Advances in Basic and Applied Sciences 6 (2025) 19-25

Advances in Basic and Applied Sciences

journal homepage: https://abas.journals.ekb.eg/

DBSCAN with cuckoo search algorithm for social networks community detection

Ramadan Babers

Faculty of Computer Studies, The Arab Open University (AOU)

ARTICLE INFO

Article history:
Received 20 September 2025
Received in revised form 25 October
Accepted 6 October 2025
Available online .. October 2025

10.21608/abas.2025.415766.1081

Keywords: Community Detection, Social Networks, DBSCAN, Cuckoo Search Algorithm, Optimization.

ABSTRACT

The discovery of community structures in large-scale, complex networks is of fundamental importance to social network analysis. This task has enormous implications for understanding information flow and the behavioral characteristics of groups. This paper introduces a new hybrid model that combines the DBSCAN algorithm with a modified Cuckoo Search Optimization (CSO)algorithm of L´evy flight. This synergy leverages DBSCAN's robust capability to isolate noise and identify dense core nodes, while the optimized CSO performs a global search for the most effective community partitions. The proposed model was rigorously evaluated against a suite of established algorithms, including the Bat Algorithm, AFSA, Multilevel, Walktrap, AKHSO, Ant-Lion Optimizer, Lion Optimization Algorithm, and standard Cuckoo Search. Experimental values of four standard social networks substantiate the better performance of our method, which obtained the top recorded results for both Modularity and NMI. These findings indeed validate that the approach not only detects communities with higher internal cohesion but also more properly mirrors the observed ground-truth structures, verifying its effectiveness and strength.

1. Introduction

Internet pervasiveness has radically changed global communications with seamless social interactions by virtue of various applications being made possible by smart electronic equipment like smart phones, smart televisions, and computers. These online spaces can themselves be envisioned as intricate networks with individual elements like users represented by nodes, while interactions or relationships amongst them are characterized by edges with sites like Twitter, LinkedIn, and Facebook being a few illustrations [1]. One critical factor to a comprehension of the functionality and architecture of these networks is community detection—sets of nodes with a larger number of connectors to one another relative to the full network and frequently with mutual attributes or functionalities [1–3].

The community discovery task has stimulated various computational techniques. These include

* **Corresponding author E-mail:** *R.Babers@aou.edu.eg*

algorithms based on random walks [4, 5], clustering techniques [6], and optimization strategies aimed at maximizing quality metrics like modularity [7]. While these new developments, it is still challenging to accurately partition communities under noise and outlying nodes. More recent surveys and advancements, including those leveraging deep learning approaches, are discussed in [34, 35].

Noting this issue, we propose a novel two-stage hybrid framework as a solution for it. At first stage, DBSCAN is applied as a preprocessing step to efficiently identify and filter out noise nodes potentially interfering with clear boundaries between communities. At the second stage, an optimized version of the Cuckoo Search Algorithm with L´evy flight for efficient global search is applied to further improve the community structure optimization in the cleaned network. The proposed framework was evaluated on four well-known social network datasets and compared against several state-of-the-art algorithms using Precision, Recall, Accuracy, and F-measure metrics.

The structure of this work is defined as follows: Section 2 provides a literature review regarding social networks, community detection measures, and preliminary algorithms applied. Section 3 provides the architecture of the new DBSCAN-CSO model. Section 4 includes experimental outcomes with comparative analysis. Section 5 discusses the findings and explores the limitations of the proposed approach. Section 6 concludes the work and briefly discusses directions for further research.

2. Literature Review

2.1. Social Network and Community Discovery Introduction

Social Network consists of sets of nodes forming interrelated objects that are joined by edges. They can mathematically be represented as $G(N, E)$, with N being the set of nodes and denoted by $|N| = n$, and E denoted by the set of relations connecting the nodes, $E \subseteq N \times N, |E| = m$.

Communities can be expressed as sets of non-empty node sets [8]. The task is to partition a set of nodes $\{x_1, x_2, \cdots, x_n\}$, into disjoint sets (communities C), expressed as $C = \{N_1, \cdots, N_{cn}\}$, where cn is the number of all the communities. The community's size is to meet the objective quality function.

$$\min\ F(S), S \in \Omega., \tag{1}$$

where $F(S)$ is an objective quality function and S is the quality measure minimized without loss in the context.

