of the friend probability when the number of common attributes grows. The user pairs who share five common attributes merely have 3% of probability to be friends. Only 0.3% of user pairs could share six common attributes, although their friend probability reach to 17.6%. The above observations imply that only using the two kinds of basic features may still be hard to predict links correctly and inspire us to explore more social features to describe user pairs' connectivity patterns.

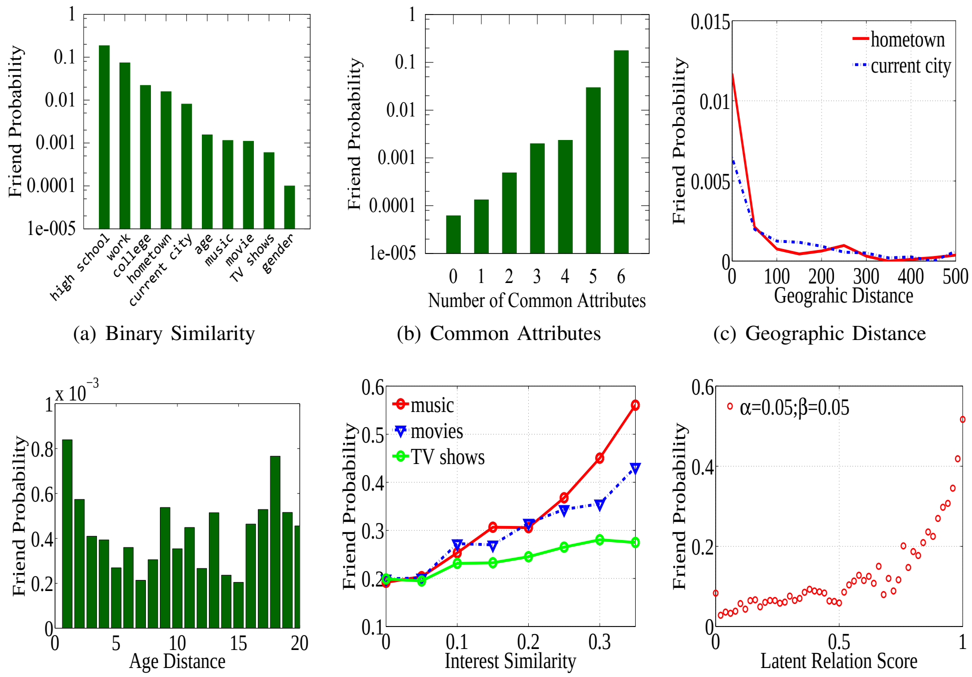
C. Derived features

In this section, we try to capture more social features from two user's attributes, which are called derived features. For different attributes, we propose three feature extraction methods, and thus get three sub-categories of derived features: *distance features*, *correlation features* and *similarity features*.

1) Attribute distance: Some existing studies indicate the homophily principle that people are more likely to link to others who are closer to them [8][11]. We attempt to use the distance function to describe the closeness in terms of the location-related attribute (e.g., current city and hometown) and age. We can calculate users' geographic distance by exploiting the location's coordinates and look further into how the distance would affect the friend probability. The absolute age difference between two users is also introduced as a distance feature.

Figure 1(c) shows the effects of geographic distance between users on the users' friend probability which holds the homophily principle both for current city and hometown. Figure 1(d) reveals that the friend probability does not correlate to the age distance. Nevertheless, the observation exhibits its rationality: people usually link to various people in different ages. For instance, a teenager may link to his parents, and a younger employee could link to an elder leader.

2) Attribute correlation: We have found that people from the same high school, workplace or college link to each other with a relatively large possibility. Besides, in reality, people from different organizations may also exhibit frequent links



(a) Binary Similarity (b) Common Attributes (c) Geographic Distance

(d) Age Distance (e) Interest Similariy (f) Latent Relation

Fig. 1: Friend probability by social features

because of the tight collaboration and relations between them. For example, TELECOM SUDPARIS as a telecommunication institute may have a very close relationship with TELECOM ORANGE LAB because of their regular project collaborations. Therefore, many of the employees from these two workplaces may know each other and establish links.

