International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 7 (2020), pp. 1525-1533
 https://dx.doi.org/10.37624/IJERT/13.7.2020.1525-1533

On The Use of Optimization for Mining Clusters in Networks

Nadia Abd-Alsabour

Cairo University, Egypt.

ORCID: 0000-0003-4080-5615

Abstract

In the real life, networks can abstract various complex systems. Organizing networks' vertices into coherent subgroups (clusters, groups, or communities) is considered one of the essential rules of complicated networks. A crucial feature of complex systems is having a community structure. An efficient methodology for tackling this vital feature is the community detection. The discovering of communities in these real-world complex networks is vital, as it helps gain strategic insights leading to crucial decisions, and realizes and discovers the dynamics of these systems in the real world. Detecting communities in networks has been recently realized as one of the major research areas in various domains such as science, physics, biology, marketing, engineering, ecology, political sciences, and economics. Meanwhile, the significance of optimization and subsequently the importance of the optimization approaches have been recently emphasized. This is because almost all hard applications and real-world ones deal with maximizing or minimizing some quantity to improve some outcome. Although many hard problems can be handled by numerous optimization approaches, there exist many factors for which applying these approaches for dealing with the community detection problem needs much research. This work researches this important research point with the use of community detection optimization approaches each represents a crucial class of optimization algorithms. In addition, the essentials of discovering clusters in networks are detailed.

Keywords: Community detection, Community structure, Networks, Optimization, Community detection approaches.

1. INTRODUCTION

Large productiveness advances in various fields have led to a huge quantity of data which can be modeled as networks having interacting components like proteins or genes. In almost each domain, the expansion of networks is prevailing because of the digital transformation of business, customers, and communities [1]. Also, since online repositories like the user-created (such as blogs) and digital libraries have been very common, the analysis of this networked data has been an expanding crucial research topic whose main point is to discover salient modules among its members [2].

Networks occur in numerous contexts. For example, Facebook is a huge social network where about a billion persons are connected by acquaintance virtually. Another well-known instance is the Internet. Other examples are in computer sciences, political sciences, physics, biology, marketing, engineering, ecology, social sciences, economics, etc. [3] - [4]. They form an effective architecture in order to typify the relations among components comprising various real-world systems. The world-wide-web, nervous systems, ecological systems, biological networks, transportation systems, collaboration networks, power systems, communication networks, disease transmission, semantic Web networks, and social relations are just a few instances which show how networks affect our daily life [5]-[7].

Recently, studying networks has resurged because of the availability and capability of storing and generating data from various systems [8]. It is considered one of the liveliest multidisciplinary domains as they are represented using networks like computer science, sociology, and complex systems [3]. These systems are called complex ones since their multiple interrelated components are able to communicate with one another and also with the environment. In addition, the way these components work is not comprehended fully. Such systems have been increased quickly due to the quick technological advances in the modern digital world [1], [5]. Instances of the complicated systems are WWW consisting of a huge number of interlinked web pages, the telecommunication networks comprising a huge number of mobile phones, the internet consisting of a huge number of interlinked routers, a power grid system is consisting of electric substations linked via transfer cables, and the human brain consisting of thousands of synaptically linked neurons [1]. As all the networks resulted in complex systems (regardless of the variety of their scope, size, nature, and source) following prevalent organizing rules, they resemble fundamentally in the structure. Organizing networks' vertices into coherent subgroups (communities) is considered one of the essential rules of complicated networks.

Usually, a graph represents a networked dataset in which the vertices represent the members (objects being interested) in the given network. These vertices are tied with one another via undirected or directed edges that represent the relations (associations) among these members (an edge links a pair of nodes) [2]-[3]. Graphs have a chief role in complex systems. Indeed, they are the preferred tool for mathematical modeling. They are found naturally in the study of many areas like sociology, biology, linguistics, physics, and computer science. These graphs can reach large sizes. When they have more than a hundred of nodes, it becomes difficult to understand their structure and to view them legibly. The seeking for highly interconnected modules will simplify representing the structure of big graphs. This is important for the end-user because it allows understanding very intuitively the modeled network [9].

A crucial feature of complex systems is having a community (cluster, module, or group) structure. An efficient methodology for tackling this vital feature is community detection [9]. The discovering of communities and analyzing functions in these complex networks is considered one of the most crucial interdisciplinary scientific challenges nowadays [10]. It is vital as it helps gain strategic insights leading to crucial decisions and realize and discover the dynamics of these systems in the real world [1]. The clusters discovering has been widely used in many fields such as understanding their social structure in various firms [2]. Also, recognizing the changes among healthy and unhealthy Governorates in a given interaction network of proteins in human cells [1].

This work researches the use of optimization approaches for tackling clusters discovering in networks. This is because there are numerous reasons for which this usage needs to be investigated, despite these important algorithms have been utilized successfully for years for tackling a variety of complex optimization problems.

