Use of Non-Topological Node Attribute Values for Probabilistic Determination of Link Formation

Abhiram Gandhe

Computer Science and Engg.
Visvesvaraya National Institute of Technology
Nagpur, India

Parag Deshpande

Computer Science and Engg.
Visvesvaraya National Institute of Technology
Nagpur, India

Abstract—Here we propose a probabilistic model for determining link formation, using Naïve Bayes Classifier on non-topological attribute values of nodes, in a social network. The proposed model gives a score which helps to determine the relationship strength in a non-formed link. In addition to Naïve Bayes Classifier, weighted Average of the Attribute value match helps to determine the friendship score of a non-formed link.

With the increase in online social networks and its influence on people, more and more individuals are getting wider and enhanced social connect. Everyone tries to connect more to explore more. In this race of more, an individual needs better and definitive tools to help them grow their network. Wider is the network more is the possibility to explore.

Here we present a novel approach for predicting a link (friendship) between two individuals (nodes) in a social network. The proposed approach uses non-topological attribute data values of both the nodes and predicts linkage possibility by applying Naïve Bayes Classifier on non-topological attribute data values of nodes in existing linkages.

A linkage possibility is expressed using one quantitative measure FSCORE. We call it friendship score (FSCORE) between two unconnected individuals. FSCORE is used to predict linkage between two nodes. Higher FSCORE means a higher possibility of linkage between two nodes.

Keywords—*Non-Topological Attribute; Link Prediction; Naïve Bayes Classifier; Weighted Average; Graph Database; Social Network; Data Mining*

I. INTRODUCTION

Online Social Networks (OSNs) have become an integral part of today's life. OSN is where everyone keeps her/his social connect. Social networks are still doing the same work like information exchange, furthering a cause, keeping up communication, guiding and developing a society, to name a few. More connected is an individual; more he can achieve out of his social connects.

In the early days greeting and meeting people in social gathering was the only way to increase your social network and influence. Today, with the acceptance and spread of online social networks, ways to connect individuals have significantly improved. Different Online Social Networks addresses different interests of an individual. Facebook and Google+ mainly exist to share information, initiate a conversation and discuss on a certain topic. Twitter a micro-blogging site helps

commenting on any issue at hand/in mind and letting the world know about it. Flickr is a photo sharing social network. LinkedIn is an online network of professionals.

With every online Social Network site, there is a new and different social network, of people in a context, created by an individual. With the increasing options and different focus areas of different networks, the data created out of these networks is diverse and huge. This data provides a great opportunity for analysts to dig the created data and interpret the future course of the network.

While there are inherent risks in use and distribution of OSNs data, there are also many potential benefits of this data. Interpretation of Social networking data along with related tools created to interpret the data can help to strengthen existing relationships and provide opportunities for creating new relationships. With better means and tools, a stronger and more connected network can be intended and created. Today networks are deploying different techniques to help a user to grow their social circle, connect with new individuals and find superior content of interest.

Every individual is on the lookout to increase his network with people of interest. As faced by every individual, there are two impediments in connecting with individuals of interest.

1) Who is the one of interest?
2) How high is the possibility of connecting with the one of interest?

Different Social Networks are engaging different ways to augment a user's arsenal to help them grow their own network. Most common ways of predicting a higher probability of connecting in a network are:

a) Individuals with maximum mutual friends are suggested a connect

b) Individuals are asked to suggest a connect between their unconnected friends

c) Unconnected individuals having multiple short length paths in the graph are suggested a connect

d) Unconnected individuals commenting on the same conversation, multiple times are suggested a connect

New and better tools are evolving at day end to provide users with better services to enhance their experience of social connect. There is a wide range of research going on in the area of suggesting connects. In research terminology, it is called as Link Prediction in Graphs. Link prediction can be used to

identify hidden links, not yet formed in an Online Social Network, in a friend suggestion mechanism.

