Genetic Algorithms for Community Detection in Social Networks

Ahmed Ibrahem Hafez[1,4], Neveen I. Ghali[2,4], Aboul ella Hassanien[3,4], Aly A. Fahmy[3]

[1]CS Dept., Faculty of Computer and Information, Minia University, Egypt
[2]Faculty of Science, Al-azhar University, Egypt
[3]Faculty of Computers and Information, Cairo University, Egypt
[4]Scientific Research Group in Egypt http://www.egyptscience.net
Email: ah.hafez@gmail.com, aboitcairo@gmail.com, aly.fahmy@gmail.com

***Abstract*— Community detection in complex networks has attracted a lot of attention in recent years. Community detection can be viewed as an optimization problem, in which an objective function that captures the intuition of a community as a group of nodes with better internal connectivity than external connectivity is chosen to be optimized. Many single-objective optimization techniques have been used to solve the problem however those approaches have its drawbacks since they try optimizing one objective function and this results to a solution with a particular community structure property. More recently researchers viewed the problem as a multi-objective optimization problem and many approaches have been proposed to solve it. However which objective functions could be used with each other is still under debated since many objective functions have been proposed over the past years and in somehow most of them are similar in definition. In this paper we use Genetic Algorithm (GA) as an effective optimization technique to solve the community detection problem as a single-objective and multi-objective problem, we use the most popular objectives proposed over the past years, and we show how those objective correlate with each other, and their performances when they are used in the single-objective Genetic Algorithm and the Multi-Objective Genetic Algorithm and the community structure properties they tend to produce.**

Keywords-community detection; genetic algorithm; social network analysis; multi-objective optimization

I. INTRODUCTION

Networks are used to model and represent many real world systems this fact had made complex network analysis a hot research area recently. Collaboration networks, the Internet, the world-wide-web, biological networks, communication and transport networks, social networks are just some examples of such complex networks, a common feature of complex networks is community structure [1], i.e. groups of nodes in the network that are more densely connected internally than with the rest of the network. Communities are supposed to play special roles in the structure-function relationship, and thus detecting communities (or modules) can be a way to identify substructures which could correspond to important functions in the network and thus uncovering or detecting this community structure is one of the most important problems in the field of complex network analysis.

Many methods have been developed for the community detection problem, using tools and techniques from disciplines like physics, biology, applied mathematics, computer and social sciences, for a recent survey see [2].

One of the most known algorithms proposed so far is the Girvan-Newman (GN) algorithm that introduces a divisive method by iteratively cutting the edge with the greatest betweenness value [3]. Some improved algorithms have been proposed [4,5]. These algorithms are based on a foundational measure criterion of community, modularity, a popular quality function proposed by Newman [3]. The larger the value is, the more accurate the community partition would be. As a consequence, the community detection becomes a modularity optimization problem. Because the search for the optimal (largest) modularity value is an NP-complete problem, many heuristic search algorithms have been applied to solve the optimization problem, such as extremal optimization, simulated annealing and genetic algorithm [2].

The rest of this work is organized as follow in section II we define the community problem as a single-objective and multi-objective optimization problem and we introduce the objective functions used in our genetic algorithm. In section III we describe our genetic algorithm and its genetic representation and operations used by the algorithm. In section IV shows the experimental results for the single-objective and multi-objective cases on real life social networks and Synthetic network; finally section V conclude our result and suggestions for future work.

II. THE COMMUNITY DETECTION PROBLEM

A. Community detection problem

A social network SN can be modeled as a graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ where $\boldsymbol{V}$ is a set of vertices, and $\boldsymbol{E}$ is a set of edges, that connect two elements in $\boldsymbol{V}$. A community $\boldsymbol{S}$ 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, given a quality measure of communities $\boldsymbol{F}(\boldsymbol{S})$, can be formulated as finding a partitioning of the nodes in k subsets that best satisfy the quality measure $\boldsymbol{F}(\boldsymbol{S})$. The problem can be viewed as an optimization problem in which one usually want to optimize the given Quality measure $\boldsymbol{F}(\boldsymbol{S})$, we will describe most popular quality measures that has been used to detect communities in the following subsection.

We can view the problem as Single-Objective optimization [13,14,10] in which one use one community quality measure as the objective function a formal definition can be found in [6], and without loss of generality we assume that all quality measures need to be minimized, so we want to find community $\boldsymbol{S}$ that max/min the quality measure $\boldsymbol{F}(\boldsymbol{S})$ according to the definition of the quality measure.

Recently researchers treat the problem as a multi-objectives optimization problem [6,15,16] where given a set of quality measures; we want to find a set of community that simultaneously optimizes each quality measure, formal definition can be found in [6]. In Multi-Objective problems there is not one unique solution to the problem, but a set of solutions are found through the use of Pareto optimality theory [7].

