GA-Net: A Genetic Algorithm for Community Detection in Social Networks

Clara Pizzuti

ICAR-CNR,
Via P. Bucci 41C, 87036 Rende (CS), Italy
pizzuti@icar.cnr.it

Abstract. The problem of community structure detection in complex networks has been intensively investigated in recent years. In this paper we propose a genetic based approach to discover communities in social networks. The algorithm optimizes a simple but efficacious fitness function able to identify densely connected groups of nodes with sparse connections between groups. The method is efficient because the variation operators are modified to take into consideration only the actual correlations among the nodes, thus sensibly reducing the research space of possible solutions. Experiments on synthetic and real life networks show the capability of the method to successfully detect the network structure.

1 Introduction

The suitability of networks to represent many real world systems has given an impressive spur to the recent research area of complex networks. Collaboration networks, the Internet, the world-wide-web, biological networks, communication and transport networks, social networks are just some examples. Networks, in general, are constituted by a set of objects and by a set of interconnections among these objects. In social networks the objects are people and the connections represent social relations, such as common interests, friendship, religion, and so on. An interesting property to investigate, typical to many networks, is the *community structure*, i.e. the division of networks into groups (also called clusters) having dense intra-connections, and sparse inter-connections. The capability of detecting the partitioning of a network in clusters can give important information and useful insights to understand how the structure of ties affects individuals and their relationships. The problem of community detection has been receiving a lot of attention and many different approaches have been proposed [10,16,18,4,20,2,11].

In this paper we propose a new algorithm, named *GA-Net*, to discover communities in networks by employing genetic algorithms. The approach introduces the concept of *community score* to measure the quality of a partitioning in communities of a network, and tries to optimize this quantity by running the genetic algorithm. All the dense communities present in the network structure are obtained at the end of the algorithm by selectively exploring the search space, without the need to know in advance the exact number of groups. Specialized variation operators allow to reduce the space of the possible solutions thus improving the convergence of the method. The main novelties of the approach can be summarized as follows. The concept of *community score*, that provides a global quality measure of a partitioning in communities, is defined. The notion

G. Rudolph et al. (Eds.): PPSN X, LNCS 5199, pp. 1081–1090, 2008.

of $safe$ individual is introduced to avoid useless computation to the genetic algorithm and specialized variation operators that generate safe individuals are employed. Unlike many existing methods, the algorithm does not require the number of communities to find. This number is automatically determined by the optimal value of the *community score*. Experiments on synthetic and real life networks show the capability of the genetic approach to correctly detect communities with results comparable with state-of-the-art approaches.

The paper is organized as follows. The next section provides the necessary background to formalize the problem and defines the quality metric. In section 3 a description of the representation adopted and the variation operators used is provided. In section 4 an overview of the main proposals in community detection algorithms is given. In section 5, finally, the results of the method on synthetic and real life data sets are presented.

2 Problem Definition

A social network $\mathcal{SN}$ can be modelled as a graph $G = (V, E)$ where V is a set of objects, called nodes or vertices, and E is a set of links, called edges, that connect two elements of V. A community (or cluster) in a network is a group of vertices having a high density of edges within them, and a lower density of edges between groups. The problem of detecting k communities in a network, where the number k is unknown, can be formulated as finding a partitioning of the the nodes in k subsets that are highly intra-connected and sparsely inter-connected. To deal with graphs, often the adjacency matrix is used. If the network is constituted by N nodes, the graph can be represented with the $N \times N$ adjacency matrix A, where the entry at position (i, j) is 1 if there is an edge from node i to node j, 0 otherwise. The problem of detecting communities in a network can then be transformed to that of finding a partitioning of A in k sub-matrices that maximize the sum of densities of the sub-matrices. A naive density measure for a sub-matrix of n rows/columns is the number of ones (i.e. interactions) it contains. The higher the number of ones, the more connected the n nodes. However, counting the number of interactions does not give any information about the interconnections among the nodes. A density measure based on volume and row/column means, allowing to detect maximal and dense sub-matrices, has been introduced in [1], and applied to find Co-clusters in sparse binary matrices. Co-clustering[13], also known as bi-clustering, differently from clustering, tries to simultaneously group both the dimensions (objects and features) of a data set. Sub-matrix identification can be considered as a special case of co-clustering in which the two dimensions represent the same concept, i.e. the nodes of the graph. In the following the density measure used in [1], specialized for adjacency matrices, is recalled, and the new concept of *community score*, that gives a global measure of the network partitioning in clusters, is defined. In the following, without loss of generality, we assume an undirected graph. This assumption implies that the adjacency matrix is symmetric.

