A Novel Method for Community Detection Using Multi-Objective Particle Swarm Optimization Algorithm

Elaf abdul-husein[1], H. K. Jabbar[2], Zainab mohammed flayh[3], Maha falih[4] and Jabbar R. Rashed [5]

[1,3]*Ministry of Education, Maysan, Iraq*

[2,5]*University of Misan/ Misan, Electrical Engineering Department, College of Engineering, Iraq*

[4]*Imam Ja'afar Al-Sadiq University, Department of computer Techniques Engineering, Information Technology College, Baghdad, Iraq.*

Abstract

In the last decade, social media has exploded in popularity, with some platforms now having tens of millions of users. Many sorts of data, such as audio, video, and text, may be found in these networks and must be investigated individually. Because to the wide variety of uses for these networks and the impact of past events and societal norms, their designers confront a number of obstacles. The existence of communities inside social networks is one of their most distinguishing characteristics. Tools for "community extraction" make it easier to investigate all components of a network and classify activities that are typical of certain groups. Although it may seem like a simple task, recognizing communities on social media is computationally difficult owing to the unknown number of groups and the varied internal density of communities. The features of a network and the goal of studying it might guide the selection or development of a particular approach for finding communities within that network. There is a distinct grouping of current community detection techniques based on their individual characteristics. Selecting the right algorithm requires knowledge of the proper categorization and features of each category. In this research, we offer an approach to social media community discovery that makes use of a multi-objective particle swarm optimization algorithm. Evaluation findings confirm the superiority of this approach.

Keywords

Social networks, communities, community detection, meta-heuristic algorithms, multi-objective particle swarm optimization algorithm, NMI criteria.

1. Introduction

In social network analysis, community discovery is a crucial step in the process of locating clusters of nodes that exhibit commonalities in terms of some defining trait [1-2]. Real-world applications like social media platforms like Facebook and Twitter, as well as web infrastructure information networks and complicated molecular networks in biological data, make extensive use of this analytic job [3]. Community discovery algorithms have historically ignored the qualities of nodes in favor of analyzing the structural aspect of social networks, or the connections between nodes. While the interactions between nodes are an important part of any social network, most actual networks also give additional information about the nodes themselves. Age, gender, and area of interest are just a few examples of

ICECI-2023: International Conference on Emerging Computational Intelligence, February 11-12, 2023, Aligarh, India,
EMAIL: Hhaider2.allamy@gmail.com (1); Haider.allamy@uomisan.edu.iq (2); Za2013_mcs@yahoo.com (3); maha.falih@sadiq.edu.iq (4) dr.jabar72@uomisan.edu.iq (5);
ORCID: 0000-0001-7647-8252 (A. 1);

CEUR Workshop Proceedings (CEUR-WS.org)

node information that may be readily seen in social network charts and used as part of the available forecasting knowledge. When characteristics are associated with each node in a social network, we call it an attributed social network. Hence, it is possible to enhance the precision of community discovery [5] using a technique that uses both the structure and features of nodes concurrently.

Real-world applications of community detection include categorizing nodes, compressing networks, analyzing similar nodes, and community structures in place of analyzing nodes individually, as well as grouping customers with similar interests on social media for efficient product recommendations and systems. Further applications of community identification include consumer segmentation, recommendation engines, outlining, labeling, analyzing networks based on who influences them, and information distribution. The difficulty of community discovery in social networks remains unsolved despite many attempts over the last several years. To some extent, community detection is an optimization issue. It is crucial to determine the needs of the extraction algorithm after gaining context knowledge, including member attributes, connection kinds, asymmetric relationships, hierarchical or overlapping communities, factor impacts on weighting, and knowing the number of communities in advance. Communities may be shown using a variety of approaches, some of which are context- or purpose-specific. As a result, examining the environment and the efficacy of each technique is essential before settling on a strategy and algorithm for locating preexisting communities [5]. Using a multi-objective particle swarm optimization approach, we provide a best practice for community discovery and social interaction. We introduce the concept of society, provide links to relevant literature, outline the approach, provide the evaluation's findings, and draw a broad conclusion.