2.2. Social Network Analysis Measures.

Social Networks has quite a few of these measures that are used for analysis, for example, Association Index, Strength, Closeness Centrality, Eigenvector centrality, Node Degree, Affinity, and Reach [9], [10].

Modularity is yet another quantity built to validate the quality and strength of partitions of networks. Modularity measure by Girvan and Newman [11], [12]. Mathematically, the Modularity can be expressed as in equation (2). Suppose ci is the community where node i is assigned. Then the fraction of graph edges that are inside the communities, i.e., that connect the nodes belonging to the same community, is:

$$\frac{\sum_{ij} A_{ij}\delta(c_i, c_j)}{\sum_{ij} A_{ij}} = \frac{1}{2m}\sum_{ij} A_{ij}\delta\left(c_i, c_j\right). \tag{2}$$

The adjacency matrix (A) with (n) nodes:

$$A_{ij} = \begin{cases} 1\,, & \text{where } i \text{ and } j \text{ connected} \\ 0. & \text{otherwise} \end{cases} \tag{3}$$

The δ -function $\delta(u, v)$ has two values: 1 where $u = v$ and 0 otherwise, while the term $\frac{1}{2m}\sum_{ij} A_{ij}$ represents the number of edges. The modularity Q defined as:

$$Q = \left(\frac{1}{2m}\right)\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta\left(c_i, c_j\right) \tag{4}$$

Normalized Mutual Information (NMI) is a measure of performance that measures the similarity between the detected communities in networks and the ground truth [13] [1]. NMI presented by A. Lancichinetti [14]. NMI calculated by equation (5).

$$I(X, Y) = -\frac{2\sum_{i=1}^{C_X}\sum_{j=1}^{C_Y} C_{ij}.log\left(\frac{C_{ij}N}{C_i}.C_j\right)}{\sum_{i=1}^{C_X} C_i.log\left(\frac{C_i}{N}\right) + \sum_{j=1}^{C_Y} C_j.log\left(\frac{C_j}{N}\right)} \tag{5}$$

X and Y represents two networks, C is the confusion matrix, C_{ij} is the number of individuals in community i in X and in community j in Y , C_X and C_Y are the number of groups in part X and in part Y , C_i and C_j are the elements of row i and column j of C and N is the total number of individuals in the networks. The output of the equation value will be 1 if two communities are similar and 0 if there are various communities.

2.3. Density-Based Spatial Clustering of Applications with Noise

To enhance the quality of found communities, we use the DBSCAN algorithm [15] due to its established efficiency for finding clusters of arbitrary shape from noise-contaminated spatial data. Although partition-based methods, DBSCAN clusters are based on the concept of density connectivity, hence emerging successful in partitioning of core, borderline, and outlying nodes.

The algorithm takes two parameters: a number of points (MinPts) and radius (ε) to be at least. Mark a node as a core node if at least MinPts nodes can be seen within its ε-neighborhood. A border node is within the ε-neighborhood of a core-node but does not itself qualify for MinPts. Each of those nodes that is not a core node nor a borderline node is labeled as noise or an outlier. Its natural capability to exclude non-core elements during data preprocessing makes DBSCAN particularly for social network analysis, for which outliers are the typical case. It is thus desirable to observe that the algorithm's success depends upon the choice of ε and MinPts judiciously [18].

2.4. Cuckoo Search Algorithm

The Cuckoo Search Algorithm, pioneered by Yang and Deb [19], is a metaheuristic optimization algorithm inspired by the brood parasitism of certain cuckoo species. The algorithm is governed by three idealized rules:

1. Every bird drops off only one egg (solution) at a random nest.
2. The nests containing superior solution eggs are kept for later generations.
3. The number of nests of the host is constant, and the host can find the alien egg with the probability $pa \in [0,1]$. When finding it, the host can abandon the nest or refuse to hold the egg.