Let $S = (I, J)$ be sub-matrix of A, where I is a subset of the rows $X = \{I_1, \ldots, I_N\}$ of A, and J is a subset of the columns $Y = \{J_1, \ldots, J_N\}$ of A.

Let a_{iJ} denote the *mean value* of the ith row of the S, and a_{Ij} the mean of the jth column of S. More formally,

$a_{iJ} = \frac{1}{|J|}\sum_{j\in J} a_{ij}$, and $a_{Ij} = \frac{1}{|I|}\sum_{i\in I} a_{ij}$

The *volume* v_S of a sub-matrix $S = (I, J)$ is the number of 1 entries a_{ij} such that $i \in I$ and $j \in J$, that is $v_S = \sum_{i\in I, j\in J} a_{ij}$.

Given a sub-matrix $S = (I, J)$, the *power mean of S of order* r, denoted as $\mathbf{M}(S)$ is defined as

$$\mathbf{M}(S) = \frac{\sum_{i\in I}(a_{iJ})^r}{|I|}$$

A measure based on volume and row/column mean, that allows the detection of maximal and dense sub-matrices, can be defined as follows. Given a sub-matrix $S = (I, J)$, let $\mathbf{M}(S)$ be the power mean of S of order r. The *score* of S is defined as $Q(S) = \mathbf{M}(S) \times v_S$. The *community score* of a partitioning $\{S_1, \ldots S_k\}$ of A is defined as

$$\mathcal{CS} = \sum_{i}^{k} Q(S_i)$$

The problem of community identification can be formulated as the problem of maximize $\mathcal{CS}$. It is worth to note that higher values of the exponent r bias the CS towards matrices containing a low number of zeroes. In fact, it amplifies the weight of the densely interconnected nodes, while reducing those of less connected in the computation of the *community score*. In the experimental result section we show that when the modular structure of the network is not well defined, higher values of r help in detecting communities.

3 Genetic Representation and Operators

In this section we give a description of the algorithm *GA-Net*, the representation adopted for partitioning the network, and the variation operators used.

Genetic representation. Our clustering algorithm uses the locus-based adjacency representation proposed in [19] and employed by [9,14] for multiobjective clustering. In this graph-based representation an individual of the population consists of N genes $g_1, \ldots, g_N$ and each gene can assume allele values j in the range $\{1, \ldots, N\}$. Genes and alleles represent nodes of the graph $G = (V, E)$ modelling a social network $\mathcal{SN}$, and a value j assigned to the ith gene is interpreted as a link between the nodes i and j of V. This means that in the clustering solution found i and j will be in the same cluster. A decoding step, however, is necessary to identify all the components of the corresponding graph. The nodes participating to the same component are assigned to one cluster. As observed in [9], the decoding step can be done in linear time. A main advantage of this representation is that the number k of clusters is automatically determined by the number of components contained in an individual and determined by the decoding step. Suppose to have the network shown in figure 1(a). It consists of eleven nodes numbered from 1 to 11. The network can be partitioned in the three groups visualized by different colors and shapes of the nodes. Out of the many possible genotypes, that showed in figure 1(b), corresponding to the optimal solution, is translated in the graph structure given in figure 1(c). Each connected component provides a grouping of nodes that corresponds to the partitioning of the network in figure 1(a).

