DOI: 10.24996/ijs.2021.62.6.32

ISSN: 0067-2904

The Influence of NMI against Modularity in Community Detection Problem: A Case Study for Unsigned and Signed Networks

Mayasa M. Abdulrahman*, Amenah D. Abbood, Bara'a A. Attea
Computer Science - Computer Engineering Department, College of Science, College of Engineering Science, University of Baghdad, Baghdad, Iraq

Received: 16/10/2020 Accepted: 30/11/2020

Abstract

Community detection is useful for better understanding the structure of complex networks. It aids in the extraction of the required information from such networks and has a vital role in different fields that range from healthcare to regional geography, economics, human interactions, and mobility. The method for detecting the structure of communities involves the partitioning of complex networks into groups of nodes, with extensive connections within community and sparse connections with other communities. In the literature, two main measures, namely the Modularity (Q) and Normalized Mutual Information (NMI) have been used for evaluating the validation and quality of the detected community structures. Although many optimization algorithms have been implemented to unfold the structures of communities, the influence of NMI on the Q, and vice versa, between a detected partition and the correct partition in signed and unsigned networks is unclear. For this reason, in this paper, we investigate the correlation between Q and NMI in signed and unsigned networks. The results show that there is no direct relationship between Q and NMI in both types of networks.

Keywords: Community Detection, Signed Networks, Unsigned Networks, Multi-Objective Algorithms, Optimization.

تأثير معامل *NMI* مقابل *Modularity* في مشكلة تحديد المجتمعات : دراسة في الشبكات غير المؤشرة و المؤشرة

مياسة محمد عبد الرحمن*, آمنة داحم عبود , براءالعطية
قسم علوم الحاسوب, كلية العلوم, جامعة بغداد, بغداد, العراق

الخلاصة

يعد اكتشاف المجتمع مفيدًا لفهم بنية الشبكات المعقدة بشكل أفضل، ويساعد في استخراج المعلومات المطلوبة من هذه الشبكات. حيث لها دورًا حيويًا في مجالات مختلفة من الرعاية الصحية إلى الجغرافية الإقليمية والاقتصاد والتفاعلات البشرية والتنقل. تتضمن طريقة اكتشاف بنية المجتمعات تقسيم الشبكة المعقدة إلى مجموعات من العقد ذات اتصالات واسعة داخل المجتمع ومتناثرة مع المجتمعات الأخرى. في الدراسات السابقة، تم استخدام مقياسين رئيسيين، النمطية Q، والمعلومات المتبادلة الموحدة NMI لتقييم التحقق من صحة وجودة هياكل المجتمع المكتشفة. على الرغم من تنفيذ العديد من خوارزميات التحسين لكشف هياكل المجتمعات ، إلا أنه من غير الواضح تأثير NMI على Modularity والعكس بالعكس بين القسم المكتشف

*Email: mayasa.abdul-rahman@coeng.uobaghdad.edu.iq

والقسم الصحيح في الشبكات الموقعة وغير الموقعة . لهذا السبب، في هذا البحث، نتحقق من العلاقة بين QوNMI في الشبكات الموقعة وغير الموقعة. تظهر النتائج أنه لا توجد علاقة مباشرة بين Q و NMI في كلتا الشبكتين.

Introduction

Technological and computational advances have paved the way for mining of huge data volumes. This mining process has deepened the understanding of the structure and function of real systems; it has also helped in discovering interesting but unknown patterns that exist in such data. Various forms of real-world data can be modelled with networks; a network serves as a powerful tool for mathematical representation of the relationships in datasets. There are different categories of networks generated from real-world datasets; they include social network, information network, biological network, and technical network. *Social network* is a network that facilitates connection and communication between people. It is not limited to just "online social networks" such as Facebook or Twitter; it also includes the network of people collaboration, co-appearance, co-authorship, and networks of communication between people, such as emails and telephones [1]. *Information network* is another category; it is also called sometimes "knowledge network". It is a network of information-carrying entities, such as the network of citations and the World Wide Web[2]. *Biological networks* are networks, such as neural networks, protein-protein interaction networks, networks of metabolic pathways, networks of blood vessels, and the food web. Several biological systems can be modelled as networks; one of the representative examples of biological networks is the network of metabolic pathways, which represents the metabolic substrates and products with directed adjoining edges. The occurrence of a known metabolic reaction on a given substrate gives rise to a given product[3, 4]. *Technical networks* are another type of networks; they are technological networks and often referred to as manmade networks; they are designed for the distribution of certain resources or commodities, such as the Internet. Some of the examples of technical networks include the electric power grid, road networks, railways network, etc. [3, 5].

Networks with only positive links are referred to as unsigned networks, while those with both positive and negative connections are referred to as signed networks. In the signed networks, the links carry more information compared to the links in unsigned networks. For instance, a positive link in an unsigned network implies just a 'relationship', but in a signed network, such link denotes a 'positive relationship', while a negative link denotes a 'negative relationship'. The relationships between parties in a signed social network may be political alliances and oppositions [6]. Ferligoj and Kramberger established the positive and negative links to capture the political alignments with positive and negative links connections, respectively [6]. In the biological field, the gene can either be repressed or enhanced by another one; the relationship between the repressed and enhanced genes can be captured as positive or negative links[7-10]. Furthermore, a certain type of lung cancer can express a certain protein that will be lacking in another subtype. Here, the relationships between the expressed and non-expressed proteins in these lung cancer subtypes can be captured as positive and negative links [11]. Kunegis *et al*. recently proved that consideration of only the positive and negative links could be helpful in finding more useful information than relying only on positive links analysis [12], which motivated analysing community detection problems in signed networks.

Since the discovery of the Community Detection (CD) problem by Girvan and Newman[13, 14], it has received a considerable attention by many researchers. However, the difficulty in this problem is that most of the existing CD methods are only capable of handling networks with no negative connections, i.e. unsigned networks[15-18], because in such networks, communities are represented as groups of nodes with dense intra-links (extensive connections within community) and less dense inter-links (sparse connections between communities). Contrarily, communities in signed networks are groups of nodes with positive dense intra links and negative less dense inter links. This means that CD methods focus merely on link density in unsigned networks but not the link signs as their clustering attributes. But in signed networks, communities depend, not only on the link density, but also on the link signs. Hence, the previous CD frameworks in unsigned networks cannot be efficiently performed in signed networks. Considering the importance of signed networks here, there is a need to develop CD methods for signed networks. However, the major problem of CD in signed networks is the ambiguous nature of the community structure owing to the presence of negative intra-community link and some positive inter-community links. With this problem, studies have focused more on the development of algorithms for recent community structure detection in order to get the best partitioning of signed networks[19, 20].

Despite the existence of numerous CD algorithms[21-23], it should be noted that the literature lacks the investigation of relations between Modularity (Q) and Normalized Mutual Information (NMI). Addressing this issue is the major contribution of this paper.

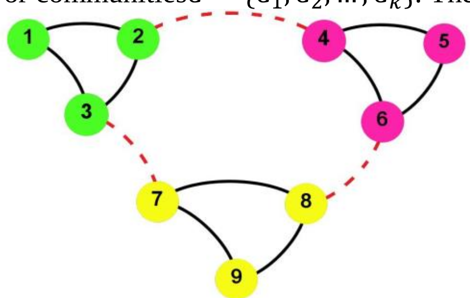
There is several evaluation measurements used for validating the generated solutions. The Modularity and NMI are two of the most used evaluation methods and sources for many publications[19, 24]. Addressing this issue is the major contribution of this paper with respect to the multi-objective optimization algorithms in signed and unsigned networks. In this way, more informative vision can be visualized on this prospective.

The rest of the paper is structured as follows. Section 2 presents an overview on the community detection problem and literature review, while section 3 presents the results, including the description on the datasets used in this study, and the main comparison between the NMI and Modularity based on several published methods. General conclusions are drawn in section 4.

Problem definition

Networks (graphs) are one of the most fundamental data structures in computer science. A network can be represented as an adjacency matrix $A \in R^{N \times N}$, where N denotes the number of nodes in the network. Here, the entry A_{ij} is one if there is connection between node i and j, and zero otherwise. A graph is used to represent the relationships of objects in a certain network. An object in a network represents a single node or vertex, and the relationship between two objects is called edge or link. The network can be used to describe the relationships among humans in social life, countries in the world, trading commodities, cities in a delivery problem, train stations or bus stops in some transportation system, connected computers in the Internet, airports in flight data set, interactions between proteins in biological systems, and so forth. Analysing such types of networks has become an immensely promising research area, and there is a lot of active research in network science, including community detection.

