Global Journal of Enterprise Information System

DOI: 10.18311/gjeis/2017/15687

Analysing Asymmetrical Associations using Fuzzy Graph and Discovering Hidden Connections in Facebook

Amita Jain[1]*, Sunny Rai[2], Ankita Manaktala[1] and Lokender Sarna[3]

[1]*Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India; amitajain@aiactr.ac.in*

[2]*Netaji Subhas Institute of Technology, New Delhi, India*

[3]*Hewlett Packard Enterprise, India*

Abstract

The fuzzy graph theory to analyse the relationship strength in Social Networks has gain significant potential in last few years and has seen applications in areas like Link Prediction, calculating Reciprocity, discovering central nodes etc. In this paper, we propose a framework to analyse and quantify the degree of strength of asymmetric relationships and predict hidden links in social networks using fuzzy logic. Till now, the work in fuzzy social relational networks has been limited to symmetric relationships. However, in this paper, we consider the scenario of asymmetric relations. The proposed approach is for web 2.0 application *Facebook*. Our contribution is three fold. First, the measurement of the strength of asymmetric relationship between nodes on the basis of social interaction using the concept of fuzzy graph. Second, a hybrid approach for prediction of missing links between two nodes on the basis of similarity of attributes of user profiles such as demographic, topology and network transactional data. Third, we perform fuzzy granular computing on attribute 'strength of relationship' and categorise into four granules namely {*socially close friends, socially near friends, socially far friends*, *socially very far friends}* based on the results of supervised learning conducted over dataset. Similarly, actual outcome for predicted links is categorised into three granules namely *Accept, Not accept* and *May be*. The proposed approach has predicted relationship strength with mean absolute error of 9.26% whereas the proposed approach for Link prediction has provided 64% correct predictions.

Keywords: Asymmetry, Computing With Words, Fuzzy Graphs, Fuzzy Sets, Fuzzy Granular Computing, Link Prediction, Reciprocity, Social Relational Networks, Strength Of Relationship

Manuscript Accepted: 05-Jan-2017; **Originality Check:** 10-Feb-2017; **Peer Reviewers Comment:** 16-Feb-2017; **Double Blind Reviewers Comment:** 02-Mar-2017; **Author Revert:** 12-Mar-2017; **Camera-Ready-Copy:** 28-Mar-2017)

Introduction

A social network is defined as a set of social actors, or nodes, or members that are connected by one or more types of relations[1]. On a Social Networking Sites (SNSs) like Facebook, Twitter etc., Friendship links or connection between two nodes are undirected and weightless i.e. it is formed through mutual agreement of both users. Friendship appears on both profiles and considered to be of equivalent strength for both nodes. It considers that either the connection exists or doesn't exit. Subsequently, strength of relationship can be represented in binary logic of '0' and '1' for existing connection and absence of connection respectively. However, this is rarely the case in real world and so in virtual world. The strength or intensity of the relationship between two nodes in a social network is a significant factor which can be used to determine future links, optimal path for information dissemination, advertisement etc. There are various approaches which are proposed to determine the strength of relationship between the nodes on the basis of similarity of nodes and transactional data exchanged between nodes. Unlike Classical Mathematics which deals with binary logic ('0' and '1') only, Fuzzy logic is a many valued logic and values between '0' and '1' are considered too. Thus, fuzzy logic considers the fact that links of varying strength may exist instead of just presence or absence of links. Therefore, strength of relationship can lie between '0' and '1'. However, the work in domain of fuzzy logic has focussed on symmetric graphs only i.e. undirected graphs where strength of relationship is symmetrical. However, the relationship between two nodes is rarely symmetric. The trust or influence of (x→y) is often not equivalent to trust or influence of (y→x). The rapid increase in connections on social networking sites further enforces this concept. On an average, there are 130 friends per profile on Facebook. Thus the assumption that all 130 friends are equally important to a user is not practical. There has to be difference in strength of relationship between any two nodes. As for evidence to this fact, a study says that median of 'actual' friends on Facebook is 50 only[6].

Our approach for quantifying the strength of relationship between two nodes is based on the 'interaction'. Unlike concept of Similarity which is relatively a static factor, Interaction is a true representative to determine and analyse the dynamic nature

*Author for correspondence

of relationship between two connected nodes over a period of time. Two nodes is said to be strongly connected if they interact regularly and weakly connected if there is very less or negligible interaction. Here, in context of *Facebook,* the term 'interaction' refers to activities such as comment, tag, like or timeline posts done by a node x on node y's page. In this paper, we incorporate concept of fuzzy logic in asymmetric socio-grams to capture the true strength of relationship which can lie between 0 to 1, thus providing more clear picture about influence and reciprocity from the perspective of (x → y) and (y → x). We categorise the computed 'strength' in four granules based on the supervised learning conducted on the data set collected and performed fuzzy granular computing for the attribute "strength of relationship".

Online Social Networks (OSNs) contributes a formidable share to a user online activity. One such OSN, Facebook was created in the year 2004 by Mark Zuckerberg and fellow students. The average time spent by an user on Facebook per visit is around 18 minutes. 70% of users say that they use Facebook daily including 45% who do so several times a day[1].It is a significant part of daily media practices of its users. According to a latest report, nearly 1 billion user worldwide access Facebook through their mobile phones at least once each month in 2015. Globally Facebook has 1.35 billion users, while daily active users are 864 million[1]. Analysis of Social relational networks has applications in online advertising[2], recommendation and E-commerce system, organization management to study the interaction of employees, covert networks[3], Web applications, Co-authorship networks[4], tracking terrorist organisation and their way of expansion etc. Due to their vast reach and easy connectivity, they have become an important way for social searches in which the motive is to find a person who might be of advantage to them in future or pass on an introduction to a specific people or organization for recommendation and to get connected[5]. "Link Prediction" is the concept which deals with finding "hidden connection" in a social network between nodes. The term 'hidden connection' refers to links which might appear in future between two nodes which are presently not connected. Further extending the concept of strength of relationship between two nodes, we introduce a method for finding the hidden connections in the social network. The link prediction problem can be stated as: Given a snapshot of a social network, can we guess which new interactions among its members are likely to occur in the near future? OSNs like Facebook has well placed algorithms to recommend these hidden links in form of "people you may know" on the home page of users.Friendships between people are not formed arbitrarily. *Homophily* is the principle which states that a contact between similar people occurs at a higher rate than among dissimilar people[6]. It emphasises about the localized flow of cultural, behavioural, genetic and material information in a network. The more similar social characteristics are, the lesser will be the network distance i.e. the number of nodes between them. OSNs also demonstrate this social phenomenon which predicts that occurrence of connection between two similar persons is more likely. People who lived or lives in the same neighbourhood, attended the same school or college or organization, having similar kinds of interests are more likely to become friends[7]. Similarly, Researchers have also validated the tendency for clustering or 'transitivity'[8]. Therefore, having higher number of mutual friends suggests higher possibility of connection in future between x and y. To quantify the degree of similarity between two profiles, we utilise profile attributes that defines the person's cultural and geographical background, education, personality, interests etc. For dealing with the problem of link prediction we incorporate three types of features a) Demographic b) Topological and c) Network transactional features. Demographic feature includes age, relationship status, home town, location, employment status, school college information etc. Topological feature (here) measures the number of shared neighbours in the social network graph. Network transactional features here include the number of groups joined by the individual. For the aggregation of all these similarity features, we use modified version of OWA(ordered weighted aggregation or 'or-and' operator) operator which is introduced by Ronald R. Yager[9]. We predict the link on the basis of value of 'aggregated similarity features' and categorise the actual outcome of predicted links in three granules namely *Accept, Not Accept* and *May be.*

