ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 9, No. 3, September 2021, pp.796-807

Pearson coefficient matrix for studying the correlation of community detection scores in multi-objective evolutionary algorithm

Amenah D. Abbood[1], Ammar A. Hasan[2] and Bara'a A. Attea[1]
[1]University of Baghdad, College of science, Iraq
[2]University of Baghdad, College of Engineering, Iraq

ABSTRACT

Assessing a community detection algorithm is a difficult task due to the absence of finding a standard definition for objective functions to accurately identify the structure of communities in complex networks. Traditional methods generally consider the detecting of community structure as a single objective issue while its optimization generally leads to restrict the solution to a specific property in the community structure. In the last decade, new community detection models have been developed. These are based on multi-objective formulation for the problem, while ensuring that more than one objective (normally two) can be simultaneously optimized to generate a set of non-dominated solutions. However the issue of which objectives should be co-optimized to enhance the efficiency of the algorithm is still an open area of research. In this paper, first we generate a candidate set of partitions by saving the last population that has been generated using single objective evolutionary algorithm (SOEA) and random partitions based on the true partition for a given complex network. We investigate the features of the structure of communities which found by fifteen existing objectives that have been used in literature for discovering communities. Then, we found the correlation between any two objectives using the pearson coefficient matrix. Extensive experiments on four real networks show that some objective functions have a strong correlation and others either neutral or weak correlations.

Keywords: Community detection, Community structure, Complex network

Corresponding Author:

Ammar A. Hasan
Department Computer Engineering, University of Baghdad
Baghdad, Iraq
Email: mr.ammaradel@coeng.uobaghdad.edu.iq

1. Introduction

Community detection in network science aids us to visualize a large scale map of many real complex network systems to understand the structures and functions of these systems. Recently, a lot of research has been introduced to define and discover communities in complex networks and very relevant in network science. Communities can be grouped of related nodes in information networks [1], scientific collaboration networks [2], biological networks (protein-protein interaction networks), transportation networks [3], the metabolic networks, etc. Although the identification of communities is an interesting and currentlyhot topic, it has remained a complex task.

The complex network can be represented as a graph where each node represents an object such as people in social networks, protein in biological networks and the connections between nodes are represented by links such as friendship or communication. The network is partitioned into groups of nodes, defined communities that have dense intra-connection and sparse inter-connection. However, the main problem with community detection is that there is no gold standard definition of community (nor should there be one), but there are many slightly different structures in networks that could be called communities.

Detecting communities in real-world networks such as social, scientific collaboration networks, biological networks, transportation networks, metabolic networks or information networks is a problem of significant interest and vital research in recent years. Typically one or two objective function scores are chosen to capture the intuition of community structures. A given network is partitioned into structure of communities and each community is a group of nodes that has strong connections within a community than the external connections

with the other communities. There are many objectives that have been optimized, either optimizing one objective function or multi objectives [4, 5, 6, 7, 8, 9, 8]. Most the existing studies show that optimizing more than one objective produce more accurate network partition than the optimizing one objective as the community structure have different properties [5, 6, 10, 11, 12]. There are many objectives that have been used in the research area such as Community Fitness [13], Conductance [14], Community Score [15], Expansion, Internal Density [16], Normalized Cut and Ratio cut [17].
The literature, however, lacks which objectives are optimized to increase the accuracy of the network division. In this study a strategy will be proposed to generate a candidate set of partitions by including all partitions in the last population that has been generated by using the SOEA together with random partitions for a given network based on the true partition. After that, these partitions are evaluated by using the fifteen most popular objective functions in the literature and then find the strength of correlations among them using the pearson coefficient matrix [18]. This strategy has a vital role to determine which two contradictory objectives could be optimized effectively using multi-objective evolutionary algorithms (MOEAs) such as Multi-Objectives Evolutionary Algorithm with Decomposition (MOEA/D) [19] and Nondominated Sorting Genetic Algorithm II (NSGA-II) [20].
he rest of this paper is organized as follows: In section 2 the basic background is described for the community detection problem. In section 3 a variety of community detection measures are presented that have been proposed in the literature. The formulation for evaluating the objective functions that have been used for community detection problem is introduced in section 4.1. Based on this, section 5 provides a description for the real world networks and a discussion of experiment results. Finally, conclusions and future directions are presented in section 6.

2. Preliminary

In the literature, the problem formulation of community detection in the social network is divided into graph definition (partitioning or clustering) and finding a global solution to reflect an optimal graph partitioning method. Mathematically, for an undirected unweighted graph, a given network is modeled into G of N vertices or nodes, and M links or edges connecting between two vertices. Generally, G = (V,E), is a representation of a group V(G) = {v_1,v_2,… ,v_N} of N vertices (i.e. N(G)= |V|) and a set E(G) of L connections between the nodes (i.e. L(G)= |E|). Each node is imposed to have certain connections with other nodes, and the degree of the node presents the total number of these connections. The structure of network G is represented by a symmetric matrix $N \times N$ denoted by adjacency matrix A. $i,j \in \{1,2,...,N\}$, an element A_{ij} =1 is in the matrix A if there exists a link between the vertices v_i and v_j,otherwise A_{ij} =0 when no neighborhood relation exists. The rows and columns of the matrix are denoted by i and j respectively. Each element in the matrix associates with a single link between pair of nodes, and its value (1 or 0) refers to the existence or not of the neighborhood relation. It is noted that there is not any neighborhood relation between a node with itself, (when $i = j$), thus the elements in the matrix existing on the main diagonal are set to zero.
Graphically, the goal of dividing the nodes and their edges into groups in a network is proposed to provide possible partitions or communities that have nodes with dense connections within its community and sparse across communities. Let C (G) is the all potential communities of the graph G. Under this assumption, a community divisions to be a set of communities, $C_i \subset$ G = {C_1, C_2,… , C_i}, where $C_i \neq \emptyset$,K is the number of communities and each C_i has a number of certain nodes denoted by n_i. Vertex degree, l(v,C_i), is the number of links between the vertex v, belongs to community C_i, and other vertices in the whole network, which is defined by:

