Presenting a Novel Method for Identifying Communities in Social Networks Based on the Clustering Coefficient

Zhihong HE[1*], Tao LIU[2]
College of Art and Design, Chongqing Vocational College of Culture and Arts, Chongqing 400067, China[1]
College of Artificial Intelligence and Big Data, Chongqing College of Electronic Engineering, Chongqing 401331, China[2]

Abstract*—In recent decades, social networks have been considered as one of the most important topics in computer science and social science. Identifying different communities and groups in these networks is very important because this information can be useful in analyzing and predicting various behaviors and phenomena, including the spread of information and social influence. One of the most important challenges in social network analysis is identifying communities. A community is a collection of people or organizations that are more densely connected than other network entities. In this article, a method to increase the accuracy, quality, and speed of community detection using the Fire Butterfly algorithm is presented, which defines the algorithm and fully introduces the parameters used in the proposed algorithm and how to implement it. In this method, first the social network is converted into a graph and then the clustering coefficient is calculated for each node. Also, the butterfly algorithm based on the clustering coefficient (CC-BF) has been proposed to identify complex social networks. The proposed algorithm is new both in terms of generating the initial population and in terms of the mutation method, and these improve its efficiency and accuracy. This research is inspired by the meta-heuristic algorithm of Butterfly Flame based on the clustering coefficient to find active nodes in the social network. The results have shown that the proposed algorithm has improved by 23.6% compared to previous similar works. The findings of this research have great value and can be useful for researchers in computer science, social network managers, data analysts, organizations and companies, and other general public.

Keywords—Social network; detection of communities; butterfly fire algorithm; clustering coefficient

I. INTRODUCTION

Today, the Internet and web services are expanding rapidly, and at the same time, virtual social networks play an important role in people's real lives [1]. In fact, social networks are interactive networks that use the Internet as a medium to create communication between people [2]. With the rapid increase in social network users, high-scale data exploration can provide a better and more effective view of the hidden potential of these networks [3]. Internet social networks, as the most important examples, for the presence of different segments of society and the exchange of ideas, thoughts, and needs have changed according to social life [4]. A social network is a social structure formed by a group of people, organizations, or other social entities [5]. This collection is connected by social relations such as information exchange, cooperation, friendship, kinship, or financial exchanges [6]. Social network analysis is an approach used to study the interaction of humans and examine the patterns, communication structure, or organization of social networks [7]. With the expansion of the use of electronic and online communication methods, the number of social networks has increased, and the importance of extracting communities from these networks in order to analyze social networks has become more important [8]. In social networks, some nodes are more connected than the entire network's nodes, which are called communities. Nodes are connected by one or more specific types of dependencies [9]. For example, financial exchanges, friendships, kinship, business, web links, disease transmission, or airline routes are examples of communication. But the resulting structure of these networks is often very extensive and complex. Analysis of social networks is the mapping and measurement of cooperation relationships among individuals, groups, organizations, and any entity that has the ability to process information and knowledge. Graph theory is usually used to display and analyze social networks. The components of graph theory are nodes and edges [10].

Groups of nodes (or members) connected by one or more different kinds of relationships are referred to as social networks [11]. In social network analysis, network structure is understood as the pattern organization of those nodes and their relationships, which helps to explain how these patterns have an impact on people's behaviors and attitudes [12]. Different communities inside the network are formed by these exchanges, links, or connections [13]. Those communities' members frequently share certain traits or interests. Comprehending and using the network effectively begin with comprehending the structure of society [14]. In the literature, clustering and community detection are frequently used interchangeably. While community detection strategies for network analysis and an emphasis on network structure are created as a function of connectivity that incorporates social interaction, clustering techniques typically concentrate on a single strategy, such as using particular traits to group nodes in a network [15]. Clustering algorithms, on the other hand, can be thought of as a workable substitute for community detection techniques, and both of them can be used to solve a variety of network analysis problems [16]. Existing paper swarm algorithm-based methods have drawbacks, including a slow rate of community discovery, among others. With increasing fitness value, the detection speed of communities slows down, which could lead to a local optimum and take some time to attain a virtually global optimum [17]. A near-global optimum is quickly reached when efficiency is decreased, which lowers the likelihood of becoming stuck in a local optimum like speed detection. If an algorithm has improved speed detection, it may be able to reach a near-global optimum more quickly and with fewer iterations. In order to detect communities in a network, a new solution for the butterfly-fire technique based on the clustering coefficient (CC-BF) is proposed in this research. In addition to identifying communities in highly populated networks, CC-BF discovers cohesive collections of nodes as anomalous networks. The primary concept behind CC-BF is that it leverages the clustering coefficient, a social network analysis indicator, to improve community detection's accuracy, quality, and speed.

Other academics have successfully used strategies similar to CC to locate community structures in complex networks. Existing studies, however, don't look at the application of the clustering coefficient in BF. The clustering coefficient-based butterfly-fire algorithm's (CC-BF) essential phases are as follows: The speed of community detection using the clustering coefficient Modularity-based efficiency assessment.

To evaluate the quality of the population, the modularity criterion has been applied as a fitness function. Prior information about the size, quantity, or structure of communities is not necessary for the butterfly-fire approach, which is based on the clustering coefficient. According to experimental findings, CC-BF outperforms several other approaches for various networks. The following are the paper's contributions:

- A novel mutation technique that exploits the identified community structure to rewire existing connections;
- Using the clustering coefficient for the speed of community detection using the Fire Butterfly algorithm, comparing eight cutting-edge methods and evaluating the suggested algorithm on 12 different kinds of small and large networks.