The paper is organised as follows. Section 2 describes previous theoretical advancements and proposed works which is further subdivided in two sections, Section 2.1 – Estimation of strength of relationship and Section 2.2 Link Prediction. Section 3 outlines the definition of fuzzy graphs and how the concept of fuzzy graphs is applied on social relational networks and combined as fuzzy social relational networks. Section 4 is again subdivided into two parts. In Section 4.1, we formulate the strength of relationships for asymmetric relationships in a social network. Section 4.2, the method to predict the hidden links is explained .In Section 5, the idea of granular computing in the strength of ties is explained. In Section 6, we describe our experiment work which is subdivided two section. Firstly, Section 6.1 deals with the data collection for Facebook activities of 75 university students. Second, Section 6.2 explains the experimental work for Link Prediction which is performed on 10 seed nodes for evaluation of obtained results. Finally we state our conclusion and future work in the end.

2. Related Work

The existing literature for estimation of 'strength of relationship' between two nodes and link prediction is elaborated in this section.

2.1 Estimation of 'Strength of Relationship'

Mark Granovetter[10] was first to introduce the "tie strength" problem. Assessment of strength of relationship between the two socially connected people is previously done and has underlying structure based on the exchange of emails, mobile calls, tweets on twitter, instant messaging, *Facebook* activities and so on. Ogatha[11] investigated the strength of social relations between individuals through the email conversation. The relationship is strong if the emails are exchanged frequently, recently and reciprocally. In[12] and[13] the tie strength of mobile phone graphs are investigated. In[14] the affinity based on phone call-detail records and how to quantify the social ties' strength between actors in groups is studied. In[15] the method for calculating the indirect as well as direct relationship strength is introduced. Srba *et al.*[16] focussed on calculation of the relationship strength by means of the interaction data and other "rate factors". Viswanath *et al.*[17] analyzed the relationships between nodes using the *Facebook's* "wall posts". Some of the previous work not only focussed on the strength prediction but also intended to categorise the relationships based on the strength they predict. In[18] the author describes how to categorise the friendship among teenagers based on the SMS, Instant messaging, telephone calls and messages on Social networking sites. In[19] the further progress has been done in categorising the social relationships into four groups based on their strengths. In [14] the social relationships are categorised in the groups based on affinity caused due to phone calls. However the above mentioned authors have considered the social relationships as symmetric. The concept of unequal relationship between two nodes on a social network was firstly introduced by Hangal et al. in[20] for twitter. Ronald R. Yager[21,22] was first to introduce fuzzy logic in the social networks but his contribution is limited to symmetric ties. In this paper we show the existence of biased friendship which is common in real world by showing the existence of varying degree of tie strength that people have on *Facebook*. Our approach takes into account the asymmetric behaviour of ties in the fuzzy social relational networks. We show the asymmetric behaviour by categorising the relationships among four granules i.e. fuzzy subsets using the idea of granular computing.

2.2 Link Prediction

Liben-Nowell and Kleinberg[23] proposed one of the earliest approaches of link prediction for social network. Those methods were developed for prediction hinged on measures considering the proximity of nodes in a network. Bliss[24] proposed the dynamic link prediction algorithm in social networks. In[25] the new algorithm was introduced which combines both content based filtering and friends of friends concept to produce new algorithm for recommending friends. In[26] a new weighted content based recommendation technique is proposed. Backstrom et al. [27] involves the idea of homophily to upgrade the link prediction model in Myspace and LiveJournal. Crandall et al. [28] study about the temporal evolution of link structure in which they incorporates both homophily and influence concepts. The aforementioned work is limited to the dilemma of link existence problem using non-fuzzy approach. In[29] a new approach which incorporates the fuzzy model concept for link prediction problem in social networks emerged. Till now, the work is limited to proposing approaches for link prediction or friend recommendation but none of them has done any experimental work to validate the efficacy of their approach. In this paper, we propose an approach for link prediction by incorporating concepts like modified OWA operator and membership functions from fuzzy logic. We categorise the outcome of recommended nodes in three categories, namely accepted, not accepted and may be. Further, we calculate the accuracy of our proposed approach on the basis of comparison between results obtained and data collected from actual users.

3. Prerequisites

To understand the fuzzy relational social networks, it is necessary to review the fuzzy graph, graph theoretic and algebraic concepts.

3.1 Fuzzy Graphs

Fuzzy logic was proposed by L. Zadeh in the year 1965 to address the realistic cases more precisely. Classical Mathematics logic which deals with 0 and 1 is effective for exact systems only and fails to accurately model imprecise system which is the case with the real systems dealing with humans. Fuzzy models the fuzzy relationship concept to represent the weighted graph that can be termed as Fuzzy Graph.

Let S be a set, then a fuzzy subset of S is a mapping $\sigma: S \to [0,1]$ which assigns to each element $x \in S$ a degree of membership, $0 \leq \sigma(x) \leq 1$. A fuzzy relation on a S is a fuzzy subset $S \times S$, i.e a mapping $\mu: S \times S \to [0,1]$ which assigns to each ordered pair of elements (x, y) a degree of membership, $0 \leq \mu(x, y) \leq 1$.