In this review we use Genetic algorithms as the optimization technique. In a standard genetic algorithm one has a set of candidate solutions to a problem, which are numerically encoded as chromosomes, and an objective function to be optimized on the space of solutions. The objective function plays the role of biological fitness for the chromosomes. One usually starts from a random set of candidate solutions, which are progressively changed through manipulations inspired by biological processes regarding real chromosomes, like point mutation (random variations of some parts of the chromosome) and crossing over (generating new chromosomes by merging parts of existing chromosomes). Then, the fitness of the new pool of candidates is computed and the chromosomes with the highest fitness have the greatest chances to survive in the next generation. After several iterations only solutions with large fitness survive.

B. *Communities Quality meseares*

In GA, The objective function plays important role in the evolution process. It is the steering wheel in the process that leads to good candidate's solution to the problem. For the community detection problem, many objective functions have been proposed to capture the intuition of communities and there is no straight way to compare these objective functions based on their definitions.

Here we state objective functions that capture this intuition and/or are popular in the community detection literature and can be potentially used for community detection. More details about the objectives in [8] and a similar comparison can be found in [8, 9].

We use Conductance, Expansion, Internal Density, Cut Ratio, Normalized Cut, Maximum-ODF, Average-ODF, and Flake-ODF as qualities measure[8]; the lower the value of those objectives the better the community structure. And we use Modularity [3], Community Score [10], Community Fitness [11] and Surprise [12] as qualities measures; the higher the value of those objectives the better the community structure.

All previous objective is parameter free i.e. their calculation depend only on the network, except for community score and community fitness both have a positive real-valued parameter that control the size of the communities.

III. Genetic Algorithms for Community Detection

In this section we describe the genetic algorithms used in this work and the genetic representation and genetic operations. In the single objective GA we employ standard single objective GA algorithms with a roulette selection function and elitism; the algorithm was implemented in a .Net environment using C#. In multi-objective case we use Non-dominated Sorting Genetic Algorithm (NSGA-II) as the multi-objective GA proposed by Deb et al. in [17] and implemented in the Genetic Algorithm and Direct Search Toolbox of MATLAB.

In both case we use the same genetic representation and genetic operations, we adopt in this experiment the same genetic representation and genetic operations proposed in [10]. We describe the various stages of the genetic algorithm in the following subsections

A. *Genetic Representation*

The algorithm uses the locus-based adjacency representation proposed in [18]. In this representation each individual chromosome consists of N genes $g_1, \ldots, g_N$ and each gene can take values j in the range $\{1, \ldots, N\}$, where $N = |V|$ is the number of nodes in the network, and a value j assigned to the i-th gene is interpreted as a link between the nodes i and j. This means that in the community structure found i and j will be in the same community. A farther decoding step is necessary to identify all communities. Advantages of this representation is that the number k of communities is automatically determined by the number of components contained in an individual and determined by the decoding step, and the decoding step can be achieved in linear time as mentioned in [19].

B. *Initialization*

A random generation of individuals could generate components that in the original graph are disconnected. Pizzuti [10] proposed the term safe initialization in which each gene i is assign to a value j from the i-th node's neighbors. Here we use the same initialization steps used in [10] then with a probability of (1 - mutationRate) of the population size, we select a η percent of the genes and for each gene i we assign it to itself and for all its neighbor we assign the value i to them as long, so node i along with its neighbor will be in the same community. We found that a value 0.5 for η achieve a good result after a series of trials.

C. *Uniform Crossover*

Given two 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 new child.

D. *Mutation*

The mutation operator that randomly change the value *j* of a *i*-th gene causes a useless exploration of the search space, so we select a m_j percent of the genes randomly and for each gene *i* we randomly change its value to *j* such that nodes *i* and *j* are neighbors as we did in the initialization step. We found that a value 0.1 for m_j achieve a good result after a series of trials

E. *Objectives*

As mention in section II, many communities' quality measures have been proposed over the last years and some of them are similar in behavior, so we try the algorithm with each quality measure as the optimization objective and compare the result to gain more insight of the definition of each objective and its properties.

IV. Experimental Results

We test the algorithm on two synthetic data sets and two real social networks, to compare the accuracy of the result community structures we use Normalized Mutual Information to measure the similarity between the true community structures and the detected ones The Normalized Mutual Information NMI is a similarity measure proved to be reliable by Danon et al. [20]. We show the averaged Normalized Mutual Information for many runs of the algorithm for each objective and compare them with each other.

Real Social Network: We test the algorithm for each objective on two real life data sets the Zachary's Karate Club and the Bottlenose Dolphins well studied in the literature (see http://www.personal.umich.edu/~mejn/netdata).The Zachary Karate Club data contains the community structure in a karate club, which is analyzed first in [21]. The network consists of 34 vertices and 78 edges. The network is divided into two groups almost of the same size. The Bottlenose Dolphin Network consists of 62 bottlenose dolphins living in Doubtful Sound, New Zealand, compiled by Lusseau [22] from seven years of dolphin's behavior. A tie between two dolphins was established by their statistically significant frequent association. The network split naturally into two large groups.

