



# FGSN: Fuzzy Granular Social Networks – Model and applications



Suman Kundu\*, Sankar K. Pal

Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, India

## ARTICLE INFO

### Article history:

Received 16 June 2014

Received in revised form 17 March 2015

Accepted 29 March 2015

Available online 3 April 2015

### Keywords:

Granular computing

Fuzzy set

Entropy

Influence maximization

Community detection

Big data

## ABSTRACT

Social network data has been modeled with several approaches, including Sociogram and Sociomatrices, which are popular and comprehensive. Similar to these we have developed here a novel modeling technique based on granular computing theory and fuzzy neighborhood systems, which provides a uniform framework to represent social networks. In this model, a social network is represented with a collection of granules. Fuzzy sets are used for defining the granules. The model is named Fuzzy Granular Social Network (FGSN). Familiar measures of networks viz. degree, betweenness, embeddedness and clustering coefficient are redefined in the context of this new framework. Two measures, namely, entropy of FGSN and energy of granules are defined to quantify the uncertainty involved in FGSN arising from fuzziness in the relationships of actors. Experimental results demonstrate the applicability of the model in two well known problems of social networks, namely, target set selection and community detection with comparative studies.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Popularity of on-line social networks like Twitter, Facebook, WhatsApp is increasing day by day. Active presence of the urban society in the e-Universe opens a new avenue of research opportunities. These networks are dynamic, large scale and complex. For a long time, sociologists and economic analysts worked in this field with off-line social network data. But, the data is now available from the on-line social networks which is characterized by large volume, velocity and variety. This forces computer science researchers to come up with new tools and algorithms to analyze these networks effectively and efficiently.

Apart from social and economical significance analysis, we can classify the research in social network analysis broadly into four groups, namely, (a) analysis of network values [9,19,52], (b) community detection [3,34,4], (c) link predictions [23,26] and (d) evolution of networks [24]. Trivial approach to analyze a network is to model it with graphs and use the networks analysis tools. Other modeling techniques to work with social network data include, statistical model, sociomatrices model, algebraic model, and agent-based model. There has been a development of game theoretic modeling of the network as well. We will discuss more on these in Section 2.

The goal of this paper is to develop a unified framework to model social networks effectively and efficiently. A social network is viewed as a collection of relations between social actors and their interactions. These actors form closely operative groups, which are often indistinguishable in the process of problem solving. This resembles the concept of granules. As

\* Corresponding author.

E-mail addresses: [suman@sumankundu.info](mailto:suman@sumankundu.info) (S. Kundu), [sankar@isical.ac.in](mailto:sankar@isical.ac.in) (S.K. Pal).

described in [51], a granule is a clump of objects (points) in the universe of discourse, drawn together, for example, by indistinguishability, similarity, proximity or functionality. Significance and merits of granular computing in data mining and knowledge discovery are adequately addressed in [38,40]. This motivates us to model a social network in granular computing framework.

In addition, the basic concepts of “conceptual similarities” ‘between nodes’, ‘cluster of nodes’, ‘relation between nodes and their interactions’, etc. do not lend themselves to precise definition, i.e., they have ill-defined boundaries. So, it is appropriate and natural if a social network is represented in terms of *fuzzy granules*. Accordingly, in the proposed model a granule is constructed around a node with fuzzy boundary. The membership function for computing the degree of belonging of a node to the said granule is determined depending upon the problem in hand. Within this framework, we have redefined some of the popularly known network measures under granular space. We have also defined the entropy of the network, which provides a measure of uncertainty arising from fuzziness, as used in the model. Finally, we use some of these new measures to solve the problem of target set selection and community detection, as examples.

The rest of the paper is organized as follows. Section 2 provides some related research work on modeling of a social network and Section 3 reports basic notions of fuzzy set theory and granular computing. In Section 4, we provide the details of fuzzy granular model of social networks (FGSN). Section 5 contains some modified social measures as applicable to the fuzzy granular model of network. Section 6 reports the entropy measures of the network. Application of FGSN on target set selection and community detection problems along with their results is given in Section 7. Finally, in Section 8, we conclude with discussions on the applicability and future prospects of the model.

## 2. Related approach

The network structure with actors and their relationships is usually modeled as graphs. In sociology, it is sometime referred as sociogram. In a sociogram, actors are represented by vertex of a graph, and relations by edges. Graphs appear naturally here as it is useful to represent how things are either physically or logically linked together.

Social network data, sometime represented in two-way matrices, is termed as sociomatrices [46]. The two dimensions of a sociomatrix are indexed by the senders (rows) and the receivers (column) of relationships. When the relation is of dichotomous type, we obtain the sociomatrix exactly same as the adjacency matrix. Sociogram and sociomatrices were first used by Moreno [29] who showed how social relationship can be pictured through these.

The same network can also be represented using the relational form. Relational algebras (also called role algebras) are used to analyze the structure of social roles by emphasizing multiple relations rather than actors. Harrison White and his students [2,47] pioneered this approach as an extension to block modeling.

Another approach to model networks is using the statistical model. The idea of statistical modeling of network is to represent the main features of the social network by a few parameters and express the uncertainty of those estimates by standard error,  $p$ -value, posterior distribution, etc. There are two approaches for statistical modeling of network, viz. model-based inference and design-based inference. When a sample is drawn from a larger graph, design-based method can be used. Example of this technique is link-tracing design [44]. On the other hand, in model-based inference it is required to construct a probability model with the assumption that the observed data can be regarded as the outcome of a random draw from this model [11,13].

Thus several models for describing social network exist starting from 1930s. Recently, the development on modeling social network problems using multi-agent theory and/or game theory has been observed. In their paper [21], Kleinberg et al. modeled a network with  $n$  distinct agents who build link to one another based on a strategic game. The payoff to an agent arises as a difference of costs and benefits. Narayanam et al. [31], on the other hand, mapped the information diffusion process of social network to the formation of coalitions in an appropriately defined cooperative game. In [16], authors modeled the user interactions of a network to explore the dynamic evolutionary process of knowledge sharing among users using agent-based computational approach. But the focus of these researches is mostly problem centric.

Fuzzy set theory has also received attention on social network analysis in recent years. In their work, Nair and Sarasamma [30] analyzed multi-modal social networks using fuzzy graphs and referred it as fuzzy social network. Later in 2008, Davis and Carley [8] used a stochastic model to identify fuzzy overlapping groups in social networks. Here they modeled the fuzzy overlapping group detection using an optimization problem. Another area where fuzzy sets have been used by different scientists is positional analysis (finding similarities between actors in the network) of social networks [10]. Instead of a general framework, these recent developments of fuzzy set theoretic approach in social network are more focused on a particular type of the network or particular application on the network. The fundamental difference of fuzzy social network and the proposed fuzzy granular social network is that in the former, individuals are treated as actors, whereas, in the latter we treat a group of individuals, i.e., a granule, as an actor.

Beside these, an attempt was made to use the concept of granular computing to model relational database for association discovery [15]. The technique is a specialized version of the general relational data mining framework which efficiently provides the search space for association discovery. Also, there were several research investigations focused on a problem oriented modeling of social network using different soft computing tools [3,22,25,45]. However, none of these techniques provides any general framework which can serve as a generic platform, similar to sociogram or sociomatrices, to analyze social network data in view of different problems in the field.

### 3. Fuzzy sets and granular computing

In this section, the basic notions of fuzzy sets and granular computing are mentioned in brief.

#### 3.1. Fuzzy sets

Traditional set theory deals with whether an element “belongs to” or “does not belong to” a set. Fuzzy set theory [50], on the other hand, concerns with the continuum degree of belonging, and offers a new way to observe and investigate the relation between sets and its members. It is defined as follows:

Let  $X$  be a classical set of objects, called the universe. A fuzzy set  $A$  in  $X$  is a set of ordered pairs  $A = \{(x, \mu_A(x)) | x \in X\}$ , where  $\mu_A : X \rightarrow M$  is called the membership function of  $x$  in  $A$  which maps  $X$  to membership space  $M$ . Membership  $\mu_A(x)$  indicates the degree of similarity (compatibility) of an object  $x$  to an imprecise concept, as characterized by the fuzzy set  $A$ . The domain of  $M$  is  $[0, 1]$ . If  $M = \{0, 1\}$ , i.e., the members are only assigned either 0 or 1 membership value, then  $A$  possesses the characteristics of a crisp or classical set.

The set of all elements having positive memberships in fuzzy set  $A$  constitutes its support set, i.e.,

$$\text{Support}(A) = \{x | \mu_A(x) > 0\} \quad (1)$$

The cardinality of the fuzzy set  $A$  is defined as

$$|A| = \sum_{x \in X} \mu_A(x) \quad (2)$$

Union and intersection of two fuzzy sets  $A$  and  $B$  are also fuzzy sets and we denote them as  $A \cup B$  and  $A \cap B$  respectively. The membership functions characterizing the union and intersection of  $A$  and  $B$  are as follows:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)), \quad x \in X \quad (3)$$

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)), \quad x \in X \quad (4)$$

#### 3.2. Granular computing

Granular computing (GrC) is a problem solving paradigm with the basic element, called granules. The construction of granules is a crucial process, as their sizes and shapes are responsible for the success of granular computing based models. Further, the inter and intra relationships among granules play an important role. A granules may be defined as the clump of elements that are drawn together, for example, by indiscernibility, similarity and functionality. Each of the granules according to its shape and size, and with a certain level of granularity may reflect a specific aspect of the problem. Granules with different granular levels may represent a system differently.

Granulation is the process of construction, representation and interpretation of granules. It involves the process of forming larger objects into smaller and smaller into larger based on the problem in hand. According to Zadeh [51], “granulation involves a decomposition of whole into parts. Conversely, organization involves an integration of parts into whole.”

One of the realizations behind GrC is that – precision is sometimes expensive and not very meaningful in modeling and controlling complex systems. When a problem involves incomplete, uncertain and vague information, it may sometimes become difficult to differentiate the individual elements, and one may find it convenient to consider granules to represent a structure of patterns evolved by performing operations on the individual patterns [12]. Accordingly, GrC became an effective framework in designing efficient and intelligent information processing systems for various real life decision-making applications. The said framework can be modeled, for example, with the principles of fuzzy sets, rough sets, neural networks, power algebra, interval analysis [41]. For further details on the significance and various applications of GrC, one may refer to [38,35,37,40,48].