To this end, Section 2 addresses the community detection in networks. Optimization and community detection is addressed in Section 3. The fourth section details the experiments carried out. Discussing the obtained results is detailed in the fifth section. The concluding marks and insights for future work are depicted in the last section.

2. THE COMMUNITY DETECTION IN NETWORKS

A module is a group of nodes having a larger chance of being connected to one another than to the nodes of the other clusters [4]. In other words, it densely links a subset of vertices in a complicated network in such a way that the connection denseness within this subset is much larger than that of the other ones. This is because of the functional or the organizational elements within a network like a community of firms having complicated managements within organizational networks, and a community of performers related to joint films within film star networks. In view of the complexity and applicability of the community discovering problem, an efficient approach to tackling it can transform the society and business via extracting unknown insights [1].

The community structure exists in most of the essential networks as their nodes are structured into groups. For instance, a collaboration scientists network at the Santa Fé Institute in New Mexico shown in Figure 1. The nodes are scientists, and the edges link coauthors. Edges are concentrated in the clusters of nodes that represent the researchers cooperating in the same field since research collaborations are so common there. Other examples are neuron networks and protein–protein interaction networks (the communities here are proteins with comparable functions) in ecosystems, websites on the Web graph specialized in comparable subjects, clusters of friends within social networks, etc. [4], [10].

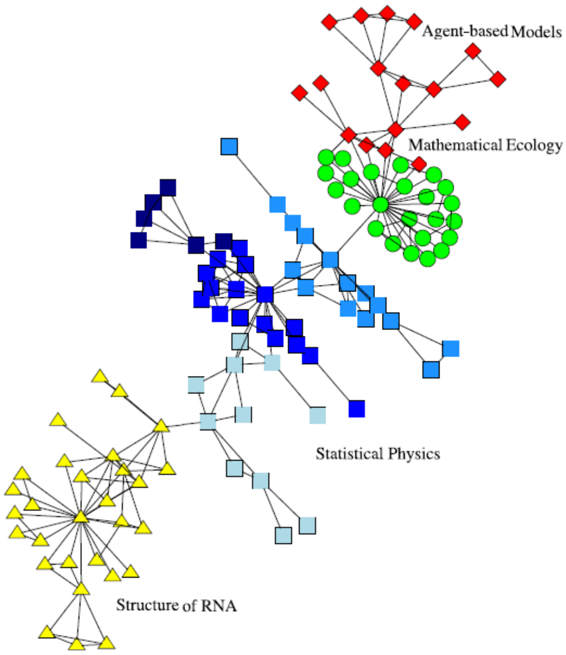

Figure 1. The collaboration network of scientists at the Santa Fé Institute [4]

The community detection in networks (also known as network clustering, or graph clustering but it is different from graph partitioning as the latter requires prior telling the no. of the modules as well as their sizes [11]) is considered one of the most widely known research fields in the modern network science (the science of networks is a new field ranging the social, natural, computer sciences, and engineering) [4]. It is crucial in realizing the functions and structures of networks, [9] and when analyzing the dynamics of numerous networks such as the biological ones. This helps recognize how these vertices are related to each other, and how the structure of these networks affects the way they work. It can provide insight into how the network is structured. It allows concentrating on distinctive areas of the graph. It helps categorize the nodes on the basis of their roles in the communities having them. For instance, we can differentiate between the nodes included in their clusters from the ones at the border of the clusters that may play a crucial part in the dynamics of disseminating procedures throughout the network and in making the communities together [4].

The modules discovering has been employed widely in various domains like biology, computer science, combinatorial optimization, and sociology [5], [8]. Its applications are reached out from subject discovering within collaborative tagging systems to occasion discovering within the social networks as well as modeling huge-scale networks in internet services [2]. Also, in various firms, it helps understand their social structure [2]. In addition, in a given interaction network of proteins in human cells, it helps recognize the changes among healthy and unhealthy Governorates [1]. Further, in social networks, it helps segment members and label nodes as well as recommendations and connection inferences [11]. Furthermore, in the internet, it helps in routing the data efficiently and deciding on the best path for the destination as well as telling how and what viruses are spreading over the internet. In addition, in the World Wide Web, it helps determine the sort of the network, its structure, and its dynamics. Besides, in the sensor networks, it helps enhance numerous administration functions, such as communication functions and energy consuming [11]. Further, in throngs' movement, it helps get insight into the moving patterns such as in Amsterdam after having the recent underground line since it helps in telling how the persons move around Amsterdam [12]. Besides, in payment networks involving millions of customers (nodes) and billions of their transactions (connections), telling the community of a particular customer is essential in various systems such as legal, risk administration, marketing, compliance, etc. [13].

The modules discovering algorithms have been utilized for determining the main terrorists within terrorist networks and developing policies for countering terrorism. These algorithms aid in placing and pricing the products in the promotion. Besides, the functional patterns of the human mind can be comprehended via the community structure of the neuronal relationships [1]. Recent applications include friend recommendation, text processing, and detecting intrusion and images [7].