2. Definition of community

As a result of the many different approaches and methods that have been created to define communities, this is not always an easy undertaking. Depending on context and purpose, what constitutes a "community" may change. For example, Yang [6] defines a community as a collection of nodes with high connection density among community nodes but low link density overall. Graphs of any directionality may benefit from this definition.

Yet, "a set of nodes that have greater connection density among themselves than other" [7] is how Porter characterizes a community. Clustering methods, Quality function algorithms, Centrality-based community recognition algorithms, Group refining algorithms, and Modular optimization algorithms are all examples of how this notion of "greater density" varies from Yang's.

There are a variety of ways to define a community, but all of them agree that there are fewer connections between communities than there are between individuals inside those communities. Specifically, most definitions demonstrate that overlapping communities have lower edge density in the overlapping section compared to the non-overlapping part. Yet Yang's definition provides a new angle. He demonstrated that the probability of a connection between nodes increases as the number of common communities increases. A node's ability to interact with other nodes increases as the number of communities to which it belongs grows. Due to this, there are more edges in the overlapping region than in the surrounding regions. With social media, for instance, one's network expands as a result of more content sharing.

Previous community definitions suffer from the restriction of creating an abnormal picture of society, which causes algorithms relying on these definitions to incorrectly recognize the overlapping section as a distinct community, or two overlapping communities, as a community.

3. Related Works

In the field of social network analysis, identifying communities is a major challenge. Despite this progress, it is still difficult to identify communities in huge social networks with many characteristics. Community identification approaches that include feature reduction and optimization strategies were developed to meet this difficulty. Several meta-heuristic algorithms have been applied to the problem of feature selection in order to optimize it with reasonable accuracy and in a reasonable amount of time. These include Ant Colony Optimization (ACO), the genetic algorithm (GA), Simulated Annealing (SA), Tabu Search (TS), and particle swarm optimization (PSO). Moreover, hybrid search tactics are used, which include both filtering processes and advanced approaches.

Recent years have seen a rise in the study of community detection across several disciplines. The challenge, however, is that the exact number of communities present in real-world networks is unclear, making this an open issue for all approaches. It's often held that a community's hub has extensive links to its surrounding nodes, and that any two hubs are spread widely apart. Also, within a community, there is more similarity between nodes than there is across communities. As a result, community detection relies heavily on data on both local and global structures.

Uncertain as to the total number of communities, a three-step technique for identifying them using both local and global data is provided in a recent paper [8]. Discovering the hub node, spreading the labels, and merging the communities are all parts of the method. Because of their higher than usual separation, central nodes may be easily spotted. When the degree of similarity between two nodes is at its highest, the nodes are given the same label, and if the increment of the modularity is positive and maximal when the two communities are joined, the two communities are merged into a single one. The three-stage approach has been shown to be effective in detecting communities in social networks via experiments and simulations on both real-world and synthetic networks.

The study of social impact in networks is crucial for comprehending the propagation of behavior. There are a lot of ideas that attempt to make sense of how and why people take up new practices and fashions. According to the conventional wisdom, popular movements are usually driven by a small number of influential people who have special talents that make them particularly effective at persuading others. Often times, these people are knowledgeable, esteemed, and connected. The LCDA community-leader recognition strategy is described in this study [9] as a novel scalable and deterministic way to recognizing communities based on their leader nodes. The method has two stages: first, a leader extraction is performed, and second, a community is identified by node similarity. When compared to the Earth Truth membership community, the algorithm produces satisfactory results.

Online social network mining has grown in importance, and so has the study of community detection. The label propagation algorithm (LPA) is a well-known choice because it is straightforward to grasp and apply, and it only takes linear time to complete its tasks. Nevertheless, one of its key drawbacks is the unreliability of its findings, since it reports unique combinations of communities in each run owing to the random selection involved. This research [10] proposes a novel approach called Balanced Link Density-based Label Propagation (BLDLP), which builds on the LPA technique and takes use of the input network's natural structure (its link density characteristic). In order to fine-tune the BLDLP technique to specific needs, a sensitivity parameter (balance parameter) is used.

