Community Detection in Signed Social Networks Using Multiobjective Genetic Algorithm

Nancy Girdhar
School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, 110067, India. E-mail: nancy.gr1991@gmail.com

K. K. Bharadwaj
School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, 110067, India. E-mail: kbharadwaj@gmail.com

Clustering of like-minded users is basically the goal of community detection (CD) in social networks and many researchers have proposed different algorithms for the same. In signed social networks (SSNs) where type of link is also considered besides the links itself, CD aims to partition the network in such a way to have less positive inter-connections and less negative intra-connections among communities. So, approaches used for CD in unsigned networks do not perform well when directly applied on signed networks. Most of the CD algorithms are based on single objective optimization criteria of optimizing modularity which focuses only on link density without considering the type of links existing in the network. In this work, a multiobjective approach for CD in SSNs is proposed considering both the link density as well as the sign of links. Precisely we are developing a method using modularity, frustration and social balance factor as multiple objectives to be optimized (M-F-SBF model). NSGA-II algorithm is used to maintain elitism and diversity in the solutions. Experiments are performed on both existing benchmarked and real-world datasets show that our approach has led to better solutions, clearly indicating the effectiveness of our proposed M-F-SBF model.

Introduction

People maintain different kinds of relationships and articulate their emotions via different actions like chatting and blocking of users on social websites, friendly and unfriendly nature of gossips at workplace; friendship and bullying in schools or colleges; cliques and disputes at nightspots like clubs; and partnerships and competitions in different organizations or corporations. We can easily see these social human relations of love-hate, like-dislike, friends-enemy, trust-distrust in social media sites through different activities of users such as liking or disliking of posts of others, rating blogs on different levels or scale, tagging users as “friends” and “foes” or “fan” and “freak” (terms used in Slashdot website). These social relations broadly classified as positive (friendly) relations including “liking,” tagging as “friend” or “fan,” rating “high” on scale and negative (antagonistic) relations (Zhang, Lo, Lim, & Prasetyo, 2013) such as “disliking,” tagging as “foe” or “freak,” rating “low” on scale etc.

Many social networking sites explicitly show the mixture of positive and negative interactions among the users by indicating the sign on link connecting the users. Some of the popular websites are (a) Tech- news related website Slashdot, where users can tag other users as “friend” and “foe” or as “fan” and “freak.” Here, the reviewers can comment on the blogs of others depending on their liking or disliking. (b) Trust-network of Epinions is the largest product review website in which users can express their views by rating items of different category like hardware, music and TV shows. Not only items, they can also rate other raters depending on their trust or distrust on that rater. These social networks contain the information about the links as well as the type of links exhibited by the users. As these networks are the extension of social networks, containing information about the sign of links besides the information of links itself, a special term has been coined for these networks called as signed social networks (SSNs; Moshirpour et al., 2013).

Exploring community structures from these networks can help us to gain unfathomable knowledge and better understanding of social ties existing among the users constituting the network (Chen, Chuang, & Chiu, 2014; Verma & Bharadwaj, 2017). CD in unsigned networks involves grouping of those nodes or vertices which are densely connected within groups and sparsely connected between groups. In

Received August 3, 2017; revised September 27, 2018; accepted October 21, 2018

 DOI: 10.1002/asi.24164

signed networks besides the information about the link density, we have information about the type of links also. CD in SSNs can be defined as the partition of the network into communities such that it has dense positive intra-community connections and negative inter-community connections (Girdhar & Bharadwaj, 2016).

We can see from the existent work in the field of community detection (Shi, Yu, Cai, Yan, & Wu, 2011), that discovering community structures in social networks implicitly have multiple objectives to be optimized. The most commonly considered objective for CD in SSNs is modularity (Amelio & Pizzuti, 2013; Anchuri & Magdon-Ismail, 2012; Newman & Girvan, 2004). Recently, some research work (Amelio & Pizzuti, 2013) has also taken into account multiple objectives such as: modularity and frustration. However, balance in signed social networks for finding community structures has not been explored yet. In our proposed work, besides considering commonly used objectives we have also considered an important facet of signed social networks that is, social balance (Awal & Bharadwaj, 2017; Zheng, Zeng, & Wang, 2015). Thus, our work includes social balance as a significant objective for signed social networks together with the main objectives modularity and frustration for community detection.

Evolutionary algorithms (EAs) are one of the most popular and best suited methods to solve multiobjective optimization (MOO) problems which are inspired from biological mechanisms of evolution and heredity. To maintain the elitism and diversity in the solutions, we are using fast and elitist Nondominated Sorting Genetic Algorithm (NSGA-II; Deb, 2011) which is proven to have less computational complexity than NSGA.

The main contributions of this paper are summarized as follows:

1. Development of multiobjective framework model based on evolutionary algorithm (EA) for community detection in SSNs considering both link density and the type of link information.
2. Design of matrix representation of chromosome with appropriate crossover and mutation operators.
3. Consideration of three fundamental properties of a network as optimization criterias:
    - Modularity
    - Frustration
    - Social Balance Factor
4. Evaluation of proposed framework using benchmarked and real-world social media datasets.

Overview on State-of-the-Art

Because of vast, heterogenic and dynamic nature of social networks, they have become very complex and difficult to analyze. To understand and "make sense" out of these network structures, community detection plays a vital role. Community can be considered as a summary of whole network and thus, make these network structures easy to visualize and comprehend. Hence, different methodologies are explored for CD in social networks and numerous aspects of community structures are investigated (Bedi & Sharma, 2016).

During the past decade various algorithms for CD have been proposed for unsigned networks that are generally classified as: Graph based (Fortunato & Barthelemy, 2007) and Density based (Ester, Kriegel, Sander, & Xu, 1996). Graph based algorithms are further categorized into two main approaches: Graph Partitioning (Shi & Malik, 2000) and Graph Clustering (Papadopoulos, Kompatsiaris, Vakali, & Spyridonos, 2012). In graph partitioning algorithms, the main idea is to partition the graph into predetermined number of groups, having higher density inside the groups and lower density between the groups. Signed social networks imbibe scalable dynamic properties which pose a limitation to the common clustering methods which require prerequisite knowledge of the size and number of communities. Other approach for CD is density-based algorithms which are designed to discover communities of arbitrary shape. Ester et al. (1996) proposed a density-based algorithm DBSCAN that requires only one parameter as input and takes object as points to form communities. To identify overlapping communities, Bhat and Abulais (2015) proposed OCMiner, a density-based algorithm which also identifies outliers in social networks.

In recent years, new approaches grounded on optimization of quality function for partition have been proposed. One of the most widely used methods for community detection is based on modularity optimization (Clauset, Newman, & Moore, 2004). A fast greedy modularity optimization method for community detection was proposed by Clauset et al. (2004). GA based methods have been used for optimization of modularity and normalized graph cuts (Pizzuti, 2008). As it is mathematically proved by Fortunato and Barthelemy (2007) that modularity optimization suffers from resolution limit, so it is important to explore other quality functions to find more reliable modules. A few algorithms proposed for CD in signed networks also considers various other objective criterias. Kunegis, Lommatzsch, and Bauckhage (2009) analyzed different aspects of signed network at global level, node level and at edge level. For network partitioning, Doreian and Mrvar (1996) proposed algorithm based on optimization of criterion function *frustration.* A study of structural balance aspect of signed networks is done by Facchetti, Iacono, and Altafini (2011). FEC algorithm proposed by Yang, Cheung, and Liu (2007), is an agent-based method for CD that can work for both signed and unsigned networks in which dense positive intra-group and negative inter-groups relations are mined. Wu, Ying, Wu, Lu, and Zhou (2011) demonstrated the impact of negative links and despite the connections among communities being dense, the observation showed that the patterns in the spectral space of the graph's adjacency matrix were significantly different from the one shown in a k-balanced signed network. Doreian and Mrvar (1996) presented an approach based on the structural balance theory and the block model (Doreian, 2008) where local search method is used to partition a signed network to minimize a predefined error function. One drawback of the scheme was the requirement to know a priori the number of groups and it only considers the sign of links not the density of links ignoring a prominent aspect of partition.

Evolutionary algorithms (EA) have been applied successfully for community detection. One of them is GA-Net (Pizzuti, 2008) which optimized only one objective function: community score. Two memetic algorithms and two EA were proposed (Li, Liu, & Liu, 2014) for CD in SSNs but all four algorithms were based on single objective framework concluding that memetic algorithms outperformed EA's. A multiobjective genetic algorithm, known as MOGA-Net is presented in (Pizzuti, 2009) which optimizes two objectives: community score and the fitness of nodes. Because it is a multiobjective approach instead of giving one best solution, it gives multiple optimum solutions which show a trade-off between these two objectives. SN-MOGA algorithm proposed by Amelio and Pizzuti (2013), in which modularity and frustration concepts were exploited for community detection. However, in this approach structural balance of network was not considered and modularity values obtained were low as compared to other multiobjective approaches. A multiobjective evolutionary algorithm for SSNs proposed in (Liu, Liu, & Jiang, 2014), is used to detect overlapping and disjoint communities in undirected signed social networks by optimizing two objective functions. Analytical results shown after investigating correlations among 11 objective functions by Shi et al. (2011) established that MOEAs optimization of negatively correlated objectives pair gives better results than optimizing single objective individually.

Preliminary

In this section we have specified preliminaries and a brief introduction to Non-dominated Sorting Genetic Algorithm (NSGA-II) used for multiobjective optimization.

Undirected Signed Social Networks

Mathematically, an undirected signed social network represented as signed graph can be defined as $G = (V, E, \beta)$ where V is the vertex set representing nodes of the network, $E \subseteq V \times V$ is edge set which represents the links of the network. Here, $\beta : E \rightarrow \{-1, 0, +1\}$ is a function which assigns +1 value to the edge if the link between two nodes is positive, −1 is assigned if the link is negative and 0 if there is no link between two nodes.

Let $A_{i,j}$ represents the weighted adjacency matrix associated with graph G i.e $A_{i,j} = \beta(i, j)$ and the number of vertices in graph G be n labeled from 0 to $n - 1$. $A'_{i,j}$ denotes the unsigned version of adjacency matrix i.e. $A'_{i,j} = |\beta(i,j)|$, $P_{i,j}$ and $N_{i,j}$ denotes the adjacency matrices corresponding to positive and negative relationships, respectively. Then, $P_{i,j}$ and $N_{i,j}$ can be calculated using Equation 1.

$$P_{i,j} = \frac{A_{i,j} + A'_{i,j}}{2},\ N_{i,j} = \frac{A'_{i,j} - A_{i,j}}{2} \quad (1)$$

Modularity measures the strength of a community partition by taking into account the degree distribution (Newman & Girvan, 2004; Kaur, Singh, Kaushal, & Sangaiah, 2016). Given a network with m edges, modularity is given by Equation 2 as follows:

$$Q = \frac{1}{2m} \sum \mathrm{i,j} \left(A_{i,j} - d_i\, d_j/2m\right) \delta\left(c_i, c_j\right) \quad (2)$$

where d_i and d_j are the degrees of node i and node j respectively, c_i and c_j are the communities of node i and node j, $\delta(c_i, c_j) = 1$, if node i and node j belong to the same community and $\delta(c_i, c_j) = 0$ otherwise.