A static network is modelled as a graph $G = (V, E)$, where V represents the set of nodes or vertices, $V(G) = \{v_1, v_2, .. v_N\}$ with $N = |V|$ and $E(G)$ represents a set of L links or edges between nodes; $L = |E(G)|$. This definition is extended for singed networks as $G = (V, E, W)$, where W represents the type of the connections, W can be formulated as $W: V \times V \rightarrow \{-1, 0, +1\}$, where A_{ij}= +1 if there is a link positive connection between v_i and v_j where $i,j \in \{1,2,\ldots, N\}$, A_{ij}= -1 if there is a negative connection between v_i and v_j, while A_{ij}= 0 otherwise.Therefore, the graph is considered undirected and unweighted; each node has some connections to other nodes, and this number of connections is the degree (deg) of the node. The adjacency matrix contains all the important information about the graph. Each row and column are indexed by a node's number, and all elements on the main diagonal in the adjacency matrix are zero as there are no connections between a node and itself [25]. Figure- 1 illustrates a graph partitioned into three communities. It also displays the matrix representation of the graph. The objective of community detection is to partition the graph, or equivalently, into a set of K clusters or communities $C = \{C_1, C_2, \ldots, C_k\}$. The number of nodes in cluster can be denoted $|C_k|$.



0	1	1	0	0	0	0	0	0
1	0	1	-1	0	0	0	0	0
1	1	0	0	0	0	-1	0	0
0	-1		0	1	1	0	0	0
0	0	0	1	0	1	0	0	0
0	0	0	1	1	0	0	-1	0
0	0	-1	0	0	0	0	1	1
0	0	0	0	0	-1	1	0	1
0	0	0	0	0	0	1	1	0

Figure 1- Graph and adjacency matrix representation of a network which consists of three communities in different colours. The Adjacency matrix has three communities in different colours, with a value of 0 if no connection exists between two nodes, 1 if there is a connection, -1 if there is a negative connection, as shown in red colour.

The problem can be considered an optimization problem that requires the optimization of a given quality measure. Therefore, there is a need for high quality algorithms, such as Evolutionary Algorithms (EA) and Swarm Intelligence, for solving this kind of optimization problem[21-23]. Thus, community detection problem could be formulated as an optimization problem, in both Single and Multi-Objectives optimization problem types. The single objective-based community detection algorithms have already achieved great theoretical and application successes[26-29]; however, they are also still associated with certain issues, such as trying to optimize only one criterion which leaves the solution applicable to a specific community structure. Thus, there is often a case of fundamental discrepancy of different algorithms performing differently on the same network. Furthermore, the single-objective optimization algorithms may not perform well when the optimization criteria are not appropriate. This is typified by the resolution limit that exists in the modularity; modularity optimization may not identify modules that are smaller than a scale even when the modules are well defined [30]. Some other single objective algorithms have similar resolution limits [31]. Additionally, many single-objective algorithms demand pre-information regarding the number of communities and this is not usually known for real networks[32-34]. To address the problems of single-objective community detection algorithms, a natural solution could be to model community detection as a multi-objective optimization problem, implying that we have to simultaneously optimize several objective functions to arrive at a comprehensive and more accurate community structure. Multi-objective approaches, in comparison to the single objective algorithms for community detection, have notable advantages. For instance, community detection with multiple criteria is more consistent with human intuition; secondly, multi-objective frameworks can find the optimal solutions that correspond to the trade-offs between the different objectives (this is in addition to the optimal solution found by the single objective frameworks). Lastly, researchers are now beginning to note that enumerating the modules in a network is a trade-off among multi-objectives. Fortunato *et al.* [26] noted that finding the maximum modularity implies searching for the ideal balance between the modular number and the value of each term[30, 33-37].

The terminologies used in this paper are summarized in Table- 1 below.

Table 1- The nomenclature and abbreviations that used in this study.

Term	Meaning
N	The number of nodes in the network
A	The adjacency matrix
n	The total number of nodes in the cluster
K	The total number of clusters in a solution or partition.
$\mathcal{C}, \mathcal{C}^*$	The detected and correct partitions
$\deg(v, C_K)$	The total number of the connections of a node v in the cluster k
$Deg(N)$	The total number of connections in the network
$Deg(C_K, N)$	The total number of connections in a cluster
Deg_{in}	The total number of internal-connections in a cluster
Deg_{in}^+	The total number of internal-positive connections in a cluster
Deg_{in}^-	The total number of internal-negative connections in a cluster
Deg_{out}	The total number of external-connections in a cluster
Deg_{out}^+	The total number of external- positive connections in a cluster
Deg_{out}^-	The total number of external-negative connections in a cluster
n_s	The number of strong nodes in a cluster
n_w	The number of weak nodes in a cluster
$\lvert C_k \rvert$	The total number of nodes in a cluster

A. Community Detection: Historical Overview

We now present a synthesis of the relevant background material. Doreian and Mrvar [38] presented the earliest and well-known signed graphs partitioning methods. This study, handled the problem of graph partitioning to find the minimum frustration for the positive and negative connections.

First, their algorithm generates random Partition $\mathcal{C}$ in the given network based on a pre-determined number of clusters K (meaning that it explores only a subset of solutions in $\Omega_K \subset \Omega$ of only K

clusters). Then, the suggested method strives to evaluate the frustration of the solution . In an attempt to minimize the frustration value, the local relocation algorithm iteratively tries to move a vertex from one community to a neighboring community or to exchange the community of two vertices belonging to two different communities. The model proposed by Doreian and Mrvar is given as follows:

$$CDN(\mathcal{C}) = minFS(\mathcal{C}) \tag{1}$$

$$S.T::FS(\mathcal{C}) = \sum_{C_k \in \mathcal{C}} a\mathrm{Deg}_{\mathrm{in}}^{-}(C_k) + (1-a)\mathrm{Deg}_{\mathrm{out}}^{+}(C_k) \tag{2}$$

where α represents a weight in the range [0,1]; if α = 0.5, then the positive and negative connections have equal value. Shi *et al*. proposed an evolutionary algorithm for handling multi-objective version of the community detection problem, which is called "Multi-Objective Community Detection" (MOCD) [39]. In their study, the main aim was to maximize the modularity; however, they handled each term in the Q equation as a separate objective. The first objective was to maximize the first term (i.e., intera-connections), which was calculated using equation (3), while the second objective was to minimize the second term (i.e., inter-connections), which was calculated using equation (4), as follows:

$$Intra(\mathcal{C}) = 1 - \sum_{v \in \mathcal{C}} \frac{|Deg_{in}(C_k)|}{Deg(N)} \tag{3}$$

$$Inter(\mathcal{C}) = \sum_{c \in \mathcal{C}} \left(\frac{\sum_{v \in c} \deg(v)}{2Deg(N)} \right)^2 \tag{4}$$

$$\therefore Q(\mathcal{C}) = 1 - Intra(\mathcal{C}) - Inter(\mathcal{C}) \tag{5}$$

Based on the work of Shi *et al*. [39], a recent study by Wu and Pan [40] was published; however, they have implemented the Memetic Algorithm (MA) for identifying multiple community structures.
The study by Attea *et al*. [19] focused on the community structure in both weak and strong connections. They proposed a novel model for community detection based on weak and strong connections in signed networks. This model was evaluated in terms of its performance against the other existing methods. The study introduced a novel multi-objective model with anti-frustration heuristic operator for signed community detection, which is formulated in Eq. (6). The experiments showed that the proposed model performed better than the other models; it introduced an anti-frustration heuristic operator which was also found to have no detrimental effects on the robustness of the detection models, as the proposed model exhibited a high level of reliability.