The process of identifying communities is crucial to the study of social networks. Several intelligent and meta-heuristic methods have been presented in recent years for detecting communities in complicated social networks. Although the label propagation algorithm (LPA) is very quick at finding patterns in large datasets, its results are too unpredictable to be used in more generalized forms of network research. To further optimize modularity measure, the authors of [11] offer AntLP, a variant of LPA that utilizes similarity indices and ant colony optimization (ACO) to cluster nodes into communities according to their local similarity. AntLP is a two-stage process. The technique starts by employing a number of similarity indices to provide weights to the edges in the input network. Second, AntLP use ACO to disseminate the labels. To test AntLP, we do many tests on popular social network datasets. According to the results of the experimental simulations, AntLP outperforms other community recognition algorithms for social networks in terms of modularity, normalized mutual information, and execution time.

Detecting communities is a difficult challenge in the field of social network analysis. Although there are a variety of algorithms dedicated to discovering social network clusters, many of them suffer from inefficiencies like excessive runtime complexity or the inability to efficiently identify groups that have common interests. With the fire propagation model serving as inspiration, a new method named the Fire Spread Community Detection Algorithm (FSA) is introduced in [12]. To locate neighborhoods close to a seed node, the FSA algorithm mimics the spread of a fire from its origin. Synthetic and real-world networks are used to compare the algorithm's modularity and conductance scores to those of top-tier community discovery techniques.

Feature selection is a major challenge in machine learning, right up there with community detection. There are many methods for selecting features; some examples are the PSO-based composite plastic coating subset selection algorithm [13], the advanced genetic algorithm [14], the PSO-based hybrid feature selection method [15], the PSO-based feature selection with multiple classifiers [16], the ACO-based feature selection [17][18], the combination of feature selection and ACO [19], the ABC multi-hive programming [20], and the multi-purpose ABC-based feature selection approach [21,22]. In order to enhance classifier performance, decrease computational complexity, and manage high-dimensional data sets, these methods use a variety of optimization methodologies to pick an optimal subset of features. The provided findings show that these techniques are successful in increasing classification accuracy while decreasing the total number of characteristics.

4. Proposed Method

The approach used in this research is meant to facilitate socializing by providing an appropriate framework for it. The Particle Swarm Optimization (PSO) algorithm is used here for its tried-and-true effectiveness. In 1991, James Kennedy and Russell Eberhart introduced the PSO algorithm at a conference [23]. The existing social models and social ties motivated their study, which resulted in the invention of a strong optimization method called the Particle Swarm Optimization (PSO) algorithm. This strategy is inspired by the group dynamics of animals like fish schools and flocks of birds.

The PSO method uses particles dispersed over the search space, each of which determines the local value of the objective function. Then, it chooses its next course of action based on past successes, present strengths, and recommendations from the best particles in the group. Once the desired result is reached, the algorithm's steps are repeated several times. The PSO technique is a general minimization approach that may be used for any issue having a point or surface as a solution in n-dimensional space.

Each particle's location is encoded as a binary value of 0 or 1 in this approach, and the i-th particle in a search space of dimension d for the position vector Xi is characterized by Equation (1). As there are a lot of particles involved, the PSO method may be more adaptable than other minimization techniques when dealing with the local minimum response issue.

$$X_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id}) \tag{1}$$

The velocity vector of the i-th particle is also defined by the vector Vi in the form of Equation (2):

$$V_i = (v_{i1}, v_{i2}, v_{i3}, ..., v_{id}) \quad (2)$$

The best position found by the i particle is defined by Equation (3) with Pi. best:

$$P_{i.\,best} = (p_{i1}, p_{i2}, p_{i3}, ..., p_{id}) \quad (3)$$

The best position of the best particle among all particles is defined by Pg. best as Equation (4):

$$P_{g.\,best} = (p_{g1}, p_{g2}, p_{g3}, ..., p_{gd}) \quad (4)$$

Equation (5) is used to update the location of each particle:

$$V_i(t) = w^* \, V_i\,(t\text{-}1) + C_1{*}rand_1 * (P_{i.\,best} – X_i\,(t\text{-}1)) + C_2{*}rand_2 * (P_{g.\,best} - X_i\,(t\text{-}1)) \quad (5)$$

Where $X_i = X_i\,(t\text{-}1) + V_i(t)$

The idea presented in the base paper [24] uses a different structure for socialization and considers a n × n matrix for n nodes.