The L´evy flight creates new candidate solutions by a random-walk process whose step sizes are selected from a heavy-tailed distribution. It balances extensive exploitation of the search space to the same extent that it balances aggressive searching. The new solution is computed by the following equation [21]:

$$x_c^{(t+1)} = x_c^t + \alpha \oplus Levy(\lambda) \qquad (6)$$

The new solution is expressed as $x_c^{(t+1)}$, x_c^t is the detected solution, and α is the step size where α > 0.

The probability of distribution that L´evy flight step is governed by:

$$Levy = t^{-\lambda}, 1 < \lambda \leq 3 \qquad (7)$$

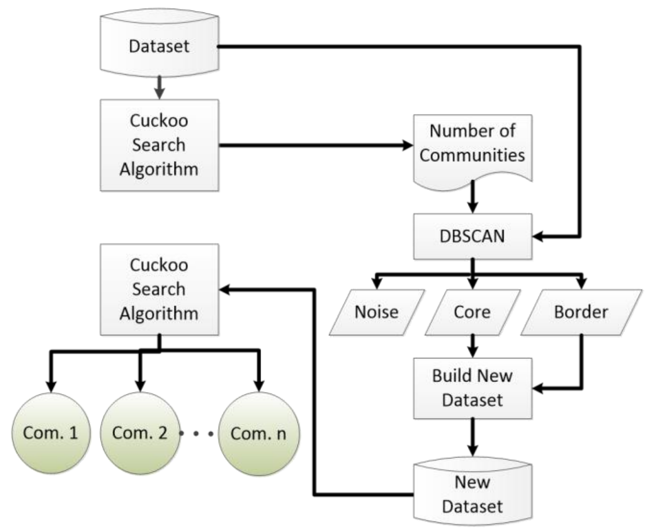

Figure 1: The proposed two-stage hybrid DBSCAN-CSO model. Stage 1: The Cuckoo Search Algorithm (CSA) provides an initial estimate of the number of communities (k). Stage 2: DBSCAN uses k to guide its parameters (ε, MinPts) to identify and remove noise/outlier nodes. Stage 3: The purified network is then processed by CSA again to obtain the final, optimized community partition.

3. The Proposed Model

The CSA employed in this research to identify the quantity of communities from the utilized datasets. Subsequently, the DBSCAN algorithm uses this estimated community count as an input parameter to cluster individuals engaged in Social Network datasets to core, border, and noise individuals by changing Eps. value to eliminate the outliers individuals in the datasets. After that, the datasets will be free of noise nodes and ready to implement CSA on it again to detect communities. The model of this research is shown in Fig. [1]. The DBSCAN Cuckoo Search Algorithm based on Algorithm 1. The locus-based adjacency encoding scheme employed in order to present individuals in CSA who are based on the genetic algorithm [22] [23].

4. Experimental Results

The developed approach was implemented for four data sets of social networks to discover the communities there. The number of communities in each data set is already defined. NMI was used [24] to compare the accuracy of the quality of the resulting groups is measured by modularity.

Algorithm 1: Cuckoo Search Algorithm and DBSCAN For Detecting Communities in Social Networks

---

Input: A network $G = (N,E)$.
Output: Group of similar nodes.
procedure Initial value for the parameters:
ϵ, MinPts, $\alpha, s,$ Max no. of iteration and #trails
Label all nodes as *core*, *border* and *outlier*.
Eliminate *outlier* nodes.
Generate a random node of n host nests $\overrightarrow{x_h}$
For each nest, calculate the fitness F_h.
Get a solution randomly by L´evy flight with
    fitness F_c.
Detect the best-solution
for each solution **do**
    if $F_c > F_h$ **then**
        Replace h by new-solution c
    end if
end for
Update the best solution.
$t \leftarrow t + 1$.
Until $t >$ *Max no. ofIteration*.
return The best-solution.
end procedure

The model was applied to four benchmark datasets: Zachary Karate Club [25], The Bottlenose Dolphin Network [26], American College Football Network [27], and the Facebook dataset [28]. The algorithm given is run with 10 iterations, and NMI and modularity are processed. The algorithm's iteration number is 10 times, and average NMI and Modularity are maintained. The parameters of CSA were as follows: maximum number of iterations was 100, trials were 10, α was 1.5, and $s = 1000$.