In the context of social networks, The fuzzy relations help in modelling the strength of relations between the members. The membership degree $\mu(x, y)$ represents the strength of relationship between $x \ and \ y$, so the membership degree can be defined as

$\mu(x, y) = \{$ (1 *if x has the strongest possible degree of relationship with y*@@$\delta \to]0,1[$ *if x is related to y to a certain extent* ... @@0 *if x i* eq 1

In case of binary or crisp relations, relationship between x and y is either 1 or 0. The researchers commonly described social relationships as binary i.e. nodes x and y can either be friends or not.

a. Following are the main properties of Fuzzy graph.

- Reflexivity: A fuzzy Relation μ on σ is reflexive if $\mu(x,x) = \sigma(x)$ for all $x \in S$
- Transitivity: A fuzzy Relation μ on σ is transitive if $\mu(x,y) \geq Max_y[\mu(x,y) \wedge \mu(y,z)]$

In this paper, we use socio-gram as directed and weighted graph. Therefore, the property of symmetry of relation between two nodes is not applicable here, i.e. $\mu(x,y) \neq \mu(y,x)$ and $w(x,y) = \mu(x,y)$ respectively. In social networks, the transitivity property deals with the concept of 'friend of a friend can be my friend. We assume the relationship strength with x to x will be the highest always that is 1.

3.2 Distance

Distance $d(x,y)$ between two users x and y in social networks can be defined as the shortest length of the path from x to y. The distance $d(x,y) = 1$ if x and y are directly connected.

3.3 Weighted Average Aggregation Operator (WA)

Ordered Weighted Average is the concept to aggregate the functions depending on the importance/weight of the features. It provides a way to "orand" the different criteria on the basis of their assigned weight. This operator helps in decision making for a problem where there is a requirement to satisfy "all" criteria and "or" at least one of the given criteria. OWA operator becomes WA operator if the two property of OWA is neglected i.e weight association w.r.t ordered positions and symmetric property but the property for "orand" operator still survives. A WA operator g of dimension n is function g: $\mathbb{I}^n \rightarrow \mathbb{I}$ (where I = [0,1])with an associated weighting vector W,

$$W=\begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} \quad (2)$$

such that

1) $W_i \in (0,1)$ (3)

2) $\sum_{i=1}^{n} W_i = 1$ (4)

3) $g(a_1, a_2 \dots\dots, a_n) = W_1 a_1 + W_2 a_2 + \cdots \dots\dots + W_n a_n$ (5)

Conveniently $g(a_1, a_2 \dots\dots, a_n)$ is denoted as $g(A)$ where A is associate argument vector. Here it is necessary to highlight the case that the weights, the W_i' s are associated with a particular element i.e. W_i is the weight associated with the th^{i} ^{i} thelement of A. For each criterion a_i , $a_i(x) \in [0,1]$

$$g(a_1, a_2 \dots\dots, a_n) = W.A \quad (6)$$

W.A symbolises multiplicative product of W and A vectors where $0 \leq g(A) \leq 1$

It should be noted that we eliminate the term "ordered" from ordered weighted average operator because here the condition of symmetry (commutative property) is not required.. Asymmetry condition implies that $g(a_1, a_2) \neq g(a_2, a_1)$ where g is weighted average operator.

4. Proposed Approach

This section is subdivided into two sections. Section 4.1 explains our approach regarding prediction of strength of relationship on the basis of interaction between two nodes in context of asymmetric relationships. Section 4.2, explains link prediction based on the concept of *Homophily* and similarity of three categories of parameters a) demographic features b) topological features and c) the network transactional features.

4.1 Prediction of Strength of asymmetric relationships

4.1.1 Interaction Vector

Granovetter and Mark [10] said "*The 'strength of a tie' are amount of time, the emotional intensity, the intimacy(mutual confiding), and the reciprocal services.*"

Among above mentioned four factors, measurement or analysis of '*emotional intensity*' *and* '*intimacy*' requires access to personal messages exchanged which are generally not publicly available. Further, demands human judgement to label what is emotional and what is not for accurate text analysis. However, '*time*' and '*reciprocal services*' data are comparatively easy to obtain, quantify and analyse. From the above statement, it is clear that similarity cannot be the criterion to determine the strength of relationship between two nodes after the formation of connection. The concept of 'Complementarity in Interaction'30 states the similar idea that in spite of being completely similar in terms of cultural and educational background (like in case of siblings), one cannot say that they will have the strongest relationship in the virtual environment. Thus, instead of relying on Similarity attributes, we consider only interaction based features i.e. Interaction Vector to determine the strength of relationship. This enables us to identify relationship between people with dissimilar profiles. Interaction Vector represents the dynamic concept and comprises of attributes relevant to quantify the strength of an existing relationship. Unlike Similarity vector which is relatively a static factor, Interaction vector is a true representative to determine and analyse the change in the strength of relationship between two connected nodes over a period of time.

For modelling of interaction vector, we use attributes derived from interaction activities to know how people maintain their social relationships. For example in *Facebook,* there are certain interaction activities like timeline posts, picture tagging, likes and comments that defines the '*closeness*' between node on the basis of efforts made i.e. time or resources spent by the profile owner. According to a survey31,32 44% of Facebook users '*like*' content posted by their '*friends*' while 29% do it several times in a day. Similarly, at least 31% users '*comments*' on other photos on a daily basis and 12% users have '*tagged*' a friend in a month duration. The approach can easily be understood by the fact that each user has a limited amount of resources (i.e. time) to use in the building up and maintaining a relationship, it is more likely that the users would spend these limited resources to those whom they find more important. Thus the formula for the computation of strength of relationship $IV_f(x, a)$ of x $towards$ a is the combination of the interaction activities done by x for a incorporated with the use of weighted average aggregation operator.

The formula contain features which denote the interaction between two nodes like wall posts, likes, comments and tagging.

$$IV_f(x,a) = \gamma_p \frac{Post(x,a)}{\sum_{y \in V} Post(x,y)} + \gamma_t \frac{tags(x,a)}{\sum_{y \in V} tag(x,y)} + \gamma_l \frac{likes(x,a)}{\sum_{y \in V} likes(x,y)} + \gamma_c \frac{comments(x,a)}{\sum_{y \in V} comments(x,y)} \quad (7)$$

Where $0 \leq IV_f\ (x) \leq 1$

Reciprocity can be termed as the "*action of returning similar acts*" 14. In social networks, the concept of reciprocity is very important in applications like privacy controls (e.g. detection of spammers, phishing), marketing of a product to a specific set of audience etc. Reciprocity is the factor which distinguishes an asymmetric relationship. In this paper we focus on the reciprocity of interaction activities that takes place in the Facebook *from perspective of both nodes* and mapping it to calculate the strength of relationship *i.e. strength(x → y) is not always equal to strength(y→x).* Suppose in case of Facebook, one person x *likes/comments* every activity done by another person y but y does not reciprocate to x. Thus supposedly node *x* is important for node *y* but not vice versa. Hence, this is a perfect example of asymmetrical relationship. This has been experimentally shown in Section 5.