$$\min MOSCD(\mathcal{C}) = \left[F_{\overline{Intra}}(\mathcal{C}), F_{inter}(\mathcal{C})\right]^T \tag{6}$$

where $F_{\overline{intra}}(\mathcal{C})$ and $F_{inter}(\mathcal{C})$ represent the score of internal and external connections, respectively, which can be calculated as follows:

$$F_{\overline{intra}}(\mathcal{C}) = \sum\nolimits_{\forall C_k \in C} \frac{\sum_{\forall i \in C_k}\left(Deg_{in}(C_k) - Deg_{in}^{-}(C_k)\right) + n_s(C_k)}{|C_k|} \tag{7}$$

$$F_{inter}(\mathcal{C}) = \sum_{\forall C_k \in C} \sum_{\forall i \in C_k} \frac{Deg_{out}(C_k) - Deg_{out}^{-}(C_k)}{Deg(N)} + n_w(C_k) \tag{8}$$

Based on definition provided by Huang *et al*. [10] of the structural similarity between vertices in a given graph, a study was proposed by Liu *et al*. [41] for handling the problem of community detection with multiple objectives. The key point of their method is the structural similarities between two neighbouring nodes in undirected graphs or networks. The performance of the multi-objective maximization model proposed by Liu *et al*. [41] was compared against the FEC proposed by Yang, Cheung, and Liu [42], and the extension provided by Blondel *et al*. [43]. The results showed higher effectiveness of the model presented by Liu et al. [41] compared to the other models. The maximization multi-objectives problem proposed by Liu *et al*. [41] is given as follows:

$$F(\mathcal{C}) = min\left(F_{in}^{+}(\mathcal{C}), F_{in}^{+}(\mathcal{C})\right) \quad (9)$$

$$Max\ F_{in}^{+}(\mathcal{C}) = \frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{u,v\in C_i\wedge(u,v)\in E}\max(S_{sig}(u,v),0)}{\sum_{u,v\in C_i\wedge(u,v)\in E}\max(S_{sig}(u,v),0)+\sum_{u\in C_i\wedge v\in C_j\wedge i\neq j(u,v)\in E}\max(S_{sig}(u,v),0)} \quad (10)$$

$$Max\ F_{out}^{-}(\mathcal{C}) = \frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{u\in C_i\wedge v\in C_j\wedge i\neq j(u,v)\in E}\min\left(S_{sig}(u,v),0\right)}{\sum_{u,v\in C_i\wedge(u,v)\in E}\min\left(S_{sig}(u,v),0\right)+\sum_{u\in C_i\wedge v\in C_j\wedge i\neq j(u,v)\in E}\min\left(S_{sig}(u,v),0\right)} \quad (11)$$

S.T::

$$S_{sig}(u,v) = \frac{\sum_{x\in\Gamma(u)\cap\Gamma(v)}\psi(x)}{\sqrt{\sum_{x\in\Gamma(u)}w^2(u,x)}\ .\sqrt{\sum_{x\in\Gamma(v)}w^2(v,x)}.} \quad (12)$$

$$\psi(x) = \begin{cases}0 & if w(u,x)<0\ and\ w(v,x)<0\\ w(u,x).w(v,x) & otherwise\end{cases} \quad (13)$$

$$\Gamma(v) = \{v\in V|(y,v)\in E\}\cup\{y\} \quad (14)$$

Amelio and Pizzuti [44] focused on the detection of community in signed networks using multi-objectives optimization models, by proposing an alternate optimization framework that relies on the maximization of signed modularity introduced by [45] and frustration minimization as define by Doreian and Mrvar [38]. The proposed model emphasized more on identifying the partitioning solutions $\mathcal{C}\in\Omega$ with low frustration, which was proposed by Doreian and Mrvar [38], shown in Eq.(2), and the high modularity structures which will be explained in Section B, Eq.(22), in owing to the importance of these measures in exceeding “the limits of random topological structures and erroneous community-assignment of positive and negative relations”.

Another work by Amelio and Pizzuti [46] is an extension of their initial work; this new work aimed at improving the final solutions achieved by their model in terms of its signed modularity. This improvement involves the movement of the positive inter-links from their communities to the adjacent communities while sustaining the increase in their Qs value. From the experimental and simulation studies on real life networks, the proposed model was found more effective than the state-of-the-art approaches, including those proposed by [20] and [41]. Recently, several studies were published based the work of Amelio and Pizzuti, such as those of Sani *et al*. [47], and Li *et al* [48].

Attea *et al*. [49] focused on CD problem reformulation as a MOO model for simultaneous detection of intra- and inter-community structures; a heuristic perturbation operator was also suggested for emphasizing the detection of the intra- and inter-community connections in order to establish a positive relation with the MOO model, which is given as follows:

$$MOCD(\mathcal{C}) = min\left(\Phi_1(\mathcal{C}),\Phi_2(\mathcal{C})\right) \quad (15)$$

where

$$Min\Phi_1(\mathcal{C}) = n^2 - \sum_{k=1}^{K}\frac{Deg_{in}(C_k)+n_S(C_k)}{|C_k|} \quad (16)$$

$$Min\Phi_2(\mathcal{C}) = \sum_{k=1}^{K}\frac{Deg_{in}(C_k)+n_w(C_k)}{\sum_{v\in C_k}\deg(v,C_k)} \quad (17)$$

The proposed community detection model also adopted the so-called MOEA/D and the perturbation operator to facilitate the identification of the overlapped community with complex networks. The proposed model was evaluated in terms of its performance by comparison against three current multi-objectives optimization frameworks. From the evaluation results, the proposed model performed better in terms of effectiveness in community detection problem compared to the other models. Several recent research papers were published based on the above explained work by Attea *et al*. [49], such as those of Abdullatiff *et al*. [50], Attea *et al*. [51], and Attea *et al*. [52].

Gong *et al*.[20] suggested a swarm-based multi-objective algorithm for handling the problem of the complex network clustering. The suggested algorithm was called "Discrete Particle Swarm Optimization Based on Decomposition" (MODPSO). The multi-objectives optimization problem in this study was mathematically formulated as follows:

$$MOOP(\mathcal{C}) = min\big(\Phi_1(\mathcal{C}), \Phi_2(\mathcal{C})\big) \tag{18}$$

where

$$Min\Phi_1(\mathcal{C}) = KKM(\mathcal{C}) = 2(n-k) - \sum_{k=1}^{K} \frac{Deg_{in}(C_k)}{|C_k|} \tag{19}$$

$$Min\ \Phi_2(\mathcal{C}) = RC(\mathcal{C}) = \sum_{k=1}^{K} \frac{Deg_{out}(C_k)}{|C_k|} \tag{20}$$

where KKM and RC represent the kernels k-means and ratio cut, respectively. Recently, several CD models were published based on [20], such as those of Rahim *et al*. [53], Tian *et al*. [54], and Zhang *et al*.[55].

A "Mixed Representation-Based Multi-Objective Evolutionary Algorithm (MR-MOEA)" model was proposed by Zhang *et al*. [56]. It was proposed for the detection of overlapping in communities. The model has a mixed individual representation for rapid encoding and decoding of overlapping communities. This mixed representation consists of candidate overlapping and no overlapping node-based representations; different individual updating strategies were also proposed for the overlapping and non-overlapping nodes. The mathematical formulation of the optimization problem was similar to that used in Gong *et al*. [20].

Another study [57] presented a new maximal clique-based MOEA, called MCMOEA, for the detection of overlapping communities. In the MCMOEA, a maximal-clique graph was introduced using a set of maximal cliques as nodes, while the links between the maximal cliques were used as edges. Then, a clique-based representation scheme was proposed based on the maximal-clique graph. Community detection problem formalization as a multi-objective clustering problem was proposed by Zhang *et al*. [56] in complex networks; the study also presented an evolutionary multi-objective technique of discovering community structures, which maximizes the intra-links within each community and minimizes the inter-links between different communities.

B. Evaluation Measurements

The quality of the partitioning obtained can be evaluated for the validation of the performance of the community detection models by using these functions:

1. Normalized mutual information (NMI; similarity measure): a measure of the similarity between the true partitions ($\mathcal{C}^*$) and the detected portion ($\mathcal{C}$) of a network in communities; let (c) be the confusion matrix with element (c_{ij}) being the number of nodes in community (i) of the partition ($\mathcal{C}$) that are also in community (j) of the partition ($\mathcal{C}^*$). The ($NMI(\mathcal{C}^*, \mathcal{C})$) is defined as:

$$NMI(\mathcal{C}^*, \mathcal{C}) = \frac{-2\sum_{i=1}^{K_{\mathcal{C}}}\sum_{j=1}^{K_{\mathcal{C}^*}} c_{ij} log\big(c_{ij} \times n/s_i + s_j\big)}{\sum_{i=1}^{K_{\mathcal{C}}} s_i log(s_i/n) + \sum_{j=1}^{K_{\mathcal{C}^*}} s_j log\big(s_j/n\big)} \tag{21}$$

where (s_i) and (s_j) are the sum of elements of community (i) in $\mathcal{C}$ and community (j) in ($\mathcal{C}^*$). It is important to mention that the value of NMI is ranged from 0 (when $\mathcal{C}^*$ and $\mathcal{C}$ are totally different) to 1 (when $\mathcal{C}^*$ and $\mathcal{C}$ are exactly the same).