To accurately describe this connectivity pattern, we construct an *attribute correlation matrix* which learns the friend probability between users with specific value combination in one attribute (i.e., high school, workplace and college). For instance, to set up a *work correlation matrix*, both of the columns and rows represent all the workplaces that users report, and the cross-cell of ith column (representing work W_i) and jth row (representing work W_j) stands for the friend probability between the employees in work W_i and the employees in work W_j, i.e., $fp_w(i,j)$. Marking the work attribute as a_w, we get the following formula:

$$\begin{aligned} fp_w(i,j) &= P(e\langle u,v\rangle \in E \mid a_w(u)=W_i \wedge a_w(v)=W_j) \\ &= \frac{|\{e\langle u,v\rangle \in E \mid a_w(u)=W_i \wedge a_w(v)=W_j\}|}{|\{(u,v) \mid a_w(u)=W_i \wedge a_w(v)=W_j \wedge u \neq v\}|} \end{aligned}$$

The numerator in the above formula is the number of friend pairs where one's work is W_i and the other's work is W_j; the denominator is the number of all possible user pairs where two users work in W_i and W_j respectively. Back to the previous example, assume W_i is TELECOM SUDPARIS and W_j is TELECOM ORANGE LAB, then $fp_w(i,j)$ is the probability that two employees from the two institutes are friends. Besides the attribute of work, we also construct such matrices for high school and college.

Note that the friend probability study relies on an aggregation number of existing users with complete required information (i.e., friendships and value on the attribute). According to the size of population and various number of distinct attribute values (e.g., the number of workplaces reported by users), the construction of *attribute correlation matrix* may take a long time. However, it is feasible as the matrix construction can be calculated off-line, and does not need to be updated frequently.

3) Interest similarity: Cosine similarity is widely used to estimate the closeness of two vectors. Hence, for the attribute with a value of vector, like favorite music, movies and TV shows where users present multiple items, we apply the cosine similarity to describe two users' interest similarity. For the detailed description about how to calculate cosine similarity between two users' interests, readers can refer to our previous work [13]. According to the Figure 1(e), we verify that users with similar interests link to each other with high probability, which is also observed by other work [10][11][14].

D. Latent relation

Both basic and derived features are constructed by only considering the new user s and the *existing user* u's attribute vectors; besides, another kind of information is still available — u's friend set. If s and u's friends are similar, the link $e\langle s,u\rangle$ will probably be created. We call the relation between u and s captured through the relations between s and the friends of u as *latent relation*. We consider the *latent attribute relation* between s and u, which is estimated by the *latent attribute links* between s and the friends of u. Specifically, one latent attribute link is created if s has a same attribute with one of u's friends.

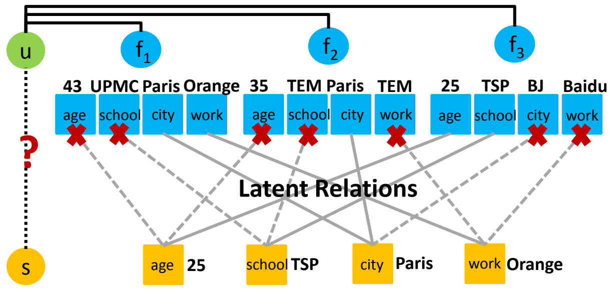


Fig. 2: Latent relations between two users

Figure 2 illustrates an example to show the latent links between a new user s and an *existing user* u, where u has three friends f_1, f_2 and f_3. We observe that s and f_1 share two attributes — work and current city — which construct two latent links between s and u. s also links to u's other friends f_2 and f_3 by various attributes. In addition, we observe many disconnections between s and u's friends on attributes, denoted by the red cross and dotted lines in the figure.

The problem then becomes how to quantify these latent links and disconnections between s and u's friends, so as to model the latent relation between s and u. Intuitively, s and u exhibit higher probability to be friends if there are more latent links and less disconnections. Therefore, we *reward* the latent relation of s and u if there is one latent link, and *punish* the latent relation if there is one disconnection. According to this idea, we estimate s and u's latent relation by $r - \alpha q$, where r equals the number of latent links, q is the number of disconnections and α is a regulator for *punish* value [16]. Accordingly, we compute a latent relation score as:

$$scr_{lr} = \frac{1}{1 + e^{-\beta(r - \alpha q)}} \tag{2}$$

where β is an exponential regulator. Figure 1(f) displays the relation between friend probability and the latent relation score when $\alpha = 0.05$, $\beta = 0.05$. It reveals that the friend probability would increase if two users exhibit more latent links and less disconnections (i.e., larger latent relation score).