Synthetic network: We use the benchmark proposed by Girvan and Newman in [1]. The network consists of 128 nodes divided into four communities of 32, 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 the communities up to value $z_{out} = 8$.

A. *Single objective community detection*

We start our study with applying single objective genetic algorithm to the community detection problem, and try each objective separately in the genetic algorithm. Here we show the result obtained for the single objective case.

First we show the result for the real life social networks Zachary's Karate Club and the Bottlenose Dolphins, In Fig.1

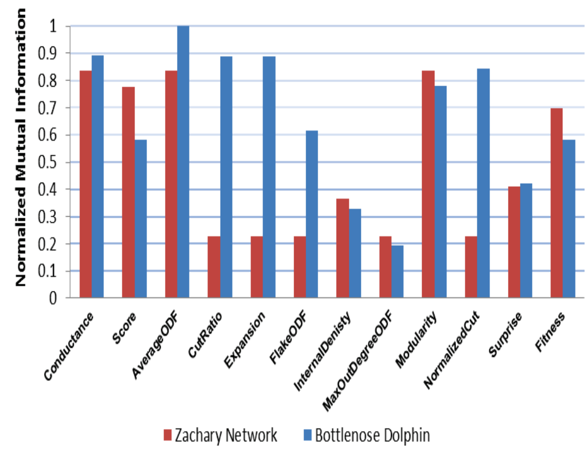
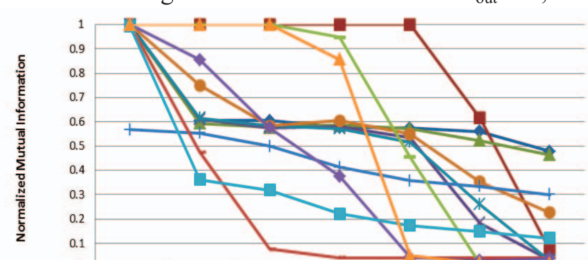


Figure 1. NMI obtained by each objective on Zachary Karate Club Network and Bottlenose Dolphin Network.

we report the average NMI over 30 runs of the GA for each objective; here we employed standard parameters for the genetic algorithm, crossover rate 0.9, mutation rate 0.4, elite reproduction 10% of the population size, population size 60, and number of iterations 35. We see that three objectives conductance, averageODF and modularity achieve an NMI value above 0.8 for both social networks. And two objective community score and community fitness achieve a good NMI value above 0.5 for both networks as well.

Second we show the result for Synthetic network, we generate 15 networks for each value of z_{out} range for 0 to 6 and run the algorithm 10 times and take the averaged NMI value for the 10 times, then we take the averaged NMI value for the 15 network for each value of z_{out}, Fig.2 show the result NMI values for the synthetic network for each objective separately. Here we employed the same values for the standard parameters for the genetic algorithm, except for the number of iteration (generation), we set it here to 400 iterations since this network is larger. We found that the community score objective achieves good result until a value for z_{out} = 5, while community fitness and modularity objectives achieve good result until a value for z_{out} = 3, only



	0	1	2	3	4	5	6
Conductance	1	0.6051	0.6051	0.5772	0.5747	0.5605	0.4779
Score	1	1	1	1	1	0.6156	0.0725
AverageODF	1	0.5939	0.5772	0.5883	0.5713	0.5237	0.4647
CutRatio	1	0.6107	0.5753	0.5809	0.5359	0.1826	0.0323
Expansion	1	0.6163	0.5828	0.5728	0.5164	0.2603	0.028
FlakeODF	1	0.7495	0.5828	0.6051	0.5489	0.3528	0.226
InternalDenisty	0.5684	0.5545	0.5006	0.413	0.3578	0.3346	0.3006
MaxOutDegreeODF	1	0.4745	0.0759	0.0403	0.0395	0.0389	0.0386
Modularity	1	1	1	0.9487	0.4554	0.0165	0.0177
NormalizedCut	1	0.8555	0.5753	0.3766	0.0426	0.0286	0.0372
Surprise	0.9938	0.3621	0.3189	0.2214	0.1739	0.1489	0.1206
Fitness	1	1	1	0.8561	0.0486	0.0189	0.0176

Figure 2. NMI for each objective on the GN Benchmark for different z_{out} range from 0 to 6

internal density objective fails to detect the correct community structure for the network when $z_{out} = 0$ in which the network consists of 4 disconnect communities, and we found that all objectives performances decrease to a value of NMI below 0.1 when z_{out} reaches value 7 except for conductance, averageODF and internal density.

a) Objectives Stabilities and similarities

Due the random nature of the genetic algorithm it may produce different solution each run, it is the responsibility of the objective function to redirect the algorithm to a local maxima or a local minima based on the definition of the objective over the solution space. A good objective function should assign the same fitness score to two solutions if they are the same or at least there is a large similarity between each other, if two completely different solutions have the same fitness score obtained by the objective function, the algorithm will tend to produce different solution each run.