Frustration can be defined as the sum of the number of positive links between different communities and negative links within the community. Based on the structural balance theory, the links constituting the unbalanced triads contributes to frustration, resulting in instability of whole network (Girdhar & Bharadwaj, 2016; Leskovec, Huttenlocher, & Kleinberg 2010).

Social Balance Factor According to structural balance theory, a triad is structurally balanced if the number of positive edges in the triad is odd (1-configuration and 3-configuration) as shown in Figure 1. For a network to be balanced, it should be divided into two clusters such that all the positive links lies within a cluster and the negative links lies between the clusters (Girdhar & Bharadwaj, 2016). Thus, to find whether a network is structurally balanced, we have to compute the social balance factor.

Multiobjective Optimization (MOO)

This type of optimization techniques deals with contradictory objectives and the "best" solution should make a compromise between them. There may exist more than one optimized solutions in MOO unlike single-objective optimization

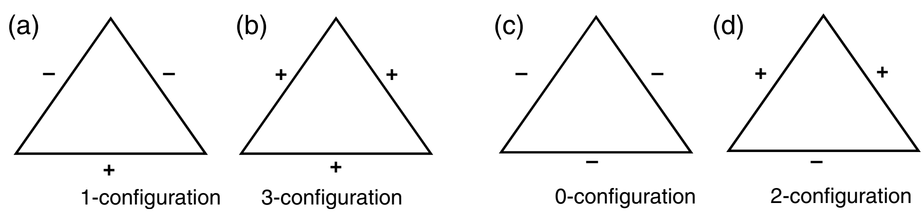

FIG. 1. Balanced triads (a and b) and unbalanced triads (c and d).

DOI: 10.1002/asi

(SOO) problems where the result shows only the best solution. Hence, in MOO there is a trade-off among different objectives, thus we are interested in all non-dominated solutions, called the *pareto front*, to have all alternatives available to determine which solution is the best. Without any further information one cannot say that a solution is better than other solution, unless it dominates the other.

Dominance (Deb, 2011) A solution **x1** is said to dominate a solution **x2** if both conditions below are true:

- The solution **x1** is no worse than **x2** in all objectives
- The solution **x1** is strictly better than **x2** in at least one objective.

NSGA-II Algorithm

One of the multiobjective optimizing techniques is Non-dominated Sorting Genetic Algorithm, popularly known as NSGA– II (Deb, 2011). NSGA-II is a fast and elitist multiobjective genetic algorithm. The ranking-based evolutionary algorithm NSGA-II combines elitism and a mechanism to distribute the solutions as much as possible. Multiobjective optimization elitism requires that some portion of the non-dominated solutions will survive. A general schema of NSGA-II algorithm is shown in Figure 2.

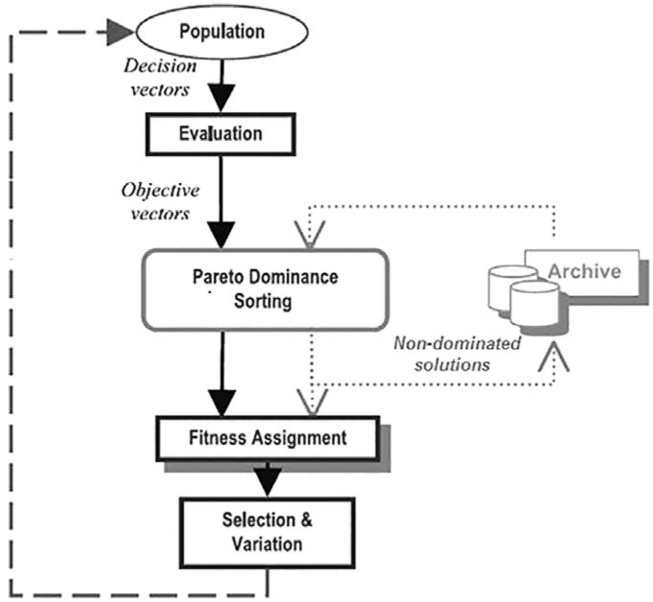

FIG. 2. General schema for NSGA-II.

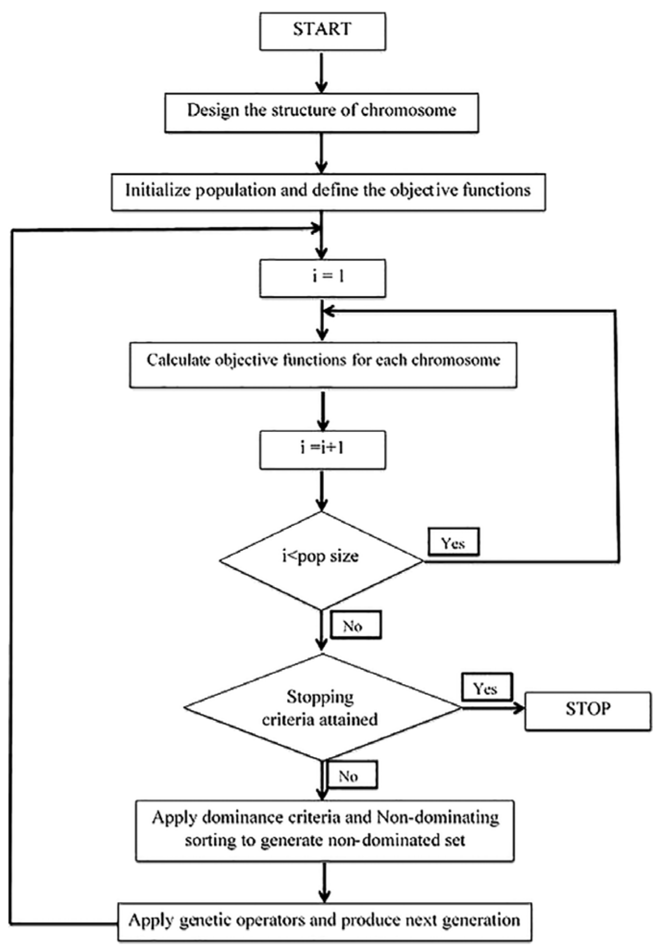

FIG. 3. Diagrammatic representation of the proposed scheme.

Proposed Scheme

This section presents our proposed multiobjective framework for community detection in undirected SSNs using GA. Our proposed scheme based on the optimization of multiobjectives: modularity (M), frustration (F), social balance factor (SBF) referred as M-F-SBF model is mainly divided into following steps:

Step 1. Initial population of chromosomes is generated randomly consisting of *N* users distributed into *k* communities, where *k* is the number of communities to be formed.

Step 2. Fitness of each chromosome is evaluated based on the objective functions.

Step 3. NSGA-II algorithm is applied to sort the population based on their dominance.

Step 4. The genetic operations are performed on the current population to generate the next generation. Repeat the steps 2 to 4 until the stopping criteria is attained. Figure 3 depicts our scheme diagrammatically and framework of the proposed algorithm is given in Algorithm 4.1. In the following sub-sections, we have explained in detail each step of the scheme.

Genetic Representation and Objective Functions

Chromosome representation and generating initial population. The representation of chromosome has a great effect on both the operators and the performance efficiency of EA. In the existing works on CD, the character string encoding (Anchuri & Magdon-Ismail, 2012; Clauset et al., 2004; Fortunato & Barthelemy, 2007; Lancichinetti, Fortunato, & Kertész, 2009), graph-based encoding (Amelio & Pizzuti, 2013; Liu, Zhong, Abbass, & Green, 2010; Pizzuti, 2008, 2009) are widely used and all of them have shortcomings, the former fails traditional crossover operator whereas the latter requires additional decoding step. To avoid above

DOI: 10.1002/asi

$M=$

	C_1	.	.	C_{k-1}	C_k
User $_1$	1	0	0	0	0
User $_2$	0	0	0	1	0
.	0	1	0	0	0
.	0	0	1	0	0
User $_{N-1}$	0	0	0	0	1
User $_N$	0	1	0	0	0

FIG. 4. Representation of encoded chromosome. [Color figure can be viewed at wileyonlinelibrary.com]

shortcomings, we have used the binary matrix representation of chromosome. The encoding of the individuals is carried out by viewing the chromosome as a matrix M of order $N * k$, where N is the number of users in the network to be clustered which is represented as rows of matrix and k is the number of communities represented by the columns of the matrix as shown in the Figure 4. Each user belongs to a community and value is set to 1 at the corresponding position in the matrix else it is set to 0. The users are distributed among the communities in such a way that every column must have at least one non-zero value that is, each community must have at least one user, so no community is empty.

The network can be partitioned in different ways and can lead to different number of communities formation, thus the number of columns of encoding matrix M is variable and one more thing worth noticing is that it represents the same partition of network irrespective of columns' order, provided there is no change in the value of columns.

Next, we have randomly generated N individuals by using the feasible representation as described above that fulfills the restriction posed by the specific task for which the optimal solution needs to be determined.

Algorithm. 4.1. Framework of the proposed algorithm

Parameter: Maximum number of generations: maxgen; Population size: popSize; Crossover probability: P_c; Mutation probability: P_m.

Input: The adjacency matrix of signed networks: A; number of communities: k.

P ← Pop_Initialization (A, popSize, k)
Repeat
F ←Evaluate_Fitness (**P**);
Communities ←NSGA-II_Sort (**P**,**F**)
$\mathbf{P}_{parent}$ ← Rol_Selection (**P**,**F**)
$\mathbf{P}_{child}$←Genetic_Operations ($\mathbf{P}_{parent}$,P_c,P_m)
P ← Update (**P**,$\mathbf{P}_{\mathrm{child}}$)
Until Termination (maxgen)
Output: Communities

Objective functions. For every chromosome in the population, it is indispensable to measure the quality of that chromosome. This refers to as measuring the fitness of the possible solution represented by that chromosome. Objective functions validate the diversion of the process toward its optimization goal by allowing best individuals to breed that leads to good solutions (Agarwal & Bharadwaj, 2015). Because our approach is based on multiobjective optimization so the fitness function employed here is the integration of three objective functions which are to be optimized together to get optimal solutions. Also, multiobjective approach results in multiple optimized solutions rather than one best solution. The three objective functions which are going to be optimized are as follows:

Objective function 1. **Modularity** For signed networks the definition of modularity is modified to take into account the contribution of positive edges inside communities and negative edges between communities (Anchuri & Magdon-Ismail, 2012).

