MOGA-Net: A MultiObjective Genetic Algorithm-Net to Find Communities in Complex Networks

1S. Thayananthporkalanchiam, 2Mrs.G. Umarani Srikanth, M.E., (PhD), 3S.Rajakuncharam.

1Department of PG Studies, S.A. Engineering College, Veeraraghavapuram, Chennai, India.
2HOD, Department of PG Studies, S.A. Engineering College, Veeraraghavapuram, Chennai, India.
3Department of PG Studies, Sree Sastha Institute of Engineering and Technology, Chennai, India.

Abstract— The problem of community detection in complex networks as a MultiObjective clustering problem, and presented an evolutionary MultiObjective approach to uncover community structure. The algorithm optimizes two objective functions able to identify densely connected groups of nodes having sparse inter connections. The method generates a set of network divisions at different hierarchical levels in which solutions at deeper levels, consisting of a higher number of modules, are contained in solutions having a lower number of communities. The cluster’s should tune the size of the communities, has been considering as same group because, the partitioning found for this value are relevant. The number of modules is automatically determined by the better tradeoff values of the objective functions. Optimization methodology define an Objective function that allows the division of a graph in sub graphs, and try to maximize this objective in order to obtain the best partitioning of the network.

Keywords— Complex networks, MultiObjective clustering, MultiObjective evolutionary algorithms.

1. INTRODUCTION
Complex Networks constitute an efficacious formalism to represent the relationships among objects composing many real world systems. Collaboration networks, the Internet, the World Wide Web, biological networks, communication and transport networks, and social networks are just some examples. Networks are modeled as graphs, where nodes represent the objects and edges represent the interactions among these objects. An important problem in the study of complex networks is the detection of community structure[1] also referred to as clustering i.e., the division of a network into groups of nodes, called communities or clusters or modules, having dense intra connections, and sparse inter connections.

Many systems of current interest to the scientific community can usefully be represented as networks. Example includes the network and the World Wide Web, social networks, citation networks, food webs and biochemical networks. Each of these networks consists of a set of nodes or vertices representing, for instance, computers or routers on the Internet or people in a social network, connected together by links or edges, representing data connections between computers, friendship between people, and so forth.

An important property of networks is the existence of a community structure. A community is a set of items that share an environment and are created when one or more entities claim an interest in the same topic. It is interesting to study and identify these groups from the sociological, economical and technological point of view. Web communities are sets of Web pages that are created by individuals or associations having a common interest. The community Web can be modeled as a graph where vertices are Web pages and hyperlinks are edges. A Web community is defined as a set of sites that have more links between members of the community than between nonmembers, and we can identify a Web community on a topic by extracting densely connected structure in the Web graph. A crawler explores the content offered for each node and extracts its information.
Community (module) structure is a common and important property of many types of networks, such as social networks and biological networks. Several classes of algorithms have been proposed for community structure detection and identification, including clustering techniques, modularity optimization, and other methods.

![An example of complex network](image1)

**Fig: 1** An example of complex network

Community detection, thus, could be formulated as a MultiObjective optimization problem and the framework of Pareto optimality can provide a set of solutions corresponding to the best compromise among the objectives to optimize. In fact, there is a tradeoff between the two above mentioned objectives because when the community structure is constituted by the overall network the number of external links is null, thus it is Minimized, however the cluster density is not high.

MultiObjective optimization is a problem solving technique that successfully finds a set of solutions when multiple and conflicting objectives must be optimized. These solutions are obtained through the use of Pareto optimality theory and constitute global optimum solutions satisfying all the objectives as best as possible. Evolutionary algorithms to solve MultiObjective optimization problems revealed successful because of their population based nature which allows the simultaneous production of multiple optima and a good approximation of the Pareto front.

A MultiObjective approach, named MultiObjective genetic algorithms for networks (MOGA-Net) [5], to discover Communities in networks by employing genetic algorithms are proposed. The method optimizes two objective functions introduced. Community detection is an important task for mining the structure and function of complex networks. Generally, there are several different kinds of nodes in a network which are cluster nodes densely connected within communities, as well as some special nodes like hubs bridging multiple communities and outliers marginally connected with a community. That revealed both efficacious in detecting modules in complex networks.