### 4. Fuzzy granular social network

A social network is viewed from the stand point of nodes (entities like persons, organizations) and their relationships. Global phenomenon of a social network always ensembles the local behaviors of individuals and their closely related neighborhoods. This motivates us to model a network in terms of granules. Quantifying this vaguely defined term “closeness”, is another concern for modeling the social network in terms of a granular neighborhood system. Fuzzy set comes naturally here to address this issue. In this section, we shall provide a model to describe a social network in terms of fuzzy granular system, and name it Fuzzy Granular Social Network (FGSN).

#### 4.1. The model

Let us consider a social network represented by a graph  $G(V, E)$ , where  $V$  is the set of all nodes (or vertex) and  $E$  represents the relationships (or edges). Let the symbol  $I$  represent the unit interval  $[0, 1]$ . A fuzzy granular neighborhood defined over

the vertex set of a social network  $G$  is a function  $\phi : V \rightarrow A(V)$ , which assigns every node  $v \in V$  to a fuzzy set  $A \in I^V$ . If  $\phi(v)$  is non empty, we call it the fuzzy neighborhood of the node  $v$ , i.e.,  $\phi(v)$  is the granule defined around the node  $v$ . Due to the complex nature of social networks a node could be a member of different such neighborhood sets reflecting its different levels of association. Let a family of fuzzy sets associated with each node  $v \in V$  be  $\Phi(v)$ .  $\Phi(v)$  represents the neighborhood sets of node  $v$ . A fuzzy granular social network is represented by a triple:

$$S = (C, \mathcal{V}, \mathcal{G}) \text{ where} \tag{5}$$

- $\mathcal{V}$  is a finite set of nodes of the network
- $C \subseteq \mathcal{V}$  is a finite set of granule representatives
- $\mathcal{G}$  is the finite set of all granules, i.e.,  $\mathcal{G} = \left\{ \bigcup \Phi(c) \mid c \in C \right\}$

#### 4.2. Directed social network

Directed social networks are represented by a graph in the same manner as in undirected network with the difference in the edge set  $E$ . Here, edges are ordered, i.e., an edge  $(a, b)$  is different than  $(b, a)$ . So, we can represent a directed graph using the fuzzy granular social network with the quadruple,

$$S = (C, \mathcal{V}, \mathcal{G}_{IN}, \mathcal{G}_{OUT}) \text{ where} \tag{6}$$

- $\mathcal{V}$  is a finite set of nodes of the network
- $C \subseteq \mathcal{V}$  is a finite set of granule representatives
- $\mathcal{G}_{IN}$  is the finite set of all granules considering only in-bound relationships
- $\mathcal{G}_{OUT}$  is the finite set of all granules considering only out-bound relationships

**Remark 1.** If one wants to capture the maximum information of the network,  $C$  should be equal to  $\mathcal{V}$ . Note that, the social network data available from online network shows Big Data characteristics. So, a model describing these kinds of networks needs to address the challenging issue of scalability. In this regard, for reducing the execution time of data analysis to a tolerable range one may restrict the number of granules either based on a threshold, decided over the cardinality of the granule, or with human intervention.  $\square$

#### 4.3. Membership functions

Based on the model described above let us consider a social network  $S = (C, \mathcal{V}, \mathcal{G})$ . Let us now define a granule  $g \in \mathcal{G}$  around a representative node  $(c \in C)$  by assigning fuzzy membership values to its neighborhood. The properties of this granule describe the network properties of node  $c$ . Any monotonically non-increasing function is suitable for capturing the properties of node  $c$  and its neighborhood. In our experiment, we use the following fuzzy membership function:

$$\mu_c(v) = \begin{cases} 0 & \text{for } d(c, v) > r \\ \frac{1}{1+d(c,v)} & \text{otherwise} \end{cases} \tag{7}$$

where  $d(c, v)$  is the distance of node  $v \in \mathcal{V}$  to the center  $c \in C$ , and  $r$  is the radius of the granule.

**Remark 2.** Distance function  $d(c, v)$  can be any metric depending upon the problem in hand. For example, one can use

1. minimum *hop*<sup>1</sup> distance from node  $c$  to  $v$ ,
2. or, minimum weighted *hop* distance, i.e.  $d(c, v) = \sum_{e \in P} \omega(e)$  where  $\omega(e)$  is the weight of the edge  $e$  in path  $P$  from  $c$  to  $v$ ,
3. or, reciprocal of the “number of paths” available between  $c$  to  $v$  in conjunction with the minimum *hop* distance.

For directed social network one can construct  $\mathcal{G}_{IN}$  and  $\mathcal{G}_{OUT}$  similarly using directional *hop* distance for calculating the membership values.  $\square$

Further, a node of the social network  $S$ , can belong to more than one granule and in such scenario, the node will have different degrees of belonging to different granules. For a node  $v$  having non-zero membership to more than a granule, membership values can be normalized using Eq. (8) such that all these normalized membership values add up to unity.

$$\tilde{\mu}_c(v) = \frac{\mu_c(v)}{\sum_{i \in C} \mu_i(v)} \text{ such that } \sum_{i \in C} \tilde{\mu}_i(v) = 1 \tag{8}$$

<sup>1</sup> Hop count refers to the number of intermediate nodes through which the information must pass between source and destination.

## 5. Social network measures

A social network, represented by a graph, encompasses social network properties. We usually analyze these by network measures like degree, betweenness, embeddedness and clustering coefficient values. This section shall provide equivalent granular measures for a social network represented by FGSN. Before describing the measures, let us describe an example network FGSN model, in the following subsection.

### 5.1. Example: Zachary karate club

Consider the friendship network of Zachary karate Club [49] shown in Fig. 1(a). This network shows the friendship relations between 34 members of a US karate club in 1970s. Fig. 1 summarizes the statistics about the data set and degree distributions of the network. Let the network be represented by  $G(V, E)$ . Our objective here is to model it by a *Fuzzy Granular Social Network*  $\mathcal{S}(\mathcal{C}, \mathcal{V}, \mathcal{G})$ . So, we need to define three sets  $\mathcal{C}$ ,  $\mathcal{V}$  and  $\mathcal{G}$ .

We first construct a portion of the network in FGSN and then provide with a general structure of this social network. Let us assume that we have two nodes in concern, node 1 and node 34, i.e.,  $\mathcal{C} = \{1, 34\}$ .  $\mathcal{V}$  is the set of all nodes in the network, so,  $\mathcal{V} = V = \{1, 2, 3, \dots, 33, 34\}$ . Let us define the set of granules  $\mathcal{G}$  around node 1 and node 34. We use the membership function described in Eq. (7) with minimum hop distance  $h(c, v)$  as the distance metric and  $r = D$ , the network diameter.<sup>2</sup> So, here  $\mathcal{G} = \{A_1, A_{34}\}$  where  $A_c = \{(v, \mu_c(v)) | v \in \mathcal{V}\}$ . These two granules are shown in Fig. 2 where darker shades of brown represent higher values of membership. As we have used normalized membership values (Eq. (8)), the nodes in less overlapping region may turn to have higher membership than the center nodes of the granules. This indicates that those nodes belong only to a fewer number of granules as compared to the centers. This is intuitively appealing as the former ones have higher possibilities of 'definitely belonging' to a granules than the later ones.

In an ideal case, i.e., for capturing maximum information of the network,  $\mathcal{C}$  should be equal to  $\mathcal{V}$ . With this configuration, the general form of the social network  $\mathcal{S}(\mathcal{C}, \mathcal{V}, \mathcal{G})$  is described as follows:

- $\mathcal{V} = \{v | \forall v \in V\}$
- $\mathcal{C} = \{c | \forall c \in \mathcal{V}\}$
- $\mathcal{G} = \left\{ A_c | \forall c \in \mathcal{C}, A_c \equiv \sum_{v \in \mathcal{V}} \tilde{\mu}_c(v) / v \right\}$

Following subsections describe different measures in the context of FGSN. The examples considered in this section assumed parameter  $r = 2$ .

### 5.2. Granular degree of a node

In the graph representation, a node's degree is measured by the number of incident nodes. The degree of a node in FGSN representation, we call it *granular degree*, is defined by the cardinality of the granule centered at the node. Here each granule is represented by a fuzzy set, so we use Eq. (2) to compute the granular degree of a node  $c$  as

$$\mathcal{D}(c) = |A_c| = \sum_{v \in \mathcal{V}} \tilde{\mu}_c(v) \quad (9)$$

In the karate club example, node 34 has a granular degree of 3.38026 and node 1 has a granular degree of 3.0044. Fig. 3(a) shows the distribution of Granular degree of karate club data.