In addition, these algorithms have been utilized for handling the clustering problem after transforming it into a community detection problem via the utilization of a dissimilarity metric [14]. Similarly, they were utilized by Bracamonte et al. [15] to cluster the tags resulted from the multimedia search. The utilized community detection algorithms do not require prior telling the number of the desired modules or their sizes. They discovered less noisy and more compact modules than that of the obtained by other methods. Besides, Mourchid et al. [16] utilized community detection approaches for dealing with the image segmentation problem and obtained good results.

For that, the modules discovering in networks has got nowadays much of consideration, and it has been advanced as a tool for demonstrating the relations between the function and the structure of the given network (representing the given organization). In the last 20 years, it has been well researched [17].

The aim of discovering community structure in networks is to get an adequate categorization where the links to the nodes in an interesting community are dense, but they are sparse to the nodes outside of this community in the case of non-overlapping groups [3] and also discovering the vertices which belong to multiple groups (known as overlapping vertices) [17] since items have usually distinctive roles and hence are members of multiple modules as in the real world, i.e., nodes are often shared between different groups in the graph representing them [9], [14], [17]. Being a member in numerous communities is highly prevalent in the real-world networks. For instance, a human can belong to numerous concerned communities in a social network, a scientist can participate in numerous communities in collaboration networks, a research paper can include several subjects in citation networks, and proteins in biological networks have various roles in the cell through participating in a few processes [6].

3. OPTIMIZATION AND COMMUNITY DETECTION IN NETWORKS

Optimization approaches play an essential role because a variety of complex problems, particularly real-world ones, are essentially optimization problems. Optimization approaches tackling such problems seek the ideal values for the variables of the given problem to get the ideal result for the given problem. Optimization approaches are categorized as exact (will discover the best solution to the optimization problem being tackled. Applying these approaches are intractable for dealing with various applications because of their running time and/or the optimization problem being handled can't be structured correctly with having all of its characteristics), or approximate (obtaining the ideal solution can not be ensured) [18]-[19].

Despite various techniques have been utilized recently for discovering the densely connected clusters in networks, this crucial problem in various disciplines is still considered an open and a hard problem and has not been tackled satisfactorily since the networks have usually complex nature [14], [20]. More particularly, in huge networks, discovering successfully their module structure is still considered a great data mining problem [21].

In general, this vital problem in numerous fields is an optimization problem for two main reasons. The first, it can be considered as a clustering problem and subsequently is an optimization problem. The second, as discovering the densely connected clusters in networks searches for maximizing the internal connection in contrast to the external ones, is considered an optimization problem. Hence, we have to properly decide on at least one suitable objective function whose optimization leads to getting the correct modules. In addition, as large modularity reflects a good module structure, the utilized optimization approaches get the desired modules via the maximization of the modularity's value over all other feasible modules of the given network, i.e., finding the densely connected clusters in networks is basically turned to be obtaining the modules having the ideal modularity's value. The modularity optimization approaches are one of the key categories of discovering the densely connected clusters in networks [5], [22]-[23].

In the last decades, modern optimization algorithms have been proposed. Nature inspired approaches like evolutionary algorithms and swarm intelligence ones are considered one of the most successful instances of these modern approaches, as they have been widely utilized in dealing with a variety of complex applications. This is a result of their advantages in comparison to the traditional algorithms like requesting less domain-specific data, which is crucial while dealing with a variety of applications for which obtaining such data is intractable. They stimulate their main aspects of nature whose systems execute tasks intelligently as they all seek the optimum which can be constructed as an optimization problem, i.e., the best result is quantified through the use of an objective function. It has been strongly believed that nature gets the best solutions for various difficult problems. Nature's concepts have been researched to introduce approaches simulating such concepts and are able to deal with a variety of optimization problems successfully [18], [24].

Recently, the utilization of swarm intelligence approaches for handling the community detection problem has been increased. This is because of the noticeable benefits of these optimization algorithms, which simulate particular ecological phenomena or biological conduct in nature [7]. There are numerous instances like the particle swarm optimization (PSO) algorithm simulating bird swarm's foraging, and the ant colony optimization algorithm taking inspiration from the ant foraging [25].

In the last decades, multi-objective optimization approaches have been a crucial research area for dealing with a variety of the optimization problems having more than one contradictive objective to be optimized. Since the desired objectives contradict, the utilized multi-objective approaches should result in the set of solutions where enhancing a single objective does not affect the others [26].

Multi-objective metaheuristics, particularly the nature-inspired ones, have been widely used. This is because they need little problem related information (the generality) and are simple to apply. Nevertheless, there exist several significant difficulties which need to be researched [26]. These are:

- their design (particularly how to combine the best members of each sub-population representing the best solutions for a single objective to get the best solutions for the contradictive objectives for the given multi-objective optimization problem. Examples are the vector evaluated genetic, and the Pareto ranking approaches),
- their scalability (their ability to deal with multi-objective problems having more than 3 objectives as these problems became very noticeable), and
- their ability to tackle the expensive objective functions (the most crucial limitation of these approaches is that these algorithms need a great number of evaluations for the fitness functions, also called the evaluation functions, or the cost functions, since they sample the search space to determine the suitable search direction as they are stochastic approaches) [26].