The first objective function employs the concept of community score to measure the quality of the division in communities of a network. The higher the community score, the denser the clustering obtained. The second defines the concept of fitness of the nodes belonging to a module and iteratively finds modules having the highest sum of node fitness, in the following referred to as community fitness. When this sum reaches its maximum value, the number of external links is minimized. Both the objective functions have a positive real valued parameter controlling the size of the communities. The higher the value of the parameter, the smaller the size of the communities found. Based on the structural connectivity information, the proposed algorithm can effectively reveal the embedded hierarchical community structure with multi resolution in large scale weighted undirected networks, and identify hubs and outliers as well. Moreover, it overcomes the sensitive threshold problem of density based clustering algorithms and the resolution limit possessed by other modularity based methods.

![An example of community structure](image2)

**Fig: 2.** An example of community structure

The example of community structure contains three different communities they are Java, PHP and J2EE. These three parts are also called as a community.
This paper is organized as follows. In the next section, the concept of community is defined and the community detection problem is formalized. Section III describes the main approaches to community detection. Section IV formulates the community detection problem as a MultiObjective optimization problem. Section V describes the method, the genetic representation adopted, and the variation operators used. In Section VI, the results of the method on synthetic and real life networks and a comparison with some of the state of the art approaches are reported.

2. COMMUNITY DETECTION

A network N can be modeled 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 (also called cluster or module) in a network is a group of vertices (i.e., a sub graph) having a high density of edges within them, and a lower density of edges between groups. This definition of community is rather vague and there is no general agreement on the concept of density. A more formal definition has been introduced in the community by considering the degree $k_i$ of a generic node $i$, defined as

$$ k_i = \sum_j A_{ij} $$

Where $A$ is the adjacency matrix of $G$. $A$ is such that an entry at position $(i, j)$ is 1 if there is an edge from node $i$ to node $j$, 0 otherwise. Let $S$ subset of $G$ be the sub graph where node $i$ belongs to, the degree of $i$ with respect to $S$ can be split as

$$ k_i(S) = k_i(S_{in}) + k_i(S_{out}) $$

Where

$$ k_i(S_{in}) = \sum_{j\in S} A_{ij} $$

is the number of edges connecting $i$ to the other nodes in $S$, and

$$ k_i(S_{out}) = \sum_{j\in S} A_{ij} $$

is the number of edges connecting $i$ to the rest of the network.

A sub graph $S$ is a community in a strong sense if

$$ k_i(S_{in}) > k_i(S_{out}), \forall i \in S. $$

A sub graph $S$ is a community in a weak sense if

$$ \sum_{i \in S} k_i(S_{in}) > \sum_{i \in S} k_i(S_{out}). $$

Thus, in a strong community, each node has more connections within the community than with the rest of the graph. In a weak community, the sum of the degrees within the sub graph is larger than the sum of degrees toward the rest of the network. In the following, we adopt the concept of weak community, thus a community is interpreted as a set of nodes having a total number of intra connections higher than the number of inter connections among different clusters.

3. MULTIOBJECTIVE GENETIC ALGORITHM

Many different algorithms, coming from different fields such as physics, statistics, data mining, and evolutionary computation have been proposed to detect communities in complex networks. The approaches adopted can broadly be classified into three different types: divisive hierarchical methods, agglomerative hierarchical methods, and optimization methods. The divisive hierarchical methods start from the complete network, detect the edges that connect different communities, and remove them. Agglomerative approaches consider each node a cluster and then merge similar communities recursively until the whole graph is obtained. Among the optimization methods [7], several approaches have been developed by using evolutionary techniques. In particular, applied genetic algorithms. Many other proposals employ MultiObjective evolutionary algorithms to partition graphs or cluster objects in metric spaces.