$$l(v, C_i) = \sum_{j}^{N} A_{ij} \quad (1)$$

For vertex $v \in C_i$, the number of its links with the others, belong to other communities, can define as:

$$l_{out}(v, C_i) = \sum_{j \notin C_i} A_{ij} \quad (2)$$

To end this, generally speaking, the total links of $v \in_i$, can divide as:

$$l(v, C_i) = l_{out}(v, C_i) + l_{in}(v, C_i) \quad (4)$$

Another assumption is explicitly specified, the links belonging to a particular community is vital for graph partitioning, and can also define as:

$$l(C_i) = l_{out}(C_i) + l_{in}(C_i) \quad (5)$$

where, $l_{out}(C_i)$ is the number of links not belonging to the community Ci (external edges),and $l_{in}(C_i)$ is the number of links belonging to the community Ci (internaledges).

Two new definitions are also considered here, strong and weak communities, [21, 22]. The community C_i is strong when:

$$l_{in}(C_i) > l_{out}(C_i), \quad \forall v \in C_i \quad (6)$$

The community C_i is named a strong partition if its internal edges > its external edges, otherwise, it is named a weak partition. This finding is confirmed when every *v*, belong to the C_i, makes the following condition:

$$\sum_{v \in C_i} l_{in}(v, C_i) > \sum_{v \in C_i} l_{out}(v, C_i) \quad (7)$$

The two summations of $l_{in}(v, C_i)$ and $l_{out}(v, C_i)$, belong to the C_i, are considered here to reflect a strong or weak community. A particular community could be strong when it has vertices having dense links (intra-connection) within the community and sparse links (inter-connection) with the others, otherwise the community is weak. This finding can be achieved if strong vertices can belong to a particular community, when the internal degree of the vertices belonging to C_i with other vertices within the community, exceeds their external degree, which is a groups of vertices belonging to other communities. This criterion supports one main scope that the communities should depend on their properties, such as degree of community or node, rather than just depending on a quality function like modularity.

3. Objective functions in community detection

One of the most popular clustering measures that have been used in the literature is the Newman-Girvan modularity *Q* [23]. It is an efficient evaluation measure for discovering strength of communities in a given network.

$$Q(C) = \sum_{i=1}^{K} \left[\frac{I_{in}(C_i)}{2L} - \left(\frac{I(C_i)}{2L} \right)^2 \right] \quad (8)$$

In 2008, Pizzuti has been used a single objective evolutionary algorithm (SOEA) for unfold community structures, arguably avoiding some of the issues associated with greedy search [15]. The model proposed in this work is the maximization of the *Community Score* (*CS*) defined as:

$$CS(C) = \sum_{i=1}^{K} \frac{1}{|C_i|} \sum_{v \in C_i} \left(\frac{I_{in}(v, C_i)}{|C_i|} \right)^r \times I_{in}(C_i) \quad (9)$$

Where, the size of the communities is regulated by *r* to increase the weight of the degree of the node that belongs to a given community. There are many alternative definitions of network partitions, where Lancichinetti et al. [13] proposed a different objective function named *Community Fitness* (*CF*), to find

communities in a network by determining higher peaks fitness histogram in a hierarchical community structure. The proposed function is defined as:

$$CF(C) = \sum_{i=1}^{K} \frac{I_{in}(C_i)}{(I_{in}(C_i) + I_{out}(C_i))^{\propto}} \quad (10)$$

where α is a positive value that regulates the scale of communities.
In 2012, Shi et al. also formulated the problem of discovering of community structures as a multi-objective minimization problem [24]. The authors reformulate Modularity (Eq. 8) into two quantitative terms to use as two objective functions. The first is the intra-connection measure:

$$Intra(C) = 1 - \sum_{i=1}^{K} \frac{I_{in}(C_i)}{2L} \quad (11)$$

While the second measures inter-connections:

$$Inter(C) = \sum_{i=1}^{K} \left[\frac{I(C_i)}{2L}\right]^2 \quad (12)$$

The Kernel K-Means (*KKM*) is a fantastic score reported in 2014 by Gong et al. [5] to find the community structures in complex networks [5]. The *KKM* works as a minimized function for creating small communities by measuring the sum of the internal degree of a particular community in a given network [25], which can be defined as:

$$KKM(C) = 2(N - K) - \sum_{i=1}^{K} \frac{I_{in}(C_i)}{|C_i|} \quad (13)$$

The Ratio Cut (*RC*) function which also works as a minimized function to cut the size of the partitions and provide balanced partitions [26]:

$$RC(C) = \sum_{i=1}^{K} \frac{I_{out}(C_i)}{|C_i|} \quad (14)$$

Another known score the Expansion (*EX*), is also used to show external links centrality, which is adopted the idea of accounting all possible external connections [16]:

$$EX(C) = \sum_{i=1}^{K} \frac{I_{out}(C_i)}{|C_i|\,(N - |C_i|)} \quad (15)$$

The Normalized cut function (*NC*) [17], is also a minimized function that is used to cut the degree of communities within a given network rather than cutting their size, which is defined as:

$$NC(C) = \sum_{i=1}^{K} \frac{I(v, C_i) - I_{in}(C_i)}{I(v, C_i)} \quad (16)$$

The Conductance function (*CO*) [14], is a cut-based function that measure the fraction of the totaledges of the nodes belonging to other communities in a given network, which is defined as:

$$CO(C) = \sum_{i=1}^{K} \frac{I_{out}(C_i)}{2\,I_{in}(C_i) + I_{out}(C_i)} \quad (17)$$