Recently, researchers have tackled the community detection in networks through the use of multi-objective approaches [10], [27], [22]. It is viewed as a modularity-based multi-objective maximization problem to be tackled using an evolutionary approach in Ying et al. [27], and a minimization problem in Mu et al. [10] where both the ratio cut and the negative ratio association are minimized (in order to overcome the resolution limitation problem, the modularity density can be utilized by separating it into the ratio cut and association). Pizzuti [23] tackled it as having 2 distinctive objective functions; the community fitness, and the community score to be handled by a genetic approach, but the obtained results need to be enhanced.

4. THE COMPUTATIONAL EXPERIMENTS

The most common utilization of optimization algorithms in this field has been for optimizing the modularity. Here, we utilize particular optimization algorithms, each of which represents a valuable class of optimization approaches.

4.1 Data

We utilized in the experiments the real-world benchmark Zachary's karate club network [28] (is called the Zachary matrix too [12]). It is a non-trivial benchmark testing network that has been widely utilized in assessing approaches to the community detection problem. The 34 members of this social network are represented in the graph by 34 vertices with 98 edges. A connection between 2 vertices depicts that the 2 members represented by this connection spend together much time than that of the club meetings. This club is partitioned into 2 partitions that are later partitioned into 2 subgroups [29].

4.2 Method

The R language [30] was used for implementing the following systems:

1. The Louvain algorithm as an instance of heuristic and greedy approaches. It is a hierarchical approach, i.e., merges recursively clusters into one vertex and performs the clustering on the condensed graphs based on maximizing the modularity score for every cluster (as it aims to the optimization of the modularity of the whole network). This measures the quality of assigning vertices to clusters, i.e., how much more densely connected the vertices in a cluster are compared to how would be in a random network [31].
2. The maximum modularity algorithm. Exhaustive approaches usually consume time and not practical since they check all the available possible solutions. Subsequently, they don't suit large and complex problems since they do not provide the best solutions within a suitable time. The main idea of the utilized exhaustive approach is to look at all the modules for the largest modularity's value [18], [24], [32].

4.3 Results

The results are shown in Table 1. The modularity's value (ranging from 0 to 1; the bigger, the better as it reveals a better module structure within the given network) obtained by each employed approach is displayed in the second column. The number of the discovered modules by each employed approach is in the third column. The members of each discovered module are in the fourth, fifth, sixth, and seventh columns.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 7 (2020), pp. 1525-1533
© International Research Publication House. https://dx.doi.org/10.37624/IJERT/13.7.2020.1525-1533

TABLE 1 THE EXPERIMENTS' RESULTS

Algorithm name	The mod. value	No. of modules	The 1st module	The 2nd module	The 3rd module	The 4th module
Heuristic alg..	0.42	4	5 6 7 11 17	1 2 3 4 8 10 12 13 14 18 20 22	24 25 26 28 29 32	9 15 16 19 21 23 27 30 31 33 34
Exhaustive alg.	0.42	4	1 2 3 4 8 12 13 14 18 20 22	5 6 7 11 17	9 10 15 16 19 21 23 27 30 31 33 34	24 25 26 28 29 32

The results of the utilized benchmark dataset using the utilized algorithms are shown in the following figures. Other results with the use of the same utilized dataset are shown in Table 2.