4.2 Link Prediction

The approaches for Link prediction can be categorized broadly into two groups. First one is based on the link structure of network i.e. shared neighbours between two nodes. Second one is based on the concept of '*homophily*' i.e. similarity of nodes[33]. In our approach we use a hybrid approach by combining both techniques and including one more attribute namely network transactional attribute (i.e. the same groups joined). Every SNS is based on a very different fundamental concept which makes it unique and suitable for a particular requirement. However, profile generation is an overlapping concept whose attributes differ according to the basic principle of the specific SNS. Along with demographic features, the personal interests and mutual friends are also the criteria which represent a profile. Thus, the attributes considered to determine similarity between two profiles of a social network is defined according to the specific OSNs (*Facebook* here). It enables a comparatively accurate prediction results in comparison to the generalized Similarity vectors.

In *Facebook*, all the members create their personal profile page which contains information regarding the views, interests, friends, location, school, college, organization, and home town which can be marked as private or public information. '*Private*' refers to access to restricted audience like friends whereas *'Public*' is information accessible to anyone. The information related to these features is available in a user profile. While recommending nodes, we are considering *distance* $d(x, y)$ to be maximally up to 334. The similarity of user's interests is highest in the direct connections and decreases as the distance between nodes increases. We take into account three features a) Demographic b) Topological features and c) Network Transactional features.

4.2.1 Demographic features

Demographic or background similarity provides the basic similarity between two nodes and refers to factors like age, sex, education, work, nationality, geographical location, extended family etc. According to a recent survey[3], 'High school and University friends' constitute the largest share and 'co-workers and groups' forms the second largest share in the average user's friend list. So, we gather that people from 'work environment (university and school for students)' are more likely to be friends rather than family members or close acquaintances. Consequently, we include aforementioned Demographic features in the prediction of missing link, given in Table 1.

Table 1. Demographic Features

Demographic profile features	
f_1	Age
f_2	High School
f_3	College
f_4	Current city
f_5	Home town
f_6	Employee in an organisation
f_7	Relationship status or extended family

These demographic features can be extracted using the tool "Wolfram Alpha Facebook Report[4]" and the Facebook application "Friend list Manager"[5,35]. We have identified 7 features to capture the profile similarity. User similarity of y with x can be represented as vector]T where $s^{(x,y)}$ is a mapping from $\mathbb{I}^n \rightarrow \mathbb{I}$ (where $\mathbb{I} = \{0,1\}$) and $s_i^{(x,y)}$ denotes similarity value for the feature *fi* between x and y. $s_i^{(x,y)}$ can be either 1 or 0.These similarities are described below:

- *Age feature*(f_1)

The people with same age or at an age gap with 1-2 years are more likely to become friends[36]. Therefore the formula to compute age coefficient of user y with x is defined as

$$Age\ coefficient(y,x) = \frac{1}{1 + (age_y - age_x)^2} \quad (8)$$

If value $0.1 < age\ coefficient \leq 1$ then *age similarity* is 1 otherwise it will be 0. The set 'age difference' has the universal space Y= {0,...,100} and the membership function to the age similarity $\mu_{age\ similarity}$ can be represented graphically

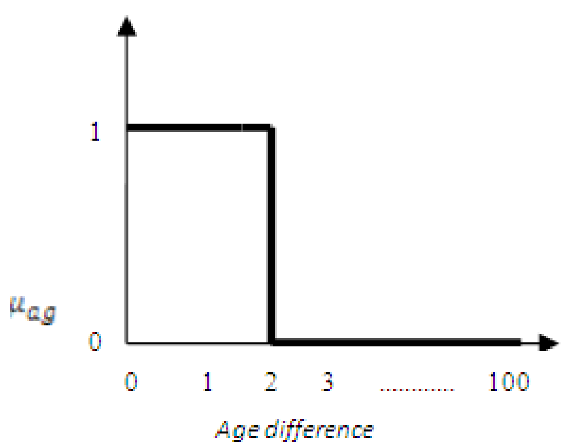

Figure 1. Graphical Representation of Membership Function "Age Similarity".

- *Features* $f_2, f_3, f_4, f_5, f_6, f_7$: If the value of the mentioned features is same or related, then value 1 is set otherwise 0 is assigned as shown in table 2.

Table 2. Demographic profile Similarity

f_2	School	{ same, different }
f_3	College	{same, different}
f_4	Current city	{ same, different}
f_5	Home town	{same, different}
f_6	Employee in an Organization	{unrelated, related}
f_7	Extended family or any relationship	{yes, no }

So the overall demographic profile similarity can be calculated using[37]

$$sim_{db}(x,y) = \frac{\text{EMBED Equation.3}}{|s^{(x,y)}|} \quad (9)$$

$|s^{(x,y)}|$ is the number of features included in $s^{(x,y)}$ i.e. 7.

4.2.2 Topological features and Network Transactional Features

Topological feature measures the number of mutual friends in the network. It measures the connectivity of the nodes in social network. This information is easily available on the user's profile itself or can be extracted using the tool "Wolfram Alpha *Facebook* Report". *Network Transactional features* help to know the number of overlapping groups joined by the user. This can be obtained using graph search available in Facebook by typing "*groups [person's name] joined*" in the graph search. The common friends are displayed with the group's name. Near about 31% of *Facebook* friends are not classified by *Facebook* users as family, co-workers, neighbours, classmates, or people from voluntary groups. The authors who conducted survey speculate that these remaining ties are predominantly dormant ties and friends-of-friends[5]. As a result we select topological feature and network transactional feature in addition with demographic feature for the computation of similarity vector. From this survey result we roughly estimate the weights associated with the each attribute in similarity vector which is explained below.