The formula works as it is for directed network as well, where we compute granular in-degree and granular out-degree as follows:

$$\mathcal{D}_{IN}(c) = |F_c| \quad \text{where } F_c \in \mathcal{G}_{IN} \quad (10)$$

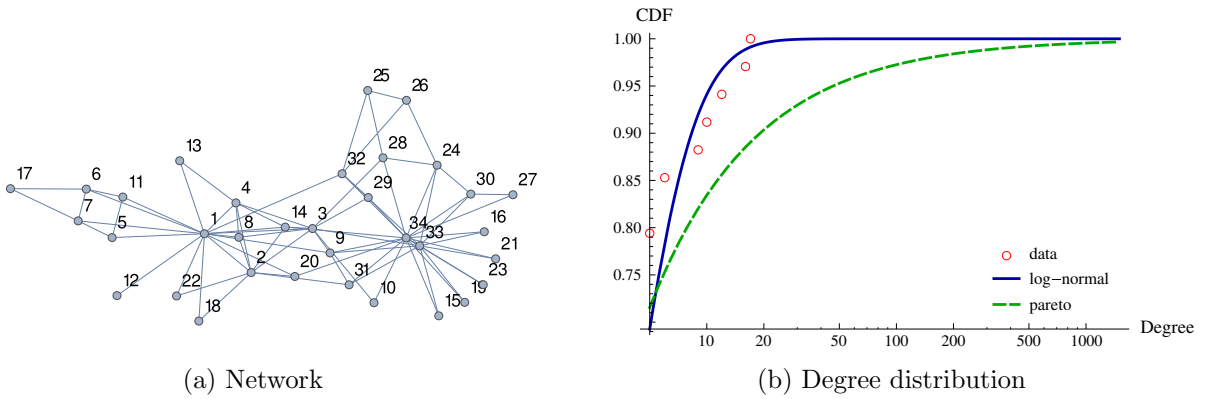
$$\mathcal{D}_{OUT}(c) = |O_c| \quad \text{where } O_c \in \mathcal{G}_{OUT} \quad (11)$$

When a granule ( $A_c$ ) around node  $c$  is represented by a crisp set, the membership  $\tilde{\mu}_c(v)$  of node  $v$  would take values only from the set  $\{0, 1\}$ . If we consider  $r = 1$  here then only nodes directly connected to node  $c$  would get membership 1. And all other nodes of the network will get zero membership. In this case, granular degree of  $c$  boils down to its conventional degree of graph representation.

### 5.3. Granular betweenness of a node

Conceptually, the betweenness of a node quantifies the number of times it is used as a bridge between two other nodes along their shortest path. In graph representation, it is measured by the ratio of the number of such shortest paths passing through a node and the total number of shortest paths in the network. With a fuzzy granular social network, granular

<sup>2</sup> Network diameter is the largest shortest path computed over all pairs of nodes in the network.



Nodes	34
Edges	78
Nodes in largest Weakly Connected Component(WCC)	34
Edges in largest WCC	78
Nodes in largest Strongly Connected Component(SCC)	34
Edges in largest SCC	78
Diameter	5
Avg. clustering coefficient	0.570638

(c) Statistics

Fig. 1. Zechary karate club.

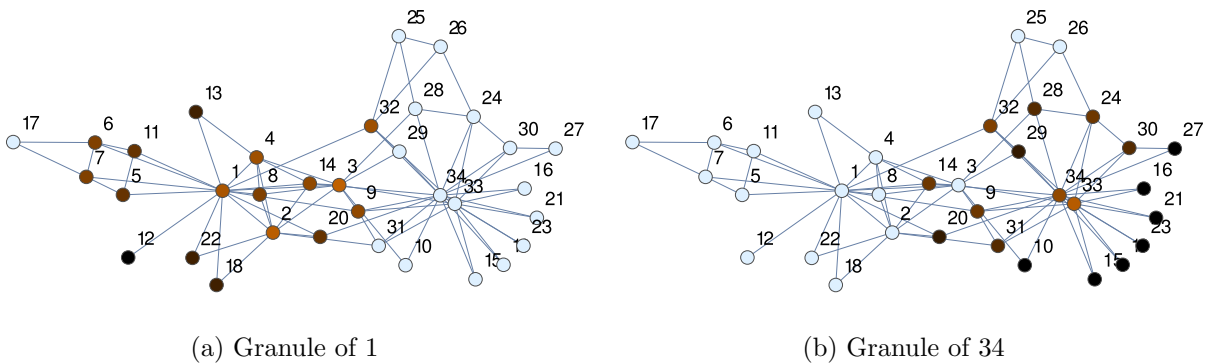


Fig. 2. Color coded granules of social network of Zechary karate club. Darker shade means higher membership value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

betweenness of a representative node  $c$  can be quantified by the sum of membership values that  $c$  possesses for all granules in the system. Using the normalized membership values, granular betweenness of  $c \in \mathcal{C}$  may be defined as

$$B(c) = \frac{1}{\max_{i \in \mathcal{C}} (\hat{\mu}_i(c))} \tag{12}$$

$B(c)$  takes values in  $[0, |\mathcal{C}|]$ . In our example network, granular betweenness of node 1 and node 34 is 9 and 9.5, respectively. The distribution of granular betweenness of karate club data is shown in Fig. 3(b).

#### 5.4. Granular embeddedness of a pair of nodes

Embeddedness, for a graphical representation of social network, is measured for an edge. It is the number of common neighbors the two end points have. Granular embeddedness, however, is differently conceptualized for FGSN because it does

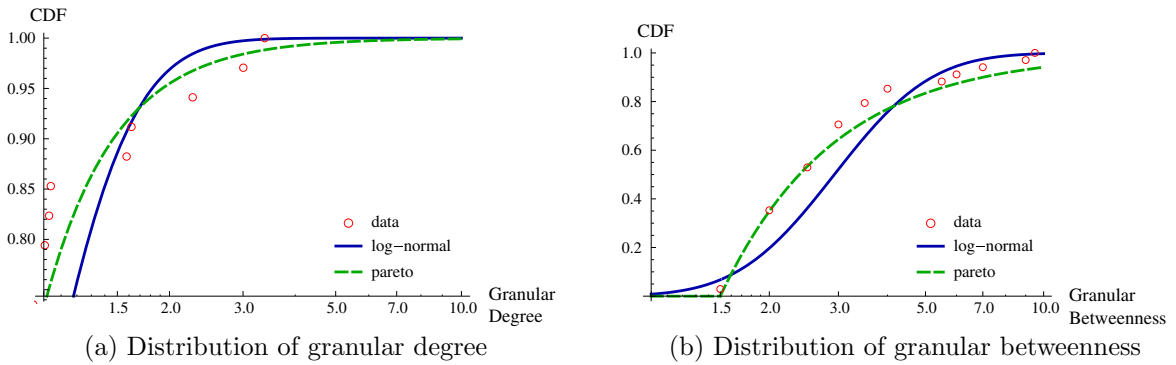


Fig. 3. FGSN of Zachery karate club.

not have a concept of edge. For a network, modeled in terms of fuzzy granules, the embeddedness for any pair of nodes defines how much a granule is embedded inside the other. It may be measured by the cardinality of the intersection of granules centered by the pair of points. Using Eqs. (4) and (2), granular embeddedness of a pair of nodes  $a$  and  $b$  is defined as

$$\mathcal{E}(a, b) = |A_a \cap A_b| = \sum_{v \in \mathcal{V}} \min(\tilde{\mu}_a(v), \tilde{\mu}_b(v)) \tag{13}$$

where  $A_a$  and  $A_b$  are the fuzzy sets representing the granules having the center nodes  $a$  and  $b$ , respectively.

In the example of karate club, the embeddedness of 1 and 34 is found to be 0.610714 when  $r = 2$ , and 0.959073 when  $r = D$ , the diameter of the network.

If we use hop distance as a distance function in measuring the similarity between nodes  $a$  and  $b$  for defining the granules, then the  $\mathcal{E}$ -value will be less when the nodes  $a$  and  $b$  are far apart physically. On the other hand, if we take, say, the retweet counts of Twitter like social network as the distance or similarity measure between nodes  $a$  and  $b$ , i.e., higher number of retweets indicates higher conceptual similarities and have lower conceptual distance between the nodes  $a$  and  $b$ , then the  $\mathcal{E}$ -value will decrease as the conceptual distance between  $a$  and  $b$  increases.

Similar to the crisp equivalence of granular degree, granular embeddedness also converges to embeddedness. For example, consider  $r = 1$  and  $\tilde{\mu} = \{0, 1\}$ . That is, a node gets membership 1 if it is an immediate neighbor of the center node and zero, otherwise. For an edge  $(a, b)$ , the granular embeddedness of its end points (Eq. (13)) is the cardinality of the intersection of the two granules centered at  $a$  and  $b$ . Since only the nodes directly connected to  $a$  (and  $b$ ) have membership of 1 for granule centered around  $a$  (and  $b$ ), the intersection of  $A_a$  and  $A_b$  contains only the common neighbors to both the nodes  $a$  and  $b$ . So the granular embeddedness of nodes  $a$  and  $b$  (which is the cardinality of the said intersection) boils down to the conventional embeddedness in graph representation.

5.5. Triadic closure and granular clustering co-efficient

Triadic closure is the property among three nodes  $a, b$ , and  $c$  such that if a strong tie exists between  $a$  &  $b$  and  $a$  &  $c$ , then there is a weak or strong tie between  $b$  &  $c$ . One of the measures of triadic closure is clustering co-efficient. Before we define the granular clustering co-efficient of FGSN, we have the following definitions.

**Definition 1 (Tie).** In a FGSN, a tie is said to be formed by two nodes  $a, b$  ( $a \neq b$ ) if they satisfy any one of the following:

$$b \in \text{Support}(A_a) \quad \text{i.e.,} \quad \tilde{\mu}_a(b) > 0 \tag{14}$$

$$a \in \text{Support}(A_b) \quad \text{i.e.,} \quad \tilde{\mu}_b(a) > 0 \tag{15}$$

where  $A_a, A_b \in \mathcal{G}$ . Let us denote a tie between node  $a$  and node  $b$  as  $\mathcal{T}(a, b)$ .

**Definition 2 (Weak and Strong Ties).** A tie  $\mathcal{T}(a, b)$  is said to be strong if it satisfies both Eqs. (14) and (15). Otherwise we say, it is a weak tie.

A tie in a graph model of social network is a physical link between two nodes  $a$  and  $b$ . In contrast, for a FGSN the tie represents a conceptual communication channel between  $a$  and  $b$ . The tie will have intensity associated with it which defines the membership of its being a physical tie of the network. A strength of a tie is the average value of the membership of one node in the other's granule and vice-verse.

**Definition 3 (Strength of a Tie).** Strength of a tie  $\mathcal{T}(a, b)$  is measured by

$$\tilde{\mu}(\mathcal{T}(a, b)) = \frac{\tilde{\mu}_a(b) + \tilde{\mu}_b(a)}{2} \tag{16}$$

Let us now define ‘Triangle’ in the network.