The paper is divided into the following sections: The related works are included in the second section. The third section provides the suggested solution. The evaluation and effectiveness of the suggested algorithm are examined in the fourth section, along with a comparison to other algorithms of a similar nature. Finally, the fifth section presents the conclusion.

II. Related Works

In addition to the structural analysis of social networks, identifying communities is also used in other issues such as customer classification, recommender systems, vertex labelling, analysis of network influencers, and information dissemination. Due to the complexity of this issue, despite the various efforts that have been made in this field in recent years, it has not yet been fully resolved, and a satisfactory answer has not been provided to solve this issue. Many community identification algorithms have been presented for analyzing social networks. Identifying communities can be considered an optimization problem; from this point of view, several methods based on the maximization of the famous modularity criterion have been presented for identifying communities [18]. Among the famous methods in this field, we can mention Newman's greedy algorithm (FN) [19]. In this method, conjunctive hierarchical clustering is used, in which groups of nodes successfully form larger communities if and only if the modularity value increases after this integration. After that, Klast and his colleagues presented the CNM method [20, 34] and showed that by using a complex data structure, they were able to reduce the computational burden of modularity in Newman's algorithm, making it usable for large networks. Shang and his colleagues also presented an improved genetic algorithm called MIGA to obtain maximum modularity [21]. Fortunato and his colleagues showed in [22, 33] that this method also has limitations apart from modularity. One of the most important limitations of modularity in this field is the limitation of separability. The problem of separability has a great impact on practical problems in the real world. Because of such issues, communities have different sizes, and this problem causes many small communities not to be recognized. Two solutions have been proposed to overcome the limitation of separability: First, other quality metrics were proposed in addition to modularity to be identified at different scales of society. For example, in a paper [23], a criterion called significance was used instead of modularity as the objective function in the Levin method, and it produced better results than using modularity. Pizzoti also introduced a standard called "community rank" [24] to guarantee the high density of communication within communities and the low density of communication between communities. The second solution to this problem is to formulate community identification as a multi-objective optimization problem. Pizzotti in [25] tries to obtain suitable communities using the NSGA-II evolutionary algorithm and two objective functions of community merit and rank. Gang and his colleagues presented the MOEA/D-Net algorithm [26], which tries to optimize two objective functions against each other. In [27], the CLAnet method is also presented, which uses learning automata to maximize modularity along with a local limiter. Wasserman and his colleagues [28, 32], as the pioneers of social network analysis, have cited "social" for social networks, which can be defined as a system of social relations described by a set of actors and their social connections. Usually, a social network can be shown in the form of a graph, which has a set of vertices (nodes) and edges (connections) so that vertices are considered as actors within the network and edges are considered as relationships between these actors. In its simplest form, a social network is a mapping of all the relevant edges between the studied vertices. Here are the gaps identified in previous studies and the suitability of the proposed algorithm for certain types of data:

1) Gaps in previous studies: Lack of Clustering Coefficient in Butterfly-Fire Algorithm (BF): The text mentions that existing studies have successfully used strategies similar to the proposed Clustering Coefficient-based Butterfly-Fire algorithm (CC-BF) for community detection. However, it notes that previous studies haven't explored the application of the clustering coefficient in the Butterfly-fire algorithm. This suggests a gap in the literature where the combination of the clustering coefficient and the BF algorithm hasn't been extensively investigated.

Community Detection Speed: The text highlights that existing swarm algorithm-based methods for community detection have limitations, including a slow rate of community discovery. This issue could lead to suboptimal results. The proposed CC-BF algorithm aims to address this gap by improving the speed of community detection. This suggests a need for faster and more efficient community detection algorithms in the existing literature.

2) Suitability for data types: The text doesn't explicitly mention the types of data for which the proposed CC-BF algorithm is more suitable. However, it does mention that CC-BF is evaluated on "12 different kinds of small and large networks." This suggests that the algorithm is designed to be versatile and applicable to a variety of network data types.

In summary, the gaps in previous studies revolve around the lack of exploration of the clustering coefficient in the context of the Butterfly-Fire algorithm and the need for faster community detection methods. The proposed CC-BF algorithm appears to be designed to address these gaps and is tested on various types of network data.

Table I shows the classification of existing studies into orderly works based on applied community detection or clustering methods.

TABLE I. CLASSIFICATION OF CASE STUDIES BASED ON COMMUNITY DETECTION METHODS OR FUNCTIONAL CLUSTERING

Tool for clustering or detecting communities	No. of studies	ID	No. of nodes (n)	Comments	Ref.
A rapid method for modularity	4	S27, s28,s45,s51	more than 5000	used to make a comparison in (S27, S45)	[16]
MapInfo algorithm	3	S,10, S13, S18	2000	used to make a comparison in (S13)	[18]
Algorithm for maximizing expectations	2	S32, S36	more than 3500	used as a comparative tool	[19]
Community detection algorithm for binary graphs	5	S16, S24, S29, S41, S47	more than 6000	used as a comparative tool	[23]
Algorithm for eigenvector label propagation	1	S03	1500	newly suggested method	[24]
Spectral clustering-based technique for left-to-right oscillation	4	S40, S46, S49, S53	more than 5000	newly suggested method	[29]
Algorithm for evaluation and identification in the community	2	S34, S45	5000	newly suggested method	[30]
Eigenvector algorithm in front	2	S22, S05	more than 4000	newly suggested method	[31]