In contrast, the density of the internal links of the community can be scored by [22]:

$$ID(C) = \sum_{i=1}^{K} 1 - \frac{I_{in}(C_i)}{|C_i|\,(|C_i| - 1)} \quad (18)$$

The average-Out Degree Fraction (*ODF*) score is another measure [27]:

$$ODF(C) = \sum_{i=1}^{K} \frac{I_{out}(C_i)}{|C_i|\, I(v, C_i)} \tag{19}$$

The Triangle Participation Ratio (*TPR*) is a quality score to measure a set of nodes that belong to a triangle in a particular community [22], which is defined by:

$$TPR(C) = \sum_{i=1}^{K} \frac{I(C_i)}{|C_i|} \tag{20}$$

A multi-objective evolutionary algorithm was proposed in 2016 by Attea et al. to simultaneously reduce two contradictory scores to detect the structure of communities in social networks [10]. Both the so-called intra-neighbor score (Q_1) in Eq.22 and the inter-neighbor score (Q_2) in Eq.23 are minimized by their model.

$$Q_1(C) = N^2 - \sum_{i=1}^{K} \frac{I_{in}(C_i) + N_{Strong}(C_i)}{|C_i|} \tag{21}$$

Where $N_{Strong}(C_i)$ represents the number of node v_i in community C_i that have $l_{in}(v_i,C_i) > l_{out}(v_i,C_i)$.

$$Q_2(C) = \sum_{i=1}^{K} \frac{I_{in}(C_i) + N_{Weak}(C_i)}{I(v, C_i)} \tag{22}$$

Where $N_{Weak}(C_i)$ represents the number of node v_i in community C_i that have $l_{in}(v_i,C_i) < l_{out}(v_i,C_i)$.

Table 1: The numbers of objective functions represented in x-axis and y-axis in Figures 1,2,3 and 4

Number	Community Score	References
1	*KKM*	[5]
2	*CS*	[15]
3	*Q*	[23]
4	*CF*	[13]
5	*TPR*	[22]
6	*EX*	[16]
7	*ODF*	[27]
8	*NC*	[17]
9	Q_1	[10]
10	Q_2	[10]
11	*Intra*	[24]
12	*Inter*	[24]
13	*CO*	[14]
14	*RC*	[26]
15	*ID*	[22]

4. Methodology

4.1. Candidate set of partitions

While various optimization functions have been reported in existing studies to identify community structure in networks, it is still unknown how well these objectives are correlated. In this paper, fifteen of these objectives are considered to discuss their performance where these objectives are already widely used to capture the intuition of communities in the literature, see Table 1. These objectives are Community Score, Community Fitness, Intra, Inter, Kernel K-means, Ratio Cut, Intra Neighbor and Inter Neighbor. In addition these objectives are the source for many community detection algorithms [28, 29, 6, 30, 31, 32, 33]. In this paper we proposed two methods for choosing candidate partitions to evaluate objective functions:

A: Single Objective Evolutionary Algorithm

In this study, we use Single Objective Evolutionary Algorithm (SOEA) to detect communities in a given network by employing Genetic Algorithm [34] for this purpose. A trial and error technique is used to determine parameter values, and then the parameter values that produced good results for the data sets are chosen. As a result, we set the crossover rate to 0.8, the mutation rate to 0.2, the population size is 100, and there are 100 generations. The accuracy of the detected partition is assessed using an external measure, the Normalized Mutual Information (NMI) [35], to approximate the similarity between the true and detected partitions. We apply SOEA on four real-world networks that have been extensively studied in the literature. The Zackary's Karate Club network Zachary [36], the Bottlenose Dolphins Lusseau [37], the American College Football network [38], and the Krebs' books on American politics [39]. Fifteen objectives are optimized for each network and with each objective we save the last population to be considered as good candidate partitions. In this case we have fifteen populations which represent good candidate partitions for each network to evaluate existing objectives. The last population could has true partition or close to the true partition. We point to the number of individuals (partitions) in the last population as N_{pop}.

B: Random partition based on true partition

Random partitions for a given real world network are generated based on the true partitions. The true partition is not always has strong communities but generally speaking it is more similar to the nature partitions of networks. Each time random nodes are migrated from its community to random communities. Candidate partitions are selected because choosing all combinations of partitions is huge and it is impossible to choose all of them. In this strategy, the total number of candidate partitions is P where $P = N_{pop} + \sum_{i=1}^{N} i$. At each iteration, i of nodes is migrated from its community to random communities. The first and second terms are calculated using the SOEA and the random partition method, see Algorithm 1.

4.2. The correlation of objective functions based on Pearson Coefficient Matrix

One of the most popular measures to find the correlation between two data sets is Pearson Coefficient Matrix. The range of Pearson Coefficient Matrix .(r) can be a value between -1 and 1, where 1 points to a strong correlation, 0 means no relationship between the two variables, and a weak correlation is shown by a value smaller than 0; that is, when one variable's value drops, so does the value of the other [40]:

$$r = \frac{\sum_{i=1}^{p} (f_1(i) - f_1)\,(f_2 - f_2(i))}{\sqrt{\sum_{i=1}^{p} \left(f_1 - f_1(i)\right)^2 (f_2 - f_2(i))^2}} \tag{23}$$

where $f1$ and $f2$ represent the mean values of two different functions in candidate partitions. In this study fifteen objective functions are evaluated based on the candidate partitions and the correlation coefficient (r) between each pair of these objectives (see Eq 23). These calculations are used to find the statistical association strength between any pair of objectives.

