INTERNATIONAL JOURNAL FOR TRENDS IN ENGINEERING & TECHNOLOGY
VOLUME 3 ISSUE 1 -JANUARY 2015 - ISSN: 2349 - 9303

Data mining Framework for Finding Friends in Online Social Network Using Minimum Information

R.Rajkumar [1]
[1] Bishop Heber College, Trichy,
rajkumarbdu@gmail.com

Dr.Anbuselvi [2]
[2]Bishop Heber College, Trichy,
r.anbuselvi@yahoo.co.in

Abstract— Online Social Network such as Face book, Twitter, LinkedIn e.tc, have become the preferred interaction, entertainment and socializing facility on the internet. However With the emergence of numerous social media sites, individuals, with their limited time, often face a dilemma of choosing a few sites over others. Users prefer more engaging sites, where they can find familiar faces such as friends, relatives, or colleagues. Link predictions method help find friends using link or content information. Unfortunately, whenever users join any site, they have no friends or any content generated. In this case, sites have no chance other than recommending random influential users to individuals hoping that users by befriending them create sufficient information for link prediction techniques to recommend meaningful friends. In this paper, we discussed find friends on a new a social media site when link or content information is unavailable. The purpose of this research paper is highlighting social forces in Online Social Network using minimum information with their latest solutions by using data mining techniques are dealt in this paper elaborately.

Keywords— Confounding, Compatibility, Homophily, Incompatibility, Influence, Social media, Social forces.

1 INTRODUCTION

NOWADAYS an average user spends less than a minute on an average site. The problem becomes more challenging for commercial sites, especially for new sites that are desperately hoping to attract users and keeping them active. This lack of interest in users was clearly observed in the early years of social media sites such as Twitter or Face book with around 60% of their users quitting within the first month [1]. As consumers of social media, we are constantly seeking "sticky" sites that keep our attentions glued to the site by providing engaging material and more importantly, showing us a familiar face. The existence of familiar faces such as our friends, relatives, and our colleagues on one site, provides a sense of comfort, piques our interest on the site, and increases the likelihood of joining it. By finding friends of individuals on social media sites, not only we increase users' engagement, but also improve user retention rates for sites, which could directly translate to more revenue for the social media site. Often, link or content information or a combination of both is used to predict and recommend friends to users. When using link information, we use the current friends of an individual to recommend new friends. For instance, we find potential friends by finding individuals that are friend-of-a-friend. That is finding individuals that are two hops away in the friendship network. We can improve recommendations by recommending individuals that are more than two hops away in the friendship network. Unfortunately, recommending friends using link information fails when prior friends are unavailable. This can happen right after a user joins a new site, as a disconnected singleton in the friendship graph. Sites such as Twitter or LinkedIn, tackle this issue by asking users to provide access to their email contacts to help recommend friends. Aside from its security and privacy concerns, this clearly requires an extra effort from the user's side, and with the short attention span of a user, provides an opportunity for the user to abandon the social media site. When using content information, friend recommendation techniques identify friends of an individual by identifying users that are highly similar to the individual in terms of the content that they generate. This content can be the profile information provided, the tweets, reviews, or blogs posted, or the products bought. However, right after a user joins a new site, he or she hasn't had the chance to complete their profile information or exhibit any activity on the site. In a sense, finding friends when no link or content information is available is a ubiquitous problem inherent to all social media sites and for each and every user, right after she joins the site. However, the approach to solving it often assumes that either link or content information is available. However, as we mentioned, when a user joins a new site, no link information or content information is available, therefore, relying on either type of information may not be feasible. In practice, sites such as Twitter address this problem by recommending individuals that have many friends such as celebrities newly joined users. Recommending friends uniformly at random from this space is extremely unlikely to find any friends. Using social forces that form friendships, we demonstrate how one can employ minimum information from individuals to significantly reduce the set of potential friends in a social media site; hence, increasing the likelihood of finding friends.

2 PROBLEM STATEMENTS

Consider a new site S with n users. When an individual joins S with no content or link information, the site has probability p = 1/n to correctly recommend a single friend and a search space of n to search for that friend. If the user has no friends on S, no method is capable of finding any friends and all attempts to recommend potential friends from S fails. However, given the enormous size of current social media sites such as Twitter and Face book, one can safely assume that the individual has some friends on the site. Let set U = {u1, u2...un} represent the set of current users on site S and un+1, the newly joined user.

2.1 Compatibility

The relationship of compatibility can algebraically be expressed as follows.

Compatible: Let X= {Set of all Users}

Let Y= {Set of all friends}

Compatible is a Function f between X and Y, f: X->Y, in such a way that:

Then ∀x in X there need not be an element y in Y such that y=f (x).

[All Users need not be friends.]

For every y there need be an x in X such that

$x=f^{-1}(y)$.