In order to calculate derived similarity or topological and network transactional similarity, Cosine Similarity is used. The overall similarity vector is calculated by using *weighted average aggregation operator.*

The overall similarity vector S_V can be calculated as

$$S_V(x,y) = \alpha_g \frac{\|g_x \cap g_y\|}{\|g_x\|_2 \cdot \|g_y\|_2} + \alpha_f \frac{\|f_x \cap f_y\|}{\|f_x\|_2 \cdot \|f_y\|_2} + \alpha_{db} Sim_{db} \quad \text{... eq. (10)}$$

Where

$\|g_x \cap g_y\|$ – number of common groups between x and y

$\|g_x\|_2 = \sqrt{g_x}$ where g_x – total number of groups joined by user x

$\|f_x \cap f_y\|$ – number of friends common between x and y

Therefore, S_V is the combined Similarity vector and moreover

$$0 \leq S_V(x) \leq 1 \quad (11)$$

Both IV_f and S_V are *weighted average aggregation operator.*

5. Granular Computing and Computing with words

Our main motive to involve fuzzy concepts is to extend the concept of social relational network with the network concepts so that the human beings can visualize social network relationship in such a way that they are explicit to both men and machines. In order to bridge the gap between man and machine conception, two very famous concepts are used namely a) *Granular computing* and b) Zadeh's fuzzy set approach for Computing with Words. Computing with Words (CW) is a technique, proposed by Lotfi A. Zadeh in 1979, in which words are used instead of numbers for computing and reasoning. This methodology allows us to employ human concepts in formal representation of the network properties. Human beings use linguistic terms for communication, logic and understanding the world where as machines are more inclined to use formal symbols. The main advantage of CW is that it maps the system with vagueness and imprecision that fits more into frame of human thinking. The importance of role played by the fuzzy sets and CW in representing the linguistic concepts has been already been evidenced in [21,38].

Granular computing is a concept, dependent on human-based thinking to facilitate high level of cooperation by providing a framework which is compatible to machine and understandable to man. It divides the system into its parts. In crisp granular computing, there are well defined components which are not possible in the real world. Therefore the concept of fuzzy granular computing arises to solve this problem. The above mentioned technology not only links the man and machine but also provides bridge between the network analyst's linguistic definition of social network concept and the formal model of network as shown in Fig.2

In the analysis of weighted social relational networks, we take an attribute "strength of tie" which has certain vocabulary associated with it. In this paper, CW uses fuzzy subsets or granules (*socially close friends, socially near friends, socially far friends*, *socially very far friends*) to formally represent the semantics of linguistic term or fuzzy set "strength of tie". Fuzzy set helps in formalizing the linguistic concepts in such a manner so that machine can compute and understand. "Strength of Relationship" has domain $I = [0, 1]$. Granule *socially close friends* can be represented as the fuzzy subset S of $[0,1]$, such that for any $y \in [0,1]$, value S(y) would indicate the degree to which y satisfies the working definition for the concept *socially close friends*. Similarly the linguistic value of other remaining granules can also be mapped in space I. The granules are defined as follows:

a. Socially close friends: These are the people in our friend list with whom we consider our relationship strongest. Quantifying them by interaction activities, we find that interaction activities and frequency is more to and from socially close friends.

b. Socially near friends: These relationships are not as strong as compared to (a). Sometimes, not always, they do the interaction. We observe intermittent frequency of interaction activities from these people. Mostly the neighbours, distant relatives, general friends, colleagues etc. come under this granule.

c. Socially far friends: These types of friends are weakly connected to each other and the frequency with which they interact is less. We seldom receive any interaction activity to and from them.

d. Socially very far friends: This category of friends belongs to dormant ties in social networks and the people we add due to similar interests or any other reason. Dormant relationship is a relationship between two individuals who have not communicated with each other for a long time. In reality, the dormant relationships are not essentially strong or weak in strength but due to some reasons they are out of touch.

Similarly the results of link prediction can be divided into three granules. These are as follows:

a. Accept (A): this link prediction granule tells that the suggested node(x) has a high similarity to the node to whom the suggestion goes(*y*). Or they have a lot of same kind of interests and they had good face to face communication. There can be senior junior relation between x and y. In short y will accept the friend request if x send it.

b. May be (MB): this granule is the outcome when y is not sure about accepting or discarding the request. There may be very less similarity but due to same interest or future benefits the y might not reject the x and in the future y can accept the request.

c. Not Accept (NA): this granule will be the outcome if y does not like x at all. Although there may be a lot of similarity but due to existence of the negative tie between $y \rightarrow x$, y will discard the request of link. Or the other reasons may be y do not know x.

6. Experiments

We analyze and validate our approach on the *actual Facebook* data. We divide the procedure of data collection in two parts.

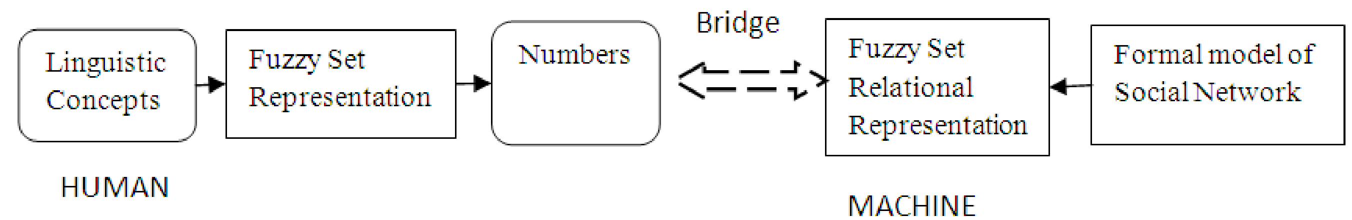

Figure 2. Paradigm for Intelligent Social Network Analysis.

Firstly, we describe data collection for estimating tie strength and extend it to show the effectiveness of interaction vector. Secondly, we repeat the prior procedure for link prediction and subsequently show the method's accuracy in predicting missing links.

6.1 Estimation of 'Strength of Relationship'

6.1.1 Data set collection

According to a survey, 71% of Online Adults (among them 87% are of age 18-29 years old) use Facebook[3]. Therefore, we collect the data from 75 students (aged between18 to 29 years old over a period of 3 months-November, 2014 to January, 2015) who visit Facebook frequently, and engage in interaction activities such as liking, tagging, commenting, sharing the posts etc. We asked them to choose at least 4 friends (75*4=300 nodes) from their friend list such that they can categorise them on Likert scale[4] on the basis of their interaction on Facebook. We define the range for Likert scale from socially close friend → socially near friend → socially far friend → socially very far friend. The participants are advised to choose friends in such a manner so that every friend belong to different categories of Likert scale. In addition, following details were also obtained.