Let a signed network G be partitioned into two communities C_1 and C_2. Considering both the positive and negative links in the network, the signed modularity (Anchuri & Magdon-Ismail, 2012) of network G, up to a constant factor, can be given as Q_{signed}, where the first two terms corresponds to the positive links in Equation 3 and the other two terms corresponds to the negative links in the network. Given a node $i \in V$, d_i^+ and d_i^- are the positive and the negative degrees of node i, respectively.

$$\begin{aligned}\text{Maximize}: Q_{signed=} &\sum_{i,j\in C_1}\left(P_{i,j}-\frac{d_i^+ d_j^+}{2m^+}\right)\\ &+\sum_{i,j\in C_2}\left(P_{i,j}-\frac{d_i^+ d_j^+}{2m^+}\right)+\sum_{i\in C_1,j\in C_2}\left(N_{i,j}-\frac{d_i^- d_j^-}{2m^-}\right)\\ &+\sum_{i\in C_2,j\in C_1}\left(N_{i,j}-\frac{d_i^- d_j^-}{2m^-}\right)\end{aligned} \tag{3}$$

$$\text{subject to}\ \forall i,j\in G\ \delta\left(c_i,c_j\right)=1$$
$$where\ i=j; else\ 0.$$

Objective function 2. **Frustration** (Doreian & Mrvar, 1996) measures the unstability in the network. Let a network G is divided into k communities $\{C_1, C_2, \ldots, C_k\}$. Given a division of the network into communities, as shown in Equation 4, frustration is sum of the number of positive edges between nodes in different communities and the number of negative edges between nodes in the same community.

$$\begin{aligned}\text{Minimize}: &F(C_1,C_2,\ldots,C_k)\\ &=\sum i,j N_{i,j}\delta(c_i,c_j)+P_{i,j}\left(1-\delta\left(c_i,c_j\right)\right)\end{aligned} \tag{4}$$

$$\text{subject to}\ \forall i,j\in G\ \delta\left(c_i,c_j\right)=1$$
$$where\ i=j; else\ 0.$$

DOI: 10.1002/asi

where c_i denotes the community to which the node i belongs. In a SSN, communities having low value of frustration imply that the people of that group are having common interests or opinions differ from other groups. Frustration implies mutually antagonistic or hostile groups with few imbalances. Therefore, minimizing frustration can be used as an objective function to detect communities (Zhang, Lo, Lim, & Prasteyo, 2013).

Objective function 3. **Social Balance Factor** According to structural balance theory (Leskovec et al., 2010), Social balance factor (SBF) of a network can be calculated as the ratio of total number of balanced triads to the total number of triads in the network as shown in Equation 5. As social balance is one of the important aspects of a SSN, thus, maximizing SBF can be used as an objective to detect communities in the network.

$$Maximize: SBF = \frac{Number\ of\ balanced\ triads}{Total\ number\ of\ triads} \quad (5)$$

Problem Specific Genetic Operators

Selection. The selection operator chooses the chromosomes from the population on which genetic crossover is to be applied. Various schemes are possible to choose the chromosomes as parents in the crossover such as roulette wheel selection, rank selection, tournament selection, Boltzman selection. The selection method used in this work is the *roulette wheel selection* (Figure 5) also known as *fitness proportionate selection* which selects the chromosome according to their fitness value.

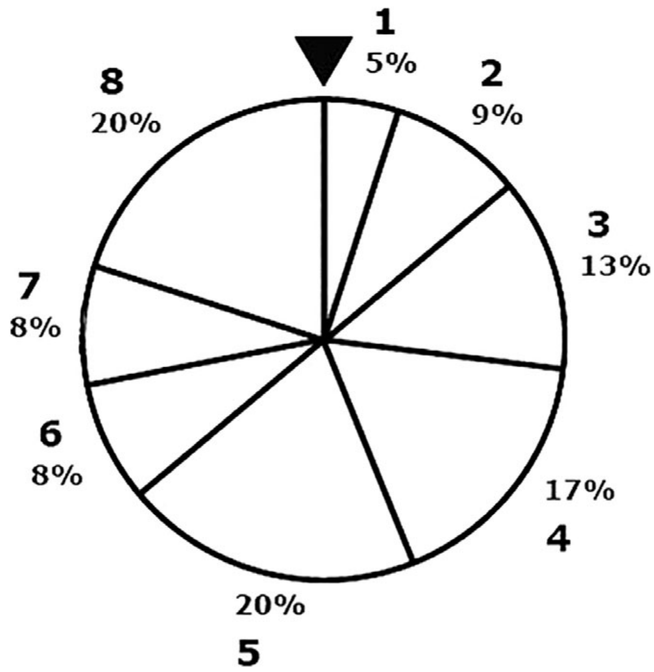

FIG. 5. Roulette Wheel for eight individuals with their probability of selection written on the circumference.

Crossover. Crossover (Handl & Knowles, 2007) is one of the basic operators used in GA that allows creation of two new individuals by allowing two parent chromosomes to exchange meaningful information. It aims to preserve and combine the best characteristics of the parents to evolve better new solutions.

In this work, from the different kinds of crossover operators used in the literature, we choose the *uniform crossover* operator in our approach, as it can generate any combination of offsprings from the two parents in a single crossover event and it is unbiased with respect to the ordering of genes unlike one-point and two-point crossover, which works only at segment level not at gene level (Handl & Knowles, 2007). Here, we have modified the uniform crossover operator accordingly to suit our problem which is described as follows:

Modified Uniform Crossover Operator: In this crossover, two parents $P1$ and $P2$ are taken, and a binary mask of size n (number of users) is randomly generated. The process of crossover is as follows:

- To generate the first offspring, first consider the mask if the value is 1, user from first parent is selected and if mask value is equal to 0, user from second parent is taken.
- To generate the second offspring, consider the mask again if the value is 0, user from first parent is selected and if mask value is 1, user from second parent is taken.

An example of uniform crossover on the encoding employed is shown in Figure 6. The good heritability of the encoding under this crossover operator can be seen.

Mutation. Mutation is a variation operator that introduces diversity by introducing a completely new member into the population and ensures that the entire search space is exploited. The real valued uniform mutation operator is tailored to fit in our application. This operator is applied directly to the encoded solutions where it changes one or more characteristics of the solutions, with a probability P_m.

***Modified Mutation Operator*:** In this, any two random communities of a chromosome are chosen, such that, each of them must have more than one user. After selection of the communities, two users are taken randomly such that one user belongs to one of the community and the second user belongs to the other community. Finally, the membership values of the corresponding users are swapped. In example shown in Figure 7, the chromosome has eight users divided into three communities $\{C1, C2, C3\}$. Let say, two random communities selected are $C1$ and $C3$. Now check each of the selected community, has more than one user, which is true as $C1$ has three users and $C3$ has 2 users. Now two users are selected randomly, say, *User 1* which belongs to $C1$ and *User 5* which belongs to $C3$. Now membership value of *User 1* and *User 5* are swapped. After mutation, *User 1* belongs to $C3$ and *User 5* belongs to $C1$.

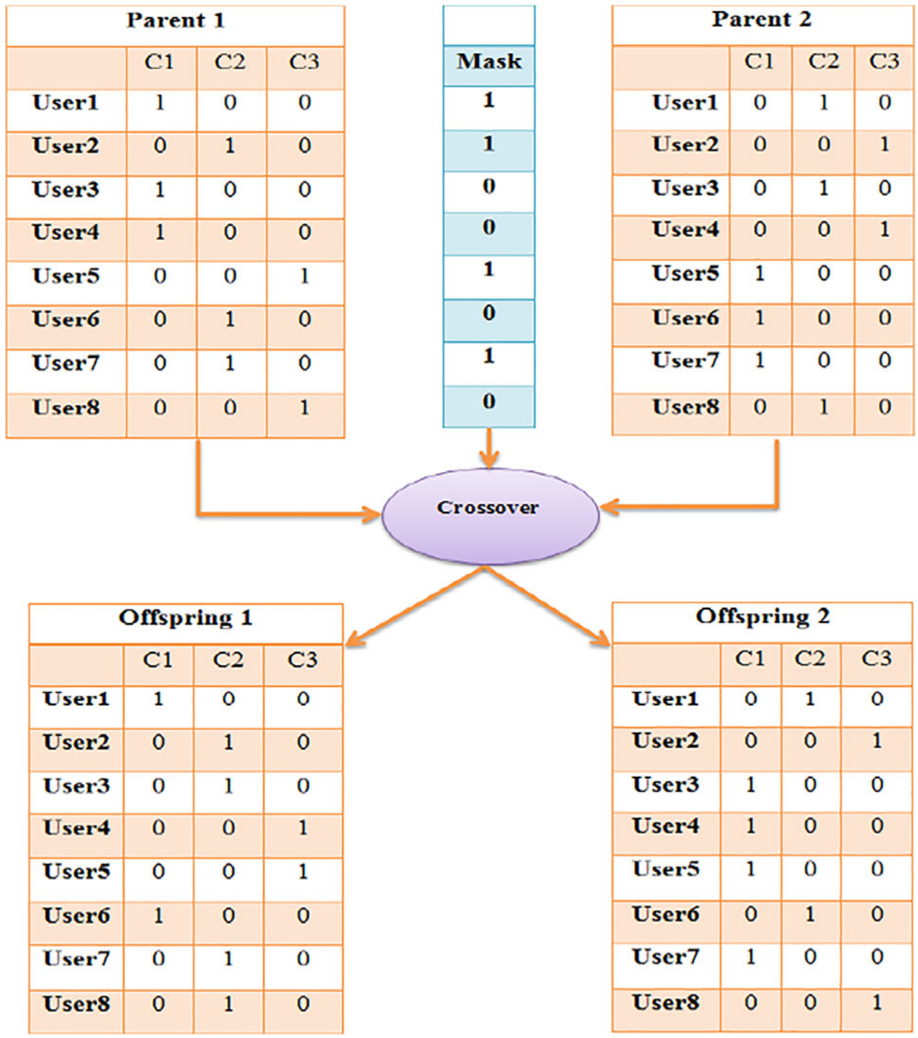

FIG. 6. Modified uniform crossover operator. [Color figure can be viewed at wileyonlinelibrary.com]