To compare the performance of proposed model, results are compared with different algorithms used to discover communities in social networks such as discrete Bat Algorithm [24], Artificial Fish Swarm Algorithm [29], Multilevel [30], Walktrap [31], A Discrete Krill Herd Swarm Optimization (AKHSO) [32], Ant Lion Algorithm [33], Lion Optimization Algorithm [34] and Cuckoo Search Algorithm.

Fig. [2] shows the average Modularity, Fitness and NMI of the proposed model before and after applying DBSCAN on the social networks we used to remove the noises nodes, as shown after eliminating the noise nodes from the datasets the Modularity, Fitness and NMI values are increased. Fig. [3] shows the Modularity of some previous algorithms used to detect communities in Social Network and the proposed model was superior. Also, Fig. [4] shows the comparison between the NMI values of some previous algorithms and the proposed one. The proposed model gets the highest values for NMI. Fig. [5] reports that the accuracy, precision, recall, and F-measure of the proposed model are neither improved nor changed by using DBSCAN to eliminate noise nodes from the datasets.

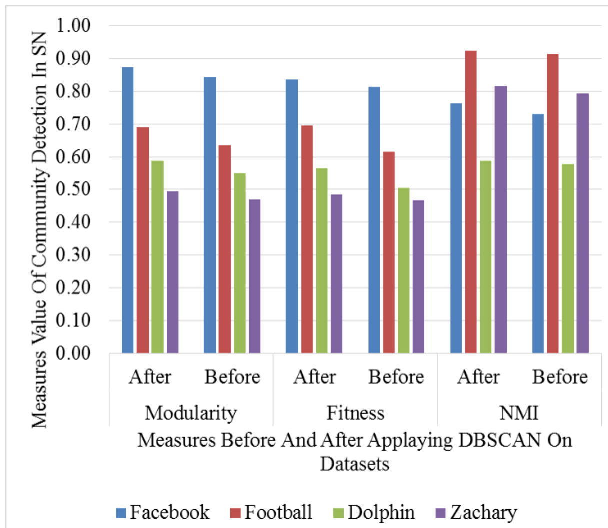

Figure 2: Comparison Between Modularity, Fitness and NMI Values of Social Networks Datasets Before and After Eliminate Noise Nodes.

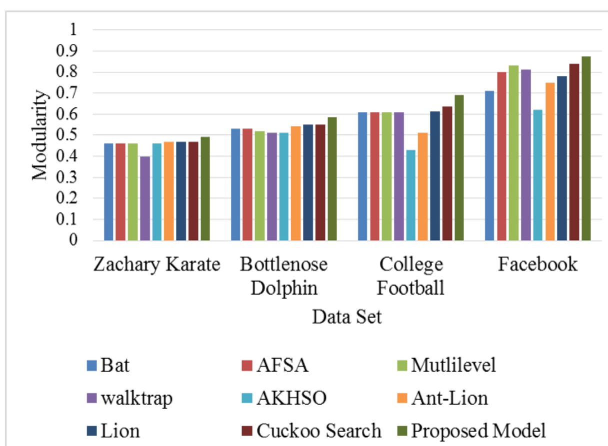

Figure 3: Comparison Between Modularity Values of Different Algorithms and The Proposed Model.

Table 1: The Number of Nodes, Edges and Communities Before and After Applying DBSCAN on Tested Datasets and Eps. Values.

Dataset	Before DBSCAN		After DBSCAN				Eps	Number of Communities	
	Nodes	Edges	Nodes	Edges	% Nodes Removed	% Edges Removed		Before DBSCAN	After DBSCAN
Facebook	3959	168486	3582	127516	9.52	24.32	5.70	156	139
Football	115	1226	106	1036	7.83	15.50	3.75	11	13
Dolphin	62	318	54	236	12.90	25.79	2.80	5	5
Zachary	34	156	30	140	11.76	10.26	1.00	4	4

Table 2: Average Number of Communities, Time, Mean, and Standard Deviation of The Tested Datasets According to Modularity and NMI Measures.