1. The low interaction activities done by them for each friend they selected and we rate their friendship according to table 3. By using Friendship Pages[6] and http://www.facebook.com/Xuser_id?and=Yuser_id(Xuser_id and Yuser_id are Facebook Profile Ids), a user can find out about the timeline posts, tags, likes and comments between any two friends of their own friend list other than the user itself.

2. Interaction activities on Facebook for given friends over the period of 3 months. For knowing the total activity done by the user, he/she can use search for activity log[1] available on the Facebook application.

3. Priority list for the low interaction activities namely "*like, comment, tag and post*" which they tend to use for the close members rather than the far members.

6.1.2 Evaluation and Results

The proposed approach contains 4 interaction activities which contribute to the calculation of strength of connection between two nodes. From the above data, we gather that interaction activities have varying priorities. So, it is not practical to assume that the contribution for every activity is equal. From the data collected for priority of interaction, we made a rough estimation of preferences. The weights are associated with each activity according to their significance and contribution in calculating tie strength.

The weights are assigned in such a way that $\sum \omega_i = \mathbf{1}$ i.e. sum of all weights should be equal to 1. From the data, we obtained that *tag* → *comment* → *timeline post* → *like* is the order assigned in decreasing priority by approximately 95% participants. It is to be noted that sometimes the user do one type of activity only, then the weight associated with that activity becomes 1. For example if the person's activity is liking the comments or images posted by others and no other activity is done by him then the weight associated with liking becomes 1. Similarly, if a user performs only two types of activities on Facebook, then we assign equal weights to both of them.

For estimation of the weights associated with the interaction vector (IV), genetic algorithm (GA) is applied on collected data. After learning the weights, we compute the "strength of connection" and apply fuzzy granular computing for dividing the social network relations into granules based on the attribute "strength of relationship". A genetic algorithm processes on the population of data collected so as to obtain the optimized solution for 'strength of relationship' problem[39,40]. A good set of weights will generate good prediction. In order to solve the problem effectively, supervised learning is performed in which we divide the whole data set into two disjoint sets, training data set (70%) and testing data set (30%). To obtain the good fitness for weights, the fitness score must be as lowest as possible. The fitness score is defined as the average difference between the actual and predicted ratings in the training data set and is given as follows:

$$fitness = \frac{\sum_{j=0}^{n_R} |r_j - pr_j|}{n_R} \tag{11}$$

Where $\mathbf{n_R}$ is the training data set cardinality for a given user. r_j is the actual rating done by a user and pr_j is the predicted rating computed through IV_f.

Testing data set are used as hidden ratings so that the interaction vector with the weights that are decoded by the fitness function would try to predict the ratings or fuzzy set for the ties. The mean absolute error(MAE) is used to check the effectiveness of the interaction vector in forecasting the correct fuzzy subset or rating for the friend as mentioned in The MAE(i) for the person i is given by the below mentioned formula.

$$MAE(i) = \frac{\sum_{j=1}^{n_i} |pr_{i,j} - r_{i,j}|}{n_i} \tag{12}$$

Where n_i is the cardinality of the test data set for person u_i. The effectiveness of interaction vector can be calculated by using the following formula

$$MAE = \frac{\sum_{i=1}^{N_T} MAE(i)}{N_T} \quad (13)$$

Where N_T is 30. Lower the value for MAE, the more accurate predictions are made by the interaction vector about '*which fuzzy subset is associated with which friend for a given user* u_i'.

A small socio-gram for collected data is shown in the figure 3. This figure not only calculates the strength but also give some idea about the different degree or type of relations exist in the social network. This snapshot also gives an idea about the biased friendship that exists in the social relational network (for node '17', node '18' is *socially close friend* but for node '18', node '17' is *socially near friend*).

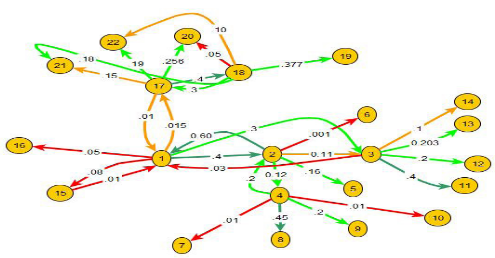

Figure 3. A snapshot of different types of friendship and the biased friendship that exists.

Table 4. Results of tie strength based on data collection

Results for 'Strength of Relationship'	
Total connections	128
True positive(correctly Predicted Strengths)	112
% of True positive	87.5
Mean absolute error	9.26 %

The results is summarized as follows. Fuzzy granular computing with granules formed (refer Section 5) for the attribute "strength of tie" (refer Fig. 4) is obtained with mean absolute error of 9.26% by applying weights $\gamma_t = .4, \gamma_p = .2, \gamma_l = .1, \gamma_c = .3$ (which provides maximum fitness) on IV_f. The user model formed is shown in fig 3. The legends for colors are mentioned in Table 3.

Table 3. Color representation and ratings used to denote different granules for attribute "Strength of Relationship"

Color of 'edge'	Type of friendship	Ratings
Dark Green →	Socially close friend	4
Light Green →	Socially near friend	3
Orange →	Socially far friend	2
Red →	Socially very far friend	1

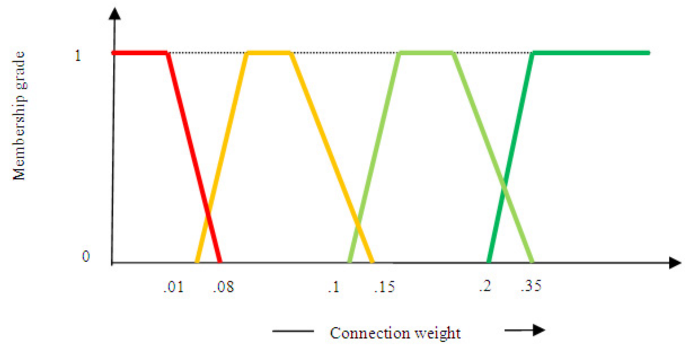

Figure 4. Fuzzy set for attribute "strength of tie".

6.2 Link prediction

6.2.1 Data collection for the link prediction

We use the data, collected for calculating strength of relationship. We identify 10 nodes (participants) and calculate the potential strength between Node *'x'* and potential connected node *'y'* in future using the approach proposed in Section. We suggest maximum 5 nodes and minimum 3 nodes to each participant on the basis of strength lying above threshold value decided experimentally. We ask participant's response whether he/she will {*accept*(A) , *not accept*(NA) or *may (MB) accept}* the recommended node. We recommend cumulatively 33 nodes to the participants. Nodes are suggested based on the concept proposed in section 4.2 and results are shown in table 6 and 7.