Objective Function. As described above, the decoding of an individual provides a different number k of components $\{S_1, \ldots S_k\}$ in which the graph is partitioned. We are interested in identifying a partitioning that optimizes the *community score* because, as already discussed in the previous section, this guarantees highly intra-connected and sparsely inter-connected communities. The objective function is thus $\mathcal{CS} = \sum_i^k Q(S_i)$

Initialization. Our initialization process takes in account the effective connections of the nodes in the social network. A random generation of individuals could generate components that in the original graph are disconnected. In fact, a randomly generated individual could contain an allele value j in the ith position, but no connection exists between the two nodes i and j, i.e. the edge (i, j) is not present. In such a case it is obvious that grouping in the same cluster both nodes i and j is a wrong choice. In order to overcome this drawback, once an individual is generated, it is $repaired$, that is a check is executed to verify that an effective link exists between a gene at position i and the allele value j. This value is maintained only if the edge (i, j) exists. Otherwise, j is substituted with one of the neighbors of i. This guided initialization biases the algorithm towards a decomposition of the network in connected groups of nodes. We call an individual generating this kind of partitioning $safe$ because it avoids uninteresting divisions containing unconnected nodes. $Safe$ individuals improve the convergence of the method because the space of the possible solutions is restricted.

Uniform Crossover. We used uniform crossover because it guarantees the maintenance of the effective connections of the nodes in the social network in the child individual. In fact, because of the biased initialization, each individual in the population is safe, that is it has the property, that if a gene i contains a value j, then the edge (i, j) exists. Thus, given two safe parents, a random binary vector is created. Uniform crossover then selects the genes where the vector is a 1 from the first parent, and the genes where the vector is a 0 from the second parent, and combines the genes to form the child. The child at each position i contains a value j coming from one of the two parents. Thus the edge (i, j) exists. This implies that from two safe parents a safe child is generated. Figure 1 on the right shows an example of crossover. Two parents, individuals A and B, and their graph-based representations are reported. Uniform crossover of A and B gives the child C.

Mutation. The mutation operator that randomly change the value j of a i-th gene causes a useless exploration of the search space, because of the same above observations on node connections. Thus the possible values an allele can assume are restricted to the neighbors of gene i. This $repaired$ mutation guarantees the generation of a safe mutated child in which each node is linked only with one of its neighbors.

Given a network $\mathcal{SN}$ and the graph G modelling it, *GA-Net* starts with a population initialized at random and $repaired$ to produce $safe$ individuals. Every individual generates a graph structure in which each component is a connected subgraph of G. For a fixed number of generations the genetic algorithm computes the fitness function of each solution member, and applies the specialized variation operators to produce the new population. In the experimental result section we show that the fitness function guides the genetic algorithm to successfully identify the best partitioning of $\mathcal{SN}$,

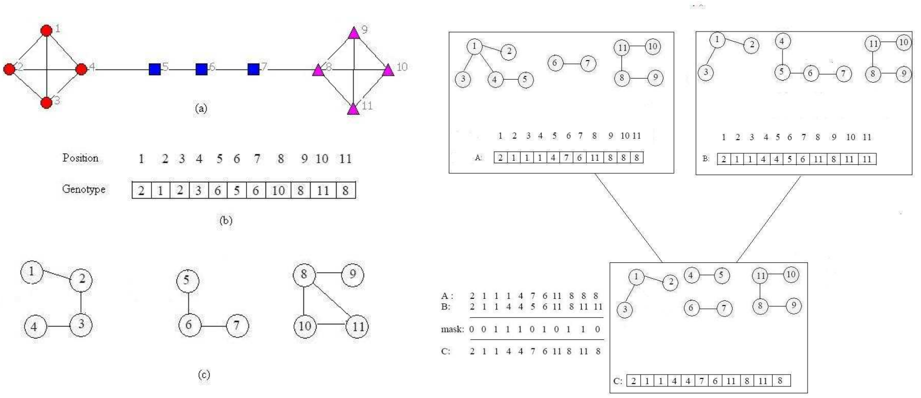

(a)
Position 1 2 3 4 5 6 7 8 9 10 11
Genotype 2 1 2 3 6 5 6 10 8 11 8
(b)
(c)
A: 2 1 1 1 4 7 6 11 8 8 8
B: 2 1 1 4 4 5 6 11 8 11 11
A : 2 1 1 1 4 7 6 11 8 8 8
B: 2 1 1 4 4 5 6 11 8 11 11
mask: 0 0 1 1 1 0 1 0 1 1 0
C: 2 1 1 4 4 7 6 11 8 11 8
C: 2 1 1 4 4 7 6 11 8 11 8