Dataset	Eps.	Average Number of Communities	Modularity			NMI		
			Time (ms)	Mean	Standard Deviation	Time (ms)	Mean	Standard Deviation
Facebook	5.70	144.20	183860.731	0.64684	0.00758	184287.010	0.70325	0.01083
Football	3.75	8.70	4867.701	0.84845	0.02548	4895.620	0.60957	0.00429
Dolphin	2.80	4.60	3277.562	0.57915	0.00766	3332.668	0.54896	0.00080
Zachary	1.00	4.00	2579.918	0.68730	0.00000	2562.486	0.46960	0.00000

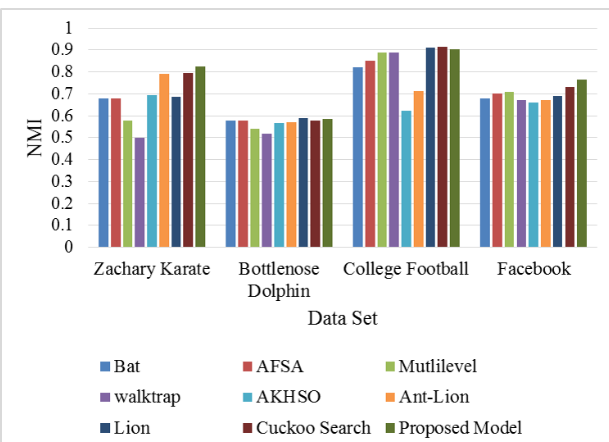

Figure 4: Comparison Between NMI Values of Different Algorithms and The Proposed Model.

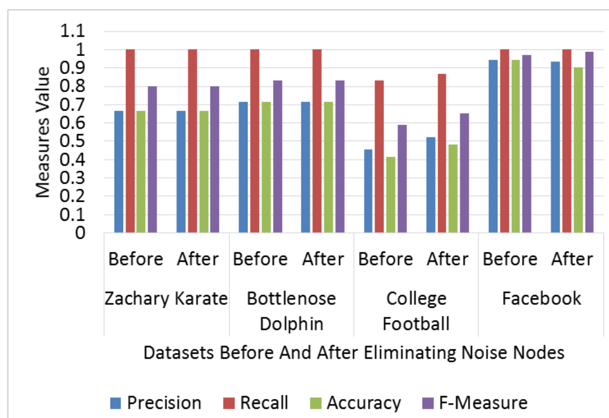

Figure 5: Recall, Accuracy and F-Measure Values of Proposed Model Before and After Applying DBSCAN on Datasets.

Table [1] shows the number of nodes and edges of each benchmark dataset before and after using DBSCAN to remove the noise nodes from the datasets. By removing those nodes, the processing time will decrease and the accuracy of communities will increase. The table shows that the number of communities detected of American College Football Network dataset increased after applying DBSCAN, which means that CSA became able to detect more communities after removing noise nodes.

The new approach is run 10 times for all tested data sets. The mean and standard deviation are obtained and shown in Table [2]. It is observed from the table that the new approach 1 is stable for all data sets tested here; the standard deviation values are zero for the smallest data set and grow when the dataset size grows, but are all closest to the zero value, thus the data sets' distribution is normally distributed.

5. Discussion and Limitations

Even as the new proposed model is demonstrated to be better than the benchmark datasets, its limitations need to be noted so as to provide a balanced perspective and guide future work.

5.1. Parameter Sensitivity:

One of its primary limitations is that the DBSCAN algorithm is too sensitive to its parameters, ε (eps) and MinPts. The noise-removal preprocessing step depends much on the selection of these parameters' values. In this study, these parameters were empirically determined for each dataset. However, for totally new or extremely large networks with no visible ground truth, such empirical tuning might be challenging and can even not lead to the best result. As mentioned in future work, this parameter choice can be made automatic, e.g., within a meta-heuristic optimization cycle, which is a crucial means of enhancing the model's generalizability and applicability.

5.2. Computational Complexity:

The computational cost of the model should also be taken into account. The Cuckoo Search Algorithm, as most population-based metaheuristics, calculates the fitness of numerous candidate solutions over numerous iterations. Node removal through DBSCAN reduces the size of the problem, but the CSO complexity may become a bottleneck in extremely large-scale networks. Alternative, more efficient encoding schemes or parallel computing schemes might be explored in future deployments to circumvent this issue.