III. SUGGESTED METHOD

The most important problem in identifying communities in social networks can be considered the speed and accuracy of identifying communities in networks, as well as the new identification methods, which, in addition to speed and accuracy, should be able to act in a way to reduce the possibility of the influence of noise on the misidentification of communities. Minimize, and with the least knowledge of the structure of networks and the number and size of associations in the network, the best number and size should be selected and identified in an excellent way. Fig. 1 shows the general design of the proposed algorithm, which is used to detect and identify active nodes in social networks. In this algorithm to detect active nodes in social networks, attention has been paid to different parts of it, which have been examined in the following:

A. Initialization

In this algorithm, it is assumed that butterflies are candidate neighbors, and the variables of the problem are the locations of the bodies in the neighborhood. Therefore, butterflies can move in a one-dimensional, two-dimensional, or multi-dimensional dimension. Since the flame-propeller optimization algorithm is a population-based algorithm, the set of propellers is a matrix of ordered n×d. In this method, in the initialization stage, a random solution is generated for each moth butterfly (k=1, 2,..., B), where B represents the number of butterflies. The active node is represented by an array of length n, where the number stored in the i index of the array shows the ID of the candidate active node that executes the task Ti. Taking into account that the butterfly will be created in the initialization step B, the initial population of solutions will be a B×n matrix.

$$M = \begin{bmatrix} m_{1,1} & \cdots & m_{1,d} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,d} \end{bmatrix} \quad (1)$$

After initialization, the function P is executed iteratively until the function T is correct. The P function is the main function that moves around the search space. As mentioned above, the inspiration for this algorithm is transverse orientation. In order to mathematically model this behavior, the position of each bullet with respect to the flame is updated using the following equation.

$$M_i = S(M_i, F_j) \quad (2)$$

F_j represents the jth flame, M_i represents the ith butterfly, and s is the spiral function for the butterfly-fire algorithm.

$$S(M_i, F_j) = D_i \times e^{bt} \times \cos(2\pi t) + F_j \quad (3)$$

In this regard, D_i refers to the distance of the i-th propeller from the j-th flame. b is a fixed number to determine the shape of the logarithmic spiral, and t is a random number in [-1,1], where D_i is obtained from Eq. (3).

$$D_i = |F_j - M_i| \quad (4)$$

Considering the above, the pseudocode of the proposed method is presented in Fig. 2.

In the general butterfly-fire method, all newly generated butterflies are accepted and kept in the next generation, while a greedy strategy is used to accept butterflies that have a good fit in our algorithm. This greedy strategy can be as follows:

$$x_{i,new}^{t+1} = \begin{cases} x_i^{t+1} & |f(x_i^{t+1}) \leq f(x_i^t) \\ x_i^t & |othewise \end{cases} \quad (5)$$

In this regard, $x_{i,new}^{t+1}$ newly produced butterfly for the next generation. and $f(x_i^{t+1})$ *and* $f(x_i^t)$ are the fitness level of butterflies x_i^t and x_i^{t+1}, respectively.

The structure of propellers in the M-dimensional space of features is shown in Fig. 3. To search in the butterfly-fire algorithm space, the features are coded as butterflies in Fig 3. A unique code is considered for each feature in all P of the license. After defining the parameters, the fitness function will be calculated.

In order to escape from the local optimum and cover more search space of each graph, instead of selecting the worst vertex in each step, τth worst vertex is selected; That is, the vertex whose rank k is calculated using eq. (3) is selected for replacement.

$$k = (1 + (n^{1-\tau} - 1)(rand)^{\frac{1}{1-\tau}} \quad (6)$$

In this regard, k refers to the active number selected from the ordered list of ranks; n is the number of graph vertices and (rand) is the random number generation function in the interval [1, 0]. The value of parameter τ is fixed, and in this research, it is determined by trial-and-error method; For the algorithm limits the search space and, in other words, looks for the answer more strictly.

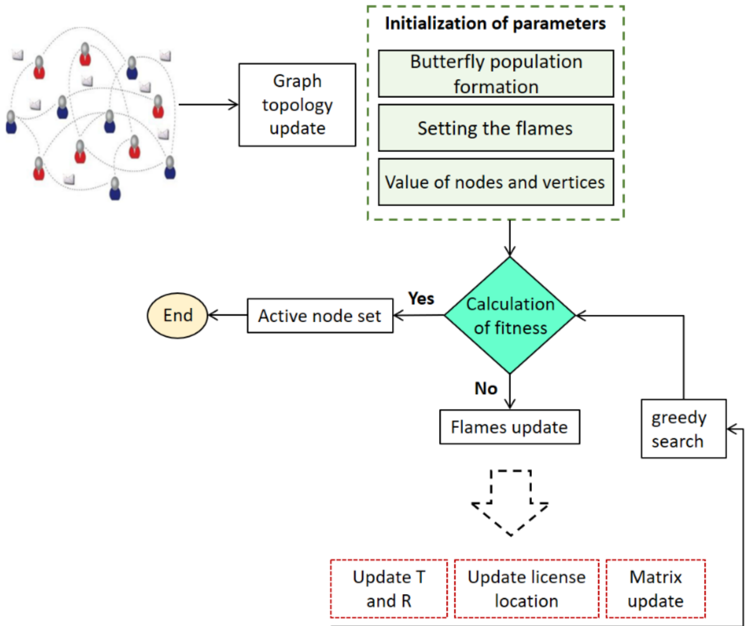


Fig. 1. Flowchart of the proposed algorithm.