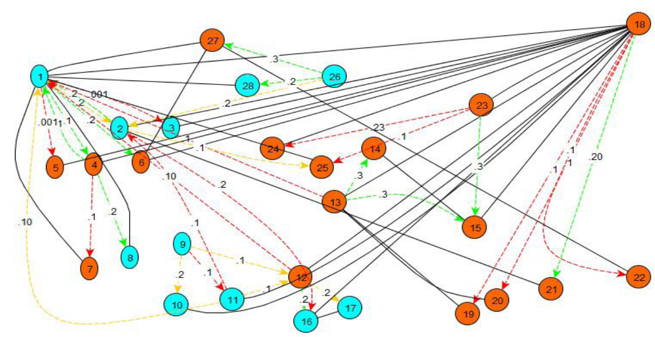

Figure 5. A snapshot for Predicted Links and their outcome.

Table 5. Color representation of prediction of missing links

Color of 'Edge'	Outcome of 'Predicted Link'
- - - - → (green)	Accept(A)
- - - - → (yellow)	May be(MB)
- - - - → (red)	Not accept (NA)
	Existing Connection

Figure 5 gives the glimpse of the collected data and predicted links where solid lines represent 'existing connection' and dashed line as 'predicted links'(legends are given in Table 5). All the blue nodes are male candidates and the orange nodes are female participants. . Fuzzy granular computing with granules formed (refer Section 5) for the attribute "strength of predicted links" is shown in Fig. 6. The results are summarized in Table 6.

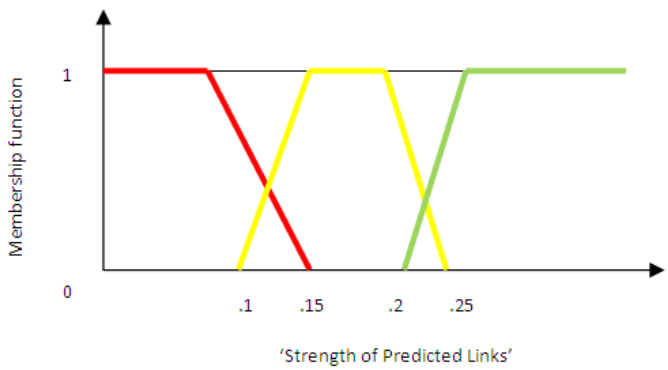

Figure 6. Granular fuzzy computing for link prediction.

From the results and data collected (Table 6), we made following inferences. The recommended links with 'strength' more than 0.1 are unlikely to be rejected (from Fig. 6). However, in given Fig. 5, we observe few '*Not accept*' links having strength higher than 0.1. On investigation about possible reasons, we found that these nodes have 'negative ties' from past. Therefore, the nodes generally won't accept request from nodes whom they don't like or have compatibility issues in past (more supposedly, a negative tie) despite of strong connection. The main reasons for accepting the friend request from the suggested node by a participant are expected '*future benefits*' from the people who are from the same domain, home town and senior junior relationships. The males are more likely to accept request from the opposite sex on the basis of '*attractiveness*'. As per results, male nodes have rejected 29.4 % of the recommended nodes and even in that, recommended male nodes are more prone to rejection as compared to female suggestions. The female nodes are more likely to reject the friend requests from male nodes whom they don't know and from the person whom they never met or have no face to face communication. The result indicates that female nodes have higher rejection rate i.e. 50%. Overall, male participants have shown low rejection ratio in comparison to female participants. The female nodes are clearer in their views about rejecting or accepting a request i.e. 6.25% for "may be" category whereas in case of male participants, 41% of recommended requests are in doubtful state of "may be". It is to be noted that all the suggestions which are made to the nodes are based on the method, proposed in the section 4.2 and the distances between nodes are considered maximally up to 3. The accuracy for the proposed approach for link prediction is about 64% on our data set shown in Table 7.

Table 7. Results for link prediction based on data collected

Results for 'Link Prediction'	
Total recommendation	33
Rejected Recommendation	13
% of Relevant1 Recommendation	64%

Table 6. Data collected from Participant's Response

Nodes(male(M))	Suggested nodes(M/F)	NA/A/MB	Actual results(from participant's response)		
			NA(M/F)	A(M/F)	MB(M/F)
1	5(2/3)	2/2/1	1/1	0/2	1/0
9	3(2/1)	1/0/2	1/0	0	1/1
11	3(2/1)	1/0/2	1/0	0	1/1
26	3(2/1)	0/2/1	0	1/1	1/0
2	3(3/0)	1/1/1	1/0	1/0	1/0
Total	17(11/6)	5/5/7	4/1(=5)	2/3(=5)	5/2(=7)
Nodes(female(F))					
4	3(2/1)	1/2/0	0/1	2/0	0
13	3(1/2)	1/2/0	1/0	0/2	0
12	3(3/0)	1/1/1	1/0	1/0	1/0
18	4(0/4)	3/1/0	0/3	0/1	0
23	3(1/2)	2/1/0	1/1	0/1	0
Total	16(7/9)	8/7/1	3/5(=8)	3/4(=7)	1/0(=1)

Error Percentage	36%

7. Conclusion and Future Work

The aim of the paper was to extend the use of fuzzy logic in asymmetric social relationships and discuss their pivotal role in modelling weighted directed social networks e.g. *Facebook*. Here we represented the investment done by user X→Y and user Y → X in the form of interaction activities which forms the basis for asymmetric strength of tie. By using the idea of CW and fuzzy granular computing, we described the important role played by the fuzzy subsets or granules (*socially close members, socially near members, socially far members* and *socially very far members)* in representation of linguistic concepts for the attribute 'strength of tie' and applied it in defining the asymmetric model of social networks. The proposed approach predicted strength of relationship with mean absolute error of 9.26%. The applications of our proposed approach for estimation of tie strength can be applied for finding the most influential path, improving global social search methodology and in extension of applications of social network database theory. By incorporating the tie strength in social networks, significant improvements can be made for resolving security issues, meaningful visualisation of networks, finding trustworthy nodes etc. We proposed a hybrid approach for link prediction with concepts from *homophily*, network transactional features and Fuzzy logic. The proposed method showed an accuracy of 64% in converting recommended nodes to actual connections. We outlined major factors for acceptance or rejection of a friend request. The proposed approach can be applied in domains of network expansion, recommender system, online advertising targeted towards specific set of audience etc. In future we will expand our concept to other OSNs like LinkedIn, twitter etc. and refine our approach by analysing possible attributes that can be considered.