This study is very effective to recognize the type of correlation between two objectives that could be used as contradictory objectives in multi-objective evolutionary algorithm to identify community structures in a given networks.

Algorithm 1: An algorithm for generating random partitions.

begin
$C^* \leftarrow \{C_1^*, C_2^*, \ldots, C_K^*\}$ Ground-truth partition
$i \leftarrow 1$ i is the number of migrated nodes ;
while $(i \leqslant N)$ **do**
while $(j \leqslant i)$ **do**
$v_j \leftarrow rand(v_1, v_2, ..., v_N)$;
$C^* \leftarrow C^*(v_j)$;
/* A new community will be generated */
$\bar{C_{v_j}} \leftarrow rand\{C_1^*, C_2^*, \ldots, C_K^*\} | \bar{C_{v_j}} \neq C^*$;
/* Change the community of node v_j */
$C(v_j) \leftarrow \bar{C_{v_j}}$;
$j \leftarrow j + 1$;
update: $\mathbf{C} \leftarrow \mathbf{C} \cup \mathbf{C}^{(i)}$;
$i \leftarrow i + 1$;

5. Experiments

This section discusses the actual correlations between objective functions that are a potential used for formulating the problem of the community detection. Fifteen objectives are considered here, which have been mostly presented in the literature. To analyze the strength of the correlation between each one with the others, experiments are carried on real networks which are Zachary Karate Club [36], Bottlenose Dolphins [37], American College Football team [38] and Krebs Books networks [39]. These networks have a well-known summery of real datasets that help to use for evaluating community detection according to the strength of correlation between the selected objectives.

Each objective considers one or more properties when it works on these networks. Thus, the calculations of each objective are individually carried on the potential candidates of the community partitions to uncover the strength of relations and find the correlations between each objective and others. Objectives that have more correlated, they are less likely to be optimized using multi-objective evolutionary algorithm because the objectives that have more correlated tend to be similar performance.

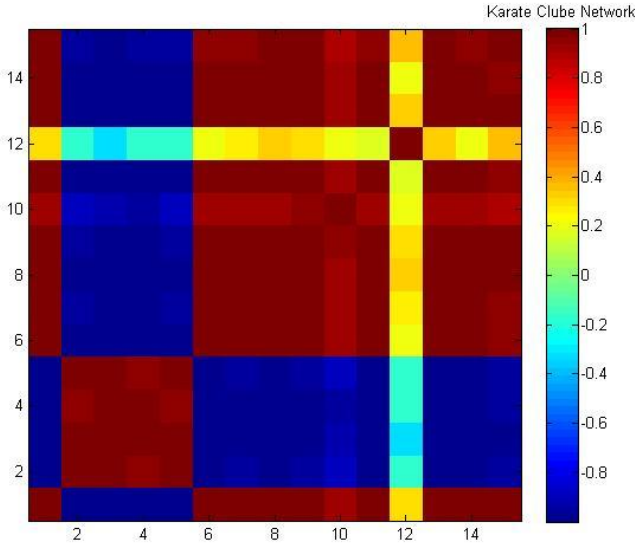

Figure 1. The correlation between the evaluation scores which have been used for evaluating Karate network partitions. The blue color points to weak correlation while the red color refers to the strong correlation.

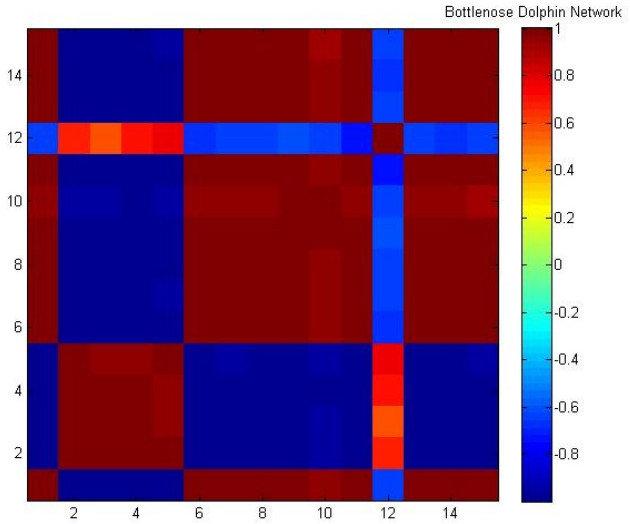

Figure 2: The correlation between the evaluation scores which have been used for evaluating Dolphin network partitions. The blue color points to weak correlation while the red color refers to the strong correlation.

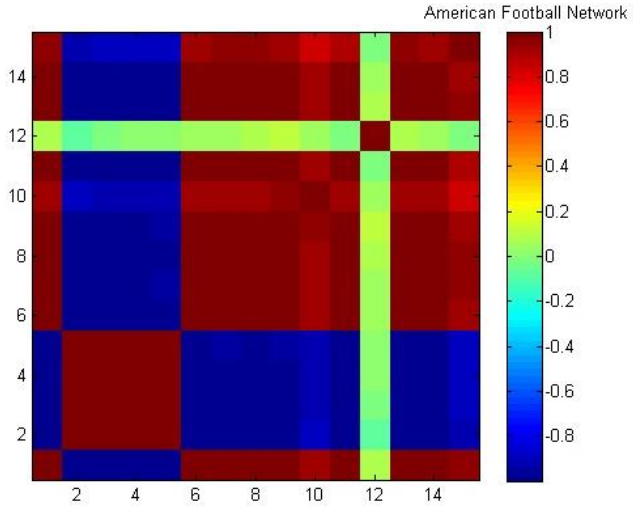

Figure 3: The correlation between the evaluation scores which have been used for evaluating Karate network partitions. The blue color points to weak correlation while the red color refers to the strong correlation.