We define a stable objective as an objective that tends to produce similar community structure over many run of the GA. In Fig.3 diagonal entries in the table show on average how much the solution obtained by an objective similar to each other in term of Normalized Mutual Information similarity measure, low value indicate that there is a lot of variation on the solutions obtained by the objective over many runs, high value indicate that solutions are much more similar to each other. We run the algorithm many time on each network Zachary's Karate Club, the Bottlenose Dolphins and the synthetic network for a value of $z_{out} = 3$, and calculate the averaged NMI for each pair of solutions for each network. We observe that community score, community fitness and modularity achieve high values for all networks which mean that they are more stable than the rest of objectives in term of that they produce similar result for different runs of the algorithm. And most of the objectives achieve high values for the two real life social networks as shown in Fig. 3.

In section II we showed many objectives and their definition and stated that they are in somehow similar in their definition, and when deployed with the genetic algorithm some of them tend to produce the same result or similar result to each other.

To understand Objectives similarities, we compared solutions obtained by each objective with the solutions obtained by the rest of the objectives and showed the averaged NMI in Fig.3 Low value (dark cells) indicate low similarity, high value (white cells) indicate high similarity.

B. Multi-objective community detection

We now move to apply multi-objective genetic algorithm to the community detection problem. We restrict our self to use only two objective in the algorithm and study the algorithm performance for each objectives pair, we exclude the surprise objective as it is computationally expensive as it requires a lot of binomial coefficients calculation so it does not scale well for large networks, also it did not achieve good results in the single objective case.

First we show the result for the real life social networks Zachary's Karate Club and the Bottlenose Dolphins. In Fig.5 we report the average NMI over 15 runs of the GA for each

Objetives	Conductance	AverageODF	CutRatio	Expansion	FlakeODF	InternalDenisty	MaxOutDegreeODF	NormalizedCut	Modularity	Surprise	Fitness	Score
Conductance	**1**	0.74	0.21	0.21	0.22	0.37	0.21	0.2	1	0.43	0.82	0.63
AverageODF	0.74	**0.98**	0.25	0.25	0.26	0.37	0.25	0.24	0.74	0.43	0.63	0.6
CutRatio	0.21	0.25	**1**	1	0.99	0.32	1	0.99	0.21	0.32	0.59	0.49
Expansion	0.21	0.25	1	**1**	0.99	0.32	1	0.99	0.21	0.32	0.59	0.49
FlakeODF	0.22	0.26	0.99	0.99	**0.97**	0.32	0.99	0.97	0.22	0.32	0.59	0.5
InternalDenisty	0.37	0.37	0.32	0.32	0.32	**0.75**	0.32	0.32	0.37	0.49	0.52	0.58
MaxOutDegreeODF	0.21	0.25	1	1	0.99	0.32	**1**	0.99	0.21	0.32	0.59	0.49
NormalizedCut	0.2	0.24	0.99	0.99	0.97	0.32	0.99	**0.97**	0.2	0.31	0.58	0.49
Modularity	1	0.74	0.21	0.21	0.22	0.37	0.21	0.2	**0.99**	0.43	0.82	0.63
Surprise	0.43	0.43	0.32	0.32	0.32	0.49	0.32	0.31	0.43	**0.43**	0.52	0.53
Fitness	0.82	0.63	0.59	0.59	0.59	0.52	0.59	0.58	0.82	0.52	**1**	0.81
Score	0.63	0.6	0.49	0.49	0.5	0.58	0.49	0.49	0.63	0.53	0.81	**0.92**

(a)

Objetives	Conductance	AverageODF	CutRatio	Expansion	FlakeODF	InternalDenisty	MaxOutDegreeODF	NormalizedCut	Modularity	Surprise	Fitness	Score
Conductance	**0.96**	0.9	0.97	0.96	0.56	0.34	0.12	0.89	0.84	0.42	0.64	0.48
AverageODF	0.9	**0.97**	0.9	0.89	0.6	0.34	0.12	0.86	0.78	0.41	0.61	0.48
CutRatio	0.97	0.9	**0.97**	0.97	0.56	0.34	0.12	0.9	0.84	0.42	0.64	0.48
Expansion	0.96	0.89	0.97	**0.97**	0.56	0.34	0.12	0.9	0.84	0.42	0.64	0.48
FlakeODF	0.56	0.6	0.56	0.56	**0.54**	0.28	0.12	0.54	0.49	0.32	0.4	0.35
InternalDenisty	0.34	0.34	0.34	0.34	0.28	**0.76**	0.12	0.37	0.41	0.57	0.54	0.64
MaxOutDegreeODF	0.12	0.12	0.12	0.12	0.12	0.12	**0.34**	0.14	0.11	0.11	0.17	0.15
NormalizedCut	0.89	0.86	0.9	0.9	0.54	0.37	0.14	**0.85**	0.79	0.43	0.64	0.5
Modularity	0.84	0.78	0.84	0.84	0.49	0.41	0.11	0.79	**0.85**	0.46	0.73	0.58
Surprise	0.42	0.41	0.42	0.42	0.32	0.57	0.11	0.43	0.46	**0.48**	0.51	0.55
Fitness	0.64	0.61	0.64	0.64	0.4	0.54	0.17	0.64	0.73	0.51	**0.79**	0.71
Score	0.48	0.48	0.48	0.48	0.35	0.64	0.15	0.5	0.58	0.55	0.71	**0.77**