**Definition 4 (Triangle).** Three nodes  $i, j$  and  $k$  of a FGSN are said to form a triangle, iff there exist three ties  $\mathcal{T}(i, j)$ ,  $\mathcal{T}(j, k)$  and  $\mathcal{T}(k, i)$ . A triangle is denoted by  $\mathcal{T}(i, j, k)$  and its membership value of being a clique in the network is

$$\tilde{\mu}(\mathcal{T}(i, j, k)) = \begin{cases} \frac{\tilde{\mu}(\mathcal{T}(i, j)) + \tilde{\mu}(\mathcal{T}(j, k)) + \tilde{\mu}(\mathcal{T}(k, i))}{\tilde{\mu}_i(i) + \tilde{\mu}_j(j) + \tilde{\mu}_k(k)} & \text{if } \tilde{\mu}(\mathcal{T}(i, j)), \tilde{\mu}(\mathcal{T}(j, k)), \tilde{\mu}(\mathcal{T}(k, i)) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

**Definition 5 (Granular Clustering Co-efficient).** Clustering co-efficient of a vertex in the graph representation quantifies how close its neighbors are for being a clique. It is measured by the ratio of cliques at the node and the total number of triples the node involves in. A triple is formed by two strong ties. Similar to the clustering co-efficient in the graph representation we now define the granular clustering co-efficient in the context of FGSN.

Let us assume that all the triangles, where a node  $i$  is involved in, are represented by the fuzzy set,

$$\delta(i) = \sum_{j, k \in \mathcal{C}} \tilde{\mu}(\mathcal{T}(i, j, k)) / \mathcal{T}(i, j, k)$$

The granular clustering co-efficient of a node  $i$  is the fraction of triangles, node  $i$  is involved in. It may be represented as

$$CC(i) = \frac{1}{\eta} \times |\delta(i)| = \frac{1}{\eta} \times \sum_{j, k \in \mathcal{C}} \tilde{\mu}(\mathcal{T}(i, j, k)) \quad (18)$$

where  $\eta$  is the number of triples that node  $i$  forms with all other nodes in the network, i.e.,

$$\eta = \binom{|\{A | A \in \mathcal{G} \text{ and } i \in \text{Support}(A)\}| - 1}{2}$$

Granular clustering co-efficient of the network is defined as

$$CC(\mathcal{S}(\mathcal{C}, \mathcal{V}, \mathcal{G})) = \frac{1}{|\mathcal{C}|} \times \sum_{i \in \mathcal{C}} \frac{|\delta(i)|}{\eta} = \frac{1}{|\mathcal{C}|} \times \sum_{i \in \mathcal{C}} \sum_{j, k \in \mathcal{C}} \frac{\tilde{\mu}(\mathcal{T}(i, j, k))}{\eta} \quad (19)$$

In the aforesaid karate club network, the granular clustering co-efficient is 0.00125334 when  $r = 2$ , and 0.00308488 when  $r = D$ , the network diameter.

## 6. Uncertainties in FGSN

Uncertainty in a social network arises due to the presence of vaguely defined closeness between nodes. In a network, relationships are not crisp, i.e., the presence of a relationship (link) in a network does not imply that both the nodes are 100% committed towards each other. Similarly, the absence of a link does not necessarily mean they are not completely following each other. In other words, each relationship has a degree of togetherness.

Let us define two measures of uncertainties in FGSN in terms of fuzziness, as follows:

### 6.1. Energy measure of a granule in FGSN

Let us consider a monotonically increasing mapping  $e : [0, 1] \rightarrow [0, 1]$  with the boundary conditions  $e(0) = 0$  and  $e(1) = 1$ . An energy measure of a granule  $A_c \in \mathcal{G}$ , denoted by  $\mathbb{E}(A_c)$ , is a function of its characterizing membership values, represented as

$$\mathbb{E}(A_c) = \sum_{x \in \mathcal{V}} e[\tilde{\mu}_c(x)] \quad (20)$$

This measure quantifies the energy associated with the granule  $A_c$ . The energy increases as the membership values of its supporting nodes increase. The energy measure of  $A_c$  reduces to its cardinality if we use the identity mapping  $e(x) = x \forall x \in \mathcal{V}$ , i.e.,

$$\mathbb{E}(A_c) = \sum_{x \in \mathcal{V}} \tilde{\mu}_c(x) = |A_c| \quad (21)$$

One can also think of a different functional for  $e$  other than the identity mapping, for example,  $e(x) = x^a, a > 0$  or  $e(x) = \sin(\frac{\pi}{2}x)$ .

## 6.2. Entropy measure of FGSN

Given a FGSN  $\mathcal{S}(\mathcal{C}, \mathcal{V}, \mathcal{G})$ , each granule  $A_c \in \mathcal{G}$  represents a fuzzy equivalence class under the attribute set  $\mathcal{C}$ . If we have  $n$  objects in the universe  $\mathcal{V}$  then the fuzzy relative frequency [28] of a granule will be

$$\rho(A_c) = \frac{|A_c|}{n} \quad (22)$$

where  $|A_c|$  is the cardinality of the granule  $A_c$ . Based on this relative frequency of granules, one can find the information gain of a FGSN through its entropy, using Shannon's logarithmic function, as

$$H(\mathcal{S}) = - \sum_{A_c \in \mathcal{G}} \rho(A_c) \log_{\beta}(\rho(A_c)) \quad (23)$$

where  $\beta$  represents the base of logarithm. Applying Eq. (22) into Eq. (23) we get

$$H(\mathcal{S}) = - \frac{1}{n} \sum_{A_c \in \mathcal{G}} |A_c| \log_{\beta} \left( \frac{|A_c|}{n} \right) \quad (24)$$

The value of  $H(\mathcal{S})$  can vary in  $[0, \log_{\beta}(|\mathcal{C}|)]$ .  $H(\mathcal{S}) = 0$  means the FGSN is least uncertain, while its value equal to  $\log_{\beta}(|\mathcal{C}|)$  signifies the highest uncertainty.

## 7. Application and results

In this section we show how the proposed fuzzy granular social network model and the different measures provided in the earlier section can provide solutions to the well known problems of social network analysis, namely, target set selection and community detection.

### 7.1. Configuration and hardware/software platforms

A social network is represented both in its graph  $(G(V, E))$  and fuzzy granular  $(\mathcal{S}(\mathcal{C}, \mathcal{V}, \mathcal{G}))$  forms. For this purpose following mapping has been used irrespective of the data set.

- $\mathcal{V} = \{v | \forall v \in V\}$
- $\mathcal{C} = \{c | \forall c \in \mathcal{V}\}$
- $\mathcal{G} = \left\{ A_c | \forall c \in \mathcal{C}, A_c \equiv \sum_{v \in \mathcal{V}} \tilde{\mu}_c(v) / v \right\}$

Membership value  $\tilde{\mu}_c(v)$  is calculated based on Eq. (8). The value of the parameter  $r$  (radius of Granule) is chosen based on the data set.

*Hardware/software platforms.* All the experiments are conducted on an "HP Compaq 8100 Elite Convertible Minitower" PC equipped with Intel Core i7 880 processor (3.07 GHz CPU speed and 8192 MB Cache) and 16 GB of DDR3 RAM. Algorithms are written and executed in Mathematica 9 under the Fedora 20 operating system.

### 7.2. Data set description

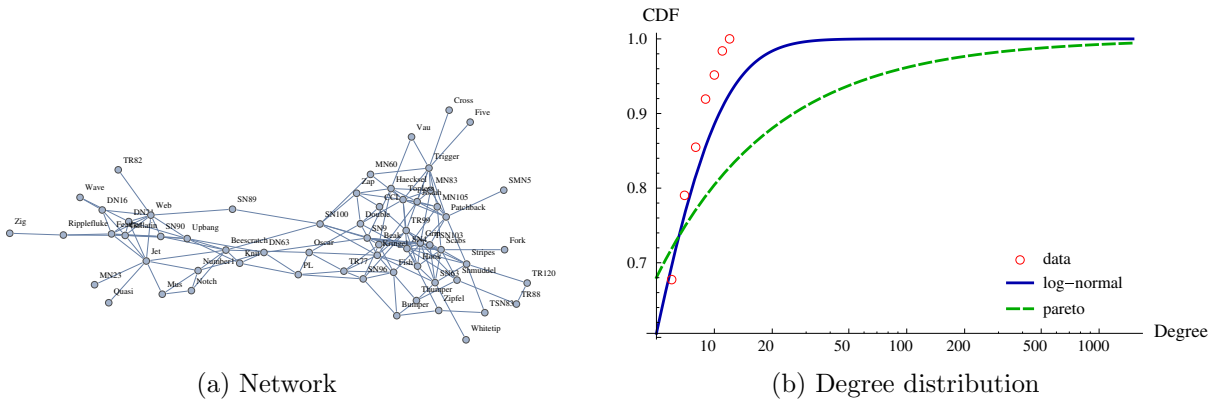
In the experiments, we use three data sets, namely Zachary karate club [49], Dolphin social network [27], and Political blog network [1]. In Section 5.1 we have described the inherent characteristics of Zachary karate club. Here we include the details of the other two data sets used before describing the results.

*Dolphin social graph.* The network of frequent associations among 62 bottlenose dolphins living in Doubtful Sound, New Zealand was collected between 1995 and 2001 by Lusseau et al. [27]. The network is an undirected graph of their interactions, and properties of the network are given in Fig. 4.

*Political blogs network.* The network of political blog was collected in 2005 by Adamic and Glance [1] which is a directed network of hyperlinks between weblogs on US politics. The network has 1490 nodes and 16,718 edges. The detail of the network is provided in Fig. 5.