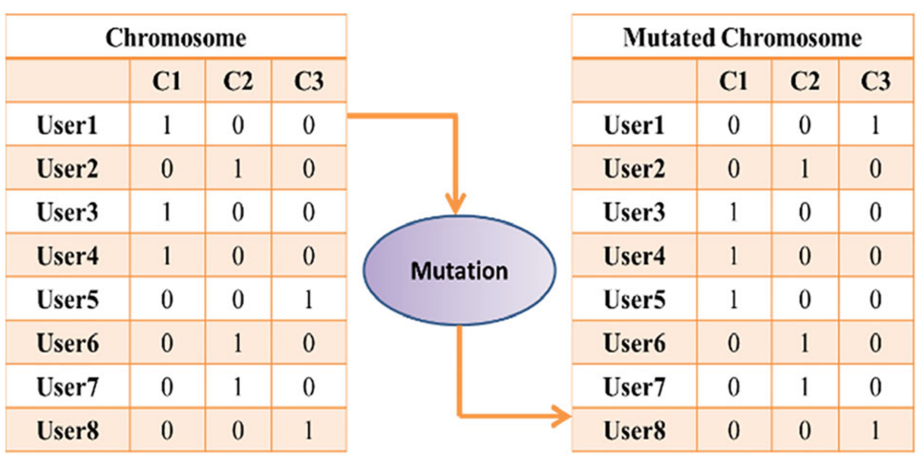

FIG. 7. Modified mutation operator. [Color figure can be viewed at wileyonlinelibrary.com]

Replacement, Elitism, and Diversity

The new population at generation $k + 1$ is obtained by replacement of the individuals in the current population at generation k by the application of genetic operators. The heuristic uses elitist strategy where the best individuals in generation k is automatically transferred to the population of generation $k + 1$ and the diversity is maintained by the crowding distance approach used in NSGA-II algorithm (Deb, 2011), which

DOI: 10.1002/asi

TABLE 1. Confusion matrix.

True Community	Predicted Community	
	P	N
P	True Positive (TP)	False Negative (FN)
N	False Positive (FP)	True Negative (TN)

ensures that the best solution encountered so far in the evolutionary process is always retained by the algorithm.

Experimental Evaluation and Results

In this section, the performance of our proposed scheme M-F-SBF is tested on both benchmarked and real-world signed networks. The description of evaluation metrics, the datasets used and experimental results are presented below:

Evaluation Metrics

For evaluation of our proposed method M-F-SBF, we used three supervised metrics: Normalized Mutual Information (NMI; Amelio & Pizzuti, 2013), F-measure (F_score; Plantié & Crampes, 2013) and Entropy (Dang & Viennet, 2012) for benchmarked datasets. As the ground truth information is not available for the real–world datasets, we have used Error Rate proposed by Yang et al. (2007), for evaluation of these networks.

Let P denotes community A and N denotes all other communities except community A. Then the confusion matrix C is given by Table 1.

- ***NMI*:** Normalized Mutual Information $NMI(A, B)$ (Amelio & Pizzuti, 2013) of a network with two partitions A and B having C as the confusion matrix shown in Table 1 is given as

$$NMI(A,B)=\frac{-2\sum_{i=1}^{c_A}\sum_{j=1}^{c_B}C_{ij}\,log\left(\frac{C_{ij}.N}{C_i.C_j}\right)}{\left(\sum_{i=1}^{c_A}C_i\,log\frac{C_i}{N}\right)+\left(\sum_{j=1}^{c_B}C_j\,log\frac{C_j}{N}\right)} \tag{6}$$

where c_A (c_B) is the number of groups in partition A (B), C_i (C_j) is the number of nodes in A_i (B_j), C_{ij} is the number of common nodes in A_i and B_j and N is the number of nodes. If both the partitions are same ie., $A = B$ then $NMI(A, B) = 1$ and if both are completely different then $NMI(A, B) = 0$. Higher the NMI, closer will be the partition to the ground truth.

- ***F-measure*:** This metric is the harmonic mean of precision and recall and also known as F_score (Plantié & Crampes, 2013). Given the confusion matrix C in Table 1, precision and recall can be computed as:

$$precision=\frac{|TP|}{|TP+FP|}\quad recall=\frac{|TP|}{|TP+FN|} \tag{7}$$

The pair-wise F_score can be computed as:

$$F\_score=\frac{2\times precision\times recall}{precision+recall} \tag{8}$$

- ***Entropy*:** Given the true community structure of a network as $C = \{C_1, C_2, C_3, \dots, C_k\}$ and the community structure achieved by the algorithm is $D = \{D_1, D_2, D_3, \dots, D_k\}$. For each D_i, entropy (Dang & Viennet, 2012) can be measured by:

$$entropy(D_i)=\sum_{j=1}^{k}Pr_i\left(C_j\right)log_2 Pr_i\left(C_j\right) \tag{9}$$

where $Pr_i(C_j)$ is the proportion of C_j data points in D_i. So, the total entropy ($entropy_{total}$) of D with n number of nodes and n_i is the number of nodes in D_i is given in Equation 10. Lesser the value of entropy, better the community structure is.

$$entropy_{total}=\sum_{i=1}^{k}\frac{n_i}{n}\,entropy(D_i) \tag{10}$$

- ***Error Rate*:** To measure the partition quality, the error rate (Yang, Cheung, & Liu, 2007) of community structure C of a signed network is given as follows:

$$Error\,(C)=\frac{F(C)}{\sum ij\left|A_{ij}\right|}\times 100\% \tag{11}$$

where $F(C)$ is defined in Equation 4. It is obvious that the smaller the value of the $error(C)$ better will be the partition quality.

Datasets.

For experimental set up, we have considered two existing social networks for benchmarking and two real-world signed social networks for validation.

Evaluation on existing benchmarked real social networks.

- *Slovene Parliamentary Party Network* (Ferligoj & Kramberger, 1996)

This network represents the relations among 10 political parties of Slovene Parliamentary established by some group of experts in 1994. The weights on the links in the network is assessed, based on the scale of −3 to +3, where the negative weight shows dissimilarity between the pair of parties and positive weight shows similarity between the two parties. The value zero (0) on the link shows that they are neither similar nor dissimilar.

- *Gahuku – Gama Subtribes Network* (Read, 1954)

This network describes the positive and negative alliances among 16 Gahuku-Gama Subtribes. This network is based on the study of cultures of highland New Guinea. According to the study, these subtribes were dispersed in a particular area and indulged in social combat with one another in 1954.

DOI: 10.1002/asi

TABLE 2. Description of datasets.

	Benchmarked networks		Real-world networks	
	Slovene Parliamentary	Gahuku-Gama Subtribes	Epinions	Slashdot
Total Nodes	10	16	500, 1000, 2000, 3000	500, 1000, 2000, 3000
Total Positive Links	36	58	9060, 14558, 30650, 51540	2006, 8068, 20534, 30404
Total Negative Links	54	58	486, 964, 2489, 4374	132, 642, 3624, 4704
Total Links	90	116	9546, 15522, 33139, 55914	2138, 8710, 24158, 35108

Evaluation on real-world signed social networks.

- *Epinions*[1]

It is the largest online product review social network. It is a general consumer review site on which users can post reviews about various products and services. These reviews are rated on the scale of 1 to 5 by different reviewers who label them by like or dislike based on their "trust" or "distrust" on the review.

- *Slashdot*[2]

This is a technology news related website founded in 1997 on which stories and blogs are submitted either by the editors or by the users. A new feature named Zoo is added on the site in 2002 which enables users to tag other users as friends or foes. This site allows users to rate other users negatively. Based on these tags, this network is considered as an example of online signed social networks where friend tags are considered as positive links and foe tags are considered as negative links.

Both the datasets are directed signed datasets, consisting of FromNode, ToNode and Sign attributes. There are 131,828 nodes and 841372 edges in Epinions whereas Slashdot has 82,144 nodes and 549,202 edges. As our work is based on undirected signed graph, so we have modified the data from the dataset such that it suits our problem. Because of memory constraint, we have conducted experiments on different partitions of Epinions $E1 = 500$, $E2 = 1000$, $E3 = 2000$, $E4 = 3000$ nodes and Slashdot $S1 = 500$, $S2 = 1000$, $S3 = 2000$, $S4 = 3000$ nodes respectively. We have considered only those nodes which are connected bi-directionally such that both edges either having positive sign or negative sign. The description of dataset created by us is given in Table 2.

Computational Experiments and Results

This section illustrates various computational experiments which show the effectiveness of our proposed scheme. We have performed experiments on population size of 50 and evolved communities of different sizes which are discussed in the later sections. To maintain elitism and diversity in the solution we have used NSGA-II algorithm (Deb, 2011). The algorithm terminates after fifty generations. A percentage α (10%) of the generation is obtained by selecting the best fitness chromosomes that are cloned into the next generation. The elitist approach ensures that the best chromosomes obtained so far are retained for next generations. To keep the number of chromosomes among generations constant, the remaining $(1 - \alpha)$ percentage (i.e. 90%) of each new generation is obtained through the genetic operators (crossover and mutation). Parents are selected randomly using roulette wheel selection algorithm. This process iterates in every generation until the termination criterion is met. In general, for GA the probability of crossover P_c is chosen very high and for mutation P_m is very small. The parameters chosen for the multiobjective genetic algorithm are mentioned in the Table 3.

Comparison Between M-F-SBF Model and Existing Methods

Here, we will investigate the comparative performance of various schemes on the basis of above-mentioned datasets and metrics. The performance of our proposed model (M-F-SBF) is compared with the following baseline approaches:

- **Modularity Maximization (MMax)**: In this method, only one objective is optimized, that is, modularity (Clauset et al., 2004) using evolutionary approach (EA). After optimizing modularity, frustration and social balance factor are computed.
- **Frustration Minimization (FMin)**: In this, frustration (Doreian & Mrvar, 1996) is optimized using evolutionary approach and then the other two objectives (modularity and social balance factor) are calculated.

Apart from above assessed methods, we have also implemented the following method for the sake of comparison.

- **Social Balance Factor Maximization (SMax)**: Similarly, this method is used to optimize social balance factor (Leskovec et al., 2010) using EA followed by the computation of other two objectives.

TABLE 3. Genetic algorithm parameters.

Parameter	Value	Description
Population Size	50	The size of chromosomes in the population at each generation
Crossover Rate (**P_c**)	0.9	Probability of crossover between two chromosomes
Mutation Rate (**P_m**)	0.1	Probability of mutation of a chromosome
Stopping Criteria (No. of Iterations)	50	Stagnation of fitness value for 50 generations

[1] http://snap.stanford.edu/data/soc-sign-epinions.html
[2] http://snap.stanford.edu/data/soc-sign-Slashdot090221.html

DOI: 10.1002/asi

TABLE 4. Performance on Slovene Parliamentary Party.