Link prediction outside the social network domain can have multiple uses like:

a) Recommendation and relevance prediction in e-commerce [3]

b) Protein Interaction prediction in Life Sciences [2]

c) Identifying hidden groups of terrorist or criminals using link prediction in the security domain [4]

The link prediction problem is relevant to different scenarios; several algorithms have been proposed in recent years to solve it. One common approach for solving **Link prediction** problem is using **supervised learning** algorithm. This approach was introduced by Liben-Nowell and Kleinberg in 2003 [6], who studied the usefulness of graph topological features by testing them on bibliographic data sets. In 2006, the work was extended to identifying hidden group of terrorist by Hasan et al [4] and since then several other researchers have implemented this approach. Most of the solutions, that these researchers proposed were tested on bibliographic or onco-authorship data sets [4], [6], [7], and [8]. In 2009, Chen et al [1] depicted several algorithms used by IBM on their internal social network, which enable its employees to connect with each other. Song et al. used matrix factorization to estimate the similarity between nodes in real life social networks such as Facebook and MySpace [9]. In 2011, W J. Cukierski et al [10] extracted 94 distinct graph features. Using a Random Forests classifier, they achieved impressive results in predicting links on Flickr datasets.

Here we are proposing a novel approach for predicting a link (friendship) between two individuals (nodes) in a social network using OSN (Online Social Network) data and predict linkage possibility by applying Naïve Bayes Classifier on attribute data values of nodes in existing linkages.

II. PROBLEM STATEMENT

Classification of links in social network can be done on different types of node data:

a) Topological Attribute Data

b) Node Interaction Data

c) Non-Topological Attribute Data

All the above types of node data can be used to classify the links. The classification helps in predicting the possibility of connection between two non-connected nodes. Most of the research to date is done on Topological Attribute data and Node interaction Data.

In this paper, we propose the **mechanism of classification based on non-**topological **attribute data**. The dataset used for experimentation will be from Facebook™. We will be using Naïve Bayes classifier for classifying the existing links and use the classification for predicting a link between two nodes.

III. METHODOLOGY

A. Online Social network

The Social Network in consideration, in this paper, is Facebook. Facebook is an online social interaction and networking service. A user above 13 years of age can create an account on Facebook. On Facebook, a user can make friends with other Facebook users. A user can post anything on her/his wall (representation of profile space) or her/his friend's wall. A user can "Like" or "comment" on posts by her/him or her/his friends. The posts, "Like" action and comments can be termed as public interaction between users. All the public interactions between users, done by a user, are available for view to all users on the timeline of the user.

Other than public interaction a user can have private interaction with a friend user. The possible ways of private communication is chatting or inbox messaging. All the private communications are confidential and are visible only to the two users between whom the interaction has taken place. These public and private interactions between users are termed as node interaction data and can be used to predict friendship.

A user stores his profile information as the time of registration with Facebook. Profile information on Facebook can range from First Name, Last Name, Date of Birth, Gender, Religion, Home Town, Current City and Relationship Status to Work, Work History and Education History information. Over the period of time, a user can update, add or delete profile information. A user can also put restrictions on visibility of this information from "Public" to "Friends Only" to "Only Me" to any other specific friends group available. Due to the selective visibility of data governed by the user and nearly all the attributes are optional there is a wide possibility of having attribute values as blank.

B. DatasetPreparation

A sample subset of Facebook was used as a dataset to work on. This Dataset was extracted from Facebook using Facebook App named "FBNetworkAnalysis". URL for this app is "https://app.facebook.com/mytestappfbabhi". The data was collected from 7637 users. In the context of this analysis, users are represented as nodes and friends are represented as two nodes on an edge. A friendship is represented as a link between two users. A user can mark a link as "Friend", "Cousin" (any other relative) or "Spouse"/"Significant Other". This Relationship is taken as name/type of the link. Link name/type is not considered in this analysis.

All the profile information made available by the user are considered as the node attributes and the analysis of the links is done using the values of the node attributes of the users (nodes) in a link (edge).

The values extracted using the Facebook application:

Date of Birth	**Gender**	Religion
Home Town	**Current City**	Relationship Status
Interested In	College/**School**	Education Year
Work Company	Work Location	Work Year
Favorite Athlete	Favorite Team	

C. Data Representation

fbgraph (finite, no multiedges, undirected)

Facebook sample data set will be represented as a graph with finite nodes and a finite number of connections. A

connection (Link) can only be established when friend request is sent by one user (Node) and accepted by another user (Node).Mutual acceptance by both the **nodes** makes the **link undirected**. There is only one link between two nodes, the link type can differ on the nodes in consideration ("Friend", "Relative" or "Spouse") so there will be no multi-edges.

fbG = (U, L) where

U (or U (fbG)) is a set of nodes

L (or L (fbG)) is a set of links, each of which is a set of two nodes (undirected)

Two nodes that are associated with a link are adjacent nodes.