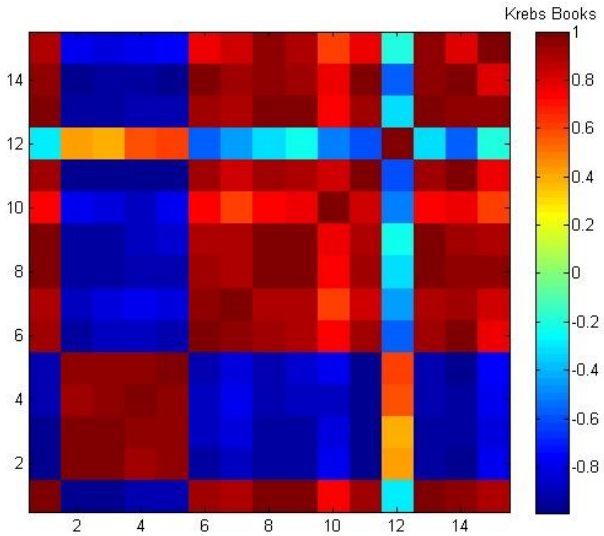

Figure 4: The correlation between the evaluation scores which have been used for evaluating Dolphin network partitions. The blue color points to weak correlation while the red color refers to the strong correlation.

To understand the trends of the correlations, the computational relations between each pair of objectives are supposed to be either strong, weak or no correlation. For a quantitative analysis, color bar is used here for reflecting correlation among the objective functions. From Figures 1,2,3,4 , in general, the objectives (*KKM, EX, ODF, NC,* Q_1*,* Q_2*, Intra, CO, RC* and *ID*) are likely to have a similar performance among them with a strong correlation (in red color) comparing with other objectives with a weak correlation (in blue color), such as *SC, Q, CF* and *TRF*. Further, the objective *'Inter'* has a moderate correlation with all other objectives. We can see the distribution of correlation strength according to the evaluated values of both objective functions from the candidate set partitions. These values are calculated using Pearson Coefficient Matrix. The correlation, in red square, shows that there is high strength correlation between the functions while the blue square represents low correlations (weak correlation). Other colored squares show a neutral or no correlations among functions. In order to produce more accurate correlation between objectives, another candidate set partitions should be generated.

Throughout our experiments, a set of group of results are presented in Tables (2- 5). The results display the correlation between each pair of objectives using Pearson Coefficient Matrix on four real world networks. The high number refers to high correlated between pair of objectives and corresponds to red color (more correlated) in Figures 1, 2, 3, 4. On the other hand, the smallest number refers to weak correlated (blue color). As a result, we avoid choosing the high correlated functions in multi-objective evolutionary algorithms as these objectives have the same or similar performance that could be naturally unsuitable to formulate multiobjective evolutionary algorithm for community detection problems.

Table 2. The correlation values of 15 objectives are found using pearson coefficient matrix for karate network

Name	KKM	CS	Q	CF	TPR	EX	ODF	NC	Q_1	Q_2	Intra	Inter	CO	RC	ID
KKM	1.000	-0.995	-0.987	-0.984	-0.980	0.998	0.996	1.000	0.994	0.921	0.985	0.076	1.000	0.998	0.943
CS	-0.995	1.000	0.988	0.985	0.985	-0.994	-0.984	-0.992	-0.984	-0.903	-0.986	-0.081	-0.992	-0.994	-0.921
Q	-0.997	0.973	1.000	0.979	0.971	-0.989	-0.985	-1.000	-0.990	-0.921	-0.985	-0.325	-1.000	-0.989	-0.981
CF	-0.977	0.967	0.979	1.000	0.966	-0.984	-0.971	-0.979	-0.977	-0.938	-0.988	-0.187	-0.979	-0.984	-0.950
TPR	-0.971	0.970	0.971	0.966	1.000	-0.979	-0.954	-0.970	-0.954	-0.888	-0.985	-0.158	-0.970	-0.979	-0.943
EX	0.990	-0.984	-0.989	-0.984	-0.979	1.000	0.988	0.988	0.983	0.923	0.995	0.206	0.988	1.000	0.956
ODF	0.983	-0.961	-0.985	-0.971	-0.954	0.988	1.000	0.986	0.985	0.937	0.980	0.264	0.986	0.988	0.955
NC	0.997	-0.971	-1.000	-0.979	-0.970	0.988	0.986	1.000	0.991	0.923	0.985	0.322	1.000	0.988	0.982
Q_1	0.992	-0.968	-0.990	-0.977	-0.954	0.983	0.985	0.991	1.000	0.956	0.978	0.309	0.991	0.983	0.978
Q_2	0.916	-0.891	-0.921	-0.938	-0.888	0.923	0.937	0.923	0.956	1.000	0.927	0.193	0.923	0.923	0.888
Intra	0.986	-0.983	-0.985	-0.988	-0.985	0.995	0.980	0.985	0.978	0.927	1.000	0.158	0.985	0.995	0.956
Inter	0.300	-0.181	-0.325	-0.187	-0.158	0.206	0.264	0.322	0.309	0.193	0.158	1.000	0.322	0.206	0.374
CO	0.997	-0.971	-1.000	-0.979	-0.970	0.988	0.986	1.000	0.991	0.923	0.985	0.322	1.000	0.988	0.982
RC	0.990	-0.984	-0.989	-0.984	-0.979	1.000	0.988	0.988	0.983	0.923	0.995	0.206	0.988	1.000	0.956
ID	0.987	-0.958	-0.981	-0.950	-0.943	0.956	0.955	0.982	0.978	0.888	0.956	0.374	0.982	0.956	1.000

Table 3. The correlation values of 15 objectives are found using Pearson Coefficient Matrix for Dolphin network