(b)

Objetives	Conductance	AverageODF	CutRatio	Expansion	FlakeODF	InternalDenisty	MaxOutDegreeODF	NormalizedCut	Modularity	Surprise	Fitness	Score
Conductance	**0.37**	0.39	0.4	0.4	0.37	0.22	0.05	0.26	0.58	0.15	0.31	0.59
AverageODF	0.39	**0.38**	0.4	0.4	0.37	0.22	0.05	0.26	0.58	0.15	0.31	0.59
CutRatio	0.4	0.4	**0.41**	0.42	0.37	0.22	0.05	0.27	0.57	0.15	0.3	0.58
Expansion	0.4	0.4	0.42	**0.41**	0.37	0.22	0.05	0.26	0.58	0.15	0.31	0.58
FlakeODF	0.37	0.37	0.37	0.37	**0.34**	0.22	0.05	0.24	0.57	0.15	0.31	0.58
InternalDenisty	0.22	0.22	0.22	0.22	0.22	**0.53**	0.05	0.16	0.43	0.28	0.57	0.43
MaxOutDegreeODF	0.05	0.05	0.05	0.05	0.05	0.05	**0.06**	0.05	0.05	0.03	0.05	0.05
NormalizedCut	0.26	0.26	0.27	0.26	0.24	0.16	0.05	**0.2**	0.35	0.11	0.21	0.36
Modularity	0.58	0.58	0.57	0.58	0.57	0.43	0.05	0.35	**0.97**	0.27	0.6	0.98
Surprise	0.15	0.15	0.15	0.15	0.15	0.28	0.03	0.11	0.27	**0.13**	0.29	0.28
Fitness	0.31	0.31	0.3	0.31	0.31	0.57	0.05	0.21	0.6	0.29	**0.65**	0.61
Score	0.59	0.59	0.58	0.58	0.58	0.43	0.05	0.36	0.98	0.28	0.61	**1**

(c)

Figure 3. Show objectives stabilities and similarities obtained by different objectives performed on:-

(a) Zachary Karate Club Network.
(b) Bottlenose Dolphin Network.
(c) GN-Benchmark for $z_{out} = 3$.

Diagonal cells correspond to objectives stabilities, other cells correspond to objective similarities .

objectives pair; here We employed standard parameters for the genetic algorithm, crossover rate 0.8, mutation rate 0.2, elite reproduction 10% of the population size, binary tournament selection function. The population size was 300, the number of generations 30. For each run we investigate the Pareto-optimal set returned by the algorithm, we select the best community structure compared to the optimal known solution of the network in term of NMI similarity measure, and select the worst community structure in the set compared to the optimal solution of the network. In Fig.5 (a) and (b) the result NMI values for the best and the worst community structure in the Pareto-optimal set for the Zachary Karate Club Network and the Bottlenose Dolphin Network; respectively when we applied multi-objective GA using two objectives, in the upper triangular of the matrix the NMI value for the best solution, in the lower triangular the NMI values for the worst solution compared to the optimal solution and the diagonal entries correspond to the single objective case. We found that community score when used with other objective achieve a good NMI result for both networks except when used with community fitness and internal density objectives. Also community fitness achieve a good NMI result for both networks except when used with community score and internal density objectives, avergeODF and conductance achieve a very promising NMI results when used with any objective for both networks.