Note that, when s reveals few attributes and most basic and derived features of s and u cannot be obtained, this latent relation score can be especially important for determining whether $e\langle s, u\rangle$ will exist.

V. Empirical evaluation

In this section, we evaluate our proposed approach on the crawled Facebook data set. We first introduce the experiment setup and then report the experiment results.

A. Experiment setup

Taking the prediction algorithm described in Section III-B, we leverage all the introduced social features in Section IV to train our new user link prediction model. In particular, we use 1) ***B***asic feature, i.e., the number of common attributes; 2) ***D***erived feature, including the distance of current city and hometown, the attribute correlation on work, high school and college; and 3) latent ***R***elation score. We call our proposed model as ***BDRlink*** model. Note that, ***BDRlink*** model does not exploit binary similarity on each attribute because it has already been involved in derived feature. For instance, if the distance of current city between two users equals 0, it indicates they are in the same city (i.e., binary similarity is 1); otherwise, binary similarity is 0.

We compare ***BDRlink*** model with three baselines — ***Blink*** model, ***Dlink*** model, and ***BDlink*** model:

- ***Blink*** model merely considers basic social feature which includes the number of common attribute and binary similarity on all attributes (i.e., current city, hometown, high school, college, work, age, gender, user's favorite music, movies and TV shows).
- ***Dlink*** model merely considers derived social feature which includes current city distance, hometown distance, age distance, high school correlation, college correlation, work correlation, music similarity, movies similarity and TV shows similarity.
- ***BDlink*** model takes into account the number of common attribute (basic feature) and all the derived features that are used to train ***Dlink***.

Note that these models are trained with users who reveal friends and more than 3 attributes. We randomly couple two users into a user pair and select one of the two users as the new user by removing his friends.

B. Evaluation results

We evaluate the proposed ***BDRlink*** model from three perspectives: 1) we compare the prediction performance of ***BDRlink***, ***Blink***, ***Dlink*** and ***BDlink*** models in terms of ROC curves with 10-fold cross validation; 2) we further carry out 'leave-one-feature-out' model comparison to investigate the influence of various social features on link prediction; 3) we evaluate the prediction performance of ***BDRlink*** by the number of available attributes from the new user, so as to inspect and verify whether the new user can derive better friends prediction if they provide more information.

1) Prediction performance comparison: We draw the ROC curves of four prediction models, shown in Figure 3. We also note the corresponding Area Under Curves (AUCs) in the legend. First of all, compared to the diagonal line (i.e., AUC= 50%) which represents the performance of random guess, all the four models with our captured social features can predict more accurately. We notice that ***Blink*** model and ***Dlink*** model exhibit equal prediction capacity as they almost achieve a same AUC of 68.8%. Additionally, the combination of basic features and derived features can slightly enlarge the AUC from 68.8% to 71%. Among the four compared model, ***BDRlink*** model generates the largest AUC and its AUC significantly outperforms the other three models by 14%, 14% and 12% respectively. It reveals that the attribute based latent relations between users not only works for the link prediction but also plays a very important role in the link prediction.

2) Leave-one-feature-out: To investigate whether the social features leveraged in ***BDRlink*** model would improve the prediction performance or not, we leverage the state-of-the-art

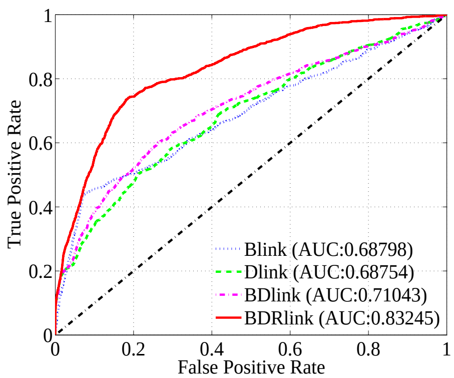


Fig. 3: ROC curves comparison

'leave-one-feature-out' strategy and remove one feature from the overall features to train additional models. Specifically, Five feature types — basic feature, distance feature, attribute correlation feature, interest similarity feature and latent relation score — are considered. Thus, we train five 'leave-one-feature-out' prediction models by taking out one of the five types of features, namely *No basic feature* model, *No distance feature* model, *No attribute correlation feature* model, *No interest similarity* model and *No latent relation score* model.