Input: graph (V, C), Initializing the population
Output: Set of the Influence node
Update the number of flames (FlameNumber)
Initializing all the parameters
Initialize the population of moths
Calculate the objective values Equations (1)
for all moths
for all parameters
update r and t
Calculate D with respect to the corresponding moth by Equations (3)
Update the matrix M with respect to the corresponding moth by Equationss (4) and (5).
Calculate the Influence of each node
end
calculate the objective values
Update j-th flames by greedy strategy as Equations (6)
Apply later flames
S the current best individual in the population.
end
Return S.
End

Fig. 2. Pseudo code of the proposed method.

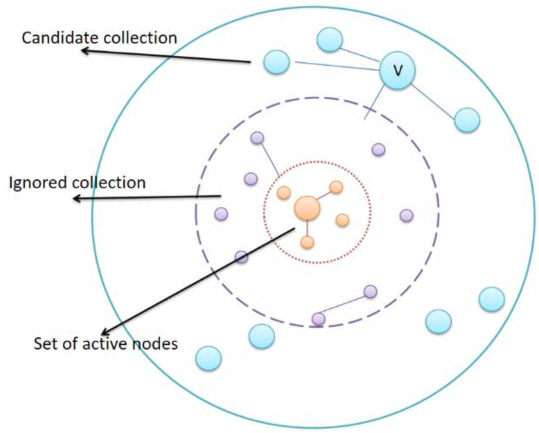


Fig. 3. The structure of nodes in the graph space.

IV. Evaluation and Simulation

Five community detection algorithms have been used to evaluate and compare the proposed method. The paper [19] presents a strategy called 2-Phase Community Detection (2PCD) to improve effective node detection algorithms in which a pre-processing step is added, and edges are weighted according to their centrality in the network topology. In the paper [4], an algorithm called Atom Stabilization Algorithm (ASA) has presented a new method for identifying influential nodes, which provides a new perspective for understanding the structure of nodes in complex networks. The algorithm presented in this paper is also used for evaluation. In this paper, the algorithms improved in research [4] and [19] have been used to evaluate the proposed method. To evaluate the results of the proposed method, the scale and quality criteria that are introduced in detail in the fourth section are used.

A. Measure Evaluation

In this section, for each of these data sets, the measure obtained by the proposed algorithm and other evaluated methods have been calculated. The relevant results are reported in Table II and Fig. 4.

From the analysis of Table II, it can be concluded that using the proposed algorithm has led to better results. The algorithm's performance has been tested on large datasets for which scaling optimization is increasingly difficult, and has shown satisfactory improvement. The proposed algorithm has improved by about 13.31% compared to the reference [19]. Compared to reference [4], it has improved significantly by 16.23.

TABLE II. The Obtained Measure

No. of Data set	2PCD	ASA	CC-BF
1	0.7846	0.5795	0.8560
2	0.7031	0.6235	0.7349
3	0.6892	0.6768	0.8296
4	0.7303	0.6138	0.7612
5	0.7865	0.6895	0.8923
6	0.8032	0.6725	0.8514
7	0.7927	0.6614	0.8225
8	0.7823	0.6424	0.8368

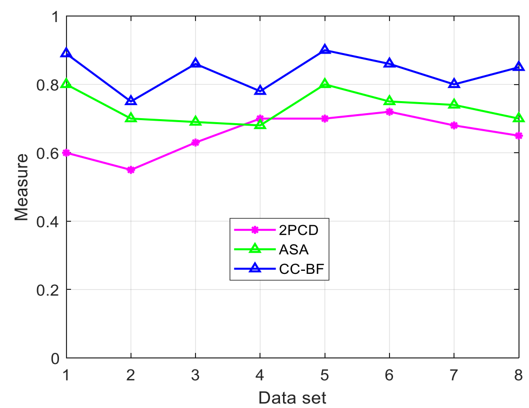


Fig. 4. Benchmark evaluation for 2PCD, ASA and CCBF methods.

According to the maximum scale obtained, the best values for parameters α1, α2 and α3 are obtained. These three parameters are used to control the relative importance of each feature. The obtained values are shown in Table III.

B. Quality Assessment

In this section, the quality of the communities created by the proposed strategy is analyzed by the NMI criterion. In this criterion, it is assumed that, according to graph G, an implicit truth is available in order to examine the communities. However, calculating NMI is challenging for real-life networks because no ground truth is usually available to assess what communities exist in G and their characteristics. Therefore, to perform experiments, a dataset should be considered in which the correct clustering of communities is specified. For this purpose, only the College Football Association of America dataset has been used in this section. The obtained results are shown in Table IV.

Quality value (NMI) ranges from 0 to 1, and higher values correspond to better algorithms. From the analysis of this table, we conclude that using the proposed algorithm has led to better results. The proposed algorithm has improved by approximately 3.7% compared to the algorithm presented in [4]. Compared to the algorithm [19], it has improved by about 4.7%. In the figure below, the computation time for 100 nodes is displayed.

TABLE III. THE BEST VALUES FOR PARAMETERS Α1, Α2 AND Α3

Parameters	Value
α1	0.64
α2	0.81
α3	0.44

TABLE IV. COMPARING THE QUALITY CRITERIA OBTAINED FOR THE PROPOSED ALGORITHM AND SIMILAR ALGORITHMS

Method	Value
2PCD	0.536508
ASA	0.679523
CC-BF	0.782158

C. Execution Time

In Fig. 5 and Table V, the results of the proposed method in this research are compared with similar methods. According to this table, it is clear that the proposed method has a real-time feature compared to similar methods.