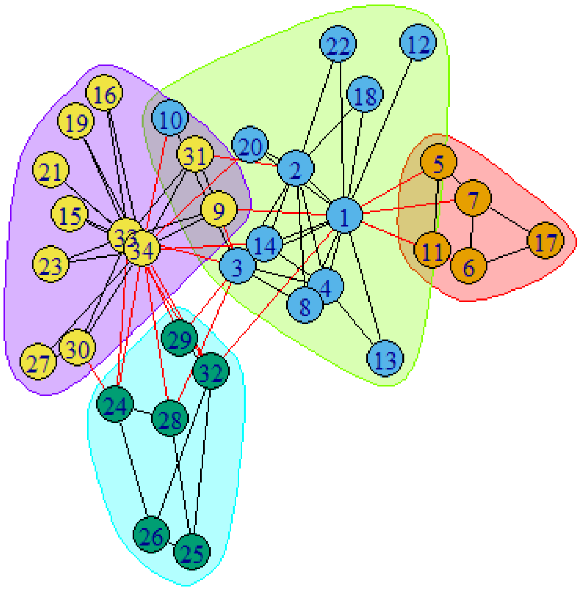

Figure 2 Using the a heuristic algorithm

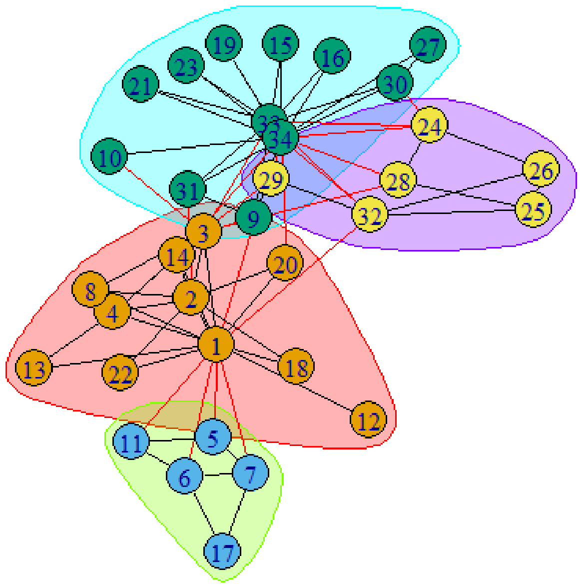

Figure 3 Using an exhaustive algorithm

TABLE 2 OTHER RESULTS FROM LITERATURE

Algorithm name	**The modularity value**
The quantum annealer approach [8]	0.42
A particle swarm optimization algorithm [29]	0.42
An ant colony optimization algorithm (1) [10]	0.42
An ant colony optimization algorithm (2) [10]	0.42
An ant colony optimization algorithm (3) [10]	0.419
An evolutionary algorithm [27]	0.42

5. DISCUSSION

Despite the fact that the exhaustive approaches are complete, utilizing them for dealing with the community detection problem is not preferred. This is because the maximization of modularity, which is the most common optimization factor (discovering the module having the largest modularity) in this field was proved to be an NP problem [33] and applying these algorithms will have large space and time complexity, particularly when dealing with a network having a large size which means it is computationally intractable. As an illustration, they could not deal with the Amazon network comprising of 334863 vertices (representing the products) and 925872 connections [21] although there are much larger networks in the real-life applications such as PayPal networks involving millions of customers with billions of their transactions [13]. Hence, these algorithms are tractable only for dealing with small networks (such as the optimal cluster approach) which is highly restrictive in so many applications.

This has motivated applying heuristic techniques for tackling the modularity optimization in order to get near optimal solutions in a suitable time. However, as it has been investigated recently [19], [34]-[35], many enhancements to these algorithms have been considered when dealing with NP problems in order to get good outcomes in reasonable time and deal with their main shortcomings. This is particularly when the solution space is huge for heuristic approaches to tackle as with the community detection problem since it involves identifying the modules number as well as discovering these modules [8]. Examples of these enhancements are:

- running these algorithms in parallel [19] (parallel computing is the concurrent utilization of multiple computing resources to tackle the given problem that is divided into components to be tackled independently & concurrently either on a group of PC's or a multi-core PC. Hence, handling particular optimization problems in parallel is intractable in numerous cases as this requires huge computational resources [36]),
- involving local search besides the utilized techniques that's not applicable in numerous scenarios as researched in [19] and in [34] (local search approaches give a vital role when being integrated into optimization approaches when dealing with many optimization applications as they prompt providing much better results. However, this is intractable when dealing with many optimization problems. This is because they won't add to the search procedure), and
- integrating them with another algorithm (despite smart integration between optimization approaches together can lead to discovering better results than using single optimization algorithms when dealing with specific applications, proposing an efficient hybrid optimization algorithm is difficult for many reasons [35] and there are numerous factors have to be evaluated to decide whether the use of a hybrid optimization algorithm is suitable for dealing with the application in hand. In addition to deciding on the methodology for integrating distinctive approaches. Besides testing and evaluating the discovered solutions [35]).

As shown in the experiments, there are numerous optimization approaches which can be applied in dealing with discovering the densely connected clusters in networks but these approaches necessitate information about the network (hence they are categorized as global approaches) and utilize the topology information and highly do not consider the node information. This does not help deal with incomplete networks [11], especially when dealing with huge networks such as the payment networks where the customers are described by hundreds of characteristics (such as the member's roles in a social network) having incomplete values. The utilized approach has to utilize both topological and node information as well as be capable of dealing with complicated, incomplete, and various characteristics (categorical and numerical) so as to discover the interested communities correctly [11], [13]. This in turn will increase the difficulty of the utilized approach's task.