Table I compares the AUCs of the five 'leave-one-feature-out' models and the ***BDRlink*** model. We observe that ***BDRlink*** model outperforms all the other models which means removing any of the used social features would decrease its prediction power. In addition, comparing the five 'leave-one-feature-out' models, we find that various social features impact the prediction performance in different degrees. For instance, removing basic features or interest similarity, the prediction performance does not fall down much; whereas latent relation score is quite sensitive to the prediction as the performance decreases obviously when it is removed.

Type of Model	**AUC**
No basic feature	0.8139
No distance feature	0.7563
No attribute correlation feature	0.7863
No interest similarity	0.8107
No latent relation score	0.7104
BDRlink	**0.8325**

TABLE I: Leave-one-feature-out comparison

3) AUC by number of available attribute: In this experiment, we aim to validate whether ***BDRlink*** can predict links more accurately if new users provide more attributes. We group the user pairs according to the number of attributes obtained from new users and test the prediction performance of ***BDRlink*** for each group in terms of AUC. Table II lists the AUC values by various attributes numbers. The results reveal that the prediction accuracy would increase if new users provide more attributes.

VI. Conclusion

This paper proposes a novel method to predict links for new users in OSNs. It leverages the attributes from new users provided at the registration phase and the profile information (attributes and links) from existing users to generate a number of effective social features. The correlation between the friend probability and these social features is investigated to select effective features for training a SVM-based link prediction model — ***BDRlink***. The empirical experiments show that the ***BDRlink*** model performs better than the other three baseline models. The leave-one-feature-out test reveals that each of the proposed social features contributes significantly to the prediction model.

#Attributes	3	4	5	6	7	8	9
AUC	0.66	0.70	0.72	0.73	0.75	0.83	0.87

TABLE II: AUC by varying number of available attributes

Acknowledgment

This work is funded by the European Union under the project SITAC (ITEA2-11020) and the project eCOUSIN (EU-FP7-318398). This works is also partially supported by the European Celtic-Plus project CONVINcE.

References

[1] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *WWW*, 2010, pp. 591–600.
[2] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *SDM06: Workshop on Link Analysis, Counter-terrorism and Security*, 2006.
[3] A. K. Menon and C. Elkan, "Link prediction via matrix factorization," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2011, pp. 437–452.
[4] L. Backstrom and J. Leskovec, "Supervised random walks: predicting and recommending links in social networks," in *WSDM*. ACM, 2011, pp. 635–644.
[5] J. Zhang, X. Kong, and P. S. Yu, "Predicting social links for new users across aligned heterogeneous social networks," in *ICDM*, 2013, pp. 1289–1294.
[6] M. Yan, J. Sang, T. Mei, and C. Xu, "Friend transfer: cold-start friend recommendation with cross-platform transfer learning of social knowledge," in *ICME*, 2013, pp. 1–6.
[7] V. Leroy, B. B. Cambazoglu, and F. Bonchi, "Cold start link prediction," in *KDD*, 2010, pp. 393–402.
[8] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: improving geographical prediction with social and spatial proximity," in *WWW*, 2010, pp. 61–70.
[9] Y. Zheng, L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, "Recommending friends and locations based on individual location history," *ACM Transactions on the Web (TWEB)*, vol. 5, no. 1, p. 5, 2011.
[10] L. M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer, "Friendship prediction and homophily in social media," *TWEB*, vol. 6, no. 2, p. 9, 2012.
[11] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social networks*, vol. 25, no. 3, pp. 211–230, 2003.
[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[13] X. Han, L. Wang, S. Park, A. Cuevas, and N. Crespi, "Alike people, alike interests? a large-scale study on interest similarity in social networks," in *ASONAM*, 2014.
[14] D. Caragea, V. Bahirwani, W. Aljandal, and W. H. Hsu, "Ontology-based link prediction in the livejournal social network." in *SARA*, vol. 9, 2009.
[15] R. Farahbakhsh, X. Han, A. Cuevas, and N. Crespi, "Analysis of publicly disclosed information in facebook profiles," in *ASONAM*, 2013, pp. 699–705.
[16] J. McAuley and J. Leskovec, "Discovering social circles in ego networks," *TKDD*, vol. 8, no. 1, pp. 4:1–4:28, 2014.