Let   n = |U| and m = |L|

The neighbor of each node u is

N (u) = {v |uvƐ L}

The degree of user u is d (u) = |N (u)|

D. *Naïve Bayes Classifier*

Naïve Bayes classifier depends on Bayes theorem

$$p(cj|d) = \frac{p\left(\frac{d}{cj}\right)p(cj)}{p(d)}$$

Where,

$p(cj|d)$: Probability of instance **d** being in class **cj**
(This is what we will be computing)

$p(d/cj)$: Probability of generating instance **d** given class **cj**
*(We can imagine that being in class **cj**, causes you to have feature **d** with some probability)*

$p(cj)$: Probability of occurrence of class **cj**
*(This is just how frequent the class **cj**, is in our database)*

$p(d)$: Probability of instance **d** occurring
*(This is just how frequent the instance **d**, is in our database)*

And

$$p(cj/d) = \frac{p(cj \cap d)}{p(d)}$$

Where,

$p(cj|d)$: Probability of instance **d** being in class **cj**

$p(cj \cap d)$: Existing Links having instance **d** in class **cj**

$p(d)$: Probability of instance **d** occurring

To simplify the task, **naïve Bayesian classifiers** assume attributes have independent distributions, and thereby estimate:

$$p(d|cj) = p(d1|cj) \times p(d2|cj) \times ....\times p(dn|cj)$$

d: instance d

d1: Value of feature 1

d2: value of Feature 2

so on and so forth …

Advantages of Naïve Bayes Classifier:

- ✓ Fast to train (single scan). Fast to classify
- ✓ Not sensitive to irrelevant features

- ✓ Handles real and discrete data
- ✓ Handles streaming data well

Disadvantages of Naïve Bayes Classifier:

- ➢ Assumes independence of features

E. *Data Analysis*

Function	Equation	Number
Number of Users		7637
Possible Friend Connections		29158066
Existing Friend Connections	P(Friend)	101904
Non Friend Pairs	P(Non Friend)	29056162
Possible Friends with Same Gender	P(Same Gender)	12682300
Friends of Same Gender	P(Same Gender ∩ Friend)	60604
Non Friends of Same Gender	P(Same Gender ∩ Non Friend)	12621696
Possible Friends with Different Gender	P(Different Gender)	9094400
Friends of Different Gender	P(Different Gender ∩ Friend)	35379
Non Friends of Different Gender	P(Different Gender ∩ Non Friend)	9059021
Possible Friends with Same Location	P(Same Location)	2524621
Friends of Same Location	P(Same Location ∩ Friend)	27013
Non Friends of Same Location	P(Same Location ∩ Non Friend)	2497608
Possible Friends with Different Location	P(Different Location)	9917945
Friends of Different Location	P(Different Location ∩ Friend)	36839
Non Friends of Different Location	P(Different Location ∩ Non Friend)	9881106
Possible Friends with Same School	P(Same School)	726648
Friends with Same School	P(Same School ∩ Friend)	17117
Non Friends with Same School	P(Same School ∩ Non Friend)	709531
Possible Friends with Different School	P(Different School)	11388855
Friends with Different School	P(Different School ∩ Friends)	36977
Non Friends with Different School	P(Different School ∩ Non Friend)	11351878
Possible Friends with Same Favorite Athlete	P(Same Favorite Athlete)	866676
Friends with Same Favorite Athlete	P(Same Favorite Athlete ∩ Friend)	4803
Non Friends with Same Favorite Athlete	P(Same Favorite Athlete ∩ Non Friend)	861873
Possible Friends with Different Favorite Athlete	P(Different Favorite Athlete)	1211065
Friends with Different Favorite Athlete	P(Different Favorite Athlete ∩ Friend)	7836
Non Friends with Different Favorite Athlete	P(Different Favorite Athlete ∩ Non Friend)	1203229