Objectives	AverageODF	Conductance	CutRatio	Expansion	Fitness	FlakeODF	InternalDenisty	MaxOutDegreeODF	Modularity	NormalizedCut	Score
AverageODF	0.84	0.84	0.84	0.84	0.84	0.84	0.83	0.84	0.84	0.84	1
Conductance	0.83	0.84	0.84	0.84	0.84	0.84	0.87	0.84	0.84	0.84	1
CutRatio	0.25	0.23	0.23	0.23	0.84	0.23	0.52	0.25	0.84	0.23	1
Expansion	0.25	0.25	0.25	0.23	0.84	0.25	0.51	0.23	0.84	0.25	1
Fitness	0.45	0.45	0.23	0.24	0.7	0.72	0.43	0.84	0.84	0.82	0.58
FlakeODF	0.24	0.26	0.24	0.24	0.23	0.23	0.57	0.23	0.84	0.25	0.82
InternalDenisty	0.14	0.14	0.15	0.14	0.38	0.2	0.37	0.51	0.92	0.39	0.6
MaxOutDegreeODF	0.24	0.23	0.24	0.23	0.23	0.23	0.14	0.23	0.84	0.23	1
Modularity	0.84	0.84	0.26	0.23	0.46	0.25	0.41	0.23	0.84	0.84	1
NormalizedCut	0.29	0.23	0.23	0.22	0.22	0.23	0.15	0.23	0.25	0.23	0.96
Score	0.58	0.6	0.23	0.23	0.44	0.23	0.43	0.23	0.59	0.23	0.84

(a) Zachary Karate Club

Objectives	AverageODF	Conductance	CutRatio	Expansion	Fitness	FlakeODF	InternalDenisty	MaxOutDegreeODF	Modularity	NormalizedCut	Score
AverageODF	1	1	1	1	0.87	0.99	0.78	0.99	1	1	0.94
Conductance	0.89	0.89	0.89	0.89	0.82	1	0.84	0.89	0.89	0.89	0.95
CutRatio	0.89	0.89	0.89	0.89	0.85	1	0.85	0.9	0.89	0.89	0.9
Expansion	0.88	0.89	0.89	0.89	0.88	0.99	0.84	0.89	0.88	0.89	0.91
Fitness	0.35	0.35	0.35	0.31	0.58	0.8	0.33	0.82	0.75	0.8	0.38
FlakeODF	0.6	0.54	0.56	0.6	0.33	0.62	0.74	0.89	1	1	0.85
InternalDenisty	0.12	0.09	0.09	0.08	0.32	0.15	0.33	0.42	0.86	0.85	0.39
MaxOutDegreeODF	0.37	0.14	0.26	0.19	0.23	0.15	0.06	0.19	0.89	0.89	0.94
Modularity	0.79	0.77	0.79	0.8	0.37	0.67	0.38	0.26	0.78	0.89	0.74
NormalizedCut	0.81	0.8	0.81	0.82	0.35	0.57	0.29	0.26	0.76	0.84	0.86
Score	0.42	0.41	0.4	0.4	0.35	0.36	0.34	0.28	0.42	0.4	0.58

(b) Bottlenose Dolphin

Figure 4 NMI values for the best and the worst community structure in the Pareto-optimal set for (a) the Zachary Karate Club Network and (b) the Bottlenose Dolphin Network, the upper triangular of the matrix show the NMI values for the best solution for each pair, the lower triangular show the NMI values for the worst solution for each pair, and the diagonal entries correspond to the single objective case.

Second we show the result for the result for the Synthetic network in Fig.6. Here we employed the same values for the standard parameters for the genetic algorithm, except for the number of iterations; we set it here to 100. We try each objectives pairs on the network for values of z_{out} range from 0 to 7, as we did before we generate deferent network for each z_{out} value and take the average result from each run. We only show the result for z_{out} values from 4 to 7 in Fig. 6, and we summarize the result for values of z_{out} from 0 to 3 as the following, for $z_{out} = 0$ all pair was able to detect the correct community structure for the network except for the internal density with community fitness which achieves NMI values of 0.6, for z_{out} values from 1 to 3, community score when applied with all objective correctly detect the optimal community structure as it did for the single objective case. Also community fitness correctly detect the optimal community structure when applied with all objective except for the internal density, also modularity achieves a good result NMI values above 0.9 when applied with other objectives except for normalized cut. For values of $z_{out} > 4$ most objectives pair's performance begins to decrease. However we found that the community fitness achieve good result until $z_{out} = 6$ when applied with other objective. And community fitness with flakeODF and avergODF achieve good result for $z_{out} = 7$.

One advantage of multi-objective GA is that it returns a set of Pareto-optimal solutions so there is no best or worst solution, it is up to us to select a solution that satisfy our need or some other criteria. For example we could run the algorithm using community fitness and flakeODF then we could select the solution with maximum network modularity from the Pareto-optimal solutions set.

V. CONCLUSION AND FUTURE WORKS

Genetic algorithm as an optimization technique works effectively for the community detection problem in single-objective and multi-objective cases, However it is performance is influenced directly by the objective function used in the optimization process. For the multi-objective case we saw that some objective work well together and other not, however it would be interesting to design new objective form those objectives we showed in this review that best suited the multi-objective genetic algorithm. The GN benchmark network is not enough to explain the performance of different objectives, so objectives that achieve good result in this review will need farther investigation on other benchmarks and on larger networks to see its performance on large scale network.