Name	KKM	CS	Q	CF	TPR	EX	ODF	NC	Q_1	Q_2	Intra	Inter	CO	RC	ID
KKM	1.000	-0.993	-0.998	-0.980	-0.975	0.995	0.989	0.999	0.996	0.954	0.989	-0.624	0.999	0.995	0.990
CS	-0.993	1.000	0.985	0.978	0.985	-0.996	-0.982	-0.994	-0.984	-0.940	-0.992	0.682	-0.994	-0.996	-0.982
Q	-0.998	0.985	1.000	0.975	0.965	-0.990	-0.985	-0.996	-0.995	-0.955	-0.982	0.585	-0.996	-0.990	-0.981
CF	-0.980	0.978	0.975	1.000	0.968	-0.982	-0.971	-0.981	-0.983	-0.972	-0.987	0.695	-0.981	-0.982	-0.970
TPR	-0.975	0.985	0.965	0.968	1.000	-0.986	-0.960	-0.977	-0.967	-0.935	-0.993	0.757	-0.977	-0.986	-0.963
EX	0.995	-0.996	-0.990	-0.982	-0.986	1.000	0.990	0.996	0.991	0.953	0.995	-0.675	0.996	1.000	0.979
ODF	0.989	-0.982	-0.985	-0.971	-0.960	0.990	1.000	0.989	0.987	0.948	0.979	-0.624	0.989	0.990	0.973
NC	0.999	-0.994	-0.996	-0.981	-0.977	0.996	0.989	1.000	0.995	0.952	0.990	-0.634	1.000	0.996	0.989

Name	KKM	CS	Q	CF	TPR	EX	ODF	NC	Q_1	Q_2	Intra	Inter	CO	RC	ID
Q_1	0.996	-0.984	-0.995	-0.983	-0.967	0.991	0.987	0.995	1.000	0.974	0.985	-0.615	0.995	0.991	0.984
Q_2	0.954	-0.940	-0.955	-0.972	-0.935	0.953	0.948	0.952	0.974	1.000	0.959	-0.651	0.952	0.953	0.936
Intra	0.989	-0.992	-0.982	-0.987	-0.993	0.995	0.979	0.990	0.985	0.959	1.000	-0.728	0.990	0.995	0.977
Inter	-0.624	0.682	0.585	0.695	0.757	-0.675	-0.624	-0.634	-0.615	-0.651	-0.728	1.000	-0.634	-0.675	-0.632
CO	0.999	-0.994	-0.996	-0.981	-0.977	0.996	0.989	1.000	0.995	0.952	0.990	-0.634	1.000	0.996	0.989
RC	0.995	-0.996	-0.990	-0.982	-0.986	1.000	0.990	0.996	0.991	0.953	0.995	-0.675	0.996	1.000	0.979
ID	0.990	-0.982	-0.981	-0.970	-0.963	0.979	0.973	0.989	0.984	0.936	0.977	-0.632	0.989	0.979	1.000

Table 4: The correlation values of 15 objectives are found using Pearson Coefficient Matrix for American Football network.

Name	KKM	CS	Q	CF	TPR	EX	ODF	NC	Q_1	Q_2	Intra	Inter	CO	RC	ID
KKM	1.000	-0.995	-0.987	-0.984	-0.980	0.998	0.996	1.000	0.994	0.921	0.985	0.076	1.000	0.998	0.943
CS	-0.995	1.000	0.988	0.985	0.985	-0.994	-0.984	-0.992	-0.984	-0.903	-0.986	-0.081	-0.992	-0.994	-0.921
Q	-0.987	0.988	1.000	0.999	0.997	-0.990	-0.976	-0.987	-0.976	-0.918	-1.000	-0.018	-0.987	-0.992	-0.892
CF	-0.984	0.985	0.999	1.000	0.997	-0.987	-0.973	-0.984	-0.972	-0.916	-1.000	0.014	-0.984	-0.989	-0.892
TPR	-0.980	0.985	0.997	0.997	1.000	-0.984	-0.966	-0.979	-0.967	-0.906	-0.997	0.011	-0.979	-0.986	-0.885
EX	0.998	-0.994	-0.990	-0.987	-0.984	1.000	0.995	0.998	0.992	0.925	0.988	0.055	0.998	1.000	0.933
ODF	0.996	-0.984	-0.976	-0.973	-0.966	0.995	1.000	0.996	0.992	0.922	0.974	0.059	0.996	0.994	0.956
NC	1.000	-0.992	-0.987	-0.984	-0.979	0.998	0.996	1.000	0.993	0.923	0.985	0.069	1.000	0.998	0.943
Q_1	0.994	-0.984	-0.976	-0.972	-0.967	0.992	0.992	0.993	1.000	0.953	0.973	0.097	0.993	0.991	0.937
Q_2	0.921	-0.903	-0.918	-0.916	-0.906	0.925	0.922	0.923	0.953	1.000	0.917	0.039	0.923	0.925	0.839
Intra	0.985	-0.986	-1.000	-1.000	-0.997	0.988	0.974	0.985	0.973	0.917	1.000	-0.011	0.985	0.990	0.893
Inter	0.076	-0.081	-0.018	0.014	0.011	0.055	0.059	0.069	0.097	0.039	-0.011	1.000	0.069	0.050	-0.030
CO	1.000	-0.992	-0.987	-0.984	-0.979	0.998	0.996	1.000	0.993	0.923	0.985	0.069	1.000	0.998	0.943
RC	0.998	-0.994	-0.992	-0.989	-0.986	1.000	0.994	0.998	0.991	0.925	0.990	0.050	0.998	1.000	0.930
ID	0.943	-0.921	-0.892	-0.892	-0.885	0.933	0.956	0.943	0.937	0.839	0.893	-0.030	0.943	0.930	1.000

Table 5: The correlation values of 15 objectives are found using Pearson Coefficient Matrix for Krebs Books network.