2. Modularity: Newman and Girvan [45] evaluated the goodness of a partition as a measure of the quality of a particular division of a network.
Network partitions with high values of Modularity have dense connections within the community and sparse connections with the others. Modularity is defined as:

$$Q(\mathcal{C}) = \sum_{k=1}^{K} \frac{|Deg_{in}(C_k)|}{2Deg(N)} - \left(\frac{\sum_{v \in C_k} deg(v, C_k)}{2Deg(N)}\right)^2 \quad (22)$$

It can be noticed that equation (21) is the difference between the number of intra-connections and the number of inter-connections for all nodes. The value of modularity ranges between (-0.5, 1) where 1 represents an accurate partition structure. The Modularity value is positive if the number of connections within the community is higher than the degree of inter-connections for all networks, whereas it is negative when each node is in one community or sometimes when the network is partitioned into very small communities and 0 when all nodes are in one community.
Gómez, Jensen, and Arenas [45] reformulated the definition of Modularity, as signed Modularity, to capture the strength of the positive and negative node connections in signed networks, while retaining the probabilistic semantics of Q.

$$Q_s(\mathcal{C}) = \frac{1}{2Deg^{+}(N) + 2Deg^{-}(N)} \sum_{v_i \in V} \sum_{v_j \in V} \left[\left(A_{i,j} - \frac{\deg^{+}(v_i, A)}{2Deg^{+}(N)} - \frac{\deg^{-}(v_i, A)}{2Deg^{-}(N)}\right)\right] \delta(C_i, C_j) \quad (23)$$

where δ represents the Kronecker delta function which is equal to 1 when (v_i) and (v_j) are in the same community C_k and 0 otherwise.

C. Multi-Objectives Evolutionary Algorithm with Decomposition (MOEA/D)

Evolutionary algorithms (EAs) are a group of stochastic optimisation techniques that mimic the natural evolution process. The use of EAs, especially evolutionary multi-objective optimisation (EMO) algorithms or MOEAs, in solving MOPs has attracted much interest over the last decade; these algorithms have recorded success in various fields, such as engineering, chemistry, biology, physics, operations research, economics, marketing, and social sciences[58-62]. This success of EAs in different fields is attributed to their two major advantages; (i) they do not need much problem features and can handle large and highly complex solution spaces; (ii) they can approximate the Pareto Front problem as their search is population-based and each solution represents a specific balance between the objectives.
Pareto Front optimization itself is a multi-objective problem. In the absence of any further information provided by the decision maker, EMO algorithms generally focus on two ultimate goals which are to minimize the distance to the Pareto Front, i.e. convergence or proximity, and to maximize the distribution over the Pareto Front, i.e. diversity. Both goals are considered when designing the components of EMO algorithms.
The use of MOEA in solving MOOPs is well documented due to its capability to establish multiple Pareto-optimal solutions in a single iteration. There are three basic goals of an MOEA; these are: (i) finding a set of objective vectors that are close enough to the PF; (ii) finding a set of objective vectors that are well distributed; and (iii) covering the entire PF. Various kinds of MOEAs have been developed to achieve these goals, but more attention has been given to the decomposition-based algorithms since the introduction of MOEA/D algorithm by Zhang and Li [63].
The MOEA/D framework relies on the decomposition approach to decompose MOOPs into several scalar optimization sub-problems which are associated with different weight vectors. A population of solutions is maintained at each generation to preserve the so far best- found solution for each sub-problem. The establishment of the neighbourhood relationships among the sub-problems in MOEA/D is based on the distance between their weight vectors. They ensure that all the neighbouring sub-problems have similar optimal solutions and each sub-problem can be optimized based on the information from a neighbouring sub-problem. Therefore, better solutions can be obtained by applying evolutionary operators to two neighbouring solutions. These sub-problems are simultaneously optimized by evolving this solution population. Various studies reported the good performance of

MOEA/D based approaches in solving multi-objective routing problems[19, 49]; therefore, it has been adopted for handling the problem of community detection when two or more objective functions are required to be maximized or minimized. The main steps of MOEA/D are summarized in the following algorithm.

MOEA/D Algorithm for Community Detection

INPUT

Population size (POP), Number of generations (GEN), Objective dimensions (Dim), Community Detection Model (CDM), Crossover Probability (P_C), Mutation Probability (P_m), Optimization Problem $f(Ob_1, Ob_2, \dots, Ob_z)$

PROCEDURE

1. DECOMPOSE the optimization problem using a scalar approach in (S) subproblems
2. INITIALIZE the weight vectors randomly
3. GENERATE an initial population via Create Population
4. ***While*** $(G < Gen)$
5. DECODE the generate solutions in the population
6. EVALUATE The generated solutions based on CDM
7. SELECT two random solutions as $Parent1$ and $Parent2$
8. CROSSOVER $Parent1$ and $Parent2$
9. MUTATE the best generated solution from the previous step
10. ***Loop (G+1)***
11. SAVE the best generate solutions in Near Pareto Optimal Set
12. RETURN Near Pareto Optimal Set

EMPIRICAL ANALYSIS

A. Datasets Description

In this section, we will describe the three well-known datasets of real-world networks with known correct partitioning:

- Zachary's karate network is the first famous network for CD algorithms; it is commonly known as Zachary's karate club network [64] with 34 members. It is partitioned into 2 communities, one with 16 members surrounding node 1 and the second with 18 members surrounding node 34, making up 78 relations.
- The Bottlenose Dolphins network [65] is the second network; it has a population of 62 bottlenose dolphins and 159 relations.
- The American football game is the third network; it was put together by Girvan and Newman [14] with 115 teams competing against each other in championship games. In this network, five nodes, namely 29, 37, 43, 91, and 111, have no positive intra links.

Signed complex networks are basically founded on the decomposition of links into positive and negative ones and this is important for surpassing the limits of unsigned networks. The performance of community detection models can be affected by the lack of correct and enough connection types among the nodes of an unsigned/signed network.

B. Discussion and Analysis

As stated previously, the main contribution of this study is to analyze the two main evaluation methods, i.e. NMI and Q, and the influence of each one on the other. In other words, the study seeks to answer the following question: Does the rise in Modularity imply a rise in NMI? or vice-versa? Although Pizzuti showed that the maximum modularity does not correspond to the correct network partition in an unsigned network [26], most of the proposed algorithms in the literature have been validated based on Modularity and NMI. Modularity has been used as an internal quality function which assigns a value for the partition, while NMI has been used as an external evaluation measure. In this section, a comparison among several multi-objective algorithms for both signed and unsigned networks is presented. The methods proposed for handling the unsinged networks are: MOCD [39], MODPSO [20], MOGA-Net [66], and MOO [49]. The methods proposed for handling the signed networks are: MOSCD [19], SN-MOGA 2013 [40] and 2016 [42], and MEAs-SN [41]. It is worth to mention that we have implemented all models and algorithms using MATLAB 2018b on a PC machine with Windows 10/64 bit. The settings for each parameter were as follows: The size of the population is 100, the number of generations is 100, the value of mutation probability P_m = 0.2, and the crossover probability C_p = 0.8. Although these methods were implemented to find the best partitions in several well-known datasets, the comparison in this study is performed based on three datasets which have been described in the previous section. Tables- 2 and 3 present the Modularity and NMI results for the above-mentioned algorithms.

Table 2- Maximum and average of NMI and Q for testing models. $\mathrm{NMI_{max}}$ is calculated from the maximum NMI between the true partition and all partitions forming the Pareto fronts of twenty runs of the algorithm. $\mathrm{NMI_{avg}}$ is the average of the twenty runs of the maximum NMI between the correct partition and partitions in a Pareto front from one run. $\mathrm{Q_{max}}$ and $\mathrm{Q_{avg}}$ are the maximum and average modularity values over the twenty runs. The testing models are evaluated on three datasets which are the Zachary's karate club, the Bottlenose Dolphins, and the American football network in unsigned networks.