REFERENCES

[1] Newman M.E.J, Girvan M.: Community structure in social and biological networks. Proc. National. Academy of Science. USA 99, 7821–7826 (2002).

[2] S. Fortunato, Community detection in graphs. arXiv:0906.0612 (2009).

Objectives	AverageODF	Conductance	CutRatio	Expansion	Fitness	FlakeODF	InternalDenisty	MaxOutDegree	Modularity	NormalizedCut	Score
AverageODF	-	0.58	0.59	0.55	1	0.58	0.44	0.49	0.62	0.52	1
Conductance	0.58	-	0.59	0.58	1	0.6	0.44	0.53	0.6	0.55	0.99
CutRatio	0.59	0.59	-	0.52	1	0.57	0.43	0.44	0.57	0.49	0.99
Expansion	0.55	0.58	0.52	-	1	0.59	0.4	0.45	0.6	0.5	0.99
Fitness	0.58	0.58	0.57	0.58	-	1	0.55	1	1	1	0.98
FlakeODF	0.58	0.58	0.55	0.57	0.72	-	0.43	0.5	0.6	0.53	1
InternalDenisty	0.04	0.04	0.03	0.03	0.52	0.04	-	0.27	0.53	0.27	0.98
MaxOutDegree	0.11	0.11	0.03	0.03	0.58	0.25	0.03	-	0.45	0.06	0.96
Modularity	0.59	0.57	0.57	0.55	0.71	0.59	0.14	0.07	-	0.54	1
NormalizedCut	0.52	0.5	0.48	0.38	0.56	0.51	0.04	0.05	0.5	-	0.99
Score	0.58	0.58	0.58	0.58	0.7	0.86	0.67	0.56	0.99	0.57	-

(a) Zout = 4

Objectives	AverageODF	Conductance	CutRatio	Expansion	Fitness	FlakeODF	InternalDenisty	MaxOutDegree	Modularity	NormalizedCut	Score
AverageODF	-	0.58	0.44	0.38	0.98	0.56	0.32	0.39	0.35	0.35	0.62
Conductance	0.57	-	0.48	0.48	0.95	0.55	0.36	0.44	0.37	0.39	0.6
CutRatio	0.29	0.38	-	0.15	0.96	0.47	0.25	0.23	0.06	0.13	0.5
Expansion	0.22	0.38	0.15	-	0.98	0.51	0.23	0.11	0.02	0.08	0.44
Fitness	0.55	0.55	0.52	0.54	-	1	0.5	0.91	0.95	0.63	0.92
FlakeODF	0.49	0.48	0.33	0.43	0.54	-	0.4	0.39	0.4	0.39	0.65
InternalDenisty	0.05	0.05	0.02	0.02	0.47	0.05	-	0.23	0.22	0.24	0.4
MaxOutDegree	0.04	0.1	0.02	0.02	0.36	0.08	0.03	-	0.08	0.05	0.43
Modularity	0.05	0.18	0.02	0.02	0.64	0.15	0.02	0.02	-	0.04	0.4
NormalizedCut	0.04	0.04	0.02	0.02	0.04	0.04	0.04	0.03	0.02	-	0.46
Score	0.57	0.56	0.38	0.3	0.63	0.48	0.13	0.03	0.21	0.02	-

(b) Zout = 5

Objectives	AverageODF	Conductance	CutRatio	Expansion	Fitness	FlakeODF	InternalDenisty	MaxOutDegree	Modularity	NormalizedCut	Score
AverageODF	-	0.57	0.27	0.27	0.81	0.55	0.26	0.29	0.29	0.3	0.32
Conductance	0.55	-	0.28	0.3	0.82	0.6	0.31	0.3	0.3	0.34	0.43
CutRatio	0.04	0.04	-	0.03	0.75	0.34	0.21	0.06	0.03	0.06	0.11
Expansion	0.04	0.03	0.03	-	0.71	0.36	0.21	0.07	0.03	0.06	0.14
Fitness	0.46	0.48	0.17	0.05	-	0.96	0.46	0.73	0.63	0.46	0.8
FlakeODF	0.49	0.55	0.07	0.08	0.52	-	0.35	0.28	0.32	0.35	0.49
InternalDenisty	0.06	0.06	0.01	0.01	0.44	0.06	-	0.24	0.21	0.23	0.23
MaxOutDegree	0.07	0.06	0.02	0.02	0.09	0.03	0.05	-	0.04	0.06	0.07
Modularity	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.01	-	0.04	0.1
NormalizedCut	0.05	0.04	0.02	0.02	0.05	0.05	0.04	0.04	0.02	-	0.15
Score	0.27	0.4	0.02	0.03	0.61	0.41	0.07	0.03	0.02	0.04	-