In addition, these approaches focus on optimizing the modularity which is considered the most common and prominent category of modules discovery approaches, has been employed successfully for dealing with numerous applications, and is still utilized in so many recent research papers as it has become the main objective function in these modern approaches [12], [14]. However, it is not the ideal measure to be optimized. This is mainly because of its severe problem which is the resolution limit problem (failing to discover small communities within big networks in comparison to the entire network although this is very crucial in various applications such as the biological systems where the relevant clusters are very small and discovering them help realize the etiology of various illnesses and the functional roles of various protein clusters in particular illness. In other words, in some cases, a module can not be identified whether it is a single one or a collection of weakly interlinked smaller modules which seriously affect practical applications). Once this problem was proved, another fact has also been proved which is the modularity based approaches are not able to identify the correct no. of modules, i.e., they do not reflect the structures of complex networks. In addition, various modules can have alike modularity, leading to difficulties in discovering the global optimum. Further, the greatest modularity relies upon the given network's size as well as its no. of communities, i.e., the correct module structure is not necessarily the one with the biggest value of the modularity. Such drawbacks result in difficulties in applying the modularity based approaches to various applications, especially the ones that have huge-scale networks. Despite that numerous adjustments to these approaches have been proposed to deal with networks having self-connections and numerous links, various enhancements for dealing with numerous sorts of networks are still required [12], [21], [37]-[38].

For that, numerous researchers focus on dealing with the main disadvantages of modularity optimization approaches such as the work of Chen et al. [39] which deals with the resolution limit problem. Other recent research goes towards adding an extra computational preprocessing step (the previous difficulties apply here too) considering the node's information to enhance the performance of the utilized community detection approach [11]. In addition, performing an additional division within the big modules discovered by these approaches, but it is hard to get the appropriate stopping criteria [37]. However, improving the performance of an algorithm involves extra computation and can affect other current functions [40], which is unaffordable in various applications.

Hence, in order to discover the correct modules, the utilized approaches do not have to utilize the modularity as their fitness function (the optimization criterion). Therefore, it would be much better to utilize the optimization algorithms efficiently to make them able to provide the best result for the problem being tackled. This should be performed early in the analysis phase when deciding on which optimization approach much suits the given optimization problem, as well as designing correctly its crucial part, which is its objective function. It plays the main role to direct the direction of the search procedure. Hence, when it is designed inappropriately, the optimization approach will not succeed especially if it is a single objective approach (since this type of optimization approaches examine only the objective function, which usually restricts the desired solution to the structural features of the modules [12] in contrast to the multi-objective ones that do not have restrictions [22]) as with the resolution limit problem in Fortunato and Barthelemy [41]. Nevertheless, Zadeh and Kobti [20] designed their proposed knowledge-based approach in such a way that this crucial part is independent of their proposed approach, i.e., the knowledge (which is extracted from the given network and updated on the basis of the current states of this network in every step of this proposed approach) is utilized to direct the search direction and discover the optimal result.

As an illustration of optimizing other factors, the optimization algorithm Infomap [42] optimizes the predicted length of a random walker trajectory instead of the modularity. Another instance is to consider the distances among the vertices and discover the modules with the minimal overall distances, such as the random walk algorithm (having low complexity) and the work of Shao et al. [21] where these authors advanced a novel approach on the basis of investigating the alterations in the spaces among the vertices. In the given network, every vertex interacts with its neighboring vertices resulting in alterations in the spaces among the vertices that will impact these interactions. This prompts a stable distribution of the spaces and subsequently detecting modules having small, arbitrary, or huge size. The work of Sani et al. [22] is a recent example. The authors made use of the concepts of Pareto-archive and Pareto advantages, and two contradictory objective functions of ant colony optimization methods. These functions are the module scoring, and the module fitness measuring the resulted module density whilst minimizing the external links. Their work has better performance and results than numerous community detection approaches. One more recent example is the work of Brummer et al. [12]. Optimizing the Kemeny measure (the average no. of the needed steps to proceed from a randomly selected node to another randomly selected one) can be utilized instead of the modularity, especially when moving from a vertex to another one in a graph (representing the given problem) reveals the movement from a state to another one such as the Markov chain graph. This is proposed to produce good results as this measure has a variety of functions that can be employed in different applications. In addition, minimizing it can be an efficient division of the given graph due to minimizing the no. of the needed transitions to go around.

On the other hand, authors such as Fortunato and Barthelemy [41] and Chen et al. [43] have supposed that utilizing the algorithms depending on a particular criterion like the modularity do not suit the community detection problem. This is because these approaches will necessarily involve searching for the modules having large modularities. Besides, the exhaustive search over all the feasible modules is impractical due to the large time cost (the time complexity of the modularity based optimization algorithms has been proved to be NP complete [21]). In addition, the modularity based optimization algorithms do not often get the best modules when handling real-life applications [41].

This has motivated recently proposing new approaches which are not based on the optimization such as the work of Chen et al. [43] having near-linear time complexity and does not involve a heuristic search. Their simple proposed work utilizes a new similarity involving common neighbors of the 2 contiguous vertices as well as their mutual exclusive degree. Initially, it combines the vertices based on this new similarity to get the first module structure, which then is adjusted based on the detected cores via the novel utilized local denseness. Finally, the discovered modules are expanded in order to get the last modules' structure.