The proposed structure of this paper is such that at first there are a series of nodes that need to be socialized. For example, if the Zichary Karate Club dataset is desired, it has 34 nodes that need to be socialized. Here, a structure is considered to be able to have the particle swarm algorithm and search space and actually search for optimal socialization. In this method, for example, it is assumed that there are 6 nodes that are named from numbers 1 to 6 and should be socialized. Initially, a 6-array array with 6 cells is considered. Here, if the number 1 is in cell 1, it means that node 1 is inside community 1. Cell 2, if the number 1 is in it, that is, nodes 1 and 2 are inside community 1. In the third cell, if the number is 2, that is, the third node is inside community 2. In general, Figure (1) shows the structure used in this paper for particles.

1	2	3	4	5	6
1	1	2	3	4	3

Figure 1: Structure used for particles for socialization

The method proposed in this paper utilizes the multi-objective particle swarm optimization algorithm, which considers two objective functions. The first goal is modularity, which is a very popular goal function and is used in most articles. However, the next goal is the main idea of this article, and it is a function called neighborhood power or neighborhood rate. For example, for a community like the Zichary Karate Club dataset, there is a community. There are bunches of nodes that are socialized; now we have to look at how good the quality of this community is per particle. It is assumed that another particle presents another socialization and puts part of the nodes in one society and another part of the nodes within another society. Now it is going to look at the suitability of different particles and find out which one is better. Modularity is used here, and secondly, this article uses a criterion called neighborhood density. This criterion states how much direct communication and communication there is between members within a community with a maximum length of, say, 4 steps. The greater the number of different paths between members of a community, the better the quality of that community. So first, all possible paths with a maximum length of 4 steps are calculated between nodes and then this is considered as a measure of the quality of that community. Because calculating path lengths in less than 4 steps is a very tedious process, first for all the data sets worked with, such as Dolphin, Football, and Karate, calculate paths and place them in a dataset that reads from the dataset when the program is run. The shorter the distance between the paths between the nodes, the better. In the method of this article, a criterion called similarity is used to calculate the similarity between communities. This criterion is calculated by Equation (6):

$$S= S+1/\text{length (paths (k)-1)} \tag{6}$$

Using the similarity criterion, how to socialize is determined.

5. Evaluate the proposed method

To evaluate the efficiency of the proposed process, MATLAB programming language was utilized. The proposed method was compared with another method called MGA [24]. The proposed method is referred to as "Proposed Method" in the results of this evaluation. The experiments were performed in an environment with the conditions listed in Table (1).

Table 1
Test system specifications

Item	Specifications
Processor	Intel Core ™ 2due CPU T6570@ 2.10 GHz 2.10 GHz
Main memory	4.0 GB
Hard Disk	320 GB
Operating system	Microsoft Windows 7 Ultimate
Programming Language	Matlab R2019b

For evaluation in this article, three datasets named Dolphin, Football and Zichary Karate Club have been use. The criterion for evaluation is a criterion called NMI, which measures the similarity of a community with the main community. In the results, all three communities compared with the final NMIs and the path to those NMIs. The results presented in the form of graphs in Figures (2-4). For example, in the Football community, the NMI obtained by the MGA comparison method is 0.5798%, but for the proposed method, it is close to 0.6%. Also, the path to obtain NMIs in 200 repetitions in the proposed method is better. In the other two Dolphin and Zichary Karate Club societies, the NMI proposed method was 0.5445% and 0.8372%, respectively, which is an improvement over the compared method. Also, the path of obtaining NMIs in 200 repetitions in the proposed method is better. According to the obtained results, it can be seen that the method of the compared article is stuck in local optimization and could not work better from one place to another, but the proposed method of this article has been able to escape to a large extent the trap of local optimization.