A. Community Detection in Networks

The community detection problem has been studied by several researchers, and a complete description of the state of the art proposals is beyond the scope of this paper. Extensive and detailed overviews of community identification methods in complex networks can be found. One of the most famous algorithms to detect communities has been presented by Newman and Girvan. The method iteratively splits the network by removing edges. The edges to be removed are chosen by using the betweenness measure. 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 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 no edges remain.
B. MultiObjective Clustering Methods

The application of MultiObjective optimization to clustering data has recently obtained an increasing interest though few proposals regard the partitioning of networks. A reference approach to MultiObjective clustering algorithms for numerical and categorical data is that proposed by Handl and Knowles and named MultiObjective clustering with automatic K-determination (MOCK) [5]. The first objective of MOCK is to minimize the overall deviation of a partitioning, i.e., the summed distances between data items and the center of the cluster they have been assigned. The second objective is the minimization of the cluster connectedness, which evaluates for each cluster data point how many of its nearest neighbors have been placed in the same cluster.

The clusters obtained are then used in a Web recommendation system for representing usage patterns. The sequences of Web pages visited by a user are represented as a weighted undirected graph where each sequence is a node, and the weight of an edge connecting two sequences is the computed similarity between the two nodes. Their algorithm uses the same representation of MOCK, but the conflicting objectives to optimize are the min-max cut and the silhouette index [5]. The former tries to optimize the intra cluster similarity and to minimize the inter sub graph similarity; the latter computes the average silhouette index of vertices belonging to the same cluster.

4. MULTIOBJECTIVE OPTIMIZATION PROBLEM

Many problems in different fields are naturally formulated with multiple objectives. In particular, the division of a network in subgroups of nodes having dense intra connections and sparse interconnections has two competing objectives. The first is to maximize the links among the nodes belonging to the same module; the second is to minimize the number of connections between the communities. Thus, the problem of community detection cannot adequately be represented as a single objective augmented with constraints. To formalize this problem as a MultiObjective clustering problem.

A. Objective Functions

Our aim is to partition a network in groups of vertices \( \{S_1, \ldots, S_k\} \) such that the density of edges within them is higher than the density of edges between the groups. To this end, we need an objective function that maximizes the number of connections inside each community, and another objective function that minimizes the number of links between the modules found.

The volume \( vS \) of a community \( S \) is defined as the number of edges connecting vertices inside \( S \), i.e., the number of 1 entry in the adjacency sub matrix of \( A \) corresponding to \( S \).

\[
vS = \sum_{i,j \in S} A_{ij}
\]

The score of \( S \) is defined as

\[
Score(S) = M(S) \times vs.
\]

Thus, the score takes into account both the fraction of interconnections among the nodes (through the power mean) and the number of interconnections contained in the module \( S \) (through the volume)[4][7]. The community score of a clustering \( \{S_1, \ldots, S_k\} \) of a network is defined as

\[
CS = \sum_{k=1}^{n} \text{score}(S_i).
\]

The first objective to maximize is then the community score \( CS \). The concept of community fitness of a module \( S \) as

\[
P(S) = \sum_{i \in S} \frac{kiin(S)}{kiin(S) + kout(S)}
\]

The second objective is thus carried out by the community fitness by summing up the fitness of all the \( S \)-modules. The parameter \( a \), that tunes the size of the communities, has been set to 1 because, as the authors observed, in most cases the partitioning found for this value are relevant [2][11]. The second objective to minimize is thus

\[
\sum_{i=1}^{k} P(s_i)
\]

In the next section, to propose MultiObjective community detection approach that optimizes both these two objectives.

5. ALGORITHM DESCRIPTION

In this section, give a description of the MultiObjective algorithm MOGA-Net, the representation adopted for partitioning the network, and the variation operators used. In the last few years many efforts have been devoted to the application of evolutionary computation to the development of MultiObjective optimization algorithms. Evolutionary algorithms, in fact, proved to be very successful to solve MultiObjective optimization problems.

A. Genetic Representation
The clustering algorithm uses the locus based adjacency representation 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 value \( j \) in the range \( \{1, \ldots, N\} \). Genes and alleles represent nodes of the graph \( G = (V, E) \) modeling a network \( N \), and a value \( j \) assigned to the \( i^{th} \) 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 separate components of the corresponding graph. The nodes participating to the same component are assigned to one cluster. 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.