Fig. 1. (a) A network modelled as a graph; (b) the locus-based representation of a genotype; (c) the graph-based structure of the genotype. On the right Uniform crossover of two individuals, their genotype, their graph-based representation, and the child generated.

converging in a few iterations to the solution. Before presenting the experiments, in the next section an overview of the main approaches to community detection is given.

4 Related Work

Many different algorithms, coming from different fields such as physics, statistics, data mining, have been proposed to detect communities in complex networks [8,10,16,18,4,20,17,2,11]. In the following we review some of the most known.

One of the most famous algorithm has been presented by Newman and Girvan [8,18]. The method is a divisive hierarchical clustering method [6] based on an iterative removal of edges from the network. The edge removal splits the network in communities. The edges to remove are chosen by using *betweenness* measures. The idea underlying the edge betweenness comes from the observation that if two communities are joined by a few inter-community edges, then all the paths from vertices in one community to vertices in the another must pass through these edges. Paths determine the betweenness score to compute for the edges. By counting all the paths passing through each edge, and removing the edge scoring the maximum value, the connections inside the network are broken. This process is repeated, thus dividing the network into smaller components until a stop criterion is reached. The criterion adopted to stop the division is the *modularity*. Given k communities, the modularity is defined as follows. Let e_{ij} be the fraction of edges in the network connecting vertices from group i to those of group j, and $a_i = \sum_j e_{ij}$. Then $M = \sum_i (e_{ii} - a_i^2)$ is the fraction of edges inside communities minus the expected value of the fraction of edges if edges fall at random without regard to the community structure. Values approaching 1 indicate strong community structure. Thus the algorithm computes the modularity of all the clusters obtained by applying the hierarchical approach, and returns as result the clustering having the highest value of modularity. An agglomerative variant of this approach is presented in [16], and a faster method version, based on the same strategy, is described in [4].

Hopcroft et al. [10] present an agglomerative hierarchical method for clustering large linked networks to identify stable or natural cluster. A cluster is deemed natural if it appears in the clustering process when a given percentage of links are removed.

Radicchi et al. [20] propose two quantitative definitions of community and an algorithm to identify communities. The quantitative definitions of community are based on the degree of a node. A subgraph is a *community in strong sense* if each node has more connections within the community than the rest of the graph. A subgraph is a *community in a weak sense* if the sum of all in-degrees in V is greater than the sum of the out-degrees. The algorithm is a divisive hierarchical method based on the concept of *edge-clustering coefficient*, defined in analogy with the node clustering coefficient [15], as the number of triangles an edge participates, divided by the number of triangles it might belong to, given the degree of the adjacent nodes. Their algorithm works like that of Newmann and Girvan, the difference being that instead of choosing to remove the edge with the highest edge betweenness, the removed edges are those having the smallest value of edge-clustering coefficient.

Approaches to community detection based on Genetic Algorithms can be found in [21,22,7]. In [21,22] the authors present a genetic algorithm that uses as fitness function the network modularity proposed by Newmann and Girvan. An individual is constituted by N genes, where N is the number of objects. The ith gene corresponds to the ith node, and its value is the community identifier of node i. They use a non standard one-way crossover operation in which, given two individuals A and B, a community identifier j is chosen at random, and the identifier j of the nodes $j_1, \ldots j_h$ of A is transferred to the same nodes of B.

A different approach is described in [7] where a random walk distance measure between graphs is integrated in a genetic algorithm to cluster social networks. The representation they use is the k-medoids where each cluster center is represented by one of the nodes of the network. Of course this means that the number k of clusters must be known in advance. The fitness function tries to minimize the sum of all the pair-wise distances between nodes.