[All friends need not be a User.]

2.2 Incompatibility

Examples: 'Americans', 'Indians', 'Koreans'

Incompatible: Let X be a set of all disjoint subsets.

X = {A, B, C…}

Let Φ and Γ. Then Φ and Γ are incompatible if $\alpha \in \Gamma \Rightarrow \alpha \notin \Phi$.

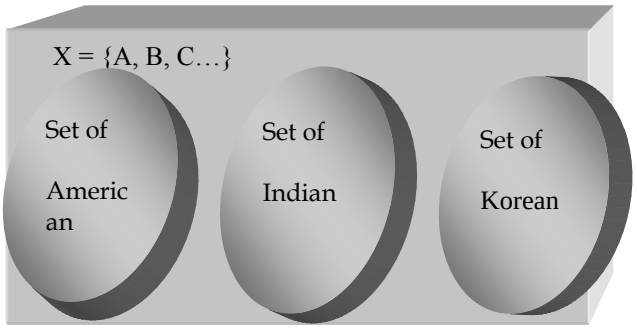

Figure-2.2.1. disjoint sub-sets

3 SOCIAL FORCES BEHIND FRIENDSHIPS

Normally, three major social forces result in friendships are homophily, confounding, and influence. Homophily is observed when similar individuals become friends. This similarity is often observed in terms of the interests of the individuals. Or their personal attributes that are unrelated to the environment they live in [4]. For instance, fans of the same movie director becoming friends are an example of friendships formed by homophily. Confounding is exhibited when friendships are formed due to the similarities in users formed by the environment they live in.

Due to confounding friends are often in close proximity or speak the same language. Finally, influence in friendships is observed when individuals form friendships due to an external factor such as the authority of an individual. For instance, befriending a public figure is due to influence. In homophily, friends are similar in terms of non-environmental attributes such as their interests. In confounding, friends are similar in terms of their environmental attributes such as their mother tongue or location. In influence, after an individual befriends an influential, though the individual can be different from the influential in terms of the environmental or non-environmental attributes, but him or her often fits well within the crowd who has already befriended the influential. Friends of the user are more likely to have friends that have the exact same attribute value. In either case, to find friends one should aim at predicting user attributes, and in our situation, from usernames. We believe that due to unique personal attributes, individuals exhibit certain behaviors. These behaviors are non-random and therefore, create information redundancies these information redundancies can be captured in terms of data features in their user- names. For instance, we expect individuals who speak the same language to have statistical language patterns observable in their usernames.

3.1 PREDICTING INDIVIDUAL ATTRIBUTES

As discussed friendships are formed by three general social forces: homophily, confounding, and influence. Our goal is to predict user attributes that are observable in user names and represent each social force. Our goal here is to demonstrate how simple user attributes that represent each social force can be predicted using only usernames. Following social forces, we measure how these attributes help better find levels of friendship.

3.1.1 HOMOPHILY – BASED FRIENDSHIPS

Homophily is observed when an individual befriends each other according to their similarities in non environmental user attributes. A major non environmental user attribute that is shown to introduce friendships is the age of the individual. One often observes that individuals in the same age range more frequently befriend each other. If the ages of individuals, represented as usernames are predicted, one expects usernames in the same age range to have higher friendship likelihoods.

3.1.2 CONFOUNDING – BASED FRIENDSHIPS

Among the many attributes that describe the environment that the users are living, we select two of the most prominent attributes are their language and location. Similar to the age of individuals, we expect users living in close proximity or sharing the same language to have a higher chance of becoming friends.

3.1.3 INFLUENCE – BASED FRIENDSHIPS

When friendships are formed according to their influence, we are assuming influential users are attracting friends. In this scenario, we can partition the users attracting other in terms of the types of friends they are attracting and compare each partition with the user for which we are searching for friends. In general, we believe the factor that is deciding in becoming a member of the crowd that has befriended an influential is how the user fits in that crowd. We assume that a user fits in a crowd when at least one member of the crowd is similar to the user in terms of some attribute. We use all three attributes predicted so far: age, location, and language to predict this similarity [2].

4 MEASURING FRIENDSHIP LEVELS

According to the social forces, we seek a procedure to discover all association rules which have at least P% support with at least Q% Confidence from the following set of friendship ID.

Table 4.1 Friendship ID with Names

Friendship ID	Name of the Friends
100	Vivin, Vedha
200	Vivin, Vedha, Priya
300	Vivin, Anantha
400	Vedha, Priya, Anantha

There are four types of Friendship ID (Vivin, Vedha, Priya and Anantha) and there are only the four transactions given in the table and interested in finding association rules with a minimum "Support" of 50% and minimum "Confidence" of 75%.