Dataset	Method	$\mathrm{NMI_{max}}$	$\mathrm{NMI_{avg}}$	$\mathrm{Q_{max}}$	$\mathrm{Q_{avg}}$
Karate	MOCD	0.8372	0.8370	0.4087	0.3952
	MODPSO	0.8372	0.8371	0.4188	0.4092
	MOGA-Net	0.8372	0.8065	0.4018	0.3832
	MOO	0.8372	0.8371	0.4188	0.4092
Dolphin	MOCD	1	0.9532	0.4678	0.4578
	MODPSO	1	0.9778	0.5199	0.4800
	MOGA-Net	0.8888	0.8125	0.4675	0.4440
	MOO	0.9033	0.8945	0.5122	0.4500
Football	MOCD	0.7224	0.7029	0.4356	0.4104
	MODPSO	0.7814	0.7291	0.4666	0.4212
	MOGA-Net	0.6207	0.5433	0.4206	0.3898
	MOO	0.8294	0.7985	0.4351	0.3998

Table 3- Maximum and average of NMI and Q for testing models. $\mathrm{NMI_{max}}$ is calculated from the maximum NMI between the true partition and all partitions forming the Pareto fronts of twenty runs of the algorithm. $\mathrm{NMI_{avg}}$ is the average of the twenty runs of the maximum NMI between the correct partition and partitions in a Pareto front from one run. $\mathrm{Q_{max}}$ and $\mathrm{Q_{avg}}$ are the maximum and average modularity values over the twenty runs. The testing models are evaluated on three datasets which are the Zachary's karate club, the Bottlenose Dolphins, and the American football network in signed networks.

Dataset	Method	$\mathrm{NMI_{max}}$	$\mathrm{NMI_{avg}}$	$\mathrm{Q_{max}}$	$\mathrm{Q_{avg}}$
Karate	MOSCD	0.8372	0.8372	0.3462	0.3462
	SN-MOGA \ 2013	0.8372	0.8166	0.3462	0.3224
	MEAs-SN	0.7329	0.6832	0.3526	0.3105
	SN-MOGA* \ 2016	0.8372	0.8276	0.3462	0.3312
Dolphin	MOSCD	1	1	0.1038	0.1038
	SN-MOGA \ 2013	0.9230	0.9230	0.1006	0.1006
	MEAs-SN	0.8042	0.8017	0.0912	0.0854
	SN-MOGA* \ 2016	1	1	0.1038	0.1038
Football	MOSCD	0.8437	0.8437	0.1077	0.1077
	SN-MOGA \ 2013	0.8028	0.7865	0.0604	0.0578
	MEAs-SN	0.7733	0.7587	0.0677	0.0567
	SN-MOGA* \ 2016	0.8215	0.8158	0.0595	0.0510

Tables- 2 and 3 show the maximum value of $\mathrm{NMI}_{\mathrm{max}}$ (calculated from the maximum NMI between the true partition and all partitions forming the Pareto fronts of twenty runs of the algorithm) and $\mathrm{NMI}_{\mathrm{avg}}$ (average of the twenty runs of the maximum NMI between the correct partition and partitions in a Pareto front from one run). $\mathrm{Q}_{\mathrm{max}}$ and $\mathrm{Q}_{\mathrm{avg}}$ are the maximum and average modularity values over the twenty runs. The testing models with higher NMI values failed to attain better Modularity, meaning that NMI is not a good proof of goodness of the predicted solutions in terms of partitioning in internal or external connections, as obtainable with Modularity. Contrarily, the methods with good Modularity exhibited unstable NMI results.

The signed version of Zachary's Karate network has only nodes 3 and 10 with equal numbers of positive intra links and negative inter ties, while the rest of the nodes have more positive intra links compared to negative inter links. Node 3 has 5 intra links and 5 inter links, while node 10 has only one intra link and one inter link. The models MOSCD, SN-MOGA \2013, and SN-MOGA* \2016 has equal results in this network, NMI = 0.8372, and same value of Q = 0.3462, as shown in Figure- 2-b, while MEAs-SN has the worst performance, with NMI = 0.7329 and Q = 0.3526, because of two incorrectly positioned nodes, 10 and 29. Similarly, the unsigned version of the same dataset attained similar results to those of the NMI. This implies that handling Zachary's Karate network with unsinged connections is similar when facing the issue with the node 10 in the singed version.

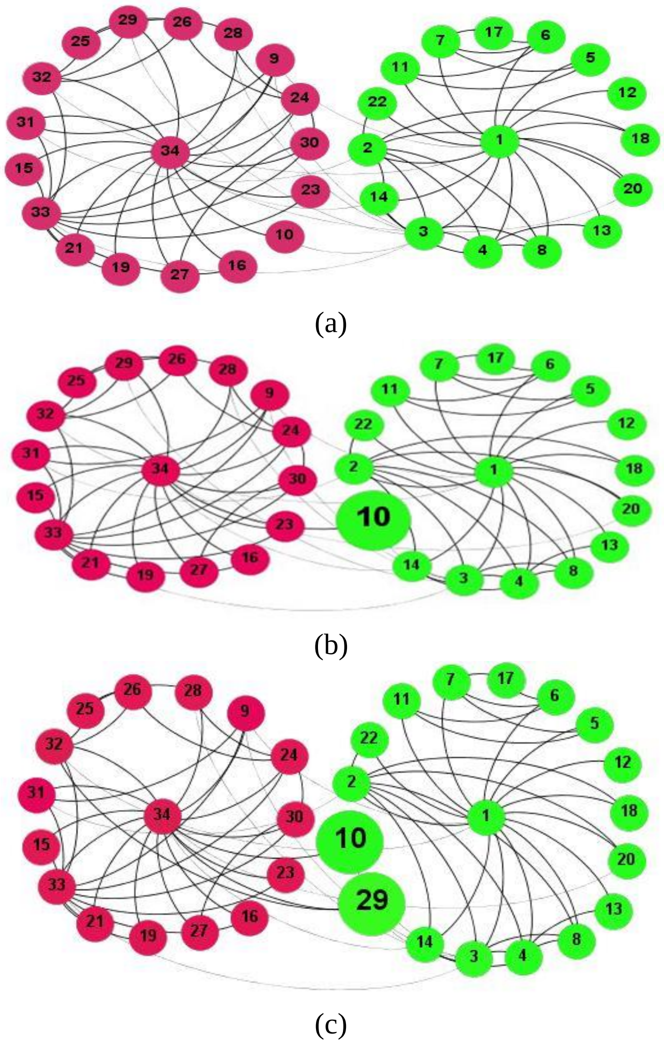
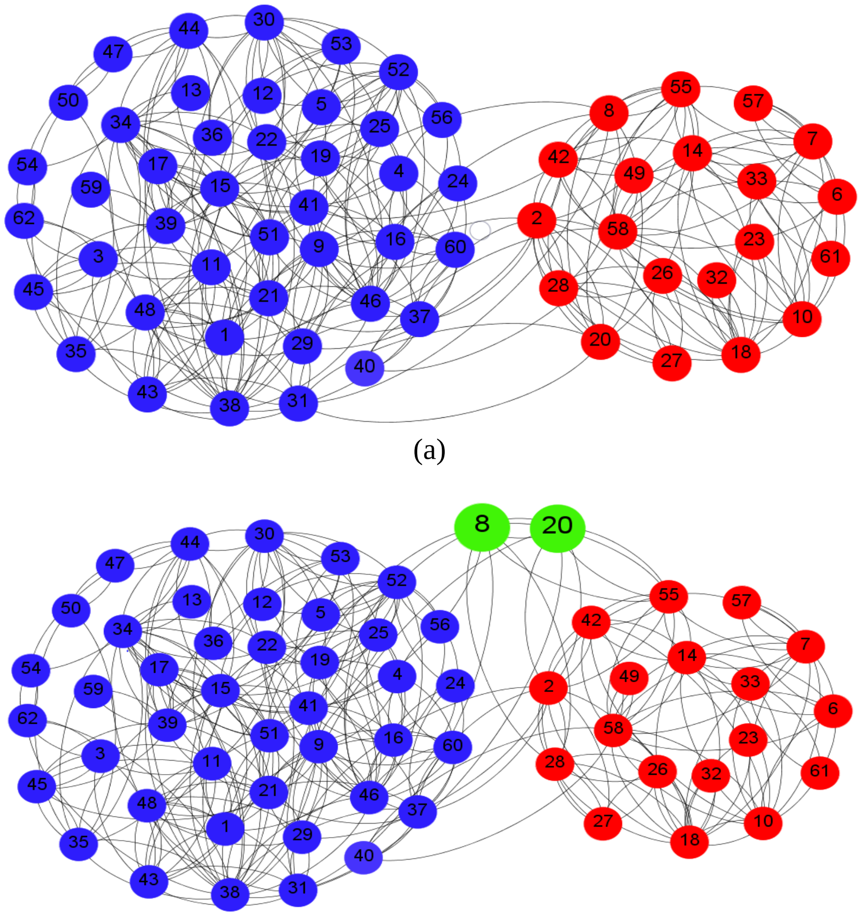