F. *Conditional Probability Bayes Rule*

Using Conditional Probability Bayes Rule on the above data:

$$p(cj/d) = \frac{p(cj \cap d)}{p(d)}$$

Same Gender:

$$P\left(\frac{Friend}{Same\ Gender}\right) = \frac{P(Friend \cap Same\ Gender)}{P(Same\ Gender)}$$

$$= \frac{60604}{12682300} = 0.0047786$$

Different Gender:

$$P\left(\frac{Friend}{Different\ Gender}\right)$$
$$= \frac{P(Friend \cap Different\ Gender)}{P(Different\ Gender)}$$

$$= \frac{35379}{9094400} = 0.0038902$$

Same Location:

$$P\left(\frac{Friend}{Same\ Location}\right) = \frac{P(Friend \cap Same\ Location)}{P(Same\ Location)}$$

$$= \frac{27013}{2524621} = 0.0106998$$

Different location:

$$P\left(\frac{Friend}{Different\ Location}\right)$$
$$= \frac{P(Friend \cap Different\ Location)}{P(Different\ Location)}$$

$$= \frac{36839}{9917945} = 0.0037144$$

Same School:

$$P\left(\frac{Friend}{Same\ School}\right) = \frac{P(Friend \cap Same\ School)}{P(Same\ School)}$$

$$= \frac{17117}{726648} = 0.0235561$$

Different School:

$$P\left(\frac{Friend}{Different\ School}\right)$$
$$= \frac{P(Friend \cap Different\ School)}{P(Different\ School)}$$

$$= \frac{36977}{11388855} = 0.0032468$$

Same Favorite Athlete:

$$P\left(\frac{Friend}{Same\ Favorite\ Athlete}\right)$$
$$= \frac{P(Friend \cap Same\ Favorite\ Athlete)}{P(Same\ Favorite\ Athlete)}$$

$$= \frac{4803}{866676} = 0.0055419$$

Different Favorite Athlete:

$$P\left(\frac{Friend}{Different\ Favorite\ Athlete}\right)$$
$$= \frac{P(Friend \cap Different\ Favorite\ Athlete)}{P(Different\ Favorite\ Athlete)}$$

$$= \frac{7836}{1211065} = 0.0064703$$

G. Attribute Value Weightage

Calculated conditional probability above shows that two individuals in a same school have a higher probability of being friends than being from the same location.

Putting this probability as weight for the attribute values of two nodes, following are the weights with the maximum weight attribute at the top.

Attribute Value Weights:

W (Same School)	0.0235561
W (Same Location)	0.0106998
W (Different Favorite Athlete)	0.0064703
W (Same Favorite Athlete)	0.0055419
W (Same Gender)	0.0047786
W (Different Gender)	0.0038902
W (Different Location)	0.0037144
W (Different School)	0.0032468

If two individuals have same school populated in the School attribute the weight of having a friendship will be **0.0235561** if school populated in the School attribute for both the individuals is different, then weight of having a friendship will be **0.0032468** instead.

If any of the individuals does not have the school attribute populated/shared, then no weight is added for school attribute in the friendship score.

H. Friendship Score (FSCORE)

If a Link is present between two nodes, a and b, then:

$$FSCORE[a, b] = 1 + FNT(a, b)$$

Where FNT (a,b) is a function non-topological attribute of a and b. If a link is not formed between two nodes, a and b, then friendship score needs to be calculated using non-topological attribute data. The FSCORE is calculated as:

$$FSCORE[a, b] = FNT(a, b)$$

If two individuals (Non Friends) have **only Same Gender** and no other attributes populated, the probability of Friendship is 0.0047786 and hence the Friendship Score for future friendship is **0.0047786.** Similarly, if two individuals **only** have **Same Location** then their Friendship Score will be **0.0106998**.

IV. MATHEMATICAL MODEL FOR DATASET WITH MISSING DATA

FSCORE or Attribute weight calculated here is with the data where there are missing attribute values in many node elements. Also in the case of the social data in consideration, all features/attributes cannot be assumed to be independent of each other. Considering features are dependent on each other

Naïve Bayesian distributed probability equation cannot be used here. Instead, we propose the use of the weighted average.

$$FNT(a, b) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} Wij \times Vij}{n}$$

Where:

i: Attributes/Features e.g. Gender (if there is no value associated with the attributes in any/both of the nodes **a** or/and **b** then that attribute will not be taken up for the calculation of FSCORE)

j: Different Attribute Values possible e.g. Same Gender, Different Gender (No gender available is also a valid possible value, but it will already be excluded from the equation because of the elimination of attributes while collating the final set of i's)