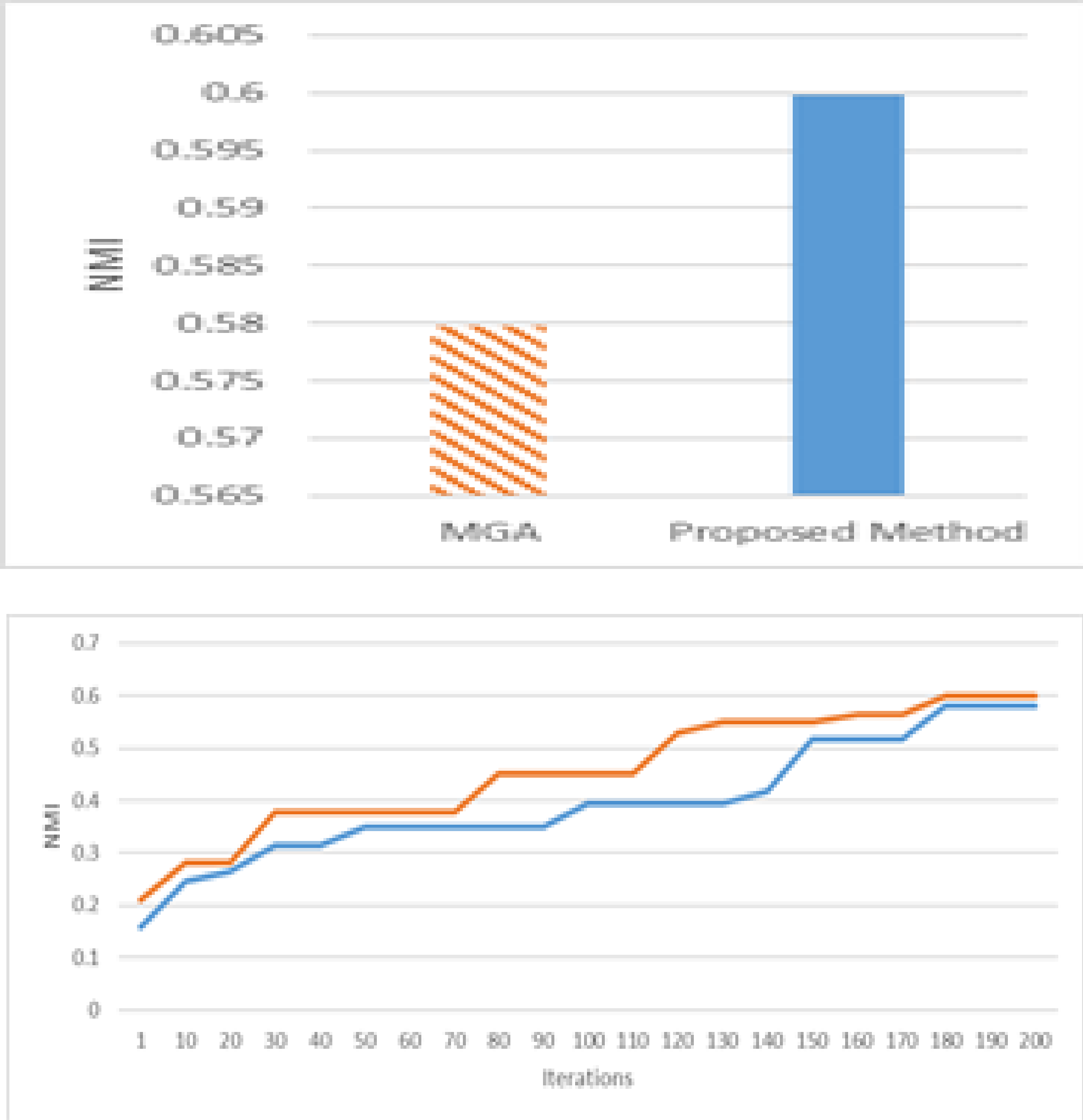
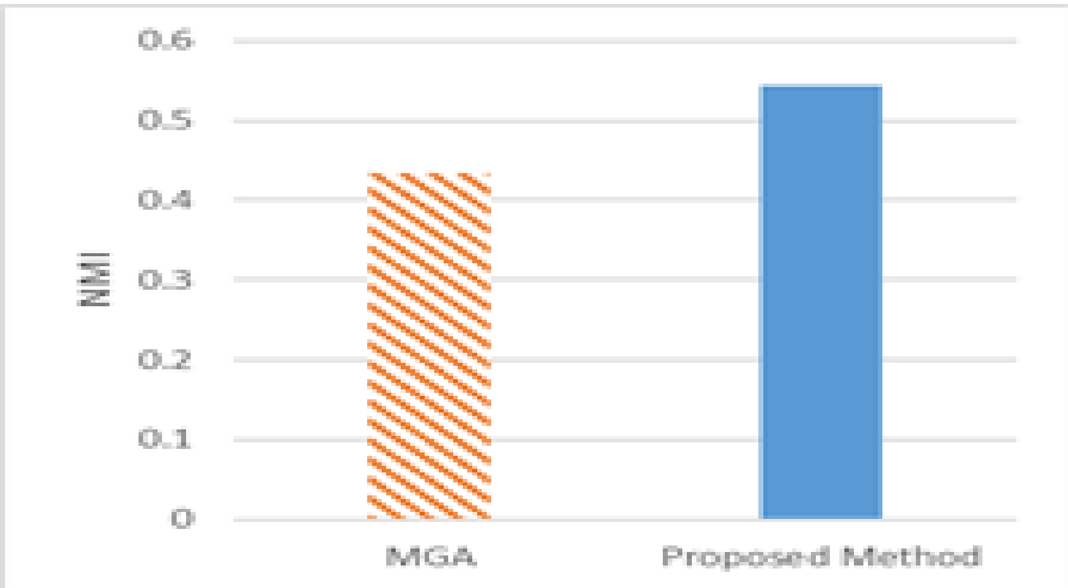