5 Experimental Results

In this section we study the effectiveness of our approach on a synthetic data set. Then we compare the results obtained by *GA-Net* with those reported by Girvan and Newman in [8,18,17] on some real-worlds networks for which the partitioning in communities is known. In both cases we show that our genetic algorithm successfully detects the network structure and is competitive with that of Girvan and Newman. The *GA-Net* algorithm has been written in MATLAB, using the Genetic Algorithms and Direct Search Toolbox 2. The experiments have been performed on a Pentium 4 machine, 1800MHz, 1GB RAM. We employed standard parameters for the genetic algorithm, crossover rate 0.8, mutation rate 0.2, elite reproduction 10% of the population size, roulette selection function. The population size was 300, the number of generations 30.

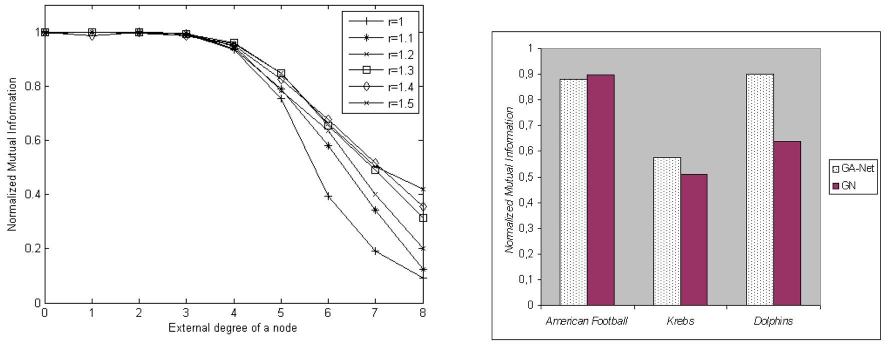
Synthetic data set. In order to check the ability of our approach to successfully detect the community structure of a network, we use the benchmark proposed by Girvan and Newan in [8]. The network consists of 128 nodes divided into four communities of 32 nodes each. Edges are placed between vertex pairs at random but such that $z_{in} + z_{out} = 16$, where z_{in} and z_{out} are the internal and external degree of a node with respect to its community. If $z_{in} > z_{out}$ the neighbors of a node inside its group are more than the neighbors belonging to the other three groups, thus a good algorithm should discover them. We generated 50 different networks for values of z_{out} ranging from 0 to 8, and used the *Normalized Mutual Information* to measure the similarity between the true partitions and the detected ones. The *Normalized Mutual Information* is a similarity measure proved to be reliable by Danon et al. [5]. Given two partitions A and B of a network in communities, let C be the confusion matrix whose element C_{ij} is the number of nodes of community i of the partition A that are also in the community j of the partition B. The normalized mutual information $I(A, B)$ is defined as :

$$I(A,B) = \frac{-2\sum_{i=1}^{c_A}\sum_{j=1}^{c_B} C_{ij} log(C_{ij}N/C_{i.}C_{.j})}{\sum_{i=1}^{c_A} C_{i.} log(C_{i.}/N) + \sum_{j=1}^{c_B} C_{.j} log(C_{.j}/N)}$$

where c_A (c_B) is the number of groups in the partition A (B), $C_{i.}$ ($C_{.j}$) is the sum of the elements of C in row i (column j), and N is the number of nodes. If $A = B$, $I(A, B) = 1$. If A and B are completely different, $I(A, B) = 0$.

Figure 2(a) shows the normalized mutual information, averaged over the 50 runs, for different values of the exponent r when the external degree z_{out} increases from 0 to 8. The figure point out that until $z_{out} \leq 5$ the algorithm is successful in detecting the true communities in almost more than 80% of cases, independently the value of r. However, as soon as the network fuzziness increases, in order to discover at least 50% of the true groups the parameter r plays an important role. In fact, the higher the number of interconnections, the more indistinguishable the network structure because communities are mixed with each other, but augmenting the value of r the algorithm is still able to identify the hidden groups.