According to the above, FSCORE for two nodes with Same Gender (SG), Same Location (SL), Same School (SS) and Same Favorite Athlete (SFA), and no other attribute value populated, in a network where there are Lots of nodes with missing attribute values is as follows. For two unconnected nodes:

$$fscore(SG, SL, SS, SFA) = FNT(SG, SL, SS, SFA)$$

$$= \frac{\begin{matrix} WSG \times VSG + WDG \times VDG + \\ WSL \times VSL + WDL \times VDL + \\ WSS \times VSS + WDS \times VDS + \\ WSFA \times VSFA + WDFA \times VDFA \end{matrix}}{4}$$

$$= \frac{\begin{matrix} WSG \times 1 + WDG \times 0 + WSL \times 1 + WDL \times 0 \\ + WSS \times 1 + WDS \times 0 + WSFA \times 1 + WDFA \times 0 \end{matrix}}{4}$$

$$= \frac{WSG + WSL + WSS + WSFA}{4}$$

$$= \frac{0.0235561 + 0.0106998 + 0.0055419 + 0.0047786}{4}$$

$$= \frac{0.0445764}{4} = 0.0111441 \tag{1}$$

Here,

WSG: Weightage for Same Gender (weight to be added to FSCORE if the two nodes under consideration have the same gender)

Similarly WDG, WSL, WDL, WSS, WDS, $WSFA$, $WDFA$ are weights for Different Gender, Same Location, Different Location, Same School, Different School, Same Favorite Athlete and Different Favorite Athlete, respectively.

VSG: Value in gender attribute for both the nodes is same. $VSG = 1$ if both the nodes under consideration have a same gender and $VSG = 0$ if gender is different for both nodes or any of the node doesn't have gender value

Similarly VDG, VSL, VDL, VSS, VDS, $VSFA$, $VDFA$ are values for Different Gender, Same Location, Different Location, Same School, Different School, Same Favorite Athlete and Different Favorite Athlete, respectively.

V. MATHEMATICAL MODEL FOR DATASET WITH NO MISSING DATA

When no nodes are missing attribute data, every attribute value matches or does not match between two nodes. In such cases calculating FSCORE can be done differently

In the Training set, number of friends with:

Same Gender + Same Location: **18154**
Same Gender + Different Location: **21524**
Different Gender + Same Location: **8379**
Different Gender + Different Location: **15240**

Friendship Score for Same Gender and Same Location:

$$fscore(SG, SL) = \frac{18154}{(18154 + 21524 + 8379 + 15240)}$$

= **0.2868066**

FSCORE for two unconnected nodes with Same Gender (SG), Same Location (SL), Same School (SS) and Same Favorite Athlete (SFA), in a network where there are no nodes with missing attribute values is as follows.

In the Training set, Number of Friends with:

SG + SL + SS + SFA = **381**
SG + SL + SS + DFA = **572**
SG + SL + DS + SFA = **612**
SG + SL + DS + DFA = **964**
SG + DL + SS + SFA = **423**
SG + DL + SS + DFA = **237**
SG + DL + DS + SFA = **447**
SG + DL + DS + DFA = **524**
DG + SL + SS + SFA = **144**
DG + SL + SS + DFA = **208**
DG + SL + DS + SFA = **212**
DG + SL + DS + DFA = **1102**
DG + DL + SS + SFA = **68**
DG + DL + SS + DFA = **83**
DG + DL + DS + SFA = **208**
DG + DL + DS + DFA = **310**

$$fscore(SG, SL, SS, SFA) = FNT(SG, SL, SS, SFA)$$

$$= \frac{381}{\begin{matrix}(381 + 572 + 612 + 964 + 423 \\ + 237 + 447 + 524 + 144 + 208 \\ + 212 + 1102 + 68 + 83 + 208 + 310)\end{matrix}}$$

= **0.0586605** -(2)

The fscore(SG, SL, SS, SFA) in 1 is different than the one in 2 due to the difference in dataset. Dataset used in equation 2 is the subset (nodes with no missing values for gender, location, school and favorite athlete attributes) of the one used for equation 1.

In the case, if complete data is available, number of permutation combination to store and update, on link formation, increases with the increase in the attributes in consideration. This becomes cumbersome to maintain and update the data of all the combinations of attributes. In the case

of n attributes, 2^n combinations need to be maintained. For excluding some of the attributes from the final set of test, attributes may lead to maintaining different combinations separately.