Figure 2 – (a) Zachary's Karate club network (correct partition) with 34 members. It is partitioned into 2 communities, one with 16 members surrounding node 1 and the second is with 18 members surrounding node 34, making up 78 relations. (b) The NMI value is 0.8372 when node 10 is in community one. (c) The NMI value is 0.7329 when nodes 10 and 29 are in community one.

Using the Bottlenose Dolphins network, all the nodes, except node 40, achieved more positive intra links than negative inter links. In the signed version, only MOSCD and SN-MOGA*\2016 attained the best NMI, where NMI = 1, and the same value of Q = 0.1038, while MEAs-SN had the worst performance with NMI = 0.8042 and Q = 0.0912. The SN-MOGA\2013 attained the results of NMI = 0.9230 and Q = 0.1006. On the other hand, the unsigned version of the dataset was also difficult to be handled by the unsinged models. However, MOCD and MODPSO reached the best results, with NMI =1, with different values of Modularity, Q = 0.4674, 0.5199 respectively, while MOO had lower NMI and higher Q than MOCD. Also, the performance of MOGA-Net was lower than that of MOCD in NMI, whereas it was higher in Q. This indicates that there is no direct relationship between Q and NMI in both these networks. i.e. the NMI does not increase with increasing Q, and vice-versa. Figure- 3 illustrates the correct and the best partitions obtained using the best signed models.



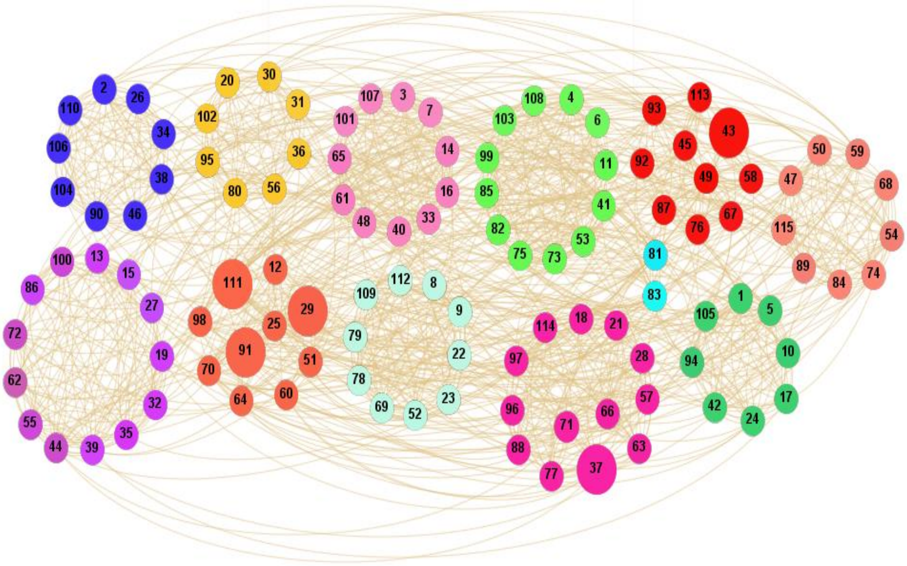
(a)

(b)

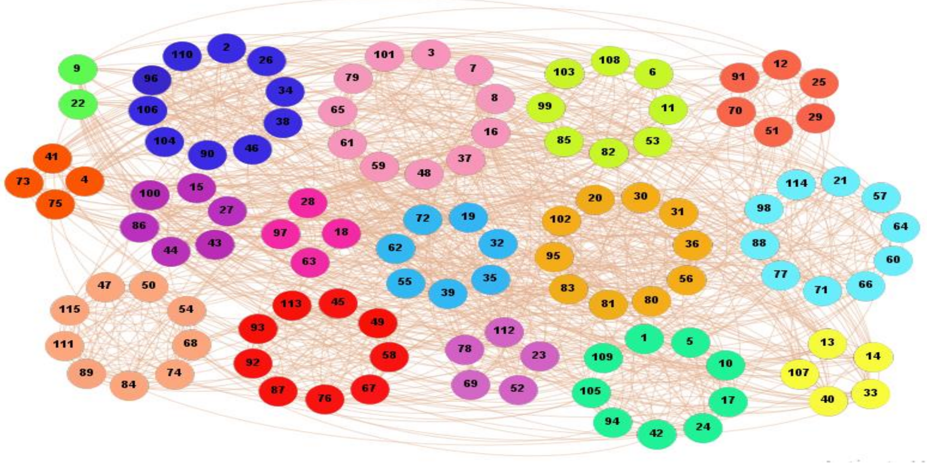
Figure 3 – (a) Bottlenose Dolphins network (correct partition) with a population of 62 bottlenose dolphins and 159 relations. (b) One of the solutions where the network consists of 3 clusters in different colors and the NMI value is 0.9230.

The last dataset tested is that of the American Football. It is more complicated than the previous two datasets, due to the large number of nodes and connections between them. Also, the correct partition consists of 12 clusters, which is a large number as compared to the previous datasets. MOSCD reached the best positions in both signed and unsigned versions, with values of NMI= 0.8437 and Q= 0.1077. For the other models, MEAs-SN had a lower value of NMI (0.7733) than that achieved by SN-MOGA\2013 (0.8028), with high values of Q=0.0677 and Q=0.0604, respectively. In the unsigned version, the MOO reached the best value of NMI=0.8294, with a low value of Q=4351. The MODPSO had a low value of NMI=0.7814, but a high value of Q=0.4666, whereas MOGA-Net had the worst performance of NMI = 0.5433, while the value Q= 0.4206 was not the lowest recorded. These results imply no direct relation or influence between these measures.

Figure- 4 illustrates the correct partition and the best obtained partition using MOCDS model based on the signed version of the dataset.



(a)



(b)

Figure 4 – (a) The American Football Dataset consisting of 12 clusters in different colors, with five nodes having no positive intra links (nodes 29, 37, 43, 91, and 111) as clarified in big circles.
(b) One of the solutions where the network consists of 16 clusters in different colors and the NMI value is 0.803.

The NMI method is important to validate how the generated solution is similar to the correct solution, regardless of the distribution of the clusters, meaning that NMI measures the quality of the solution in terms of its comparison with the correct or the optimal solution. On the other side, the Modularity measures the quality of the solution based on the quantity of the internal connection as compared to the external connections. The main difference between these two methods is clear and it helps to evaluate the community detection algorithms from two different sides - the quality and the similarity to the correct solution. Therefore, a community detection algorithm with good Modularity, i.e. quality, does not mean that it can generate solutions close to the correct solution, and vice-versa. It is worth to mention that NMI is useful for networks with known correct partitions, while Modularity does not require a correct partition. Figures- 5 and 6 illustrate the comparison between these two methods for unsigned and signed networks.