### 7.3. Target set selection problem

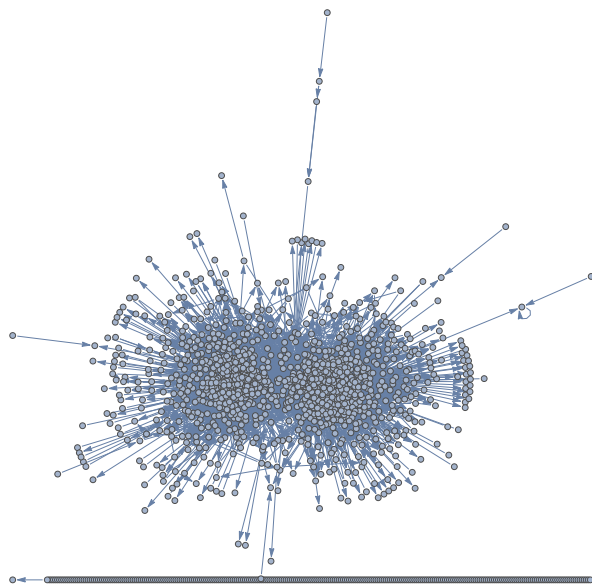
Domingos and Richardson [9,43] were the first to study this problem from the angle of the algorithmic aspect, and later other researchers [17,18,20] studied the problem as a discrete optimization problem. Here we shall describe the optimization problem as applicable to FGSN, readers may refer to [17,18,20,45,6] for problem statement based on graph representation.



Nodes	62
Edges	159
Nodes in largest Weakly Connected Component (WCC)	62
Edges in largest WCC	159
Nodes in largest Strongly Connected Component (SCC)	62
Edges in largest SCC	159
Diameter	8
Avg. clustering coefficient	0.258958

(c) Statistics

Fig. 4. Dolphin social graph.



Nodes	1490
Edges	16718
Nodes in largest WCC	1222
Edges in largest WCC	16717
Nodes in largest SCC	1
Edges in largest SCC	0
Diameter	10
Avg. clustering coefficient	0

(b) Statistics

Fig. 5. Political blogs network.

Let us consider an influence function  $\sigma : 2^V \rightarrow \mathbb{N}$ , defined for a social network  $S(\mathcal{C}, \mathcal{V}, \mathcal{G})$ , such that given a set of initial active nodes  $K \in 2^V$ ,  $\sigma(K)$  returns the expected number of active nodes at the end of information cascade. The problem of target set selection is to find the  $k$  number of influential nodes for which influence in  $S$  is maximum. So, this is a maximization problem defined as

$$\begin{aligned} & \max_K \quad \sigma(K) \\ & \text{subject to} \quad |K| = k, \quad k > 0 \end{aligned}$$

In the experiments, we simulated independent cascade model [17] of information diffusion using Monte Carlo process. We compared the total influence of  $k$  nodes selected according to their granular degrees with those of the following three algorithms:

- High Degree Heuristics: Top  $k$  nodes are selected based on the degree (for directed network in-degree) of the nodes.
- Random:  $k$  nodes are selected randomly in the network.
- Diffusion Degree Heuristics [36]: Top  $k$  nodes are selected based on the diffusion degree score of the network.

7.3.1. Results

We first selected the top- $k$  nodes (that is, the centers of the granules), from a given FGSN, according to the descending order of granular degree value. We refer this algorithm as Granular degree heuristic. Then we pass these top  $k$  nodes, as the set of seeds, in the Monte Carlo simulation of information diffusion (independent cascade model). The output of the simulation process represents the number of total nodes influenced due to the input seed set. We have varied the value of  $k$  from 1 to 15. We also repeated the experiments for all the comparative algorithms. These results are reported in Fig. 6. Here X-axis shows the value of  $k$  and the Y-axis presents the total number of nodes influenced. As the Monte Carlo process is a stochastic process, we executed each experiment for 10,000 trials and reported here the average values. It is clear, from the figure that, for Zachery karate club and Dolphin social network, results obtained with the proposed measure outperform those obtained by other algorithms for most values of  $k$ . This signifies that the set of seeds selected using the proposed method on FGSN is able to determine the superior top  $k$  influential nodes. For Political blogs, the performance is at par with the High Degree Heuristic, superior to random and inferior to Diffusion Degree Heuristics.

Fig. 7 shows the effect of the granule radius ( $r$ ) on the performance. We experimented with  $r$  varying from 1 to 5, and the results are reported for three different data sets. For all three cases, the best results are obtained for  $r = 2$ . Other values of  $r$  provide poorer performance. Radius of 1, i.e.,  $r = 1$  considers only immediate neighbors of granule as its supporting nodes. For, target set selection problem, this does not characterize a node's indirect influence. Hence, it produces poor performance

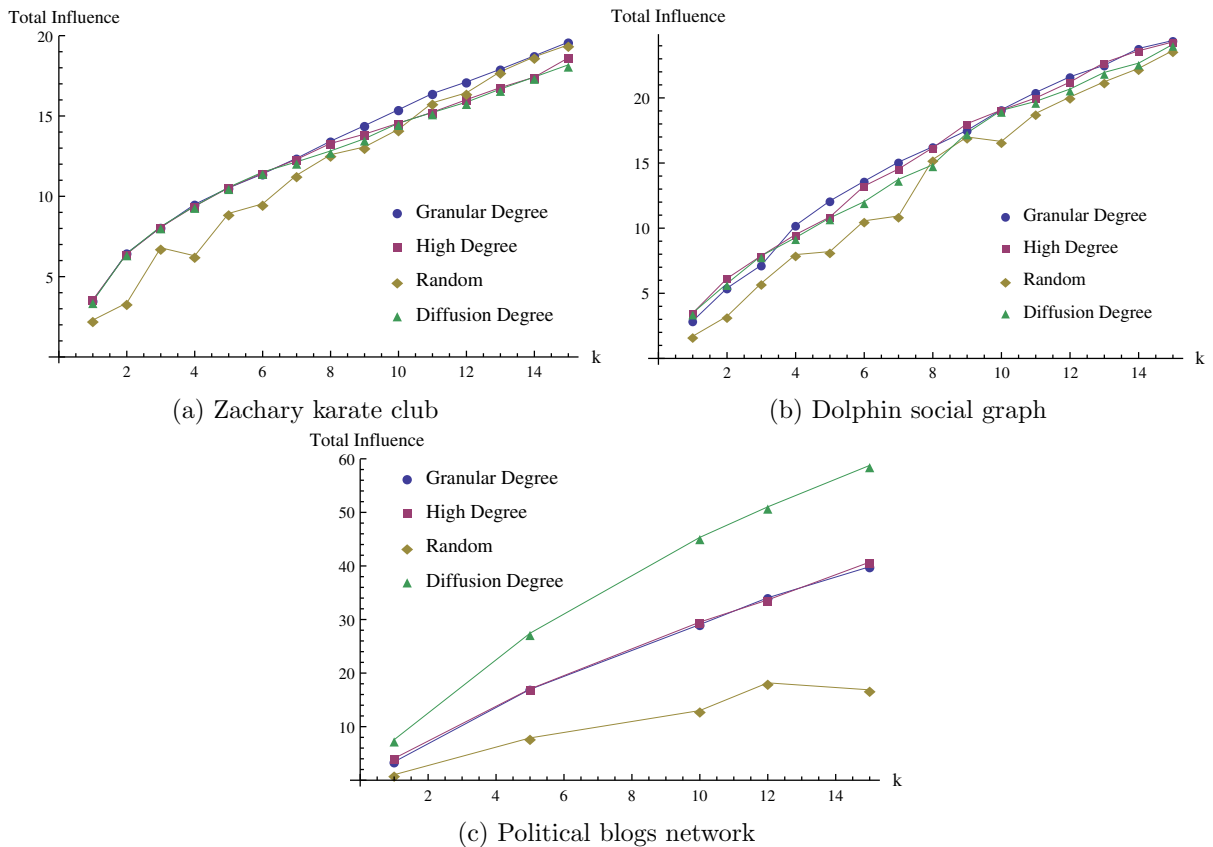


Fig. 6. Variation of total influence with  $k$  for different algorithms ( $r = 2$ ).

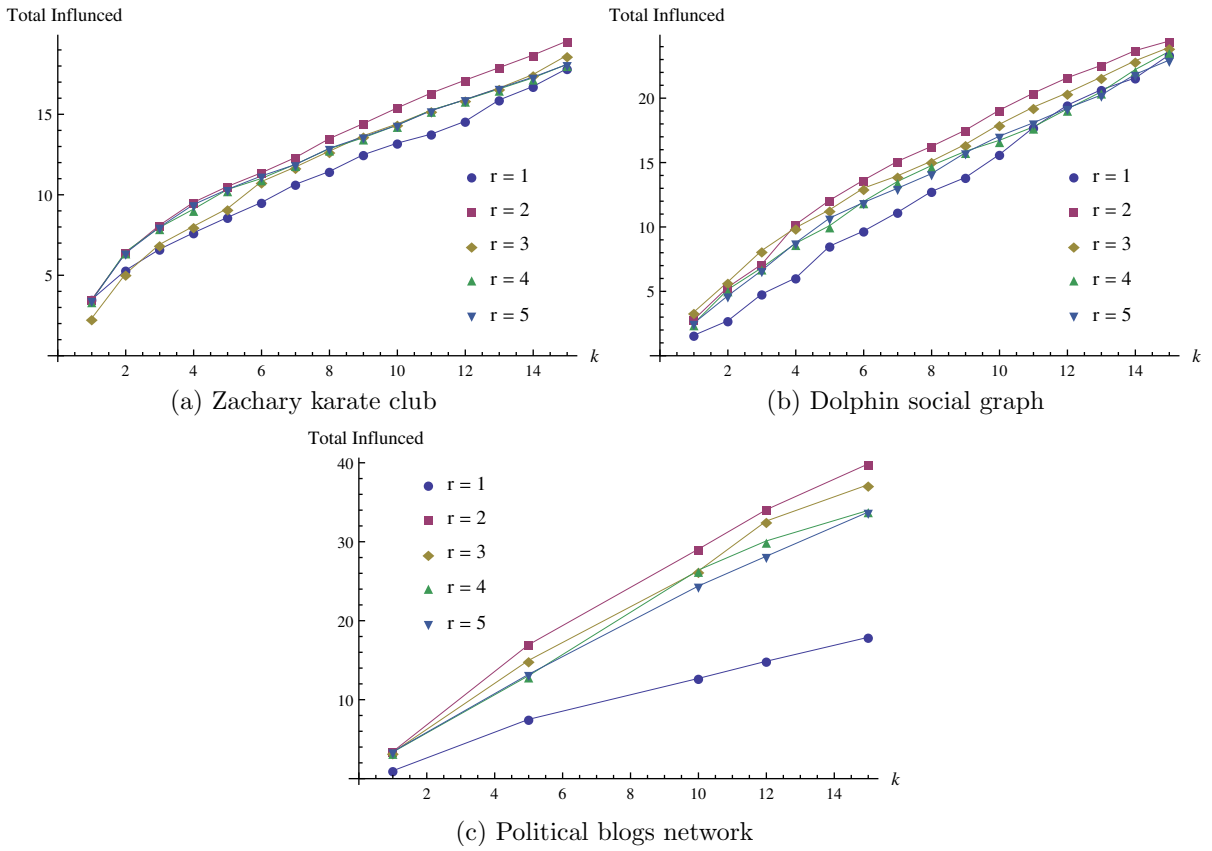


Fig. 7. Variation of total influence with  $k$  for different values of  $r$  corresponding to granular degree.

as compared to those of  $r = 2$ . On the other hand, a value of  $r > 2$  theoretically considers the indirect influences in greater extent, and due to the stochastic nature of information diffusion, this causes more and more uncertainty into the system. Thus it wrongly identifies the top- $k$  nodes, deteriorating the performance.