5.3. Network Type Scope:

Static unweighted networks were the scope of the model being tested. Most social networks in the real world are dynamic networks whose composition evolves with time, with weighted edges representing the strength of relationships. These aspects are not considered in the proposed framework explicitly. Extending the model to operate in dynamic networks using incremental clustering techniques, or in weighted networks through the modification of the density in DBSCAN and the fitness function in CSO, is a significant and valuable area of future research.

Conclusion

This research has introduced and validated a novel hybrid model for community detection in SN. The model innovatively combines the Cuckoo Search Algorithm (CSA) with DBSCAN preprocessing. An initial application of CSA identifies a preliminary community structure, the count of which informs the DBSCAN parameterization to effectively filter out noise and outlier nodes. A subsequent execution of CSA on the purified network then determines the final, optimized community partition.

The proposed model was subjected to rigorous comparative analysis against a range of established algorithms, including the Bat Algorithm, AFSA, Multilevel, Walktrap, AKHSO, Ant-Lion Optimizer, Lion Optimization Algorithm, and standard Cuckoo Search across four benchmark datasets. The results, tested by Modularity and NMI, consistently demonstrated the superiority of our approach. The significant finding is that the elimination of noise individuals through DBSCAN directly contributes to more accurate and robust community detection, thereby enhancing the performance of swarm intelligence algorithms in the complex domain of social network analysis. As explained in Section 5, future research will address the automatic optimization of parameter DBSCAN, scalability for extensive networks, and model extension to weighted and dynamic network structures.

Conflict of interest

The author confirms that there are no conflict of interests.

Funding

The research received no fund.

References

[1] A. Mahajan, M. Kaur, Various approaches of community detection in complex networks: A glance, *International Journal of Information Technology and Computer Science (IJITCS)*, **8**(4), 35 (2016).

[2] S. Bansal, S. Bhowmick, and P. Paymal, Fast community detection for dynamic complex networks, in: Complex Networks, Springer, 196–207 (2011).

[3] T. Alzahrani, K. Horadam, Community detection in

bipartite networks: Algorithms and case studies, in: Complex Systems and Networks, Springer, 25–50 (2016).

[4] M. Rosvall, C. T. Bergstrom, Maps of random walks on complex networks reveal community structure, *Proceedings of the National Academy of Sciences*, **105**(4),1118–1123 (2008).

[5] W. Wang, D. Liu, X. Liu, and L. Pan, Fuzzy overlapping community detection based on local random walk and multidimensional scaling, *Physica A: Statistical Mechanics and its Applications*, **392**(24), 6578–6586 (2013).

[6] J. Huang, H. Sun, Q. Song, H. Deng, and J. Han, Revealing density-based clustering structure from the core-connected tree of a network, Knowledge and Data Engineering, *IEEE Transactions*, **25**(8), 1876–1889 (2013).

[7] A. Medus, G. Acuna, and C. Dorso, Detection of community structures in networks via global optimization, *Physica A: Statistical Mechanics and its Applications*, **358** (2), 593–604 (2005).

[8] M. Wang, C. Wang, J. X. Yu, and J. Zhang, Community detection in social networks: an in-depth benchmarking study with a procedure-oriented framework, *Proceedings of the VLDB Endowment*, **8**(10), 998–1009 (2015).

[9] M. Salama, M. Panda, Y. Elbarawy, and A. E. Hassanien, A. Abraham, Computational social networks: Security and privacy, in: Computational Social Networks, Springer, 3–21 (2012).

[10] W. Maharani, A. A. Gozali, Collaborative social network analysis and content-based approach to improve the marketing strategy of smes in Indonesia, *Procedia Computer Science*, **59**, 373–381 (2015).

[11] M. Arab, M. Afsharchi, Community detection in social networks using hybrid merging of sub-communities, *Journal of Network and Computer Applications*, **40**, 73–84 (2014).

[12] M. E. Newman, M. Girvan, Finding and evaluating community structure in networks, *Physical review E,* **69**(2), 026113 (2014).