Time and memory restrictions are increased in real-time applications. In the suggested method, a search is conducted for each active pattern to discover all active nodes with it before allowing any active users into the system. Each search's outcome is saved as a cluster. Upon receiving an alarm in real-time correlation, it is simple to locate earlier connected alerts by searching in the corresponding cluster. As a result, each node's processing time is cut down. The test results attest to the accuracy of the performance and the method's time-saving effectiveness. The method's benefit is that it executes in 5.3126 seconds on a home computer, proving it is real-time. Investigations have been conducted for the suggested method based on the source [23] for an online social network. The network of Twitter's "followership" is represented by this data. Network nodes represent users, and a link is created between two users when one of them is "followed". This network has 24256 nodes and 44135 edges. The average CC is 0.412, and the average network degree is 3.05. Table VI displays each network's kind and size.

In order to assess the consistency of CC-BF, each network in this study had CC-BF implemented 12 times, and the maximum, mean, and standard error of modularity values for those 12 implementations were determined. Fig. 6 displays the mean modularity values and standard errors.

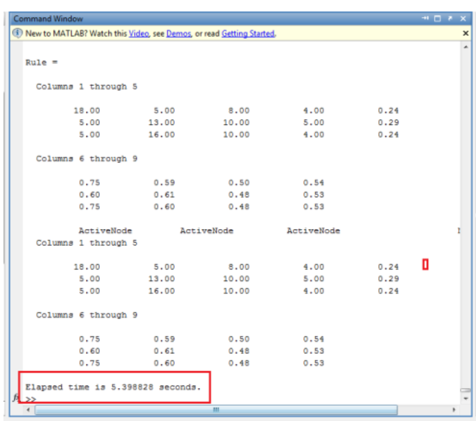


Fig. 5. The computational time for 100 nodes.

TABLE V. COMPUTING TIME FOR 100 NODES

Method	Second(s)
2PCD	8.469144
ASA	7.562415
CC-BF	5.398828

For instance, CC-BF gives steady findings with 0% standard error for the highly sparse Facebook network, as shown in Fig. 6, and significantly beats all other methods. It is modular in terms of size. Cora performs poorly and generates a low modularity rating for this network. On the other hand, methods other than dolphin provide numbers for this network's modularity that are nearly identical. Jazz performs better than all other approaches for SCN, whereas CC-BF performs better than Ecoil, Protein, PGP, Cora, and Polbook.

Additionally, the Reality Mining dataset [24], which the Harvard Media Lab supplied, has been chosen to test the suggested methodology for the infection rate of users in social networks. 200 Harvard students' contact, proximity, location, and activity data from the academic years 2006–2008 are included in the reality mining dataset.

In each test on this dataset, 0.03% of users are infected with a worm and worm propagation is simulated over three days. The CC-BF method's infection rate in each test is contrasted with the social-based 2PCD and ASA approaches. The percentage of infected users over non-infected users is used to compute the infection rate. Similar techniques yield 200 (for a static network), 300 (for a dynamic network), and 350 (for a network) clusters, respectively. The warning threshold β is established at 3, 15, and 25%, respectively, for each value of m.

For stability, each experiment is run 2,000 times. The experiment's findings for three distinct values of m and β are displayed in Fig. 7, 8, and 9. First, it is evident that in order to meet the anticipated infection rate, more people must be patched the longer one waits (, the higher the warning threshold). For instance, with m = 200 clusters and an anticipated contamination rate of 0.4, we should allocate patches to more than 20% of the users when β= 12% and to users who would be sent when β= 3%. When β= 25%, approximately 94% of all users have influence.

The second finding demonstrates that, in the static version of the social network depicted in Fig. 7, the proposed method outperforms the social-based methods 2PCD and ASA in terms of contamination rate. For instance, the proposed method's contamination rate is 7% to 12% lower than existing algorithms with comparable goals. The number of clusters m changes when new users join the network and form new social connections, and the infection rate of the social-based technique is updated using the cluster sizes of 300 and 350 as well as the warning threshold. β = 3%, 15% and 25%, respectively. Fig. 8 and 9 illustrate how the suggested method, which has the ability to quickly and adaptively update the network community structure, achieves a higher infection rate than competing methods while also having much lower computational costs and execution times.