6. CONCLUSIONS AND FUTURE WORK

The success of the optimization approaches depends mainly on the proper design as well as the proper usage of them, i.e., they are utilized for optimizing which factor of the tackled optimization problem. This was researched by this article through different optimization approaches (representing crucial classes of the optimization algorithms) to tackle discovering the densely connected clusters in networks.

Continuing the recent important research directions which have been proposed for improving the approaches tackling the community detection problem, many research issues should be considered in the future. Firstly, the optimization of various factors of the community detection problem should be studied with the utilization of a variety of networks' types and sizes in the future. In addition, in the near future, optimizing the performance of the utilized approaches such as tuning their parameters is considered. Also, other approaches for discovering densely connected clusters in networks should be considered in the future with the use of a variety of datasets. In addition, as the existence of overlapping clusters increases in the real-life applications, the use of overlapped networks should be utilized in the future.

REFERENCES

[1] S. Gupta, and P. Kumar, An overlapping community detection algorithm based on rough clustering of links, Data & Knowledge Engineering, Data & Knowledge Engineering, 125, January 2020, Article 101777, Available at: https://doi.org/10.1016/j.datak.2019.101777, 2020. Last visited on 1-2-2020.

[2] T. Yang, R. Jin, Y. Chi, and S. Zhu, Combining Link and Content for Community Detection: A Discriminative Approach, Proceedings of KDD'09, June 28–July 1, Paris, France, 2009.

[3] X. Ding, J. Zhang, J. Yang, and Y. Shen, An Autonomous Divisive Algorithm for Community Detection Based on Weak Link and Link-Break Strategy, Mathematical Problems in Engineering, 2018, Article ID 2942054, (12 pages), 2018.

[4] S. Fortunato, and D. Hric, Community detection in networks: A user guide, Physics Reports, 659: 1–44, 2016.

[5] X. Zhang, Z. Ma, Z. Zhang, Q. Sun, and J. Yan, A Review of Community Detection Algorithms Based on Modularity Optimization. Journal of Physics: Conference Series, 1069, 012123, 2018.

[6] A. Amelio, and C. Pizzuti, Overlapping Community Discovery Methods: A Survey. In Social Networks: Analysis and Case Studies (p. 105-125). Springer, Vienna, 2014.

[7] Y. Feng, H. Chen, T. Li, and C. Luo, A novel community detection method based on whale optimization algorithm with evolutionary population, Applied Intelligence, 50:2503-2522. Available at: https://doi.org/10.1007/s10489-020-01659-7, and https://link.springer.com/article/10.1007/s10489-020-01659-7, 2020. Last visited on 1-2-2020.

[8] C. F. A. Negre, H. Ushijima-Mwesigwa, and S. M. Mniszewski, Detecting multiple communities using quantum annealing on the DWave system, PLoS ONE, 15(2), e0227538, 2020.

[9] D. Rhouma, L. and B. Romdhane, An efficient algorithm for community mining with overlap in social networks, Expert Systems with Applications, 41(9): 4309–4321, 2014.

[10] C. Mu, J. Zhang, Y. Liu, R. Qu, and T. Huang, Multi-objective ant colony optimization algorithm based on decomposition for community detection in complex networks, Soft Computing, 23:12683–12709, 2019. Available at: https://doi.org/10.1007/s00500-019-03820-y. Last visited on 1-2-2020.

[11] A. Bhih, P. Johnson, and M. Randles, An optimisation tool for robust community detection algorithms using content and topology information, The Journal of Supercomputing, 76:226–254, 2020.

[12] J. Brummer, E. Dugundji, and D. V. Leeuwen, Optimizing Community Detection Using the Kemeny Constant. Available at: https://beta.vu.nl/nl/Images/werkstuk-brummer_tcm235-897735.pdf, 2018. Last visited on 1-2-2020.

[13] C. Zhe, A. Sun, and X. Xiao, Community Detection on Large Complex Attribute Network. In The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '19), August 4-8, 2019, p. 2041–2049, Anchorage, AK, USA, 2019.

[14] L. M. Naeni, H. Craig, R. Berretta, and P. Moscato, A Novel Clustering Methodology Based on Modularity Optimization for Detecting Authorship Affinities in Shakespearean Era Plays, PLoS ONE, 11(8), e0157988, 2016.

[15] T. Bracamonte, A. Hogan, and B. Poblete, Applying community detection methods to cluster tags in multimedia search results, In 2016 IEEE International Symposium on Multimedia (ISM) (p. 467-474), IEEE, 2016.

[16] Y. Mourchid, M. El Hassouni, and H. Cherifi, Image segmentation based on community detection approach, International Journal of Computer Information Systems and Industrial Management Applications, 8: 195–204, 2016.

[17] C. Shi, Y. Cai, D. Fu, Y. Dong, and B. Wu, A link clustering based overlapping community detection algorithm, Data & Knowledge Engineering, 87: 394–404, 2013.

[18] N. Abd-Alsabour, Nature as a Source for Inspiring New Optimization Algorithms, Proceedings of the 9th International Conference on Signal Processing Systems, p. 51-56, ACM, 2017.