[13] J. Huang, H. Sun, Y. Liu, Q. Song, and T. Weninger, Towards online multiresolution community detection in large-scale networks, *PloS one,* **6**(8), e23829 (2011).

[14] A. Lancichinetti, S. Fortunato, and F. Radicchi, Benchmark graphs for testing community detection algorithms, *Physical review E*, **78**(4), 046110 (2008).

[15] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, *Kdd*, **96**, 226–231 (1996).

[16] S. Chakraborty, N. Nagwani, and L. Dey, Performance comparison of incremental k-means and incremental dbscan algorithms, *arXiv preprint,* arXiv:1406.4751.

[17] Y. M. ElBarawy, R. F. Mohamed, and N. I. Ghali, Improving social network community detection using dbscan algorithm, Computer Applications & Research (WSCAR), 2014 World Symposium on, IEEE, 1–6 (2014).

[18] T. You, B.-C. Shia, and Z.-Y. Zhang, Community detection in complex networks using density-based clustering algorithm, *Journal of Complex Networks*, **5**(3), 457–473 (2017).

[19] X.-S. Yang, S. Deb, Cuckoo search: recent advances and applications, *Neural Computing and Applications*, **24**(1), 169–174 (2014).

[20] A. R. Yildiz, Cuckoo search algorithm for the selection of optimal machining parameters in milling operations, *The International Journal of Advanced Manufacturing Technology,* **64**(1-4), 55–61 (2013).

[21] X.-S. Yang, S. Deb, Cuckoo search via l´evy flights, in: Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on, IEEE, 210–214 (2009).

[22] C. Shi, Y. Wang, B. Wu, and C. Zhong, A new genetic algorithm for community detection, Complex Sciences, Springer, 1298–1309 (2009).

[23] C. Pizzuti, Community detection in social networks with genetic algorithms, in: Proceedings of the 10th annual conference on Genetic and evolutionary computation, ACM, 1137–1138 (2008).

[24] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, Comparing community structure identification, *Journal of Statistical Mechanics: Theory and Experiment*, **2005**(09), P09008 (2005).

[25] W. W. Zachary, An information flow model for conflict and fission in small groups, *Journal of anthropological research*, 452–473 (1977).

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

[27] M. Girvan, M. E. Newman, Community structure in social and biological networks, *Proceedings of the national academy of sciences*, **99**(12), 7821–7826 (2002).

[28] J. J. McAuley, J. Leskovec, Learning to discover social circles in ego networks., *NIPS*, **2012**, 548–56 (2012).

[29] E. A. Hassan, A. I. Hafez, A. E. Hassanien, and A. A. Fahmy, Community detection algorithm based on artificial fish swarm optimization, Intelligent Systems', Springer, 509–521 (2015).

[30] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, *Journal of statistical mechanics: theory and experiment*, **2008**(10), P10008 (2008).

[31] P. Pons, M. Latapy, Computing communities in large networks using random walks, Computer and Information Sciences-ISCIS 2005, Springer, 284–293 (2005).

[32] K. Ahmed, A. I. Hafez, A. E. Hassanien, A discrete krill herd optimization algorithm for community detection, 11th International Computer Engineering Conference (ICENCO), IEEE, 297–302 (2015).

[33] R. Babers, N. I. Ghali, A. E. Hassanien, and N. M. Madbouly, Optimal community detection approach based on ant lion optimization, 11th International Computer Engineering Conference (ICENCO), IEEE, 284–289 (2015).

[34] R. Babers, A. E. Hassanien, and N. I. Ghali, A nature-inspired metaheuristic lion optimization algorithm for community detection, 11th International Computer Engineering Conference (ICENCO), IEEE, 217–222 (2015).

[35] S. Yassine, S. Kadry, and M.A. Sicilia, Detecting communities using social network analysis in online learning environments: Systematic literature review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **12**(1), p.e1431 (2022).

[36] X. Su, S. Xue, F. Liu, J. Wu, J. Yang, C. Zhou, W. Hu, C. Paris, S. Nepal, D. Jin, and Q. Z. Sheng, A comprehensive survey on community detection with deep learning. *IEEE transactions on neural networks and learning systems*, **35**(4), 4682-4702 (2022).