All the combinations of the friends that are in stock and find which of these combinations are frequent, and the association rules are discussed that have the "Confidence" from these frequent combinations [6]. The four Friendship IDs and all the combinations of these four Friendship IDs and their frequencies of occurrence are in the transaction database in the table.

Table 4.2-The List of all Friendship Sets and their Frequencies

Friendship Sets	Frequency
Vivin	3
Vedha	3
Priya	2
Anantha	2
(Vivin, Vedha)	2
(Vivin, Priya)	1
(Vivin, Anantha)	1
(Vedha, Priya)	2
(Vedha, Anantha)	1
(Priya, Anantha)	1
(Vivin, Vedha, Priya)	1
(Vivin, Vedha, Anantha)	0
(Vivin, Priya, Anantha)	0
(Vedha, Priya, Anantha)	1
(Vivin, Vedha, Priya, Anantha)	0

Given the required minimum Support of 50%, we find the Friendship Sets that occur in at least two transactions, Such as Friendship IDs are called frequent.

Table 4.3 - at least Two Transactions

Friendship Sets	Frequency
Vivin	3
Vedha	3
Priya	2
Anantha	2
Vivin, Vedha	2
Vedha, Priya	2

In order to determine the Two Friendship Sets (Vivin, Vedha) and (Vedha, Priya) lead to association rules with required Confidence of 75%.

Every 2 –Friendship Sets (A,B) can lead to Two rules A→B and B→A.

If both satisfy the required Confidence as defined earlier, Confidence of A→B is given by the Support for "A and B together" divided by the Support for A" (i.e.) AB÷A×100

So there are four possible rules and their Confidence as follows:

Table 4.4 -Four possible rules and their Confidence

Vivin → Vedha with Confidence of 2÷3×100=67%
Vedha → Priya with Confidence of 2÷3×100=67%
Vedha → Priya with Confidence of 2÷3×100=67%
Priya → Vedha with Confidence of 2÷2×100=100%

Therefore only the last rule Priya → Vedha has Confidence above the minimum 75% required and qualifies.

5 RELATED WORK TO THE BEST OF OUR KNOWLEDGE

Our study is the first to help find friends when no link or content information is available. However, one can find similar unsupervised link prediction studies in the existence of link or content information that are applicable in our case. Assuming usernames are content generated by users; one can compute the similarity between individuals and the similarity between their friends. In this case, well-established link prediction methods that use node similarity or neighborhood similarity such as the common neighbors are applicable. Note that when using contents generated by users, it is common to assume large collections of documents, with thousands of words, available for each user, whereas for usernames, the

information available is limited to one word. Our technique employs the knowledge of how social forces influence friendships and additional information such as age, language, and location that represent these social forces to reduce friendship search space, helping better predict future friends.

6 CONCLUSIONS AND FUTURE WORK IN THIS PAPER

We propose an approach for finding friends when link or content information is unavailable. This problem is ubiquitous to all social media sites since when a user joins a new site; he or she has no friends or has not generated any content. Under these constraints, sites are often forced to recommend randomly chosen influential friends, hoping that users by adding these friends create sufficient information for link prediction techniques for further recommendations. Friendships in social media are often formed due to three social forces namely, homophily, confounding, and influence. We show how minimal information available on all social media sites that is usernames can be employed to determine friendships due to these forces. In particular, we employed usernames to predict personal attributes such as age, location, and language that in turn can be used to find friends and measure the effect of each social force [3]. This suggests that individuals have more tendencies to befriend others with similar friends (influence), than those who are more similar to them (homophily) or live in a common environment (confounding).Using social forces the research paper has investigated the levels of Friendship Calculation. Our work opens the door to many interesting applications. Studying addition of other information or analyzing how combining this approach with traditional link prediction can further improve the performance of link prediction are examples of the many areas that can benefit from the results of this study. Future work also includes analyzing these possibilities and discovering how these social forces can be combined to further improve friend finding performance. While we demonstrated that all social forces are helpful in finding friends, the comparison of forces can be influenced by the performance of the classifiers. We leave verifying our findings with labeled data in which age, location, or language is known as another part of our future work.

REFERENCES

[1] P. Cashmore. 60% of Twitter users quit within the First month, 2009.
[2] L.Tang and H.Liu, Community Detection and Mining in social media. Morgan & Claypool Publishers, 2010.
[3] Mohammed-Ali Abbasi Measuring User Credibility in Social Medid, 2012.
[4] Jiliang Tang, HujiGao. Exploiting Homophily Effect for Trust Prediction, 2013.
[5] P.Gundecha, S.Ranganath “A Tool for Collecting Provenance Data in Social Media”, 2013.
[6] Anirudh Kondaveeti, George Runger,”Extracting Geographic Knowledge from Sensor Intervention Data Using Spatial Association Rules”, China 2011- June 28- July 1.