Execution time (in seconds) of different algorithms for 1000 runs is shown in Table 1. As expected, the random selection method needs lowest time for all the data sets. Diffusion degree heuristics, on the other hand, takes longest time for all the cases. The proposed Granular degree heuristic requires much lower execution time as compared to diffusion degree for all the data sets. For Zachary karate club and Dolphin social graph, it is almost as fast as the high degree heuristics. For Political blog network, however, the proposed algorithm takes longer time compared to high degree heuristics.

We mention that the time taken for the initial modeling of FGSN is seen to be  $3.61$ ,  $12.54$  and  $7.09 \times 10^3$  for the Zachary karate club, Dolphin social graph and Political blog data respectively. Once the modeling is complete, algorithms for other tasks of network analysis can be formulated.

### 7.3.2. Computational complexity

The algorithm of target set selection using granular degree runs on fuzzy granular social networks. When analyzing the computational complexity here, we need to consider the time complexity of conversion of a social network from graph representation to fuzzy granular representation. This is done by assigning membership values to the neighborhood nodes which can be done by traversing the graph in a breadth-first manner. The worst case complexity of the same is  $O(|V| + |E|)$ , where  $|V|$  is the number of nodes and  $|E|$  is the number of edges in the network.

**Table 1**  
Execution time (in seconds) of different target set selection algorithms for 1000 runs.

Algorithms	Data sets		
	Zachary karate club	Dolphin social graph	Political blogs network
Granular Degree Heuristics	0.311	0.52	27.26
High Degree Heuristics	0.3	0.48	3.7
Random Selection	0.2	0.2	0.5
Diffusion Degree Heuristics	12.19	16.532	$9.29 \times 10^4$

In the next step, we measure the granular degree of each granule, for which the complexity is  $O(|C|)$ , and select the top  $k$  nodes. The final selection of top  $k$  nodes requires  $O(kn)$  time.

Thus the total time complexity of the algorithm to find  $k$ -top influential nodes using the proposed method is  $O(|V| + |E| + |C| + kn)$ .

#### 7.4. Community detection

Another important problem in social network analysis is community detection. It is very useful because nodes belonging to the same community also possess some common characteristics and thus provide better insights of the network. We can define communities within networks as subsets of nodes, which are more densely connected as compared to the rest of the network [7]. The study of community detection has a long history, and it is closely related to the graph partitioning in graph theory and hierarchical clustering in sociology. One of the highly cited investigations in this area of hierarchical clustering is by Newman and Girvan [33]. Their techniques discover natural groups based on different similarity measures or strength of connections. They also provided a measure, namely *modularity*, for evaluating the detected communities. It is defined as

$$Q = \sum_i (e_{ii} - a_i^2) \quad \forall i \quad (25)$$

where  $e_{ii}$  is the observed fraction of edges within the group  $i$  and  $a_i^2$  is the expected fraction of edges within the same group  $i$ . If  $e_{ij}$  is the fraction of edges in the network that connects community  $i$  to  $j$ , then  $a_i = \sum_j e_{ij}$ .  $Q$  can vary between  $[-1/2, 1)$  and a value close to 1 indicates stronger community structure.

In this section, we define granular modularity and a new community detection algorithm in the context of FGSN, and apply the same on the aforesaid data sets to detect the community structures.

##### 7.4.1. Granular modularity

In FGSN we represent a network with a collection of granules. A node has membership to more than one granule. Intuitively, a granule has two-way relations, inwards and outwards. For a granule in question, the “inward” relation concerns with the membership of nodes with respect to that granule, whereas the “outward” means node membership to other granules. With this notions we define granular modularity as

$$Q(A_c) = \frac{\sum_{v \in \text{Support}(A_c)} \sum_{u \in \text{Support}(A_c)} \tilde{\mu}_v(u) - \sum_{v \in \text{Support}(A_c)} \sum_{u \notin \text{Support}(A_c)} \tilde{\mu}_v(u)}{|V|} \quad (26)$$

The first part of the numerator, i.e.,  $\sum_{v \in \text{Support}(A_c)} \sum_{u \in \text{Support}(A_c)} \tilde{\mu}_v(u)$ , corresponds to inwards connections and signifies how tightly the members of the granule are oriented. And the second part, i.e.,  $\sum_{v \in \text{Support}(A_c)} \sum_{u \notin \text{Support}(A_c)} \tilde{\mu}_v(u)$  indicates how much the members of the concerned granule are coupled to other granules in the network. We measure the granular modularity of the network as a summation of granular modularity of all the individual granules in the network. Accordingly, the granular modularity of a social network represented in FGSN is

$$Q = \sum_{A_c \in \mathcal{G}} Q(A_c) \quad (27)$$

Granular modularity varies between  $[-0.5, 1)$  and it increases when granules have more inward memberships than outwards, i.e., the granules become more modular.

##### 7.4.2. Algorithm

Given a FGSN, i.e., a collection of granules, as input, the algorithm initially considers each granule as a community. Then it finds similar granules and merges them to obtain a larger community. To identify which granules would be merged, we use granular embeddedness measure. The algorithm runs as follows:

###### Step 1 Initialize:

- (a)  $\mathcal{F} = \mathcal{S}(C, \mathcal{V}, \mathcal{G})$ ;
- (b)  $\mathbb{C} \leftarrow \emptyset$  (Set of Community);  $\mathbb{S} \leftarrow \emptyset$  (Set of Centers);
- (c)  $\mathbb{P} \leftarrow$  All possible pairs of  $C$
- (d)  $i \leftarrow 0$  (Index of  $\mathbb{P}$  during iteration)

###### Step 2 Find Granular Embeddedness for all values of $\mathbb{P}$

###### Step 3 Order $\mathbb{P}$ based on the descending values of their Granular Embeddedness

###### Step 4 $(p, q) \leftarrow \mathbb{P}[i]$

###### Step 5 If $p \notin C$ as well as $q \notin C \forall C \in \mathbb{C}$ then

- (a) create new community  $C' \leftarrow \{p, q\}$  and store in  $\mathbb{C}$
- (b) Select cluster center between  $p$  and  $q$  based on the higher value of Granular Degree and store in  $\mathbb{S}$
- (c) Go to Step 9

###### Step 6 If $p \notin C, \forall C \in \mathbb{C}$ and $q \in \mathbb{S}$ then

- (a) Select  $C \in \mathbb{C}$  for which  $q$  is a member and put  $p$  in the same community
- (b) Replace  $q$  with  $p$  in  $\mathbb{S}$  if Granular Degree of  $p$  is higher than that of  $q$
- (c) Go to Step 9

**Step 7** If  $q \notin C, \forall C_i \in \mathbb{C}$  and  $p \in \mathbb{S}$  then

- (a) Select  $C \in \mathbb{C}$  for which  $p$  is a member and put  $q$  in the same community
- (b) Replace  $p$  with  $q$  in  $\mathbb{S}$  if Granular Degree of  $q$  is higher than that of  $p$
- (c) Go to Step 9

**Step 8** Skip the pair if  $p, q \in C$ , for any  $C \in \mathbb{C}$

**Step 9**  $i \leftarrow (i + 1)$ ; if  $i <$  the length of  $\mathbb{P}$  then Go to Step 4 otherwise Stop.

After getting the output from the above algorithm, we merge the granules of the same community using the union operation of fuzzy sets, and if require, repeat the same algorithm with the said reduced number of granules. Let us name this granular embeddedness based community detection algorithm as GranE-FGSN.

7.4.3. Results

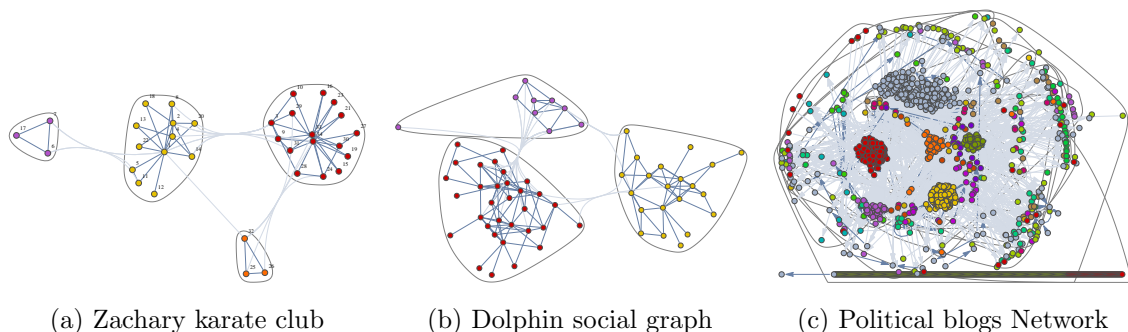
We executed the proposed GranE-FGSN on all the three data sets. We compared the output with three popular community detection algorithms viz. centrality based method [14] (referred as Cen-CD), modularity optimization based community detection method [32] (referred as Mod-CD) and  $k$ -clique percolation method (CPM-CD) [39].