8. References

1. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.
2. Bagherjeiran A, Bhatt RP, Parekh R, Chaoji V. Online Advertising in Social networks. In Handbook of social network technologies and applications. New York: Springer, 2010; 1:651–89.
3. Krebs V. Mapping networks of terrorist cells, In Connections. 2002; 24(3):43–52.
4. Cheong F, Corbitt BJ. A Social Network Analysis of the Co-Authorship Network of the Pacific Asia Conference on Information Systems from 1993 to 2008. In proceedings of PACIS 2009, Paper 23, Indianapolis, IN. 2009 Nov.
5. SanthiThilagam P. Applications of Social Networks. In Handbook of social network technologies and applications. New York: Springer, 2010; 1:637–50.
6. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: Homophily in social networks. Annual Review of Sociology. 2001; 415–44.
7. Crosnoehjy R. Friendships in childhood and adolescence: the life course and new directions. Social Psychology Quarterly. 2000; 63:377–91.
8. Burk WJ, Steglich C, Snijders TAB. Beyond dyadic interdependence: actor-oriented models for co-evolving social networks and individual behaviours. International Journal of Behavioral Development. 2007; 31(4):397–404.
9. Yager RR. On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems, Man and Cybernetics. 1988;18(1):183–90.
10. Granovetter MS. The strength of weak ties. American journal of sociology. 1973; 1360–80.
11. Ogata H et al. Computer supported social networking for augmenting cooperation. Computer Supported Cooperative Work (CSCW). 2001; 10(2):189–209.
12. Nanavati AA et al. On the structural properties of massive telecom call graphs: findings and implications. Proceedings of the 15th ACM international conference on Information and knowledge Management. ACM. 2006.
13. Onnela J-P et al. Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences. 2007; 104(18):7332–6.
14. Zhang H, Dantu R, Cangussu JW. Socioscope: Human relationship and behavior analysis in social networks. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans. 2011; 41(6):1122–43.
15. XLin X, Shang T, Liu J. An Estimation Method for Relationship Strength in Weighted Social Network Graphs. Journal of Computer and Communications. 2014; 2(04):82.
16. Srba I, Bielikova M. Tracing Strength of Relationships in Social Networks. Web Intelligence/IAT Workshops. 2010.
17. Viswanath B et al. On the evolution of user interaction in facebook. Proceedings of the 2nd ACM workshop on Online social networks. ACM. 2009.
18. Van Cleemput K. Friendship type, clique formation and the everyday use of communication technologies in a peer group: A social network analysis. Information, Communication and Society. 2012; 15(8):1258–77.
19. Kumar A, Rao T, Nagpal S. Using Strong, Acquaintance and Weak Tie Strengths for Modeling Relationships in Facebook Network. Contemporary Computing. Springer Berlin Heidelberg. 2012; 188–200.
20. Hangal S et al. All friends are not equal: Using weights in social graphs to improve search. 2010.
21. Yager RR. Concept representation and database structures in fuzzy social relational networks. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans. 2010; 40(2):413–9.
22. Yager RR. Social network database querying based on computing with words. Flexible Approaches in Data, Information and Knowledge Management. Springer International Publishing. 2014; 241–57.

23. Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology. 2007; 58(7):1019–31.
24. Bliss CA et al. An evolutionary algorithm approach to link prediction in dynamic social networks. Journal of Computational Science. 2014; 5(5):750–64.
25. Deng Z et al. Personalized Friend Recommendation in Social Network Based on Clustering Method. Computational Intelligence and Intelligent Systems. Springer Berlin Heidelberg. 2012; 84–91.
26. Garcia R, Amatriain X. Weighted content based methods for recommending connections in online social networks. Workshop on Recommender Systems and the Social Web. 2010.
27. Backstrom L et al. Group formation in large social networks: membership, growth, and evolution. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. 2006.
28. Crandall D et al. Feedback effects between similarity and social influence in online communities. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. 2008.
29. Bastani S, Jafarabad AK, Fazel Zarandi MH. Fuzzy Models for Link Prediction in Social Networks. International Journal of Intelligent Systems. 2013; 28(8):768–86.
30. Miller N et al. Similarity, contrast, and complementarity in friendship choice. Journal of Personality and Social Psychology. 1966; 3(1):3.
31. Aaron Smith 6 new facts about Facebook Pew Internet Science and Tech. 2014 Feb 3. Available from: http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/ accessed on January 14, 2015.
32. Social Networking Fact Sheet. Available from: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/ accessed on January 14, 2015.
33. Al Hasan M et al. Link prediction using supervised learning. SDM'06: Workshop on Link Analysis, Counter-terrorism and Security. 2006.
34. Wang D, King I, Leung KS. Like Attracts Like!–A Social Recommendation Framework Through Label Propagation. 2011.
35. Liu Y et al. Simplifying friend list management. Proceedings of the 21st International Conference Companion on World Wide Web. ACM. 2012.
36. Han X et al. Alike people, alike interests? A large-scale study on interest similarity in social networks. 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE. 2014.
37. Cantador I, Castells P. A multilayered ontology-based user profiles and semantic social networks for recommender systems. 2nd International Workshop on Web Personalisation Recommender Systems and Intelligent User Interfaces in Conjunction with 7th International Conference in Adaptive Hypermedia. 2006.
38. Zadeh LA. Fuzzy logic= computing with words. IEEE Transactions on Fuzzy Systems. 1996; 4(2):103–11.
39. Golberg DE. Genetic algorithms in search, optimization, and machine learning. Addionwesley. 1989.
40. Al-Shamri MYH, Bharadwaj KK. Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. Expert Systems with Applications. 2008; 35(3):1386–99.

Citation:

Amita Jain, Sunny Rai, Ankita Manaktala and Lokender Sarna
"Analysing Asymmetrical Associations using Fuzzy Graph and Discovering Hidden Connections in Facebook",
Global Journal of Enterprise Information System. Volume-9, Issue-1, January-March, 2017. (http://informaticsjournals.com/index.php/gjeis)

Conflict of Interest:

Author of a Paper had no conflict neither financially nor academically.