Real-life data set. We now show the application of *GA-Net* on three real-world networks, the *American College Football*, *Krebs' books on American politics*, *Bottlenose*



(a) (b)

Fig. 2. (a): Normalized mutual information obtained by *GA-Net* on the synthetic data set for different values of the exponent r. (b): Comparison of GA-Net and Girvan and Newman's (denoted GN) algorithms relative to Normalized Mutual Information for American College Football, Krebs'political books, Dolphins data sets.

Dolphins, well studied in the literature and compare our results with those obtained by Girvan and Newman in [8,18,17]. The American College Football network [8] comes from the United States college football. The network represents the schedule of Division I games during the 2000 season. Nodes in the graph represent teams and edges represent the regular season games between the two teams they connect. The teams are divided in conferences. The teams on average played 4 inter-conference matches and 7 intra-conference matches, thus teams tend to play between members of the same conference. The network consists of 115 nodes and 616 edges grouped in 12 teams. The network of political books was compiled by V. Krebs. The nodes represent 105 recent books on American politics brought from Amazon.com, and edges join pairs of books frequently purchased by the same buyer (unpublished http://www.orgnet.com/). Books were divided by Newman [17] according to their political alignment (conservative or liberal), except for a small number of books (13) having no clear affiliation. The last example is the social network of 62 bottlenose dolphins living in Doubtful Sound, New Zealand, compiled by Lusseau [12] from seven years of dolphins behavior. A tie between two dolphins was established by a their statistically significant frequent association. The network split naturally into two large groups, the number of ties being 159.

For each network we computed the normalized mutual information on the base of the results reported by the authors. Regarding *GA-Net*, we run it 10 times and computed the average normalized mutual information over these 10 runs. The results are reported in figure 2(b). The figure clearly shows the very good performance of *GA-Net* with respect to Girvan and Newman's approach. In fact, on the American College Football network, *GA-Net* obtained an average normalized mutual information of 0.8825 over the 10 runs, with a worst value of 0.8417, and a best value of 0.9031, while the result of [8] was 0.8957. For the Krebs' network Newman [17] obtained 0.5107, while our approach 0.5756. Finally, on the dolphin network Girvan and Newman [18] obtained a normalized mutual information of 0.64. *GA-Net* over 10 runs obtained an average value of 0.8992. It is worth to point out that for 7 out of the 10 runs, *GA-Net* misplaced just node 40, which is connected to only two nodes: node 37 which belongs to the first group, and node 58 that belongs to the second group, thus the membership to one of the two communities is indistinguishable without adding semantics to the kind of link interconnecting the dolphins. In one over the 10 runs *GA-Net* obtained the exact partitioning of dolphins. To conclude figure 3 reports the result of running our algorithm on *Zackary's Karate Club Study.* This network was generated by Zachary [23], who studied the friendship of 34 members of a karate club over a period of two years. During this period, because of disagreements, the club divided in two groups almost of the same size. The figure has been reproduced by using the $NetDraw$ software [3]. We found four groups, depicted in the figure by different colors and shapes of the nodes. However, the two small communities are each a subgroup of the two effective communities. Thus our algorithm is able to detect more compact interactions. For example, as pointed out on the figure, the small community on the bottom left is characterized by five nodes each of which is connected to the bigger community only through the friendship to node 1, while among these five nodes a tighter connection exists. Girvan and Newman in [8] found the two groups in which the karate club divided. Node 3, however, was misplaced. A similar result is reported in [21]. In this latter approach the node 10 is

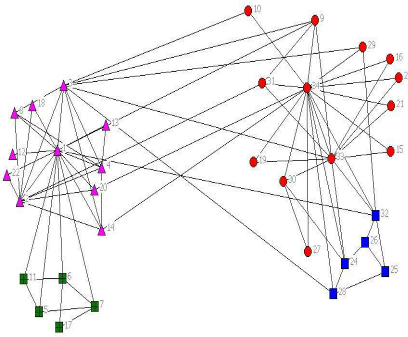


Fig. 3. Community structure found by the genetic based method *GA-Net*

misplaced. Our approach, on the contrary, is able to correctly classify both these nodes. The results obtained show the capability of genetic algorithms to effectively deal with community identification in networks.