	MMax	FMin	SMax	M-F model	F-SBF model	M-SBF model	M-F-SBF model
NMI	0.0399	0.1577	0.2783	0.7586	0.6190	0.7428	**1**
F_Score	0.3902	0.5714	0.5714	0.8372	0.7805	0.8275	**1**
Entropy	1	0.8900	0.7635	0.4236	0.3900	0.4012	**0**

Unlike above discussed three methods where objectives are optimized one at a time, in the following methods we have performed pair-wise optimization of objectives. Because it includes the optimization of multiple objectives simultaneously, dominance criteria is used to find the better solutions. In these methods, NSGA-II algorithm (Deb, 2011) is used for optimization.

- **Modularity and Frustration Optimization (M-F model)**: In M-F model (Amelio & Pizzuti, 2013), we have optimized modularity and frustration using NSGA-II algorithm. After optimizing both objectives, the third objective, that is, social balance factor is then computed.

In addition to above discussed methods, following methods have also been investigated for the purpose of assessment.

- **Frustration and SBF Optimization (F-SBF model)**: In F-SBF model, optimization of frustration (Doreian & Mrvar, 1996) and social balance factor (Leskovec et al., 2010) simultaneously is done using NSGA-II algorithm and after that value of modularity is calculated.
- **Modularity and SBF Optimization (M-SBF model)**: This approach also uses the NSGA-II approach to optimize multiple objectives of social balance factor (Leskovec et al., 2010) and modularity (Clauset et al., 2004). Once these objectives are optimized then the value of frustration is computed.

To avoid any random error, for each network the algorithm is executed five times and on varying community sizes. At each run, the solutions having the best value of objective are selected and the corresponding values of other objectives are computed. First, we provide the results of benchmarked datasets. Because the ground truth information is available for Slovene Parliamentary Party (Ferligoj & Kramberger, 1996) and Gahuku-Gama Subtribes (Read, 1954), the community size of 2 and 3 are kept as given in the original partition. The results obtained by performing experiments on these datasets are discussed below in detail.

Results Based on Benchmarked Datasets

From Table 4 and Table 5 we can easily observe that our proposed model M-F-SBF consistently outperforms all other methods (single objective optimization and pair-wise objectives optimization) for the benchmarked datasets of Slovene Parliamentary Party (Ferligoj & Kramberger, 1996) and Gahuku-Gama Subtribes (Read, 1954), respectively.

Also, the model has achieved the best possible values for NMI, F_score and Entropy as illustrated in Figure 8 and Figure 9.

Next, we have tried to establish the efficiency of our proposed algorithm by performing the experiments on large datasets of Epinions and Slashdot, and because of the unavailability of the ground truth information for both the datasets, we have used error rate as metric, already been used by other researchers (Amelio & Pizzuti, 2013; Yang et al. 2007). We have shown results on larger datasets for different sizes of Epinions $E1$, $E2$, $E3$, $E4$ and Slashdot $S1$, $S2$, $S3$, $S4$ respectively. Also, we have conducted experiments on varying community size (k) chosen empirically as 3, 5, 7, 11 and 13.

Results Based on Real-World Datasets

We investigate the performance of our proposed model on real-world datasets based on the above-mentioned error rate (Yang et al., 2007) metric only (because of the lack of ground truth information). Table 6, shows the values of error rate obtained for $E1$ network for different community sizes that are considerably less compared to other models, clearly indicating the effectiveness and better performance of our proposed model. Likewise, from Figure 10, we can easily observe that for other larger networks of $E2$, $E3$, $E4$ also, our model withstand its performance and has shown considerably reduced error rate values for varying community sizes.

From Table 7, the value of error rate achieved for $S1$ network by F-SBF model for community size 13 is little less compared to our model, still the results are comparable. Similarly, results obtained from experiments conducted on larger datasets of $S2$, $S3$, $S4$ shown in Figure 11 endorse that M-F-SBF has shown consistent performance on these large datasets as well, thus, overall our proposed model M-F-SBF has shown significantly better results compared to all other comparative models.

It is to be noted that there is still no general measure to validate and compare methods for signed networks. Also,

TABLE 5. Performance on Gahuku-Gama Subtribes.

	MMax	FMin	SMax	M-F model	F-SBF model	M-SBF model	M-F-SBF model
NMI	0.0597	0.3617	0.0597	0.5003	0.4607	0.5003	**1**
F_score	0.2703	0.4390	0.2703	0.6341	0.5652	0.6341	**1**
Entropy	1.4538	1.0103	1.4538	0.8049	0.8262	0.8049	**0**

DOI: 10.1002/asi

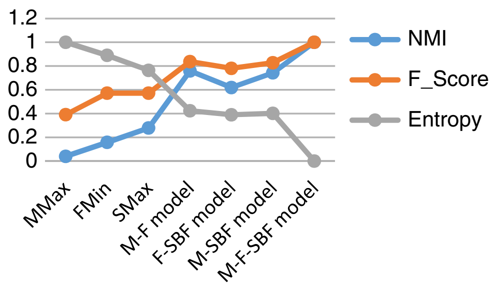

FIG. 8. Evaluation on Slovene Parliamentary Party based on different metrics. [Color figure can be viewed at wileyonlinelibrary.com]

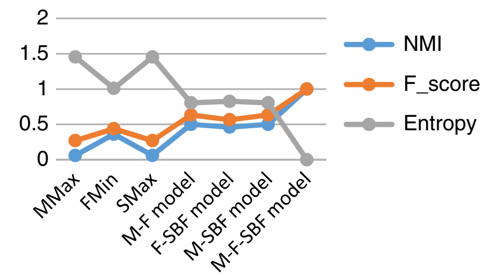

FIG. 9. Evaluation on Gahuku-Gama Subtribes based on different metrics. [Color figure can be viewed at wileyonlinelibrary.com]

as pointed out by the authors (Amelio & Pizzuti, 2013; Yang et al., 2007), the error function considers only the sign of the links, and completely disregards the edge density. To validate the effectiveness of the proposed M-F-SBF model we have given detailed analysis of the experiments performed on real-world datasets of *E*1 (Figure 12) and *S*1 (Figure 14) networks on varying community sizes. Next, for other larger networks of Epinions (*E*2, *E*3, *E*4) and Slashdot (*S*2, *S*3, *S*4), we have shown results in Figure 13 and Figure 15 respectively for the best value of *k* chosen empirically. The following section explains why our proposed model has performed better even if the error value is little compromised in the above-mentioned case of Slashdot dataset.

Detailed Analysis of Results Based on Epinions Network

As it can be seen from Figure 12(a), except a very comparable SBF value to others, M-F-SBF has performed better than all other methods in this case. Though the value of modularity has a drop in M-F-SBF compared to MMax, but the other two objectives frustration and SBF have given better values as shown in Figure 12(b). Also, frustration values of FMin and M-F-SBF are comparable, still other two objectives modularity and SBF have shown better results. In Figure 12(c), the values of different objectives are well optimized in M-F-SBF with a little less value of SBF than SMax which is still comparable considering the fact that in M-F-SBF multiobjective optimization is being performed, which results in trade-off among the values of the objectives being optimized, whereas in the case of SMax only one objective is being optimized, leaving the other two unattended. The results of M-F-SBF in Figure 12(d) are comparable to all other methods. Also, the frustration value is best optimized in M-F-SBF while sustaining the values of other objectives too. Again the results in Figure 12(e) show that M-F-SBF model preserves the trade-off among optimization of all the three objective values, the value of modularity is little compromised in M-F-SBF compared to MMax, but frustration in MMax is poorly optimized compared to most of the methods. Similarly, as shown in Figure 12(f), M-F-SBF has produced good values of modularity and SBF and performed well for frustration as compared to other methods. Thus, we observe that our proposed method has shown considerably better performance as compared to different baseline methods. Further, Figure 13 shows the results achieved at the best value of *k* obtained empirically for larger networks of *E*1 ($k = 3$), *E*3 ($k = 5$) and *E*4 ($k = 7$) that also reaffirms that our proposed model M-F-SBF has shown considerably better performance as compared to different baseline methods.

Detailed Analysis of Results Based on Slashdot Network

As evident from Figure 14, for Slashdot network also, purposed Scheme M-F-SBF has performed well compared to other methods. By using M-F-SBF, we are able to maintain the trade-off among values of all the three objectives rather than focusing on a particular objective as can be seen in Figure 14(a). Further in Figure 14(b), the value of SBF in SMax is better than M-F-SBF, but then in SMax frustration is merely optimized in this case. It is evident from Figure 14(c) that M-F-SBF has performed better considering all the three objectives. Although, SBF values are good in MMax, but frustration is poorly optimized compared to M-F-SBF. As shown in Figure 14(d), the values of modularity and frustration are best optimized by M-F-SBF compared to all the other methods. Also, the value of SBF is better than others and quite comparable

TABLE 6. Performance on *E*1 network.

	MMax	FMin	SMax	M-F model	F-SBF model	M-SBF model	M-F-SBF model
3	20.25979	16.3524	22.87869	17.50471	16.99141	20.01886	**14.83344**
5	15.18961	7.374817	17.53614	11.86885	9.710874	14.25728	**6.98722**
7	13.54494	5.541588	12.58119	6.599623	5.290174	11.10413	**3.415043**
9	10.88414	4.347371	10.97842	4.211188	3.257909	7.311963	**2.336057**
11	6.557721	3.268385	7.521475	3.195056	2.556044	5.813953	**1.393254**
13	7.594804	1.697046	9.930861	3.865493	1.309449	7.029122	**0.942803**