TABLE VI. EMPIRICAL AND PRACTICAL DATE SETS ARE UTILIZED TO ASSESS CC-BF

Data set	Type	Edge	Nodes	Avg. ccpppp	Avg. degrees	Avg. p.length
BA-1000	Social	74.1	32.3	0.5586	4.3586	2.2876
FF-256	Online-Social	151.05	58.9	0.28785	4.87255	3.18915
LFR256	Collaboration	582.35	109.25	0.38285	10.12795	2.3826
GR512	Collaboration	418.95	99.75	0.4636	7.98	2.92505
GR1000	Social	5209.8	188.1	0.60135	26.31215	2.12325
BA-512	Social	493.05	397.1	0.19855	2.35885	4.58185
E.coli	Collaboration	2604.9	1509.55	0.8341	3.27845	5.53185
WS-512	Synthetic	2831.95	2743.6	0.76285	1.9608	3.67365
Dolphins	Synthetic	8310.6	3537.8	0.2071	4.4631	4.99605
Football	Online-Social	13771.2	4979.9	0.65265	5.25445	5.74655
BA-256	Online-Social	23100.2	10146	0.418	4.275	7.11075
WS-256	Biological	31189.45	22201.5	0.3154	2.679	0.59945
Facebook	Synthetic	1945.6	243.2	0.3287	15.2	2.5764
Protein	Social	1050.7	243.2	0.51775	6.76115	3.477
FF-512	Online-Social	1983.6	486.4	0.5434	7.2656	3.6784
FF-1000	Synthetic	3802.85	950	0.53485	7.6722	3.9064
Twitter	Synthetic	242.25	243.2	0.95	1.8924	5.47295
WS-1000	Collaboration	485.45	486.4	0.95	1.8962	6.1123
NetScience	Synthetic	949.05	950	0.95	1.8981	7.08795
Scientific Co.	Biological	780.9	243.2	0.54815	6.1009	3.53305
PGP	Biological	3156.85	486.4	0.57665	12.331	3.2566
Jazz	Biological	10719.8	950	0.56905	21.4396	2.98015
GR256	Social	1216	243.2	0.48355	9.5	3.28415
Karate	Collaboration	2432	486.4	0.46835	9.5	3.76675
Polbooks	Online-Social	4750	950	0.47405	9.5	4.2617

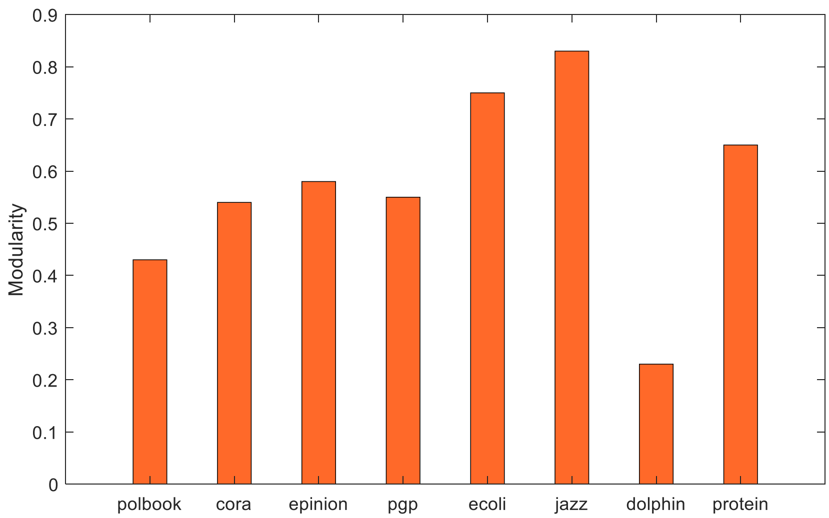
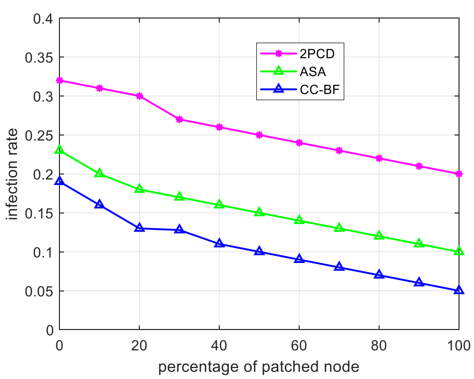
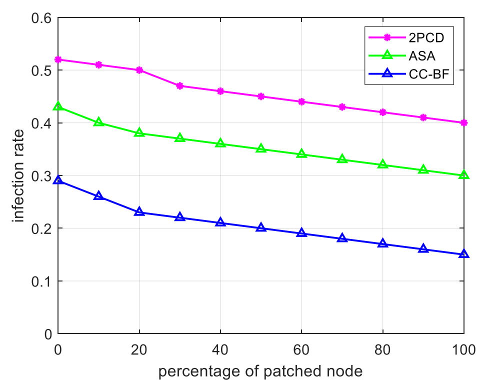


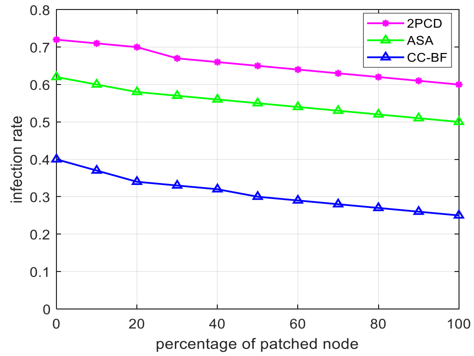
Fig. 6. Standard errors and modular mean values for 12 CC-BF runs across several networks.





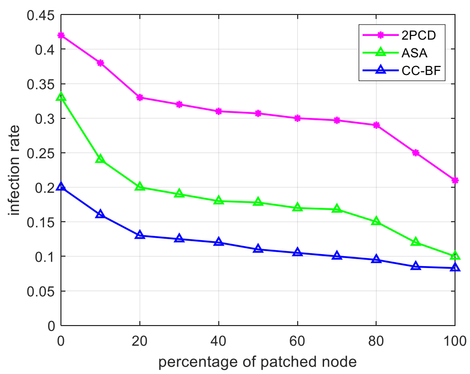
β = 3%

(b) β = 15%

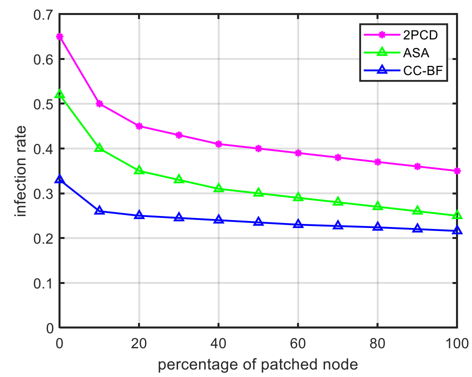


(c) β = 25%

Fig. 7. A static network with k = 200 clusters and infection rates.