Figure 2: NMI review between the proposed and compared methods in the Football dataset



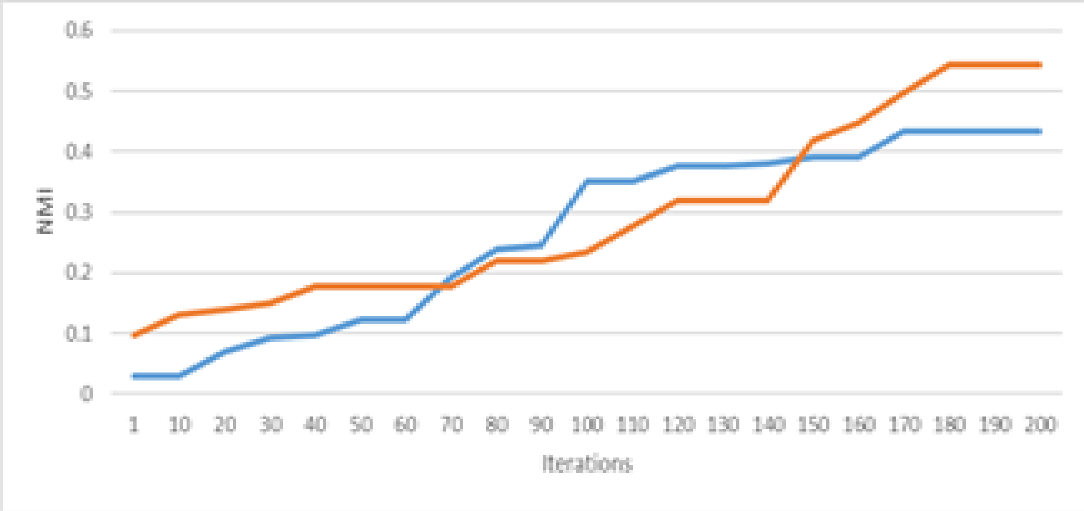
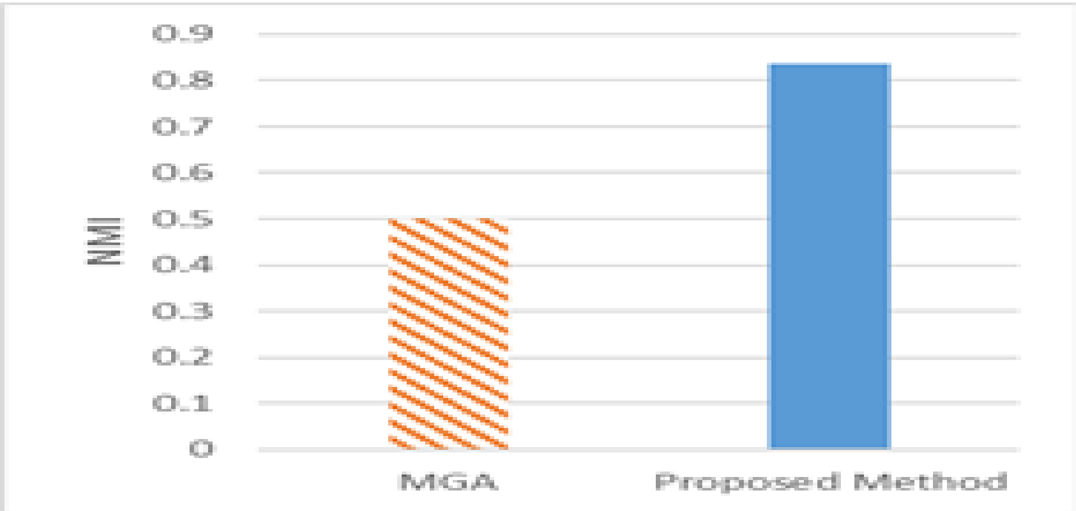