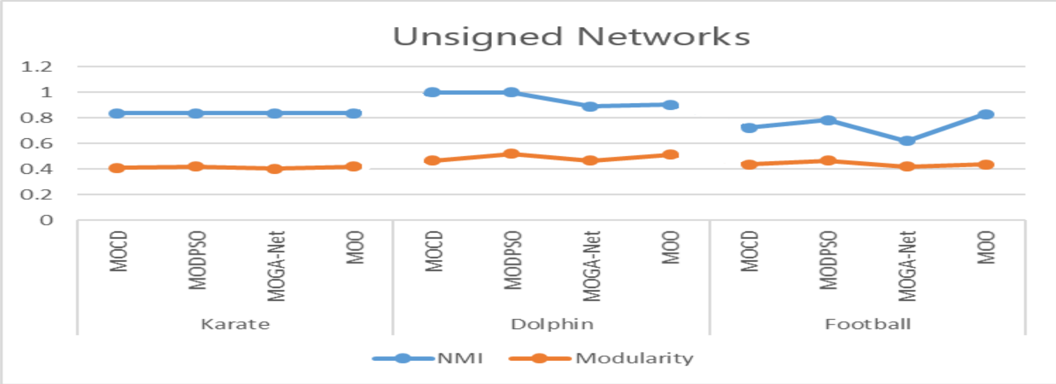


Figure 5 – The performance of NMI against Modularity; the NMI (blue line) and Modularity (orange line) over twenty runs on three datasets which are the Zachary's karate club, the Bottlenose Dolphins, and the American football, for unsigned networks.

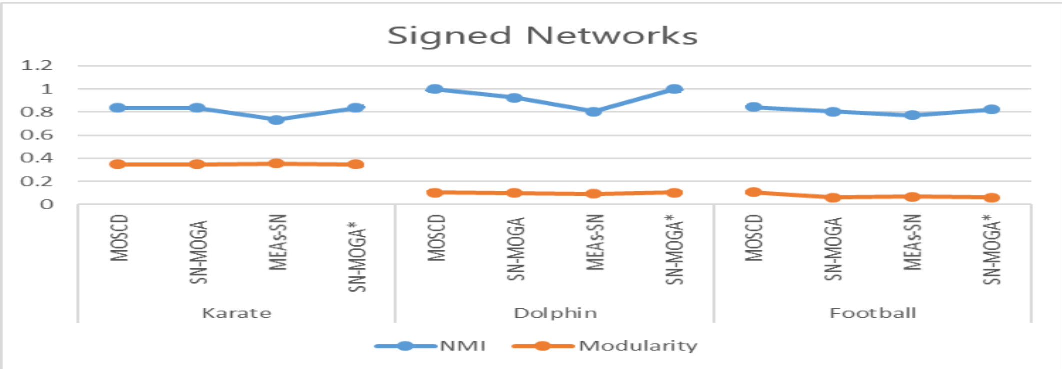


Figure 6 – The performance of NMI against Modularity; the NMI (blue line) and Modularity (orange line) over twenty runs on three datasets, which are the Zachary's karate club, the Bottlenose Dolphins, and the American football, for signed networks.

It can be seen from Figures- 5 and 6 that the NMI (blue line) and Modularity (orange line) were more stable in the signed networks than the unsinged networks. In the singed networks, the quality of solutions is measured based on the type of the connection between the nodes: Positive and Negative Connections. On the other hand, the unsigned networks do not require checking the types of the connections between the nodes. Moreover, all the models have the same performance in terms of NMI when they are applied on Zachary's Karate dataset in unsinged networks. However, they have a slightly difference performance in terms of Modularity. In the second dataset, which is the Dolphin, only the first two models have reached the best performance, i.e. NMI = 1, while MOGA-Net had the worst performance. It can be noticed that MOCD had worse results than MODPSO in terms of modularity; however, it had the best NMI value. Finally, all models have unstable performance when they are implemented on the Football network dataset. MOO is ranked as the best model in terms of NMI and Modularity. These results indicate no direct relation or influence between these measures.

Conclusions

In this paper, we investigated the relation between two measures, NMI and Q, for both signed and unsigned networks, as these two evaluation measures are commonly used for validating the community detection algorithms. The comparison was performed based on several proposed community detection models in the literature. The outcomes showed no direct relation or influence between these measures. For future works, this study opens up avenues for studying and analysing different evaluation methods for different case studies, especially those without known correct partitioning.

References

1. L. C. Freeman,**1996**. "Some antecedents of social network analysis," *Connections*, **19**(1): 39–42.
2. R. Albert, H. Jeong, and A.-L. Barabási, **1999**."Diameter of the World-Wide Web," *Nature*, **401**(6749): 130–131, Sep.
3. D. J. Watts and S. H. Strogatz, **1998**."Collective dynamics of 'small-world' networks," *Nature*, **393**(6684): 440–442, Jun.
4. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, **2000**."The large-scale organization of metabolic networks," *Nature*, **407**(6804): 651–654, Oct.
5. L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, **2000**. "Classes of small-world networks.," *Proc. Natl. Acad. Sci. U. S. A.*, **97**(21): 11149–52, Oct.
6. J. Qiu and Z. Lin, **2011**."A framework for exploring organizational structure in dynamic social networks," *Decis. Support Syst.*, **51**(4): 760–771, Nov.
7. M. P. S. Bhatia and P. Gaur, **2008**."Statistical approach for community mining in social networks," in *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*, 2008, pp. 207–211.
8. P. M. Zadeh and M. S. Moshkenani, **2008**."Mining Social Network for Semantic Advertisement," in *2008 Third International Conference on Convergence and Hybrid Information Technology*, 2008, pp. 611–618.
9. S. Fortunato, **2010**. "Community detection in graphs," *Phys. Rep.*, **486**(3–5): 75–174, Feb.
10. B. Yang, J. Huang, D. Liu, and J. Liu, **2009**. "A Multi-agent based decentralized algorithm for social network community mining," in *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009*.
11. B. Furht, **2010.** *Handbook of Social Network Technologies and Applications*. Boston, MA: Springer US.
12. T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin, **2011**."Detecting communities and their evolutions in dynamic social networks - A Bayesian approach," *Mach. Learn*.
13. N. P. Nguyen, T. N. Dinh, Y. Xuan, and M. T. Thai, **2011**. "Adaptive algorithms for detecting community structure in dynamic social networks," in *Proceedings - IEEE INFOCOM*.
14. M. Girvan and M. E. J. Newman, **2002**. "Community structure in social and biological networks," *Proc. Natl. Acad. Sci.*, **99**(12): 7821–7826, Jun.
15. T. Falkowski and M. Spiliopoulou, **2009**."Community Analysis in Dynamic Social Networks," *Thèse*.
16. J. Li, W. K. Cheung, J. Liu, and C. H. Li, **2009**. "On Discovering Community Trends in Social Networks," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pp. 230–237.
17. T. Sharma, A. Charls, and P. K. Singh, **2009**. "Community Mining in Signed Social Networks - An Automated Approach," *Proc. 2009 Int. Conf. Comput. Eng. Appl.*, 2009.
18. I. Guy, M. Jacovi, E. Shahar, N. Meshulam, V. Soroka, and S. Farrell, **2008**. "Harvesting with SONAR - The value of aggregating social network information," in *Conference on Human Factors in Computing Systems - Proceedings*.
19. B. A. Attea, H. M. Rada, M. N. Abbas, and S. Özdemir, **2019**. "A new evolutionary multi-objective community mining algorithm for signed networks," *Appl. Soft Comput.*, **85**: 105817, Dec.
20. M. Gong, Q. Cai, X. Chen, and L. Ma, **2014**. "Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition," *IEEE Trans. Evol. Comput.*, 2014.

21. X. S. Yang, S. Deb, Y.-X. Zhao, S. Fong, and X. He, **2017**.“Swarm intelligence: past, present and future,” *Soft Computing*.

22. S. Q. Salih and A. A. Alsewari, **2020**.“A new algorithm for normal and large-scale optimization problems: Nomadic People Optimizer,” *Neural Comput. Appl.*, **32**(14): 10359–10386, Jul.

23. P. Agarwal and S. Mehta, **2014**. “Nature-inspired algorithms: state-of-art, problems and prospects,” *Int. J. Comput. Appl.*, **1009**(14): 14–21.

24. J. Ma, J. Liu, W. Ma, M. Gong, and L. Jiao, **2014**. “Decomposition-Based Multiobjective Evolutionary Algorithm for Community Detection in Dynamic Social Networks,” *Sci. World J.*, **2014**: 1–22.

25. H. M. Rada and B. A. Attea, **2017**.“Understanding the Role of Positive and Negative Relations for Community Detection Problem in Signed Networks: A New Perspective,” *Iraqi J. Sci.*, vol. **58**(4B), Nov.

26. C. Pizzuti, **2008**.“GA-Net: A genetic algorithm for community detection in social networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

27. A. Lancichinetti, S. Fortunato, and J. Kertész, **2009**. “Detecting the overlapping and hierarchical community structure in complex networks,” *New J. Phys.*, **11**(3): 033015, Mar.