6 Conclusions

The paper presented a genetic algorithm for detecting communities in social networks. The approach introduces the concept of community score, and searches for an optimal partitioning of the network by maximizing the community score. All the dense communities present in the network structure are obtained at the end of the algorithm by selectively exploring the search space, without the need to know in advance the exact number of groups. The concept of community score, though simple, revealed very efficacious. In fact, experiments on synthetic and real life networks showed the capability of the genetic approach to correctly detect communities with comparable results with state-of-the-art approaches. Future research will aim at applying multi-objective optimization to improve quality results.

References

1. Angiulli, F., Cesario, E., Pizzuti, C.: A greedy search approach to co-clustering sparse binary matrices. In: Proceedings of the 18th International Conference on Tools with Artificial Intelligence (ICTAI 2006), pp. 363–370 (2006)
2. Arenas, A., Diaz-Guilera, A.: Synchronization and modularity in complex networks. European Physical Journal ST 143, 19–25 (2007)
3. Borgatti, S.P.: Netdraw 1.0: Network visualization software. Harvard: Analytic technologies (2002)
4. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E 70, 066111 (2004)
5. Danon, L., Díaz-Guilera, A., Duch, J., Arenas, A.: Comparing community structure identification. Journal of Statistical Mechanics, P09008 (2005)
6. Dubes, R.C., Jain, A.K.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)

7. Firat, A., Chatterjee, S., Yilmaz, M.: Genetic clustering of social networks using random walk. Computational Statistics and Data Analysis 51, 6285–6294 (2007)
8. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. National. Academy of Science. USA 99, 7821–7826 (2002)
9. Handle, J., Knowles, J.: An evolutionary approach to multiobjective clustering. IEEE transactions on Evolutionary Computation 11(1), 56–76 (2007)
10. Hopcroft, J.E., Khan, O., Kulis, B., Selman, B.: Natural communities in large linked networks. In: Proc. International Conference on Knowledge Discovery and Data Mining (KDD 2003), pp. 541–546 (2003)
11. Lozano, S., Duch, J., Arenas, A.: Analysis of large social datasets by community detection. European Physical Journal ST 143, 257–259 (2007)
12. Lusseau, D.: The emergent properties of dolphin social network. Biology Letters, Proc. R. Soc. London B (suppl.) (2003)
13. Madeira, S.C., Oliveira, A.L.: Biclustering algorithms for biological data analysis: A survey. IEEE Transactions on Computational Biology and Bioinformatics 1(1), 24–45 (2004)
14. Makate, N., Miki, M., Hiroyasu, T., Senda, T.: Multiobjective clustering with automatic k-determination fot large-scale dara. In: Proc. of the Int. Genetic and Evolutionary Computation Conference (GECCO 2007), pp. 861–868 (2007)
15. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)
16. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Physical Review E 69, 066133 (2004)
17. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA 103, 8577–8582 (2006)
18. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E 69, 026113 (2004)
19. Park, Y.J., Song, M.S.: A genetic algorithm for clustering problems. In: Proc. of 3rd Annual Conference on Genetic Algorithms, pp. 2–9 (1989)
20. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. Proc. Natl. Acad.Sci. USA (PNAS 2004) 101(9), 2658–2663 (2004)
21. Tasgin, M., Bingol, A.: Communities detection in complex networks using genetic algorithms. In: Proc. of the European Conference on Complex Systems (ECSS 2006) (2006)
22. Tasgin, M., Herdagdelen, A., Bingol, A.: Communities detection in complex networks using genetic algorithms. oai:arXiv.org:0711.0491v1 [physics.soc-ph] (2007)
23. Zachary, W.W.: An information flow model for conflict and fission in small groups. Journal of Anthropological Research 33, 452–473 (1977)