Figure 3: NMI review between the proposed and compared methods in the Dolphin dataset



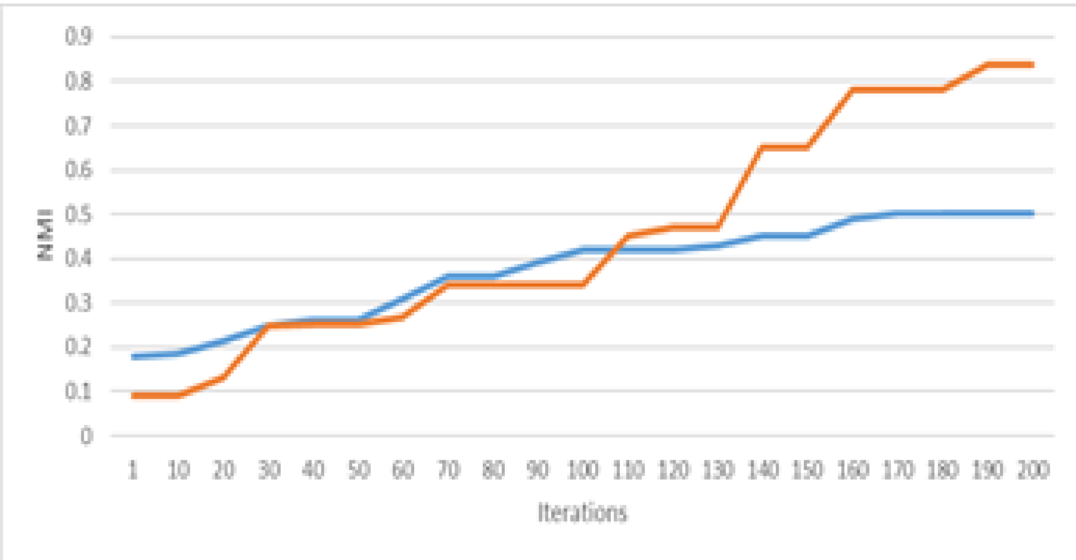


Figure 4: NMI review between the proposed and compared methods in the Zichary Karate Club dataset

6. Conclusion

The ability to discover communities inside a network is crucial for studying the structural and operational features of large-scale systems. Social scientists and practitioners of information extraction and retrieval, for example, might benefit from the knowledge gathered by community detection. Insights regarding the activities and interactions of groups of users may be gleaned from social media, making it a valuable resource. Due to their size and volatility, these networks provide significant hurdles for data mining techniques. In order to get insight into the structure and organization of a network, one might do "community identification," which includes locating clusters of nodes that are more linked to one another than other nodes. Individuals on social networks often group together into tight-knit groups with many connections between members. The structure and interactions within a social network may change over time, a phenomenon known as "dynamic evolution." Using the multi-objective particle swarm optimization technique, this research reveals the best approach to socializing and identifying communities. Extensive testing on the Dolphin, Football, and Zachary Karate Club datasets show that the suggested technique outperforms the benchmark.