[19] N. Abd-Alsabour, Local search for parallel optimization algorithms for high dimensional optimization problems. MATEC Web of Conferences, 210, 04052, 2018.

[20] P. M. Zadeh, and Z. Kobti, A Multi-Population Cultural Algorithm for Community Detection in Social Networks, In The 6th International Conference on Ambient Systems, Networks, and Technologies (ANT 2015), Procedia Computer Science, 52: 342 – 349, 2015.

[21] J. Shao, Z. Han, Q. Yang, and T. Zhou, Community Detection based on Distance Dynamics, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2015 Aug 10 (p. 1075-1084), 2015.

[22] N. S. Sani, M. Manthouri, and F. Farivar, A multi-objective ant colony optimization algorithm for community detection in complex networks, Journal of Ambient Intelligence and Humanized Computing, 11:5–21, 2020.

[23] C. Pizzuti, A multi-objective genetic algorithm for community detection in networks, Proceedings of the 21st IEEE international conference on tools with artificial intelligence, Newark, New Jersey, USA, p. 379–386, 2009.

[24] N. Abd-Alsabour, On tackling real-life optimization problems, The International Journal on Advanced Science, Engineering and Information Technology, 9(2): 640-647, 2019.

[25] R. Liu, J. Liu, and M. He, A multi-objective ant colony optimization with decomposition for community detection in complex networks, Transactions of the Institute of Measurement and Control, 41(9): 2521–2534, 2019.

[26] C. A. C. Coello, S. G. Brambila, J. F. Gamboa, M. G. C. Tapia, and R. H. Gómez, Evolutionary multiobjective optimization: open research areas and some challenges lying ahead, Complex and Intelligent Systems, Complex & Intelligent Systems, 6:221–236, 2020.

[27] W. Ying , H. Jalil, B. Wu, Y. Wu, Z. Ying, Y. Luo,

and Z. Wang, Parallel Conical Area Community Detection Using Evolutionary Multi-Objective Optimization, Processes, 2019, 7, 111. Available at: https://doi.org/10.3390/pr7020111, 2019. Last visited on 1-2-2020.

[28] W. W. Zachary, An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33(4):452-473, 1977.

[29] A. Abdollahpouri, S. Rahimi, S. M. Majd, and C. Salavati, A Modified Particle Swarm Optimization Algorithm for Community Detection in Complex Networks, In: A. Holzinger et al. (Eds.): CD-MAKE 2018, LNCS 11015, p. 11–27, Springer, Cham, 2018.

[30] R: A Language and Environment for Statistical Computing [http://www.R-project.org]. R Foundation for Statistical Computing, Vienna, Austria.

[31] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment, 2008(10), Article ID P10008, 2008.

[32] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, On modularity clustering, IEEE Transactions on Knowledge and Data Engineering, 20(2): 172–188, 2008.

[33] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, Maximizing modularity is hard. Available at: https://arxiv.org/abs/physics/0608255, 2006. Last visited on 1-2-2020.

[34] N. Abd-Alsabour, Investigating the influence of adding local search to search algorithms, Proceedings of the 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), p. 145-150, IEEE, 2017.

[35] N. Abd-Alsabour, Hybrid Metaheuristics for Classification Problems, In: S. Ramakrishnan, (Ed). Pattern Recognition - Analysis and Applications, InTech, 2016.

[36] N. Abd-Alsabour, Parallel Evolutionary Algorithms and High Dimensional Optimization Problems, Journal of computers, 13(11):1265-1271, 2018.

[37] H. Zhang, Modeling the Relationship between Links and Communities for Overlapping Community Detection, Doctoral dissertation, The Chinese University of Hong Kong, Hong Kong, 2018.

[38] B. Tripathi, S. Parthasarathy, H. Sinha, K. Raman, and B. Ravindran, Adapting Community Detection Algorithms for Disease Module Identification in Heterogeneous Biological Networks, Frontiers in Genetics, March 2019,10, Article 164, 2019.

[39] M. Chen, K. Kuzmin, and B. K. Szymanski, Community Detection via Maximization of Modularity and Its Variants, IEEE Transactions on Computation Social Systems,1(1): 46-65, 2014.

[40] W. Stallings, Operating Systems: Internals and Design Principles, 9th Edition, Pearson, 2018.

[41] S. Fortunato, and M. Barthelemy, Resolution limit in community detection, Proceedings of the National Academy of Sciences of the United States of America, 104(1): 36–41, 2007.

[42] M. Rosvall, and C. T. Bergstrom, Maps of information flow reveal community structure in complex networks, Proceedings of the National Academy of Sciences of the United States of America, 105(4): 1118–1123, 2008.

[43] M. Chen, Z. Yang, X. Wen, M. Leng, M. Zhang, and M. Li, Effectively Detecting Communities by Adjusting Initial Structure via Cores, Complexity, 2019, Article ID 9764341, 20 pages, 2019.