DOI: 10.1002/asi

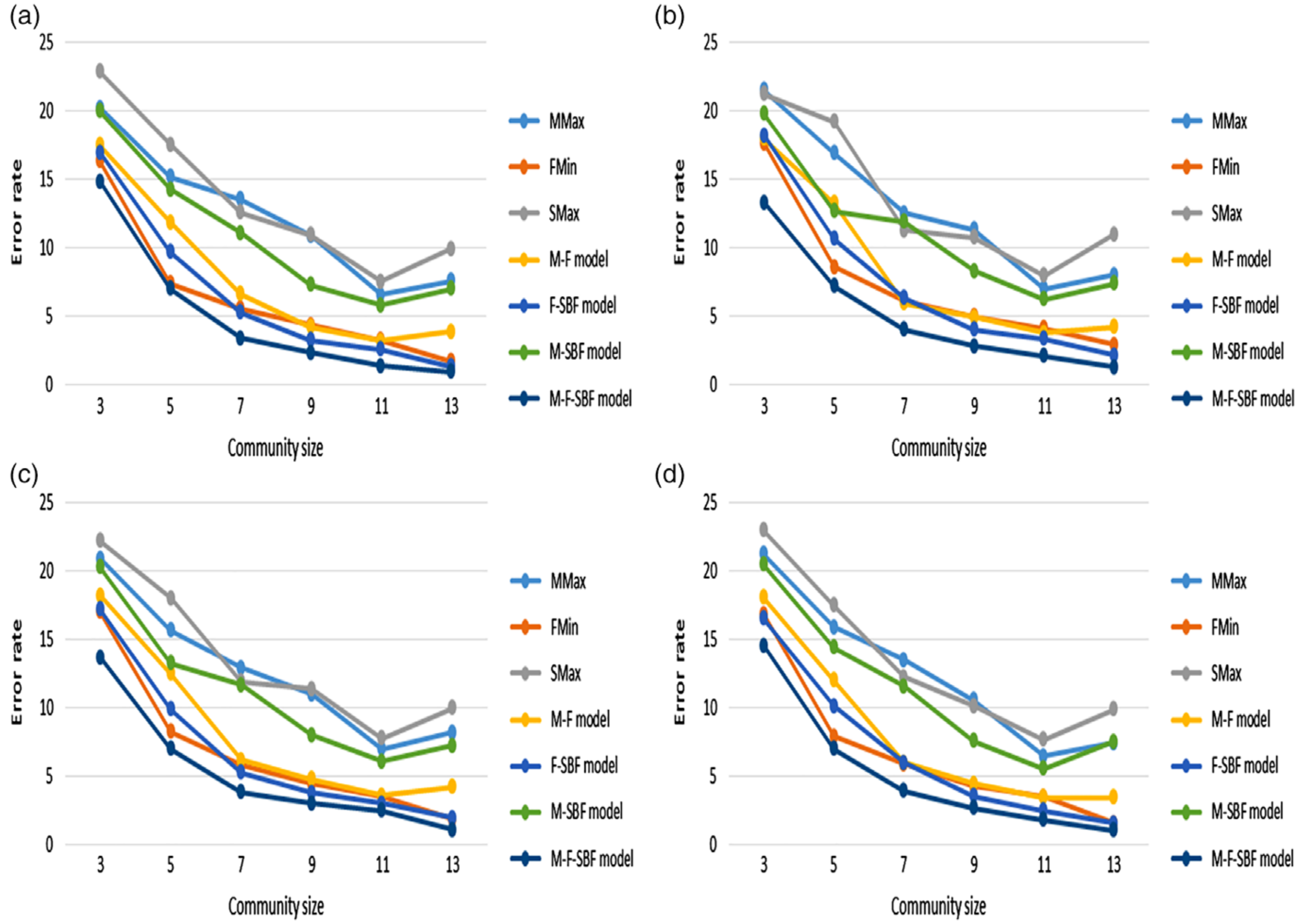

FIG. 10. Evaluation based on error rate for different Epinions networks (a) *E*1, (b) *E*2, (c) *E*3, (d) *E*4. [Color figure can be viewed at wileyonlinelibrary.com]

to M-SBF model. Considering the results shown in Figure 14(e), frustration is best optimized in M-F-SBF as compared to other methods. Although the value of modularity is compromised if equated to MMax, but then both frustration and SBF values for M-F-SBF have outperformed the MMax. Now considering the case in Figure 14 (f), M-F-SBF has comparable values of modularity and SBF compared to MMax but the frustration has better value in M-F-SBF and although the value of modularity is slightly less in M-F-SBF if matched with SMax, but then the other two objectives frustration and SBF have better values in M-F-SBF, especially frustration is optimized exceptionally well in M-F-SBF. Furthermore, in this case F-SBF and MMax both have shown good results but M-F-SBF has performed better than F-SBF in terms of two objectives and has shown better result for frustration compared to MMax. Likewise, results achieved for larger networks of *S*2 (k = 5), *S*3 (k = 9) and *S*4 (k = 11) shown in Figure 15 also restate consistent and better performance of M-F-SBF model as compared to other comparative models. Overall, we can say that, our proposed model has performed better in most of the cases and in some cases the value of objective is compromised, but still compared to what M-F-SBF achieves, the trade-off is better.

TABLE 7. Performance on *S*1 network.

	MMax	FMin	SMax	M-F model	F-SBF model	M-SBF model	M-F-SBF model
3	27.40879	15.481759	17.6333	15.060804	15.247895	17.91394	**13.14312**
5	24.69598	9.4013096	22.96539	8.2787652	8.0449018	19.22357	**7.623948**
7	27.36202	5.05145	8.278765	5.2853134	4.4434051	8.606174	**4.303087**
9	24.74275	2.3386342	7.390084	2.8063611	3.2273152	8.840037	**2.291862**
11	26.52011	2.3386342	7.343312	2.8063611	3.2740879	5.472404	**2.245089**
13	9.214219	2.2450889	18.84939	2.8531338	**1.8241347**	5.799813	2.011225

DOI: 10.1002/asi

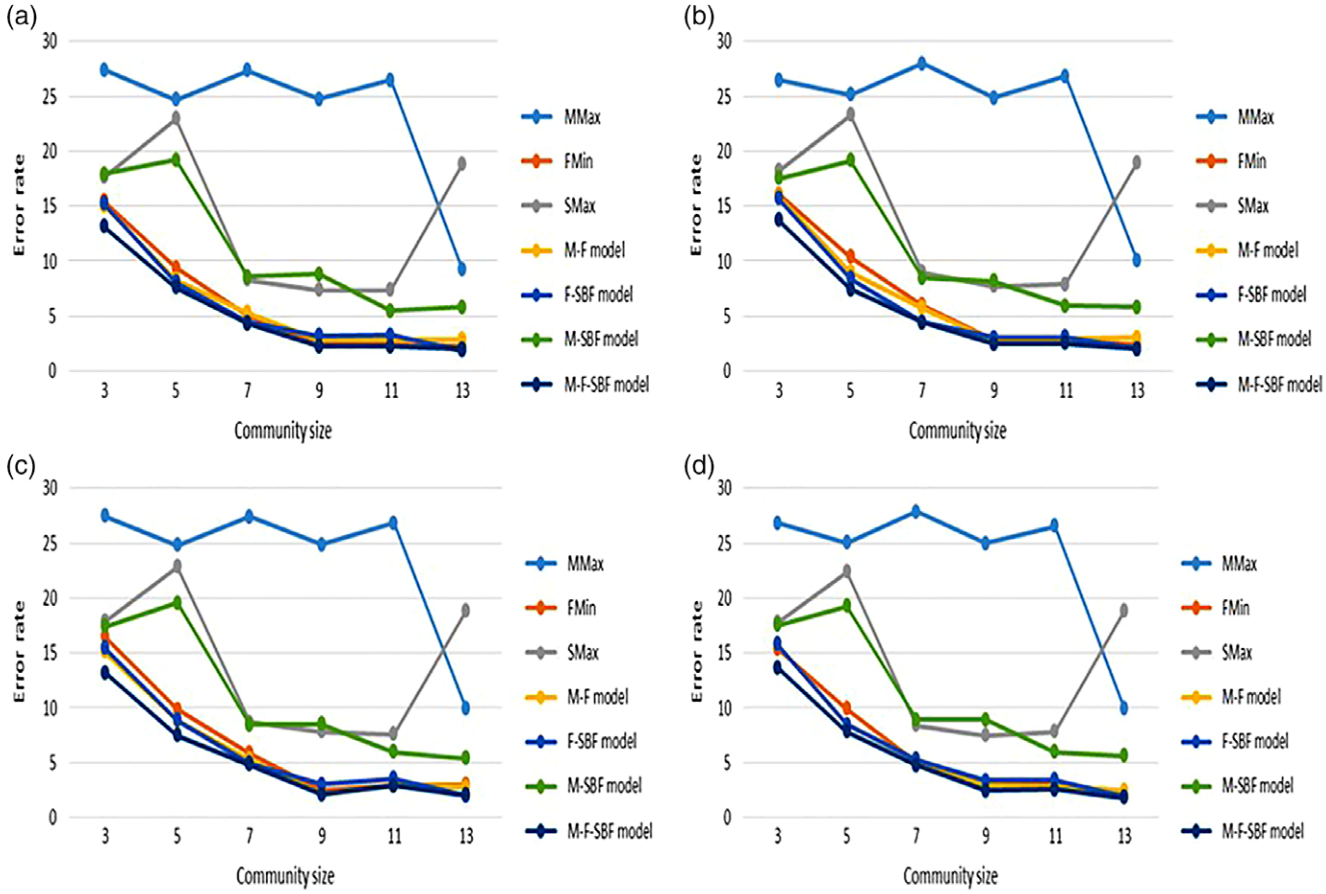

FIG. 11. Evaluation based on error rate for different Slashdot networks (a) $S1$, (b) $S2$, (c) $S3$, (d) $S4$. [Color figure can be viewed at wileyonlinelibrary.com]

Discussion

From the detailed analysis of the results on both benchmarked and real-world datasets given in the earlier section (Experimental Evaluation and Results), we can conclude that our proposed model M-F-SBF is able to achieve better trade-off among the values of all the three optimized objectives with comparable results with the best value of objective in some cases. Besides, in most of the cases in comparison to baseline methods, M-F-SBF results are superior giving maximum optimization of objectives in some cases, while maintaining the balance among optimization of all the objectives, which clearly establishes the superiority of our proposed model.

Our experimental results reaffirm that optimization of multiple objectives give better results than optimizing single objective individually which is also established by other authors (Shi et al., 2011).

Moreover, optimal number of communities can be estimated in the real-world datasets by analyzing results presented in the Experimental Evaluation and Results section. For example, Figure 16(a) shows values of modularity achieved for various networks ($E1$, $E2$, $E3$, and $E4$) for different values of k (3, 5, 7, 9, 11, and 13). It is clear from this Figure 16(a) that for $E1$ network, empirically selected optimal value of $k = 3$ has achieved the higher modularity value compared to the modularity obtained for other values of k(5, 7, 9, 11, and 13). Likewise, for larger varying partitions of Epinions networks ($E2$, $E3$, and $E4$), modularity values achieved at empirically selected k (3, 5, and 7 respectively) are better compared to the modularity obtained for other k values.

Similarly, results for Slashdot dataset, as shown in Figure 16(b), we can clearly see that for $S1$ network, empirically selected optimal value of $k = 3$ has achieved better modularity compared to other values for k (5, 7, 9, 11, and 13). Correspondingly, results achieved for modularity for larger varying partitions of Slashdot network ($S2$, $S3$, and $S4$) at empirically selected optimal value of k (5, 9, and 11 respectively) are better than for other k values.

This shows that our chosen approach to select optimal value of k empirically shows coherence with the work presented by (Fortunato & Barthelemy, 2007). According to which, modularity is considered as one of the criteria to evaluate the goodness of partitions of a network. Further, the congruency and consistency in the results achieved by our proposed scheme on varying larger partitions of Epinions ($E1$, $E2$, $E3$, and $E4$) and Slashdot ($S1$, $S2$, $S3$, and $S4$) networks validates its applicability on different kind of real-world social networks.