(c) Zout = 6

Objectives	AverageODF	Conductance	CutRatio	Expansion	Fitness	FlakeODF	InternalDenisty	MaxOutDegree	Modularity	NormalizedCut	Score
AverageODF	-	0.43	0.2	0.22	0.8	0.37	0.24	0.21	0.21	0.22	0.22
Conductance	0.4	-	0.22	0.22	0.6	0.36	0.24	0.23	0.24	0.25	0.25
CutRatio	0.02	0.02	-	0.02	0.54	0.22	0.21	0.05	0.02	0.03	0.07
Expansion	0.02	0.02	0.02	-	0.54	0.26	0.21	0.04	0.02	0.03	0.08
Fitness	0.36	0.32	0.02	0.02	-	0.76	0.44	0.53	0.44	0.39	0.5
FlakeODF	0.28	0.29	0.02	0.03	0.34	-	0.25	0.14	0.25	0.3	0.25
InternalDenisty	0.04	0.04	0.02	0.02	0.42	0.04	-	0.21	0.2	0.21	0.22
MaxOutDegree	0.04	0.07	0.02	0.02	0.07	0.02	0.03	-	0.05	0.05	0.07
Modularity	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	-	0.03	0.06
NormalizedCut	0.03	0.03	0.02	0.02	0.03	0.03	0.03	0.03	0.02	-	0.08
Score	0.05	0.1	0.02	0.02	0.19	0.04	0.08	0.03	0.02	0.03	-

(d) Zout = 7

Figure 5. NMI values for the best and the worst community structure in the Pareto-optimal set for GN benchmark for z_{out} values from 4 to 7, the upper triangular of the matrix show the NMI values for the best solution for each pair, the lower triangular show the NMI values for the worst solution for each pair.

[3] Newman M.E.J, Girvan M. Finding and evaluating community structure in networks. Physics Review E, 2004, 69(2):026113.

[4] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. Defining and identifying communities in networks. PNAS,2004, 101(9): 2658-2663.

[5] Clauset A, Newman M.E.J, Moore C. Finding community structure in very large networks. Physical Review E, 2004, 70(6): 06611.

[6] Shi C, Zhong C, Yan Z, Cai Y, Wu B. A multi-objective optimization approach for community detection in complex network. In Proc. CEC2010, Barcelona, Spain, Jul. 18-23, 2010,pp.1-8.

[7] M. Ehrgott. Multicriteria Optimization. Springer,Berlin, 2nd edition, 2005.

[8] J. Leskovec, K. J. Lang, and M. W. Mahoney. Empirical comparison of algorithms for network community detection. In WWW2010, pages 631-640,2010.

[9] Chuan Shi, Philip S. Yu, Yanan Cai, Zhenyu Yan, Bin Wu "On selection of objective functions in multi-objective community detection", Proceedings of the 20th ACM international conference on Information and knowledge management,pp2301-2304,2011.

[10] Pizzuti C. Ga-net: a genetic algorithm for community detection in social networks. In Proc. of the 10th Intenational Conference on Parallel Problem Solving from Nature (PPSN 2008), pages 1081–1090, 2008.

[11] Andrea Lancichinetti, Santo Fortunato, and Janos Kert´esz. Detecting the overlapping and hierarchical community structure of complex networks.arXiv:0802.1281v1 [physics.soc-ph], 2008.

[12] Aldecoa R, Marín I (2011) Deciphering Network Community Structure by Surprise. PLoS ONE 6(9): e24195. doi:10.1371/journal.pone.0024195.

[13] Tasgin M, Bingol H. Community detection in complex networks using genetic algorithm. arXiv:cond-mat/0604419,2006.

[14] Shi C, Wang Y, Wu B, Zhong C. A new genetic algorithm for community detection. Complex Sciences, 2009, 5(1): 1298-1309.

[15] Pizzuti C. A multi-objective genetic algorithm for community detection in networks. In Proc. ICTAI 2009, Newark, USA, Nov. 2-4, 2009, pp.379-386.

[16] Rohan A , Bi-Objective Community Detection (BOCD) in Networks using Genetic Algorithm. Communications in Computer and Information Science, 2011, Vol.168(1), 5-15.

[17] Deb Kalyanmoy, Agrawal Samir, Pratap Amrit, Meyarivan T. (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II Lecture Notes in Computer Science, 1917/2000 . pp. 849-858.

[18] Y. Park, M. Song, A genetic algorithm for clustering problem, in: ACGP98, 1998,pp. 568–575.

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms. MIT Press, Cambridge, MA, 2001.

[20] L. Danon, A. Daz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. Journal of Statistical Mechanics, P09008, 2005.

[21] Zachary, W.W. "An information flow model for conflict and fission in small groups", Journal of AnthropologicalResearch, 33:452-473, 1977.

[22] Lusseau, D. : The emergent properties of dolphin social network. Biology Letters, Proc. R. Soc. London B (suppl.) (2003)