Name	KKM	CS	Q	CF	TPR	EX	ODF	NC	Q_1	Q_2	Intra	Inter	CO	RC	ID
KKM	1.000	-0.967	-0.968	-0.898	-0.908	0.923	0.899	0.993	0.985	0.733	0.922	-0.293	0.993	0.949	0.904
CS	-0.967	1.000	0.981	0.937	0.960	-0.948	-0.873	-0.946	-0.941	-0.790	-0.970	0.434	-0.946	-0.980	-0.784
Q	-0.968	0.981	1.000	0.953	0.955	-0.892	-0.809	-0.956	-0.943	-0.811	-0.972	0.379	-0.956	-0.947	-0.826
CF	-0.898	0.937	0.953	1.000	0.947	-0.884	-0.781	-0.904	-0.878	-0.885	-0.978	0.568	-0.904	-0.93	-0.778
TPR	-0.908	0.960	0.955	0.947	1.000	-0.926	-0.819	-0.904	-0.865	-0.798	-0.990	0.604	-0.904	-0.967	-0.758
EX	0.923	-0.948	-0.892	-0.884	-0.926	1.000	0.965	0.918	0.894	0.724	0.925	-0.566	0.918	0.988	0.778
ODF	0.899	-0.873	-0.809	-0.781	-0.819	0.965	1.000	0.903	0.882	0.617	0.821	-0.443	0.903	0.927	0.818
NC	0.993	-0.946	-0.956	-0.904	-0.904	0.918	0.903	1.000	0.977	0.734	0.920	-0.325	1.000	0.943	0.939
Q_1	0.985	-0.941	-0.943	-0.878	-0.865	0.894	0.882	0.977	1.000	0.778	0.886	-0.238	0.977	0.919	0.895
Q_2	0.733	-0.790	-0.811	-0.885	-0.798	0.724	0.617	0.734	0.778	1.000	0.840	-0.513	0.734	0.781	0.603
Intra	0.922	-0.970	-0.972	-0.978	-0.990	0.925	0.821	0.920	0.886	0.840	1.000	-0.586	0.920	0.971	0.776
Inter	-0.293	0.434	0.379	0.568	0.604	-0.566	-0.443	-0.325	-0.238	-0.513	0.586	1.000	-0.325	-0.556	-0.209
CO	0.993	-0.946	-0.956	-0.904	-0.904	0.918	0.903	1.000	0.977	0.734	0.920	-0.325	1.000	0.943	0.939
RC	0.949	-0.980	-0.947	-0.937	-0.967	0.988	0.927	0.943	0.919	0.781	0.971	-0.556	0.943	1.000	0.798
ID	0.904	-0.784	-0.826	-0.778	-0.758	0.778	0.818	0.939	0.895	0.603	0.776	-0.209	0.939	0.798	1.000

6. Conclusion

This paper presents the correlation between two objectives to determine whether they have pairwise strong or weak correlation. This will, in turn, help scholars how to select a pair of contradictory objectives to properly

define the community detection problem as a multi-objective optimization problem. The proposed method includes two steps. First, a set of candidate partitions is generated and evolved using SOEA. This is associated with a random strategy to generate candidate partitions based on the true partition by migrating nodes from their communities to random communities. Second, the most popular fifteen objectives provided in the literature are evaluated with respect to these partitions to find the pairwise correlation between these objectives using Pearson Coefficient Matrix. The investigation of the correlations is very important since more weakness in the correlations means more contradiction in their semantics. As a future work, it would be interesting to use two objectives that have less correlation in the MOEA based community detection. Another direction for the future is to investigate another candidate partitions that could be used to find the correlations between any two objectives as it is difficult to consider all combinations of partitions.

References

[1] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener," Graph structure in the web", *Computer networks*, 33(1):309– 320, 2000.

[2] Mark EJ Newman, "The structure of scientific collaboration networks", *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.

[3] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani, "The architecture of complex weighted networks: Measurements and models", In *Large scale structure and dynamics of complex networks: from information technology to finance and natural science*, pages 67–92, World Scientific, 2007.

[4] Clara Pizzuti, "A multiobjective genetic algorithm to find communities in complex networks", *Evolutionary Computation, IEEE Transactions on*, 16(3):418–430, 2012.

[5] Maoguo Gong, Qing Cai, Xiaowei Chen, and Lijia Ma, "Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition", *Evolutionary Computation, IEEE Transactions on*, 18(1):82–97, 2014.

[6] Peng Wu and Li Pan, "Multi-objective community detection based on memetic algorithm", *PloS one*, 10(5):e0126845, 2015.

[7] Huda M Rada and Bara'a A Attea," Understanding the role of positive and negative relations for community detection problem in signed networks: A new perspective", *Iraqi Journal of Science*, 58(4B):2222– 2235, 2017.

[8] A Attea Bara'a, Amenah D Abbood, Ammar A Hasan, Clara Pizzuti, Mayyadah Al-Ani, Suat Özdemir, and Rawaa Dawoud Al-Dabbagh", A review of heuristics and metaheuristics for community detection in complex networks: Current usage, emerging development and future directions", *Swarm and Evolutionary Computation*, page 100885, 2021.

[9] Mayasa M Abdulrahman, Amenah Dahim Abood, and Baraa A Attea, "An enhanced multi-objective evolutionary algorithm with decomposition for signed community detection problem", In *2020 2nd Annual International Conference on Information and Sciences (AiCIS)*, pages 45–50. IEEE, 2020.