DOI: 10.1002/asi

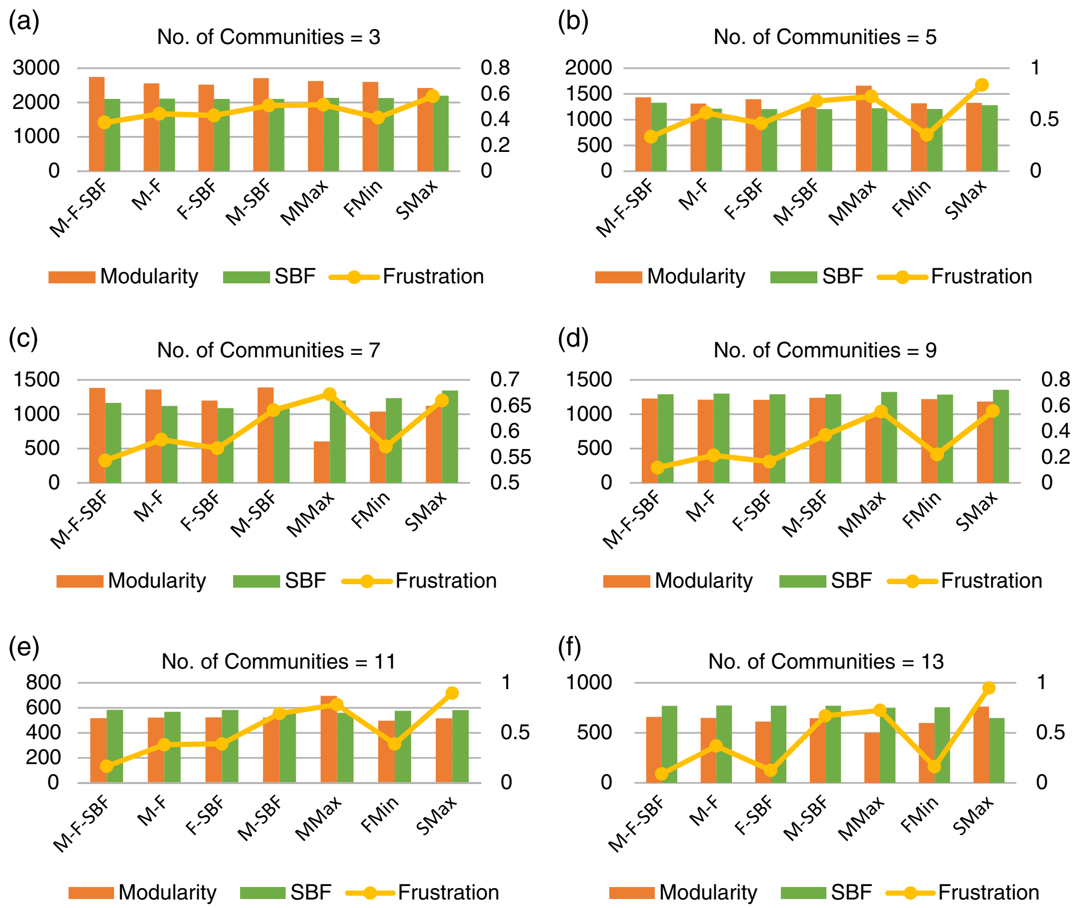

FIG. 12. Experimental results on $E1$ showing the values of Modularity, Frustration and Social Balance Factor for varying number of communities. [Color figure can be viewed at wileyonlinelibrary.com]

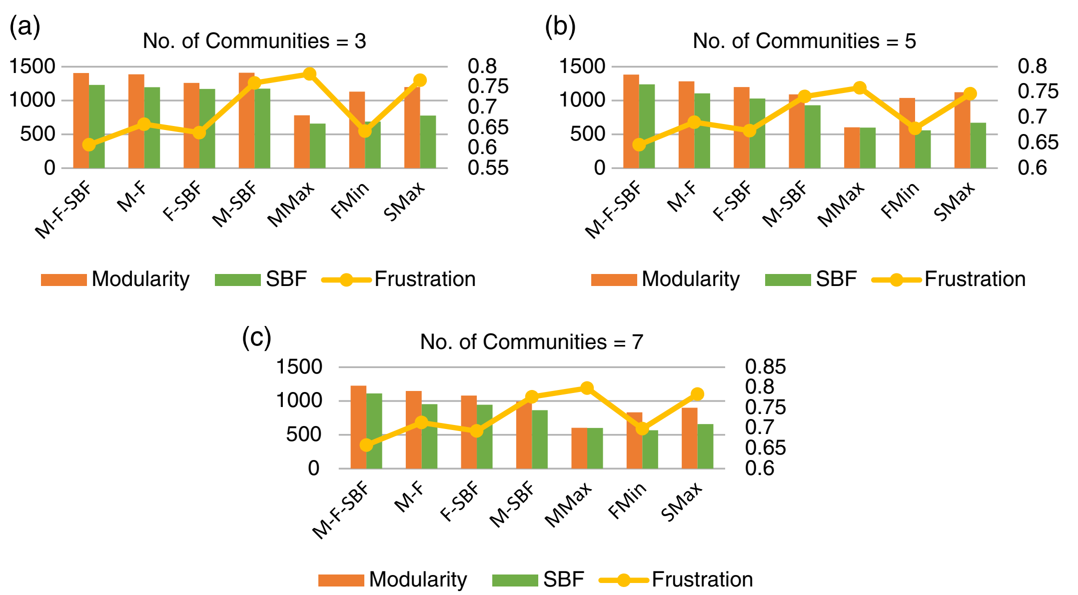

FIG. 13. Experimental results of different Epinions networks showing the values of Modularity, Frustration and Social Balance Factor for (a) $E2$ ($k = 3$), (b) $E3$ ($k = 5$), (c) $E4$ ($k = 7$). [Color figure can be viewed at wileyonlinelibrary.com]

We can conclude that our proposed model to form communities is the hybridization of all the baseline methods and results obtained through experiments on both small and large networks demonstrate that the proposed hybrid scheme performs considerably better than the component schemes.

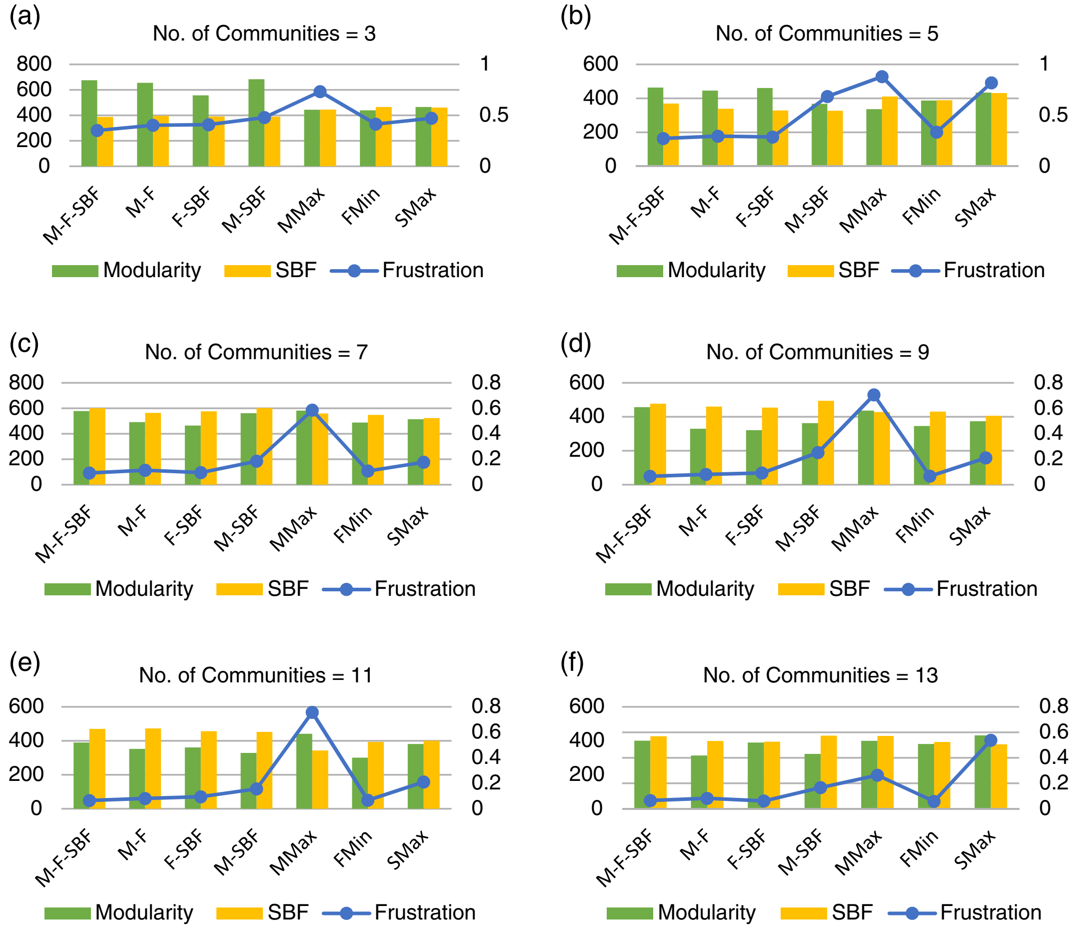

FIG. 14. Experimental results on *S*1 showing the values of Modularity, Frustration and Social Balance Factor for varying number of communities. [Color figure can be viewed at wileyonlinelibrary.com]