The numbers of communities detected by different algorithms are listed in Table 2. For convenience, pictorial representations of the community structures are shown in Fig. 8. In Zachary karate club, the proposed GranE-FGSN is seen to detect four communities which is one less than that obtained by Cen-CD and CMP-CD methods, and one more than by Mod-CD algorithm. In case of Dolphin social graph it is three for GranE-FGSN compared to 4, 5, and 20 for Mod-CD, Cen-CD and CPM-CD respectively. However, for the Political blogs network, variation of the numbers of detected communities by different algorithms is very prominent. In addition, there are very large number of communities detected with less than 10 nodes, unlike the other two data sets. These phenomena can be justified in terms of the clustering co-efficient. For example, Political blogs network have zero clustering co-efficient (Fig. 5(b)) unlike Zachary karate club (Fig. 1(c)) and Dolphin social graph (Fig. 4(c)), signifying that the Political blogs network does not possess prominent community structure. This may be the reason of the erratical behavior of the number of detected communities (see the last column of Table 2).

We now check the goodness of these identified communities with modularity ( $Q$ ) and granular modularity ( $\mathcal{Q}$ ) measures. We compute the modularity ( $Q$ ) value for each of the output community structures using Eq. (25). Table 3 reports the same for different data sets. Modularity values found for different algorithms are positive. This indicates that the communities found by these algorithms have higher number of edges within their own community as compared to other communities. It is also evident that, the modularity of proposed GranE-FGSN is higher than CPM-CD for all the data sets. In case of

**Table 2**  
Number of communities detected by different algorithms on different data sets.

Algorithms	Data sets		
	Zachary karate club	Dolphin social graph	Political blogs network
GranE-FGSN	4	3	99 (17 with 10 or more node)
Mod-CD	3	4	280 (3 with 10 or mode node)
Cen-CD	5	5	456 (2 with 10 or more node)
CPM-CD	5	20	496 (1 with 10 or more node)



**Fig. 8.** Community structures.

**Table 3**  
Modularity value of different algorithms.

Algorithms	Data sets		
	Zachary karate club	Dolphin social graph	Political blogs network
GranE-FGSN	0.37	0.47	0.25
Mod-CD	0.38	0.49	0.43
Cen-CD	0.40	0.52	0.42
CPM-CD	0.22	0.37	0.002

Dolphin social graph and Zachary karate club, the proposed one provides results comparable to Mod-CD. Cen-CD is the best for Zachary karate club and Dolphin social graph although it takes largest time (Table 5).

Granular modularity of three data sets for different methods are listed in Table 4. In Dolphin social graph, granular modularity is highest for the proposed GranE-FGSN. On the other hand, for Zachary karate club it is higher than Cen-CD, comparable to Mod-CD and lower than CPM-CD.

Table 5 shows the execution time of different algorithms. Results, shown in seconds, correspond to 1000 executions of each of the algorithms. From the table it is evident that the proposed method (including the time taken for converting a social network into FGSN) takes significantly lesser time compared to Cen-CD algorithm, although it needs more time than Mod-CD and CPM-CD.

#### 7.4.4. Computational complexity

Similar to the target set selection problem, the algorithm of community detection also runs on fuzzy granular social networks. The conversion of a social network to a FGSN takes  $O(|V| + |E|)$  time (see Section 7.3.2). In the next step, we measure the granular embeddedness for all the pairs of granules in FGSN. This requires  $\binom{|C|}{2}$  time. Then we order these values, which requires  $O\left(\binom{|C|}{2} \times \log_2\left(\binom{|C|}{2}\right)\right)$  using the best sorting algorithm available. Finally, an iteration over this sorted list is performed to get the community structure. So, the total time complexity of the algorithm in a worst case scenario is  $O\left(|V| + |E| + 2 \times \binom{|C|}{2} + \binom{|C|}{2} \times \log_2\left(\binom{|C|}{2}\right)\right)$ .

#### 7.5. Entropy

The entropy values with varying  $r$  are plotted in Fig. 9. It is evident from the figures that the lowest entropy for Zachary karate club and Dolphin social graph corresponds to  $r = 2$  and  $r = 3$  respectively, and as  $r$  approaches to  $D$ , the diameter of the network, the entropy approaches to its maximum value. Political blog network, on the other hand, is a directed social network. In addition to that it is a blog network depicting the web links as a relation. So, here a link can be established from a later published blog to an earlier one, but the reverse is not possible. In this network, we found a completely reverse trend

**Table 4**  
Granular modularity value of different algorithms.

Algorithms	Data sets		
	Zachary karate club	Dolphin social graph	Political blogs network
GranE-FGSN	0.73	0.81	0.49
Mod-CD	0.74	0.79	0.92
Cen-CD	0.67	0.75	0.87
CPM-CD	0.80	0.59	0.92

**Table 5**  
Execution time (in seconds) of different community detection algorithms for 1000 runs.

Algorithms	Data sets		
	Zachary karate club	Dolphin social graph	Political blogs network
GranE-FGSN	44.23	174.35	$3.624 \times 10^5$
Mod-CD	2.00	3.05	$1.39 \times 10^3$
Cen-CD	56.15	244.94	$2.62 \times 10^7$
CPM-CD	2.30	5.42	$1.47 \times 10^4$

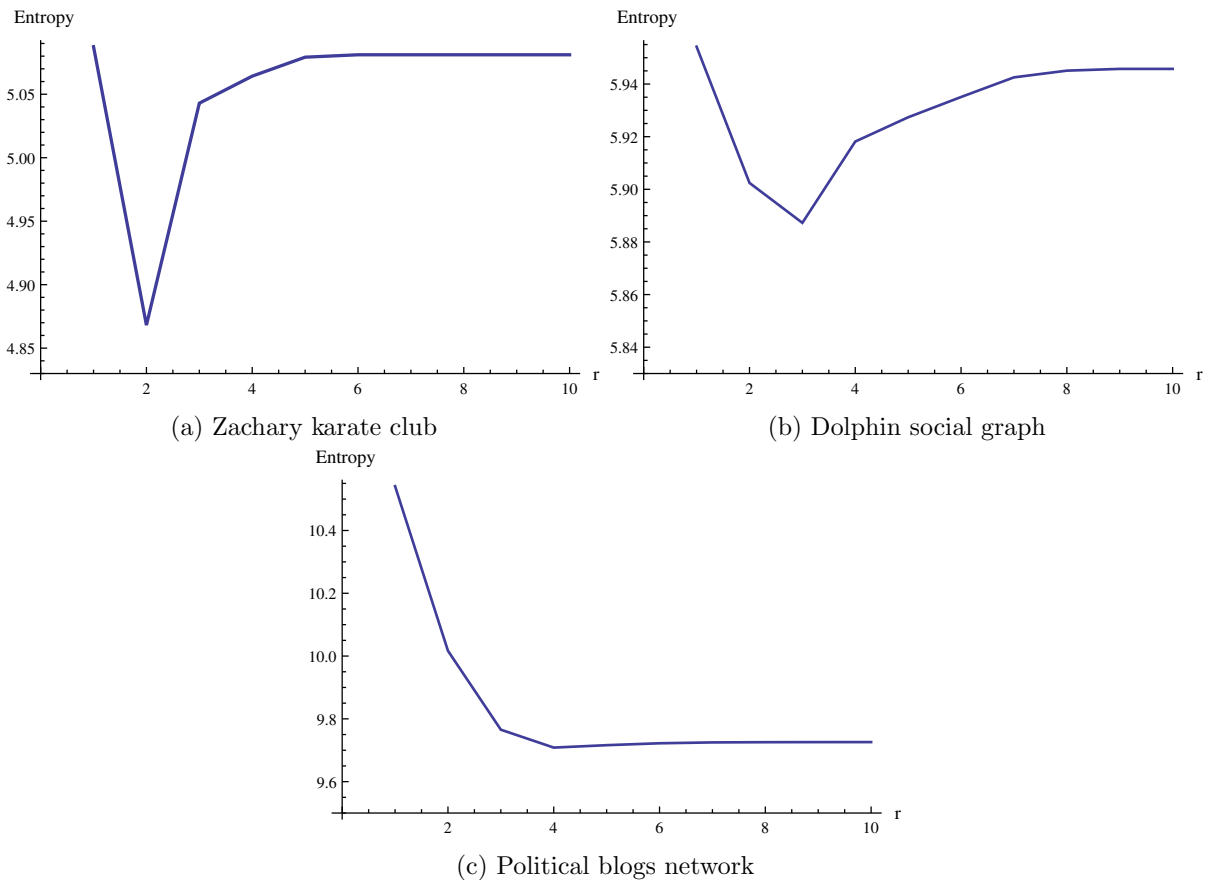


Fig. 9. Variation of network entropy with  $r$ .

compared to that in the undirected networks viz., karate club and Dolphin social graph. Here the entropy sharply falls as  $r$  increases from 1 to 4 and then attains a minimum value as the granule radius  $r$  approaches the value of  $D$ .

## 8. Discussions and conclusions

We presented a novel modeling technique based on fuzzy granular theory to describe a network in terms of granules, and name it Fuzzy Granular Social Network (FGSN). Here we expressed each granule using a fuzzy set. Based on this model, we defined some new measures, e.g., granular degree, granular betweenness and granular clustering co-efficient of a node; granular embeddedness of a pair of nodes; and granular clustering co-efficient and granular modularity of a FGSN. In addition, we defined entropy of FGSN to compute the uncertainties involved in the model.

Experimentally, it is shown that the model provides a generic platform to analyze social networks. In this regard, two well known problems were studied on three real life social networks. Result for target set selection problem shows improvement in solutions compared to their counterpart of graph representation for most of the cases and comparable results are found for community detection problem. Although, some of the algorithms available in the domain might provide better solutions as compared to the proposed methodology, this novel approach of FGSN will open a new avenue and provide directions on using the established granular computing theory and other efficient data mining techniques into the demanding dynamics of social networks and related problems with a scope of newly defined measures and efficient algorithms. Entropy, defined here over FGSN reflects well the uncertainties involved in the aforesaid tested networks. For undirected networks, as the radius ( $r$ ) of granules increases, entropy initially falls to a minimum value and then increases. It means, the minimum uncertainty is attained only when the size of the granules could appropriately characterize the overlapping nature of the communities. Other values of  $r$  either under characterize (leading towards crisp granularity) or over characterize the overlapping. However, this is not true, as expected, for the directed Political blogs network, where the network forms a tree-like structure. In the context of entropy measure in neighborhood systems, one may refer to [5,42].