28. M. Tasgin and H. Bingol, **2006**.“Community Detection in Complex Networks using Genetic Algorithm,” Apr.

29. C. Shi, Y. Wang, B. Wu, and C. Zhong, **2009**. “A New Genetic Algorithm for Community Detection,” in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 2009, pp. 1298–1309.

30. S. Fortunato and M. Barthelemy, **2007**.“Resolution limit in community detection,” *Proc. Natl. Acad. Sci.*, **104**(1): 36–41, Jan.

31. J. M. Kumpula, J. Saramäki, K. Kaski, And J. Kertész, **2007**. “Limited Resolution and Multiresolution Methods in Complex Network Community Detection,” *Fluct. Noise Lett.*, **07**(03): L209–L214, Sep.

32. S. Q. Salih, A. A. Alsewari, B. Al-Khateeb, and M. F. Zolkipli, **2018**.“Novel Multi-Swarm Approach for Balancing Exploration and Exploitation in Particle Swarm Optimization,” in *In Proceesdings of 3rd International Conference of Reliable Information and Communication Technology 2018 (IRICT 2018)*, 2018, pp. 196–206.

33. C. Shi, C. Zhong, Z. Yan, Y. Cai, and B. Wu, **2020**.“A multi-objective approach for community detection in complex network,” in *IEEE Congress on Evolutionary Computation*, 2010, pp. 1–8.

34. A. I. Hafez, E. T. Al-Shammari, A. ella Hassanien, and A. A. Fahmy, **2014**. “Genetic Algorithms for Multi-Objective Community Detection in Complex Networks,” 2014, pp. 145–171.

35. U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, **2008**.“On Modularity Clustering,” *IEEE Trans. Knowl. Data Eng.*, **20**(2): 172–188, Feb.

36. B. A. Attea and W. A. Hariz, **2015**. “A Heuristic Multi-objective Community Detection Algorithm for Complex Social Networks,” *Iraqi J. Sci.*, **56**(4): 3533–3545.

37. H. S. Khoder and B. A. Attea, **2016**.“Community Tracking in Time Evolving Networks: An Evolutionary Multi-objective Approach,” *Iraqi J. Sci.*, **57**(4A): 2539–2548.

38. P. Doreian and A. Mrvar,**1996.** “A partitioning approach to structural balance,” *Soc. Networks*, .

39. C. Shi, Z. Yan, Y. Cai, and B. Wu, **2012**. “Multi-objective community detection in complex networks,” *Appl. Soft Comput.*, **12**(2): 850–859, Feb.

40. P. Wu and L. Pan, **2015**. “Multi-objective community detection based on memetic algorithm,” *PLoS One*.

41. Chenlong Liu, Jing Liu, and Zhongzhou Jiang, **2014**. “A Multiobjective Evolutionary Algorithm Based on Similarity for Community Detection From Signed Social Networks,” *IEEE Trans.*

Cybern., **44**(12): 2274–2287, Dec.

42. B. Yang, W. Cheung, and J. Liu, **2007**. "Community Mining from Signed Social Networks," *IEEE Trans. Knowl. Data Eng.*, 19(10): 1333–1348, Oct.

43. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, **2008**. "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, **2008**(10): P10008, Oct.

44. A. Amelio and C. Pizzuti, **2013**. "Community mining in signed networks," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, 2013, pp. 95–99.

45. S. Gómez, P. Jensen, and A. Arenas, **2009**. "Analysis of community structure in networks of correlated data," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys*.

46. A. Amelio and C. Pizzuti, **2016**."An evolutionary and local refinement approach for community detection in signed networks," *Int. J. Artif. Intell. Tools*.

47. N. Shahabi Sani, M. Manthouri, and F. Farivar, **2020**. "A multi-objective ant colony optimization algorithm for community detection in complex networks," *J. Ambient Intell. Humaniz. Comput*.

48. Q. Li, Z. Cao, W. Ding, and Q. Li, **2020**. "A multi-objective adaptive evolutionary algorithm to extract communities in networks," *Swarm Evol. Comput.*, **52**: 100629, Feb.

49. B. A. Attea, W. A. Hariz, and M. F. Abdulhalim, **2016**. "Improving the performance of evolutionary multi-objective co-clustering models for community detection in complex social networks," *Swarm Evol. Comput.*, **26**: 137–156, Feb.

50. A. H. Abdulateef, B. Attea, and A. Rashid, **2019**. "Heuristic modularity for complex identification in protein-protein interaction networks," *Iraqi J. Sci.*, **60**(8): 1846–1859.

51. B. A. Attea and Q. Z. Abdullah, **2018**."Improving the performance of evolutionary-based complex detection models in protein–protein interaction networks," *Soft Comput.*, **22**(11): 3721–3744, Jun.

52. B. A. Attea and H. S. Khoder, **2016**."A new multi-objective evolutionary framework for community mining in dynamic social networks," *Swarm Evol. Comput*.

53. S. Rahimi, A. Abdollahpouri, and P. Moradi, **2018**. "A multi-objective particle swarm optimization algorithm for community detection in complex networks," *Swarm Evol. Comput.*, **39**: 297–309, Apr.

54. Y. Tian, S. Yang, and X. Zhang, **2019**."An Evolutionary Multiobjective Optimization Based Fuzzy Method for Overlapping Community Detection," *IEEE Trans. Fuzzy Syst.*, 2019.

55. X. Zhang, K. Zhou, H. Pan, L. Zhang, X. Zeng, and Y. Jin, **2018**."A Network Reduction-Based Multiobjective Evolutionary Algorithm for Community Detection in Large-Scale Complex Networks," *IEEE Trans. Cybern*.

56. L. Zhang, H. Pan, Y. Su, X. Zhang, and Y. Niu, **2107**."A Mixed Representation-Based Multiobjective Evolutionary Algorithm for Overlapping Community Detection," *IEEE Trans. Cybern.*, **47**(9): 2703–2716, Sep.

57. X. Wen, W.-N. Chen, Y. Lin, T. Gu, H. Zhang, Y. Li, Y. Yin, and J. Zhang, **2016**."A Maximal Clique Based Multiobjective Evolutionary Algorithm for Overlapping Community Detection," *IEEE Trans. Evol. Comput.*, pp. 1–1.

58. J. Xiong, R. Wang, and J. Jiang, **2019**."Weapon Selection and Planning Problems Using MOEA/D with Distance-Based Divided Neighborhoods," *Complexity*, **2019**: 1–18, Nov.

59. A. Trivedi, D. Srinivasan, K. Sanyal, and A. Ghosh, **2016**. "A Survey of Multiobjective Evolutionary Algorithms based on Decomposition," *IEEE Trans. Evol. Comput.*, pp. 1–1.

60. W. Jing, Z. M. Yaseen, S. Shahid, M. K. Saggi, H. Tao, O. Kisi, S. Q. Salih, N. Al-Ansari, and K.-W. Chau, **2019**."Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions," *Eng. Appl. Comput. Fluid Mech.*, **13**(1): 811–823.

61. T. Hai, A. Sharafati, A. Mohammed, S. Q. Salih, R. C. Deo, N. Al-Ansari, and Z. M. Yaseen. **2020**. “Global Solar Radiation Estimation and Climatic Variability Analysis Using Extreme Learning Machine Based Predictive Model,” *IEEE Access*, **8**: 12026–12042.

62. S. Q. Salih, **2019**.“A New Training Method Based on Black Hole Algorithm for Convolutional Neural Network,” *J. Southwest Jiaotong Univ.*, **54**(3): 1–10.

63. Q. Zhang and H. Li, **2007**.“MOEA/D: A multiobjective evolutionary algorithm based on decomposition,” *IEEE Trans. Evol. Comput.*.

64. W. W. Zachary, “An Information Flow Model for Conflict and Fission in Small Groups,” *J. Anthropol. Res.*, **33**(4): 452–473, Dec.

65. D. Lusseau, **2003**.“The emergent properties of a dolphin social network,” *Proc. R. Soc. B Biol. Sci.*

66. C. Pizzuti, **2012**.“A Multiobjective Genetic Algorithm to Find Communities in Complex Networks,” *IEEE Trans. Evol. Comput.*, **16**(3): 418–430, Jun.