What we propose for this is an equation of approximation.

$$p(A \cap B \cap C) = p(A|B)\, p(B|C)\, p(C)$$

On the same data set used in equation 2:

$$p(SG \cap SL \cap SS \cap SFA)$$

$$= p(SG|SL)\, p(SL|SS)\, p(SS|SFA)\, p(SFA)$$

$$= \frac{2529}{4195} \times \frac{1305}{2116} \times \frac{1016}{2495} \times \frac{2495}{6495}$$

$$= 0.6028605 \times 0.6167297 \times 0.4072144 \times 0.3841416$$

$$= \mathbf{0.0581603} \qquad\qquad\qquad\text{-(3)}$$

We consider a relation of two attributes and use the relations in a link to the next attribute in consideration. What we have done in above example is have a probability of Same Gender for Same Location with Probability of Same Location for Same School and Probability of Same School for Same Favorite Athlete along with Probability of a link having Same Favorite Athlete.

This is done on the data set which has complete data and no Missing Values. The nearness of the FSCORE in equation 2 and 3 confirms approximation works well with proposed formula.

VI. CONCLUSIONS AND RECOMMENDATION

A relationship is made on different parameters and we have tried to quantify the parameters for relationship building, depending on an existing link/relationship data as stated in the paper. Deriving a possibility of a relationship (FSCORE) can be analyzed using the proposed model in this paper.

FSCORE is an effective way of predicting the possibility of relationship/link between two nodes using Non-topological attribute values of nodes. Significance and weight, of non-topological attributes, is determined by the already existing links and recurrence of a value pattern for these non-topological attributes in existing links.

FSCORE can be used to calculate the cost of connecting to a distant node in a graph. FSCORE can provide a measure of strength between two unconnected nodes in order to make decisions or predictions in a different set of problems in a graph network. FSCORE can also provide a factor to help identify/quantify connected nodes. FSCORE can be used to compare and rate a relation of connected or unconnected nodes stronger or weaker to other relations.

In a graph network, if a link of reference is to be invoked or an optimized path for traversal has to be identified, then FSCORE can provide a quantitative value for analysis between two connected or unconnected node. FSCORE can be used as a relationship cost parameter in similar Graph Network problems. FSCORE is calculated using non-topological attribute values between nodes and can be coupled with topological attribute data to improve the prediction possibility.

REFERENCES

[1] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, "Make new friends, but keep the old: recommending people on social networking sites," in Proceedings of the 27th international conference on Human factors in computing systems, ser. CHI '09. NewYork, NY, USA: ACM, 2009, pp. 201–210. (*references*)

[Online]. Available:http://doi.acm.org/10.1145/1518701.1518735

[2] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic block models for relational data with application to protein-protein interactions," Proceedings of International BiometricSociety-ENAR Annual Meetings, 2006.

[3] Z. Huang, X. Li, and H. Chen, "Link prediction approach to collaborative filtering," Proceedings of the 5th ACM/IEEE-CS joint conference onDigital libraries, 2005.

[4] M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," SDM Workshop on Link Analysis, Counterterrorism and Security, 2006.

[5] M. A. Hasan and M. J. Zaki, Social Network Data Analytics, C. C.Aggarwal, Ed. Springer, 2011.

[6] D. Liben-Nowell and J. Kleinber, "The link-prediction problem for social networks," Journal of the American Society for Information Science andTechnology, vol. 58, no. 7, 2007.

[7] J. R. Doppa, J. Yu, P. Tadepalli, and L. Getoor, "Chance-constrained programs for link prediction," In Proceedings of Workshop on AnalyzingNetworks and Learning with Graphs at NIPS Conference, 2009.

[8] H. R. Sa and R. B. C. Prudencio, "Supervised learning for link prediction in weighted networks," III International Workshop on Web and TextIntelligence, 2010.

[9] H. H. Song, T. W. Cho, V. Dave, Y. Zhang, and L. Qiu,"Scalable proximity estimation and link prediction in online social networks," in Proceedings of the 9th ACM SIGCOMMconference on Internet measurement conference, ser. IMC '09. NewYork, NY, USA: ACM, 2009, pp. 322–335.

[10] [Online]. Available:http://doi.acm.org/10.1145/1644893.1644932

[11] W. J. Cukierski, B. Hamner, and B. Yang, "Graph-based features for supervised link prediction," International Joint Conference on NeuralNetworks, 2011.