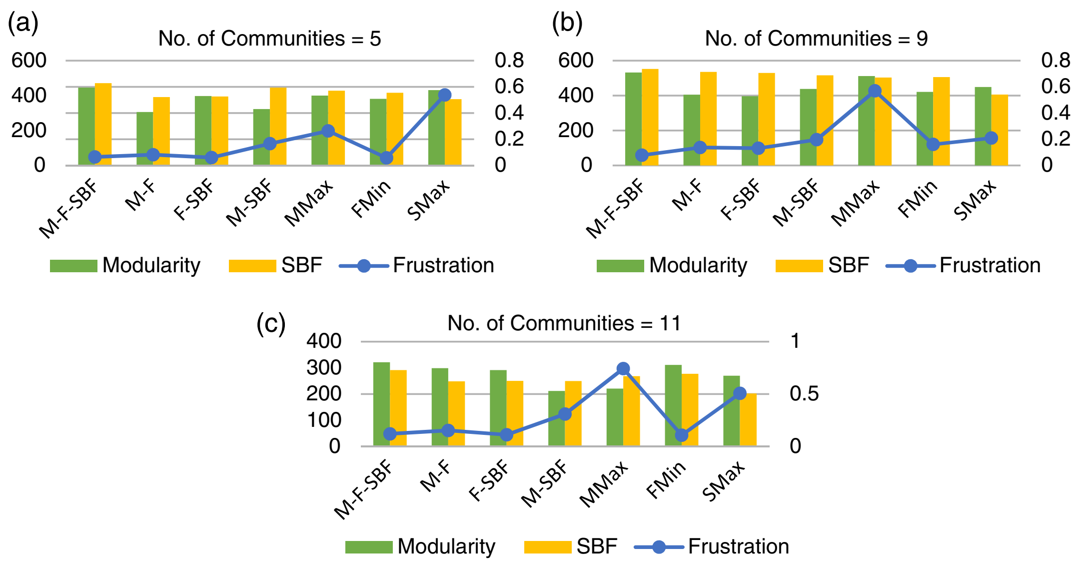

FIG. 15. Experimental results of different Slashdot networks showing the values of Modularity, Frustration and Social Balance Factor for (a) *S*2 (k = 5), (b) *S*3 (k = 9), (c) *S*4 (k = 11). [Color figure can be viewed at wileyonlinelibrary.com]

Conclusion and Future Direction

In this work, we have addressed successfully the problem of community detection in signed social network using multiobjective optimization approach. Our focus was on evolving better communities through evolutionary approach by considering both the signs of the link and link density. In

DOI: 10.1002/asi

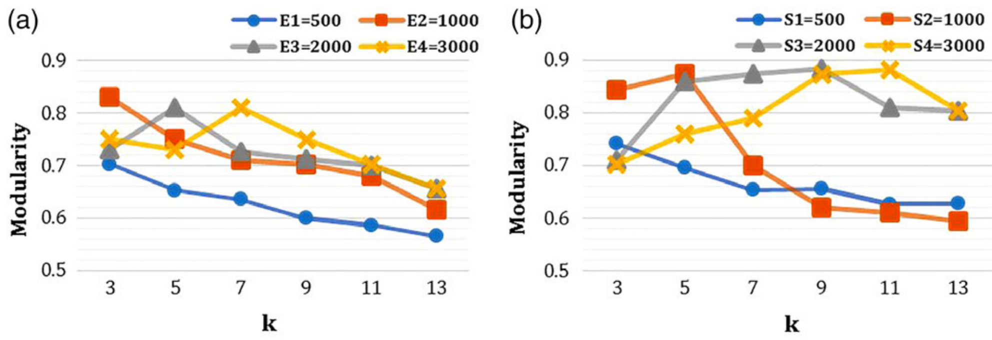

FIG. 16. Modularity obtained on varying k (3, 5, 7, 9, 11, and 13) on (a) Epinions dataset ($E1$, $E2$, $E3$, $E4$) and (b) Slashdot dataset ($S1$, $S2$, $S3$, $S4$). [Color figure can be viewed at wileyonlinelibrary.com]

our proposed scheme we have optimized the three main properties of signed social networks – modularity (Newman & Girvan, 2004), frustration (Doreian & Mrvar, 1996) and social balance factor (Leskovec et al., 2010). As in multiobjective optimization there is a trade-off among objectives thus resulting in multiple solutions rather than one best solution, so it covers wider solution space of the problem. As NSGA-II algorithm (Deb, 2011) is incorporated, it maintains the elitism and the diversity in the solutions. The experiments performed and the analysis of results has clearly established the efficiency and effectiveness of our proposed model.

As currently we worked on undirected signed social networks, handling directed signed networks would be an important future research direction. In future work, we will consider parallel implementation of the developed algorithms to make our proposed scheme scalable to handle large datasets. Other future research directions would be detection of overlapping communities (Awal & Bharadwaj, 2017; Papadopoulos et al., 2012), by incorporating trust, distrust and reputation in community detection models as quality measures of communities (Amelio & Pizzuti, 2013). Further this work can be extended by exploiting interaction patterns for accurate interpretation of interactions and link strength between the users (Chen et al., 2014).

References

Agarwal, V., & Bharadwaj, K.K. (2015). Predicting the dynamics of social circles in ego networks using pattern analysis and GA K-means clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(3), 113–141.

Amelio, A., & Pizzuti, C. (2013, August). Community mining in signed networks: a multiobjective approach. Paper presented at the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Niagara, Ontario, Canada.

Anchuri, P., & Magdon-Ismail, M. (2012, August). Communities and balance in signed networks: A spectral approach. Paper presented at the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), Istanbul, Turkey.

Awal, G. K., & Bharadwaj, K. K. (2017). Leveraging collective intelligence for behavioral prediction in signed social networks through evolutionary approach. Information Systems Frontiers. Advance online publication. https://doi.org/10.1007/s10796-017-9760-4.

Bedi, P., & Sharma, C. (2016). Community detection in social networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 6(3), 115–135.

Bhat, S.Y., & Abulais, M. (2015). OCMiner: A density-based overlapping community detection method for social networks. Intelligent Data Analysis, 19(4), 917–947.

Chen, Y.L., Chuang, C.H., & Chiu, Y.T. (2014). Community detection based on social interactions in a social network. Journal of the Association for Information Science and Technology, 65(3), 539–550.

Clauset, A., Newman, M.E., & Moore, C. (2004). Finding community structure in very large networks. Physical Review E, 70(6), 066111.

Dang, T. A., & Viennet, E. (2012). Community detection based on structural and attribute similarities. Paper presented at the Sixth International Conference on Digital Society, Valencia, Spain.

Deb, K. (2011). Multi-objective optimisation using evolutionary algorithms: An introduction. In L. Wang, A.H.C. Ng, & K. Deb (Eds.), Multi-objective evolutionary optimisation for product design and manufacturing (pp. 3–34). London: Springer.

Doreian, P. (2008). A multiple indicator approach to blockmodeling signed networks. Social Networks, 30(3), 247–258.

Doreian, P., & Mrvar, A. (1996). A partitioning approach to structural balance. Social Networks, 18(2), 149–168.

Ester, M., Kriegel, H.P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. KDD, 96(34), 226–231.

Facchetti, G., Iacono, G., & Altafini, C. (2011). Computing global structural balance in large-scale signed social networks. Proceedings of the National Academy of Sciences, 108(52), 20953–20958.

Ferligoj, A., & Kramberger, A. (1996). An analysis of the Slovene parliamentary parties network. Developments in Statistics and Methodology, 12, 209–216.

Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. Proceedings of the National Academy of Sciences, 104(1), 36–41.

Girdhar, N., & Bharadwaj, K. K. (2016). Signed social networks: a survey. Paper presented at the International Conference on Advances in Computing and Data Sciences, Singapore.

Handl, J., & Knowles, J. (2007). An evolutionary approach to multiobjective clustering. IEEE Transactions, Evolutionary Computation, 11(1), 56–76.

Kaur, S., Singh, S., Kaushal, S., & Sangaiah, A.K. (2016). Comparative analysis of quality metrics for community detection in social networks using genetic algorithm. Neural Network World, 26(6), 625–641.

Kunegis, J., Lommatzsch, A., & Bauckhage, C. (2009, April). The slashdot zoo: Mining a social network with negative edges. Paper presented at the 18th International Conference on World Wide Web, Madrid, Spain.

Lancichinetti, A., Fortunato, S., & Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics, 11(3), 033015.

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010, April). Signed networks in social media. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA.

Li, Y., Liu, J., & Liu, C. (2014). A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks. Soft Computing, 18(2), 329–348.

Liu, C., Liu, J., & Jiang, Z. (2014). A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks. IEEE Transactions, Cybernetics, 44(12), 2274–2287.

Liu, J., Zhong, W., Abbass, H.A., & Green, D.G. (2010, July). Separated and overlapping community detection in complex networks using multiobjective evolutionary algorithms. Paper presented at the IEEE Congress on Evolutionary Computation, Barcelona, Spain.

Moshirpour, M., Chelmis, C., Prasanna, V., Saravanan, M., Karthikeyan, P., Arathi, A., ... & Mohammad, H. (2013). Advances in Social Networks Analysis and Mining (ASONAM). Niagara Falls, Canada: IEEE Computer Society's Conference Publishing Services.

Newman, M.E., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E, 69(2), 026113.

Papadopoulos, S., Kompatsiaris, Y., Vakali, A., & Spyridonos, P. (2012). Community detection in social media. Data Mining and Knowledge Discovery, 24(3), 515–554.

Pizzuti, C. (2008). Ga-net: A genetic algorithm for community detection in social networks. In G. Rudolph, T. Jansen, N. Beume, S. Lucas, & C. Poloni (Eds.), International Conference on Parallel Problem Solving from Nature (pp. 1081–1090). Berlin, Heidelberg: Springer.

Pizzuti, C. (2009, November). A multi-objective genetic algorithm for community detection in networks. Paper presented at the 2009 21st IEEE International Conference on Tools With Artificial Intelligence, Newark, NJ.

Plantié, M., & Crampes, M. (2013). Survey on social community detection. In N. Ramzan, R. van Zwol, J.-S. Lee, K. Clüver, & X.-S. Hua (Eds.), Social media retrieval, computer communications and networks (pp. 65–85). London: Springer-Verlag.

Read, K.E. (1954). Cultures of the central highlands, New Guinea. Southwestern Journal of Anthropology, 10(1), 1–43.

Shi, C., Yu, P.S., Cai, Y., Yan, Z., & Wu, B. (2011, October). On selection of objective functions in multi-objective community detection. Paper presented at the 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. IEEE Transactions, Pattern Analysis and Machine Intelligence, 22(8), 888–905.

Verma, A., & Bharadwaj, K.K. (2017). Identifying community structure in a multi-relational network employing non-negative tensor factorization and GA K-means clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7(1), 1–22.

Wu, L., Ying, X., Wu, X., Lu, A., & Zhou, Z.H. (2011). Spectral analysis of k-balanced signed graphs. In J. Zhexue Huang, L. Cao, & J. Srivastava (Eds.), Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 1–12). Berlin, Heidelberg: Springer.

Yang, B., Cheung, W.K., & Liu, J. (2007). Community mining from signed social networks. IEEE Transactions, Knowledge and Data Engineering, 19(10), 1333–1348.

Zhang, K., Lo, D., Lim, E.P., & Prasetyo, P.K. (2013). Mining indirect antagonistic communities from social interactions. Knowledge and Information Systems, 35(3), 553–583.

Zheng, X., Zeng, D., & Wang, F.Y. (2015). Social balance in signed networks. Information Systems Frontiers, 17(5), 1077–1095.

DOI: 10.1002/asi