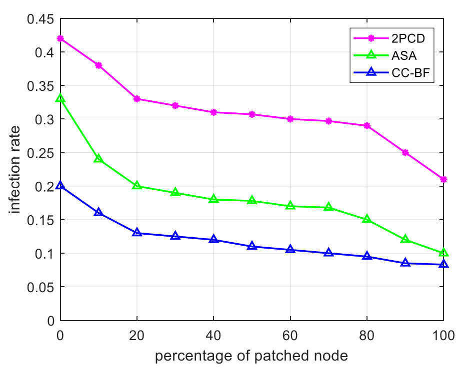


(b) β = 3%

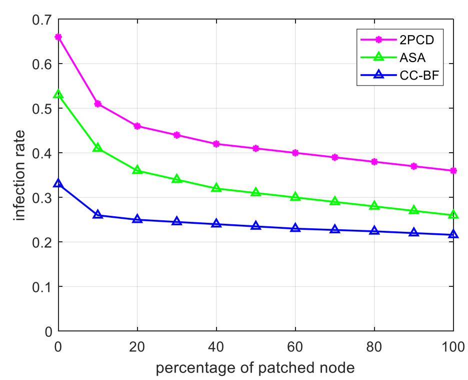


(b) β = 15%

Fig. 8. A static network with k = 300 clusters and infection rates.



(b) β = 3%



(b) β = 15%

Fig. 9. A static network with k = 350 clusters and infection rates.

V. CONCLUSION

Investigating community recognition in social networks is an important issue in many fields and disciplines, such as marketing and information technology. Community detection in social networks can be considered a graph clustering problem, where each set corresponds to a cluster in the graph. The goal of conventional community detection methods is to partition a graph so that each node belongs to exactly one cluster. A community can be defined as a group of individuals close to each other compared to other entities in the dataset. The proposed algorithm is checked and compared with the evaluated algorithms. This comparison is done according to the introduced criteria. To provide more accurate and complete results, five data sets have been used. Examining the results shows that the proposed method has led to better results. The performance of the algorithm has been tested on large datasets, for which scaling optimization is increasingly difficult and has shown satisfactory improvement. Among the advantages of NMI compared to other related methods, we can mention the competition with the latest technologies in the accuracy and scale of data, the ability to execute and implement the system in the real world, and the ability to perform all the steps of the proposed system online. The results show that the proposed algorithm has improved by 23.6% compared to previous similar work. In order to improve the quality of detecting effective nodes in future works, it is possible to mention the discovery of the semantic relationship between nodes and the virtual networks between them. Hidden connections between nodes are introduced as virtual associations. Discovering virtual associations between nodes can help algorithms based on participation in the detection process. Also, tagging the content of messages sent between nodes can help to discover nodes.

REFERENCES

[1] A. Biswas, S. Khandelwal, and B. Biswas, "Community detection in networks using atom stabilization algorithm," 2017: IEEE, pp. 89-93.

[2] P. Chunaev, "Community detection in node-attributed social networks: a survey," Computer Science Review, vol. 37, p. 100286, 2020.

[3] M. Mazza, G. Cola, and M. Tesconi, "Modularity-based approach for tracking communities in dynamic social networks," arXiv preprint arXiv:2302.12759, 2023.

[4] N. Dakiche, F. B.-S. Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," Information Processing & Management, vol. 56, no. 3, pp. 1084-1102, 2019.

[5] R. Djerbi, R. Imache, and M. Amad, "Communities' Detection in Social Networks: State of the art and perspectives," 2018: IEEE, pp. 1-6.

[6] M. Contisciani, H. Safdari, and C. De Bacco, "Community detection and reciprocity in networks by jointly modelling pairs of edges," Journal of Complex Networks, vol. 10, no. 4, p. cnac034, 2022.

[7] Trik, M., Pour Mozafari, S., & Bidgoli, A. M. (2021). An adaptive routing strategy to reduce energy consumption in network on chip. Journal of Advances in Computer Research, 12(3), 13-26.

[8] Yanwei Zhao, Ben Niu, Guangdeng Zong, Xudong Zhao, Khalid H. Alharbi. Neural network-based adaptive optimal containment control for non-affine nonlinear multi-agent systems within an identifier-actor-critic framework, Journal of the Franklin Institute. 360 (12), pp.8118-8143, 2023.

[9] S. Guo, X. Zhao, H. Wang, N. Xu, Distributed consensus of heterogeneous switched nonlinear multiagent systems with input quantization and dos attacks, Applied Mathematics and Computation 456 (2023) 128127.

[10] Khezri, E., Zeinali, E., & Sargolzaey, H. (2023). SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks. Plos one, 18(4), e0282031.

[11] Arefanjazi, H., Ataei, M., Ekramian, M., & Montazeri, A. (2023). A Robust Distributed Observer Design for Lipschitz Nonlinear Systems With Time-Varying Switching Topology. Journal of the Franklin Institute.