### B. Initialization

The initialization process takes into account the effective connections of the nodes in the network. A random individual is generated. However, if in the \( i^{th} \) position there is an allele value \( j \), but the edge \( (i, j) \) does not exist, the individual \( j \) is substituted with one of the neighbors of \( i \).

### C. Uniform Crossover

MOGA-Net uses a standard uniform crossover operator. First, a crossover mask of length \( N \), i.e., the number of nodes, is randomly generated. Each value on the mask is either 0 or 1. An offspring is generated by selecting from the first parent the genes where the mask is a 0, and from the second parent the genes where the mask is a 1. The main motivation of using uniform crossover is to guarantee the maintenance of the effective connections of the nodes in the network in the child individual.

### D. Mutation

The mutation operator that randomly changes the value \( j \) of an \( i^{th} \) 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 \( i \).

### E. Model Selection

MultiObjective clustering returns the set of Pareto optimal solutions. Each of these solutions corresponds to a different tradeoff between the two objectives and thus to diverse partitioning of the network consisting of various numbers of clusters. This gives a great chance to analyze several clustering at different hierarchical levels. However, a criterion should be established to automatically select one solution with respect to another. To this end, we adopt the concept of modularity, introduced by Newman and Girvan [3][13]. Modularity is the most used and known function to assess the quality of a partitioning obtained by a clustering method.

### F. Edge Connection

Edge connection is a weight of the node can be calculated by the node number of edges connected to the node within the same community. Maximum number of the edge connection will give the better solution. Fig.3 shows the edge connection.

### G. Hierarchical solution

The Pareto optimal solutions produce the hierarchical organization of the network, where the solutions with a higher number of clusters are included in solution having a lower number of communities. Fig.4 shows the hierarchical structure which is used to give the better solution.

### 6. CONCLUSION

The proposed formalization of the problem of community detection in complex networks as a MultiObjective clustering problem, and presented an evolutionary MultiObjective approach to uncover community structure. The objective of this method is maximizes the intra connections inside each community and minimizes inter connections between different communities. A main characteristic of the algorithm is that it automatically affords a network partitioning without the need of knowing a priori the precise number of clusters. This is particularly useful in all those applications where no information about the group division is available. The MultiObjective approach has the advantage, with respect to single objective approaches, to contemporarily optimize multiple criteria and to provide, not a single partitioning, but a set of solution, each corresponding to a different number of clusters, constituting the best tradeoff between the competing objectives. The
future work of this project assigning rating to packet,
Rating based suggested process.

ACKNOWLEDGMENT

I must thank, first and foremost, my Guide
Mrs. UMARANI SRIKANTH M.E.,(Ph.D) Professor,
Head of the Department, Department of PG studies in
Engineering, without whose guidance and patience,
this dissertation would not be possible and staffs of
the Department of PG studies in Engineering for their
consistent encouragement and ideas.

REFERENCES

structure in networks,” Proc. Natl. Acad. Sci. USA,
community detection in social networks” in Proc.
evaluating community structure in networks” Phys.
4. A. Lancichinetti, S. Fortunato, and F. Radicchi.
arXiv:0805.4770v2
5. C. A. Coello Coello, G. B. Lamont, and D. A. Van
Veldhuizen, “Evolutionary Algorithms for Solving
MultiObjective Problem’s in Proc. ACM SAC, 1999,
pp. 551–557.
interpretation and validation of cluster analysis,”
7. J. Handl and J. Knowles, “An evolutionary approach
to MultiObjective”
8. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E.
Lefevre, “Fast unfolding of communities in large
9. K. Faceli, A. C. P. L. F. de Carvalho, and M. C. P. de
Souto, “MultiObjective clustering ensemble,” Int. J.
Hybrid Intell. Syst., vol. 4.
10. S. Fortunato and M. Barth´eleyne, “Resolution limit
USA, vol. 104.
11. J. Handl and J. Knowles, “An evolutionary approach
Comput., vol. 11.
“Communities detection in complex networks using