[10] Bara'a A Attea, Wisam A Hariz, and Mayyadah F Abdulhalim, "Improving the performance of evolutionary multi-objective coclustering models for community detection in complex social networks", *Swarm and Evolutionary Computation*, 26:137–156, 2016.

[11] Bara'a A Attea and Qusay Z Abdullah, "Improving the performance of evolutionary-based complex detection models in protein–protein interaction networks", *Soft Computing*, pages 1–24, 2018.

[12] Attea Bara'a A Abdulateef, Amenah HH and Ahmed N Rashid, "Heuristic modularity for complex identification in protein-protein interaction networks", *Iraqi Journal of Science*, 60(8):1846–1859, 2019.

[13] Andrea Lancichinetti, Santo Fortunato, and János Kertész, "Detecting the overlapping and hierarchical community structure in complex networks", *New Journal of Physics*, 11(3):033015, 2009.

[14] Ravi Kannan, Santosh Vempala, and Adrian Vetta, "On clusterings: Good, bad and spectral", *Journal of the ACM (JACM)*, 51(3):497–515, 2004.

[15] Clara Pizzuti, "Ga-net: A genetic algorithm for community detection in social networks", In *Parallel Problem Solving from Nature–PPSN X*, pages 1081–1090. Springer, 2008.

[16] Yen-Chuen Wei and Chung-Kuan Cheng, "Towards efficient hierarchical designs by ratio cut partitioning", In *1989 IEEE International Conference on Computer-Aided Design. Digest of Technical Papers*, pages 298–301. IEEE, 1989.

[17] Jianbo Shi and Jitendra Malik, "Normalized cuts and image segmentation", *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[18] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen, "Pearson correlation coefficient", In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[19] Qingfu Zhang and Hui Li. Moea'd, "A multiobjective evolutionary algorithm based on decomposition", *Evolutionary Computation, IEEE Transactions on*, 11(6):712–731, 2007.

[20] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-ii", Evolutionary Computation, IEEE Transactions on, 6(2):182–197, 2002.

[21] Tsuyoshi Murata and Sakiko Moriyasu, "Link prediction of social networks based on weighted proximity measures", In IEEE/WIC/ACM International Conference on Web Intelligence (WI'07), pages 85–88. IEEE, 2007.

[22] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi, "Defining and identifying communities in networks", Proceedings of the National Academy of Sciences of the United States of America, 101(9):2658–2663, 2004.

[23] Mark E J Newman and Michelle Girvan, "Finding and evaluating community structure in networks", Physical review E, 69(2):026113, 2004.

[24] Chuan Shi, Zhenyu Yan, Yanan Cai, and Bin Wu, "Multi-objective community detection in complex networks", Applied Soft Computing, 12(2):850–859, 2012.

[25] Leonardo Angelini, Stefano Boccaletti, Daniele Marinazzo, Mario Pellicoro, and Sebastiano Stramaglia, "Identification of network modules by optimization of ratio association", Chaos: An Interdisciplinary Journal of Nonlinear Science, 17(2):023114, 2007.

[26] Lars Hagen and Andrew B Kahng, "New spectral methods for ratio cut partitioning and clustering", IEEE transactions on computer-aided design of integrated circuits and systems, 11(9):1074–1085, 1992.

[27] Gary William Flake, Steve Lawrence, and C Lee Giles, "Efficient identification of web communities", In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 150–160, 2000.

[28] Naeem Shahabi Sani, Mohammad Manthouri, and Faezeh Farivar, "A multi-objective ant colony optimization algorithm for community detection in complex networks", Journal of Ambient Intelligence and Humanized Computing, 11(1):5–21, 2020.

[29] Qi Li, Zehong Cao, Weiping Ding, and Qing Li, "A multi-objective adaptive evolutionary algorithm to extract communities in networks", Swarm and Evolutionary Computation, 52:100629, 2020.

[30] Shadi Rahimi, Alireza Abdollahpouri, and Parham Moradi, "A multiobjective particle swarm optimization algorithm for community detection in complex networks", Swarm and Evolutionary Computation, 39:297–309, 2018.

[31] Xingyi Zhang, Kefei Zhou, Hebin Pan, Lei Zhang, Xiangxiang Zeng, and Yaochu Jin, "A network reduction-based multiobjective evolutionary algorithm for community detection in large-scale complex networks", IEEE transactions on cybernetics, 2018.

[32] Ye Tian, Shangshang Yang, and Xingyi Zhang, "An evolutionary multiobjective optimization based fuzzy method for overlapping community detection", IEEE Transactions on Fuzzy Systems, 2019.

[33] Ali Reihanian, Behrouz Minaei-Bidgoli, and Hosein Alizadeh, "Topicoriented community detection of rating-based social networks", Journal of King Saud University-Computer and Information Sciences, 28 (3):303–310, 2016.

[34] David E Golberg, "Genetic algorithms in search, optimization, and machine learning", Addion wesley, 1989(102):36, 1989.

[35] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas, "Comparing community structure identification", Journal of Statistical Mechanics: Theory and Experiment, 2005(09):P09008, 2005.

[36] WayneWZachary, "An information flow model for conflict and fission in small groups", Journal of anthropological research, pages 452–473, 1977.

[37] David Lusseau, "The emergent properties of a dolphin social network", Proceedings of the Royal Society of London B: Biological Sciences, 270(Suppl 2):S186–S188, 2003.

[38] Michelle Girvan and Mark EJ Newman, "Community structure in social and biological networks", Proceedings of the national academy of sciences, 99(12):7821–7826, 2002.

[39] Mark E J Newman, "Modularity and community structure in networks", Proceedings of the national academy of sciences, 103(23): 8577–8582, 2006.

[40] K Pearson, "VII. Note on regression and inheritance in the case of two parents", proceedings of the royal society of London, 58, 240-242, 1895.