Further, the networks we experimented with, are very good representatives of different horizons of social networks viz. friendship network, web blog network, directed and undirected network. Thus, the FGSN model can easily be applied to

social networks of different shapes and sizes by just varying the membership function or by modifying the parameters. In addition, one can effectively manage big social network data using local analysis by reducing the size of nodes in concern.

## Acknowledgments

The authors acknowledge the Department of Science and Technology, Govt. of India for funding the Center for Soft Computing Research at Indian Statistical Institute. S.K. Pal acknowledges the J.C. Bose National Fellowship and INAE Chair Professorship.

## References

- [1] L.A. Adamic, N. Glance, The political blogosphere and the 2004 US election: divided they blog, in: Proc. of the 3rd International Workshop on Link Discovery, LinkKDD '05, ACM, Chicago, Illinois, 2005, pp. 36–43.
- [2] S.A. Boorman, H.C. White, Social structure from multiple networks. II. Role structures social structure from multiple networks, *Am. J. Sociol.* 81 (6) (1976) 1384–1446.
- [3] S. Chen, Y. Li, Dynamic grade on the major hazards using community detection based on genetic algorithm, in: 2009 International Conference on Signal Processing Systems, IEEE, Singapore, 2009, pp. 713–717.
- [4] S. Chen, G. Wang, W. Jia, k-FuzzyTrust: efficient trust computation for large-scale mobile social networks using a fuzzy implicit social graph, *Inform. Sci.* (2014).
- [5] Y. Chen, K. Wu, X. Chen, C. Tang, Q. Zhu, An entropy-based uncertainty measurement approach in neighborhood systems, *Inform. Sci.* 279 (2014) 239–250.
- [6] F. Cicalese, G. Cordasco, L. Gargano, M. Milanič, U. Vaccaro, Latency-bounded target set selection in social networks, *Theoret. Comput. Sci.* 535 (2014) 1–15.
- [7] L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, Comparing community structure identification, *J. Stat. Mech: Theory Exp.* 2005 (09) (2005) P09008.
- [8] G.B. Davis, K.M. Carley, Clearing the FOG: fuzzy, overlapping groups for social networks, *Soc. Networks* 30 (3) (2008) 201–212.
- [9] P. Domingos, M. Richardson, Mining the network value of customers, in: Proc. of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco, CA, 2001, pp. 57–66.
- [10] T.-F. Fan, C.-J. Liao, T.-Y. Lin, Positional analysis in fuzzy social networks, in: Proc. of 2007 IEEE International Conference on Granular Computing (GRC 2007), IEEE, Silicon Valley, 2007, pp. 423–428.
- [11] O. Frank, Estimation and sampling in social network analysis, in: R.A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science*, Springer, New York, 2009, pp. 8213–8231.
- [12] A. Ganivada, S. Dutta, S.K. Pal, Fuzzy rough granular neural networks, fuzzy granules, and classification, *Theoret. Comput. Sci.* 412 (42) (2011) 5834–5853.
- [13] K.J. Gile, M.S. Handcock, Respondent-driven sampling: an assessment of current methodology, *Sociol. Methodol.* 40 (1) (2010) 285–327.
- [14] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, *Proc. Natl. Acad. Sci. USA* 99 (12) (2002) 7821–7826.
- [15] P. Hoňko, Association discovery from relational data via granular computing, *Inform. Sci.* 234 (2) (2013) 136–149.
- [16] G. Jiang, F. Ma, J. Shang, P.Y. Chau, Evolution of knowledge sharing behavior in social commerce: an agent-based computational approach, *Inform. Sci.* 278 (2014) 250–266.
- [17] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, NY, 2003, p. 137.
- [18] D. Kempe, J. Kleinberg, E. Tardos, Influential nodes in a diffusion model for social networks, *Automata, Lang. Program.* 3580 (2005) 1127–1138.
- [19] Y.A. Kim, R. Phalak, A trust prediction framework in rating-based experience sharing social networks without a Web of Trust, *Inform. Sci.* 191 (2012) 128–145.
- [20] J. Kleinberg, Cascading behavior in networks: algorithmic and economic issues, in: N. Nisan, T. Roughgarden, E. Tardos, V.V. Vazirani (Eds.), *Algorithmic Game Theory*, Cambridge University Press, 2007, pp. 613–632.
- [21] J. Kleinberg, S. Suri, E. Tardos, T. Wexler, Strategic network formation with structural holes, in: Proc. of the 9th ACM Conference on Electronic Commerce – EC'08, ACM Press, New York, USA, 2008, pp. 284–293.
- [22] M. Lahiri, M. Cebrian, The genetic algorithm as a general diffusion model for social networks, in: Proc. of the 24th AAAI Conference on Artificial Intelligence, Atlanta, Georgia, 2010, pp. 494–499.
- [23] J. Leskovec, D. Huttenlocher, J. Kleinberg, Predicting positive and negative links in online social networks, in: Proc. of the 19th International Conference on World Wide Web – WWW '10, ACM Press, New York, USA, 2010, pp. 641–650.
- [24] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: densification and shrinking diameters, *ACM Trans. Knowl. Discov. Data* 1 (1) (2007) 2.
- [25] L. Li, S. Li, X. Chen, A new genetics-based diffusion model for social networks, in: Proc. of 2011 International Conference on Computational Aspects of Social Networks (CASoN), IEEE, Salamanca, Spain, 2011, pp. 76–81.
- [26] C. Lin, R. Xie, X. Guan, L. Li, T. Li, Personalized news recommendation via implicit social experts, *Inform. Sci.* 254 (2014) 1–18.
- [27] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, S.M. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, *Behav. Ecol. Sociobiol.* 54 (4) (2003) 396–405.
- [28] P. Maji, S.K. Pal, Fuzzy-rough sets for information measures and selection of relevant genes from microarray data, *IEEE Trans. Syst., Man, Cybernet. Part B* 40 (3) (2010) 741–752.
- [29] J.L. Moreno, H.H. Jennings, *Who Shall Survive? A New Approach to the Problem of Human Interrelations*, Nervous and Mental Disease Monograph Series, Nervous and Mental Disease Publishing Co., New York, 1934.
- [30] P.S. Nair, S.T. Sarasamma, Data mining through fuzzy social network analysis, in: Proc. of the 26th International Conference of North American Fuzzy Information Processing Society, IEEE, San Diego, California, 2007, pp. 251–255.
- [31] R. Narayanam, Y. Narahari, A Shapley value-based approach to discover influential nodes in social networks, *IEEE Trans. Autom. Sci. Eng.* 8 (1) (2011) 130–147.
- [32] M. Newman, Fast algorithm for detecting community structure in networks, *Phys. Rev. E* 69 (6) (2004) 066133.
- [33] M. Newman, M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* 69 (2) (2004) 1–15.
- [34] G.K. Orman, V. Labatut, The effect of network realism on community detection algorithms, in: Proc. of the 2010 International Conference on Advances in Social Networks Analysis and Mining, IEEE, Odense, Denmark, 2010, pp. 301–305.
- [35] S.K. Pal, Granular mining and rough-fuzzy pattern recognition: a way to natural computation, *IEEE Intell. Inform. Bull.* 13 (1) (2012) 3–13.
- [36] S.K. Pal, S. Kundu, C.A. Murthy, Centrality measures, upper bound, and influence maximization in large scale directed social networks, *Fundam. Inform.* 130 (3) (2014) 317–342.
- [37] S.K. Pal, S.K. Meher, Natural computing: a problem solving paradigm with granular information processing, *Appl. Soft Comput. J.* 13 (9) (2013) 3944–3955.
- [38] S.K. Pal, P. Mitra, *Pattern Recognition Algorithms for Data Mining*, CRC Press, Boca Raton, 2004.

- [39] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, *Nature* 435 (7043) (2005) 814–818.
- [40] W. Pedrycz, *Granular Computing: Analysis and Design of Intelligent Systems*, CRC Press, Boca Raton, 2013.
- [41] W. Pedrycz, A. Skowron, V. Kreinovich (Eds.), *Handbook of Granular Computing*, Wiley, Sussex, 2008.
- [42] Y. Qian, J. Liang, W. Wu, C. Dang, Information granularity in fuzzy binary GrC model, *IEEE Trans. Fuzzy Syst.* 19 (2) (2011) 253–264.
- [43] M. Richardson, P. Domingos, Mining knowledge-sharing sites for viral marketing, in: *Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, Edmonton, Alberta, 2002, pp. 61–70.
- [44] M. Spreen, Rare populations, hidden populations, and link-tracing designs: What and Why?, *Bull. Méthodol. Sociol.* 36 (1992) 34–58.
- [45] C. Wang, L. Deng, G. Zhou, M. Jiang, A global optimization algorithm for target set selection problems, *Inform. Sci.* 267 (2014) 101–118.
- [46] S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.
- [47] H.C. White, S.A. Boorman, R.L. Breiger, Social structure from multiple networks. I. Blockmodels of roles and positions, *Am. J. Sociol.* 81 (4) (1976) 730–780.
- [48] J. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, *IEEE Trans. Cybernet.* 43 (2013) 1977–1989.
- [49] W. Zachary, An information flow model for conflict and fission in small groups, *J. Anthropol. Res.* 33 (4) (1977) 452–473.
- [50] L. Zadeh, Fuzzy sets, *Inform. Control* 8 (1965) 338–353.
- [51] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets Syst.* 90 (1997) 111–127.
- [52] T. Zhu, B. Wang, B. Wu, C. Zhu, Maximizing the spread of influence ranking in social networks, *Inform. Sci.* 278 (2014) 535–544.