7. References

[1] W. Zhang, R. Zhang, R. Shang, J. Li, and L. Jiao, "Application of natural computation inspired method in community detection," Physica A: Statistical Mechanics and its Applications, vol. 515, pp. 130-150, 2019/02/01/ 2019.
[2] C. Li, J. Bai, Z. Wenjun, and Y. Xihao, "Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment," Information Processing & Management, vol. 56, no. 1, pp. 91-109, 2019/01/01/ 2019.

[3] X. Zhao, J. Liang, and J. Wang, "A community detection algorithm based on graph compression for large-scale social networks," Information Sciences, vol. 551, pp. 358-372, 2021/04/01/ 2021.
[4] V. Moscato and G. Sperlì, "A survey about community detection over On-line Social and Heterogeneous Information Networks," Knowledge-Based Systems, vol. 224, p. 107112, 2021/07/19/ 2021.
[5] Rostami, Mehrdad, and Mourad Oussalah. "A Novel Attributed Community Detection by Integration of Feature Weighting and Node Centrality." (2021)
[6] M. Granovetter, The strength of weak ties, Am. J. Sociol. 78 (1973) 1360–1380.
[7] G.W. Flake, S. Lawrence, C.L. Giles, Efficient identification of web communities, in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’00, ACM, New York, NY, USA, 2000, pp. 150–160 doi:10.1145/347090.347121, http://doi.acm.org/10.1145/347090.
[8] You, X., Ma, Y., & Liu, Z. (2020). A three-stage algorithm on community detection in social networks. Knowledge-Based Systems, 187, 104822.
[9] Ahajjam, S., El Haddad, M., & Badir, H. (2018). A new scalable leader-community detection approach for community detection in social networks. Social Networks, 54, 41-49.
[10] Jokar, E., & Mosleh, M. (2019). Community detection in social networks based on improved label propagation algorithm and balanced link density. Physics Letters A, 383(8), 718-727.
[11] Hosseini, R., & Rezvanian, A. (2020). AntLP: ant-based label propagation algorithm for community detection in social networks. social networks, 7, 10.
[12] Pattanayak, H. S., Sangal, A. L., & Verma, H. K. (2019). Community detection in social networks based on fire propagation. Swarm and evolutionary computation, 44, 31-48.

[13] Unler A, Murat A, Chinnam RB. mr2PSO: a maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classifcation. Inf Sci. 2011;181(20):4625–41.
[14] Wenzhu Y, Daoliang L, Zhu L. An improved genetic algorithm for optimal feature subset selection from multi-character feature set. Expert Syst Appl. 2011; 38:2733–40.

[15] Yan C, et al. A novel hybrid feature selection strategy in quantitative analysis of laser-induced breakdown spectroscopy. Anal Chim Acta. 2019; 1080:35–42.
[16] Xue Y, et al. Self-adaptive parameter and strategy based particle swarm optimization for large-scale feature selection problems with multiple classifers. Appl Soft Comput. 2020; 88:106031.
[17] Tabakhi S, Moradi P, Akhlaghian F. An unsupervised feature selection algorithm based on ant colony optimization. Eng Appl Artif Intell. 2014; 32:112–23.
[18] Moradi P, Rostami M. Integration of graph clustering with ant colony optimization for feature selection. Knowl Based Syst. 2015; 84:144–61.
[19] Liu Y, et al. A classifcation method based on feature selection for imbalanced data. IEEE Access. 2019; 7:81794–807.
[20] Arslan S, Ozturk C. Multi Hive Artificial Bee Colony Programming for high dimensional symbolic regression with feature selection. Appl Soft Computing. 2019; 78:515–27.
[21] Zhang Y, et al. Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm. Expert Syst Appl. 2019; 137:46–58.
[22] Wang X-H, et al. Multi-objective feature selection based on artificial bee colony: an acceleration approach with variable sample size. Appl Soft Comput. 2020; 88:106041
[23] Aje, O. F., & Josephat, A. A. (2020). The particle swarm optimization (PSO) algorithm application–A review. Global Journal of Engineering and Technology Advances, 3(3), 001-006.
[24] Chen, K., & Bi, W. (2019). A new genetic algorithm for community detection using matrix representation
[25] method. Physica A: Statistical Mechanics and its Applications, 535, 122259.