[12] A. A. Paracha, J. Arshad, and M. M. Khan, "SUS You're SUS!—Identifying influencer hackers on dark web social networks," Computers and Electrical Engineering, vol. 107, p. 108627, 2023.

[13] Fabin Cheng, Ben Niu, Ning Xu, Xudong Zhao, and Adil M. Ahmad. Fault Detection and Performance Recovery Design With Deferred Actuator Replacement Via A Low-Computation Method, IEEE Transactions on Automation Science and Engineering, DOI: 10.1109/TASE.2023.3300723, 2023.

[14] Khezri, E., Zeinali, E., & Sargolzaey, H. (2022). A novel highway routing protocol in vehicular ad hoc networks using VMaSC-LTE and DBA-MAC protocols. Wireless Communications and Mobile Computing, 2022.

[15] Chen Cao, Jianhua Wang, Devin Kwok, Zilong Zhang, Feifei Cui, Da Zhao, Mulin Jun Li, Quan Zou. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Research.2022, 50(D1): D1123-D1130.

[16] FanghuaTang, Huanqing Wang, Liang Zhang, Ning Xu, Adil M.Ahmad. Adaptive optimized consensus control for a class of nonlinear multi-agent systems with asymmetric input saturation constraints and hybrid faults. Communications in Nonlinear Science and Numerical Simulation, 126: 107446, 2023.

[17] Sai Huang；Guangdeng Zong；Huanqing Wang；Xudong Zhao；K. H. Alharbi. Command Filter-Based Adaptive Fuzzy Self-Triggered Control for MIMO Nonlinear Systems With Time-Varying Full-State Constraints. International Journal of Fuzzy Systems, 2023, https://doi.org/10.1007/s40815-023-01560-8.

[18] E. Jokar, M. Mosleh, and M. Kheyrandish, "Overlapping community detection in complex networks using fuzzy theory, balanced link density, and label propagation," Expert Systems, vol. 39, no. 5, p. e12921, 2022.

[19] Khezri, E., & Zeinali, E. (2021). A review on highway routing protocols in vehicular ad hoc networks. SN Computer Science, 2, 1-22.

[20] S. Tschiatschek, A. Singla, M. Gomez Rodriguez, A. Merchant, and A. Krause, "Fake news detection in social networks via crowd signals," 2018, pp. 517-524.

[21] M. Trik, H. Akhavan, A.M. Bidgoli, A.M.N.G. Molk, H. Vashani, S.P. Mozaffari, A new adaptive selection strategy for reducing latency in networks on chip, Integration. 89 (2023) 9–24.

[22] Khalafi, M., & Boob, D. (2023, July). Accelerated Primal-Dual Methods for Convex-Strongly-Concave Saddle Point Problems. In International Conference on Machine Learning (pp. 16250-16270). PMLR.

[23] C.-C. Ni, Y.-Y. Lin, F. Luo, and J. Gao, "Community detection on networks with Ricci flow," Scientific reports, vol. 9, no. 1, pp. 1-12, 2019.

[24] Behzad, A., Wakkary, R., Oogjes, D., Zhong, C., & Lin, H. (2022, April). Iterating through Feeling-with Nonhuman Things: Exploring repertoires for design iteration in more-than-human design. In CHI Conference on Human Factors in Computing Systems Extended Abstracts (pp. 1-6).

[25] M. M. D. Khomami, A. Rezvanian, M. R. Meybodi, and A. Bagheri, "CFIN: A community-based algorithm for finding influential nodes in complex social networks," The Journal of Supercomputing, vol. 77, pp. 2207-2236, 2021.

[26] M. Samiei, A. Hassani, S. Sarspy, I.E. Komari, M. Trik, F. Hassanpour, Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare, J Cancer Res Clin Oncol. (2023) 1–15.

[27] H. Li, Q. Shang, and Y. Deng, "A generalized gravity model for influential spreaders identification in complex networks," Chaos, Solitons & Fractals, vol. 143, p. 110456, 2021.

[28] Wakkary, R., Oogjes, D., & Behzad, A. (2022). Two Years or More of Co-speculation: Polylogues of Philosophers, Designers, and a Tilting Bowl. ACM Transactions on Computer-Human Interaction, 29(5), 1-44.

[29] M. Trik, A.M.N.G. Molk, F. Ghasemi, P. Pouryeganeh, A hybrid selection strategy based on traffic analysis for improving performance in networks on chip, J Sens. 2022 (2022)

[30] M. Alassad and N. Agarwal, "Contextualizing focal structure analysis in social networks," Social Network Analysis and Mining, vol. 12, no. 1, p. 103, 2022.

[31] un, J., Zhang, Y., & Trik, M. (2022). PBPHS: a profile-based predictive handover strategy for 5G networks. Cybernetics and Systems, 1-22.

[32] A. R. Costa and C. G. Ralha, "AC2CD: An actor–critic architecture for community detection in dynamic social networks," Knowledge-Based Systems, vol. 261, p. 110202, 2023.

[33] M. R. HabibAgahi, M. A. M. A. Kermani, and M. Maghsoudi, "On the Co-authorship network analysis in the Process Mining research Community: A social network analysis perspective," Expert Systems with Applications, vol. 206, p. 117853, 2022.

[34] M. Alassad and N. Agarwal, "A Systematic Approach for Contextualizing Focal Structure Analysis in Social Networks," 2022: Springer, pp. 46-56.