International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
***Article Received**: 01 October 2022 **Revised**: 12 November 2022 **Accepted**: 29 November 2022*

Deep Learning in Social Networks for Overlappering Community Detection

Pranavati Bajrang Jadhav[1], Dr Vijaya Babu Burra[2]
[1]Ph.D Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP , 522302, India,Orcid: 0000000313062836
Email: pranavati.jadhav@gmail.com
[2]Professor, Department of CSE, Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP , 522302, India. Orcid: 0000000301398025
Email: vijay_gemini@kluniversity.in

Abstract— The collection of nodes is termed as community in any network system that are tightly associated to the other nodes. In network investigation, identifying the community structure is crucial task, particularly for exposing connections between certain nodes. For community overlapping, network discovery, there are numerous methodologies described in the literature. Numerous scholars have recently focused on network embedding and feature learning techniques for node clustering. These techniques translate the network into a representation space with fewer dimensions. In this paper, a deep neural network-based model for learning graph representation and stacked auto-encoders are given a nonlinear embedding of the original graph to learn the model. In order to extract overlapping communities, an AEOCDSN algorithm is used. The efficiency of the suggested model is examined through experiments on real-world datasets of various sizes and accepted standards. The method outperforms various well-known community detection techniques, according to empirical findings.

Keywords- Social Network; Community Detection; Neural Computing; Auto Encoder.

I. INTRODUCTION

Numerous real-world systems, including the social Networks, Internet, cooperation networks, proteins-proteins interaction network, citation networks, biochemical networks and metabolic networks can be shown as networks, where nodes stand in for persons and edges for the connections or interactions between those individuals. A key characteristic of networks is community structure, if a network can be easily separated into node groups with denser internal connections and sparser external connections, then it is said to have community structure [M. Girvan and M. E. J. Newman (2002)]. As social networks developed to sizes that were far too large to examine manually, it became more crucial to create accurate and efficient computational algorithms that could highlight key aspects of a network. [M. K. Goldberg, *et al*. (2011)] Understanding the formation and operation of massive networks requires an understanding of communities. The community detection has drawn a lot of attention in recent years as it advances our knowledge of the underlying structure of many real-world networks and holds up the prospect of a wide range of useful applications for government, industry, academics, and biomedical firms. From the identification of functional modules within social networks, it is easy to the study of the behavior of communities. [Zhou Z, *et al*. (2017)] Understanding the pertinent literature and investigating the research insights from the Complex Networks, whose complexity is rising, is worthwhile to study in depth. Since identifying community structure is a difficult issue, several strategies, including modularity optimization, dynamic label propagation, statistical inference, spectral clustering, information-theoretic methods, and topology-based methods, have been developed in the past ten years. There are many different kinds of social media networks, and it is crucial to both study and analyses them. These social media platforms serve as a key source of intelligence since them effectively and correctly encode both online actions and participant input. The community types depend upon the interaction within the nodes internally and externally [Fortunato, S. and Hric, D (2016)]. The disjoint community has stronger bonds with predefined nodes and sparser relation with outsider nodes [Pranavati J.and Dr. Vijaya Babu B (2019)]. The overlapping community, a node may simultaneously be a member of multiple communities, because a node may be present in multiple communities at once. Overlapping groups have traditionally been justified more intuitively than quantitatively. The lack of a generally accepted definition of what constitutes a community, which allows communities to overlap, and a quantitative analysis of large-scale social networks that show a sizable number of communities that have non-trivial overlap with deep learning techniques are the two issues we address in this paper.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
Article Received: 01 October 2022 Revised: 12 November 2022 Accepted: 29 November 2022

II. LITERATURE REVIEW

Perhaps the most active area of the new interdisciplinary science of complex systems is the contemporary study of networks. Networks are a common representation for complicated systems. Links represent the interactions between the nodes, which serve as the system's fundamental building elements in this instance. [Du N, *et al*. (2007)]. Since identifying community structure is a difficult issue, several strategies, including modularity optimization, dynamic label propagation, statistical inference, spectral clustering, information-theoretic methods, and topology-based methods, have been developed in the past ten years. Listed important methods as to start with **graph partition** by Girvan and Newman's, made proposal for GN benchmark [Newman M. (2004)]. It was the initial method of community discovery for graphs at the time. The concept of modularity has been developed to assess the effectiveness of community algorithms. According to the value of their betweenness, which reflects the number of shortest paths between pairs of nodes that pass through the link, connections are iteratively deleted in this **hierarchical divisive** algorithm [Tianxi Li, Lihua, *et al*. (2014)]. The link elimination process comes to an end in its most widely used implementation when the resulting partition's modularity reaches its maximum. A well-known quality function called modularity by Newman and Girvan calculates the goodness of a partition by comparing it to a null model, which is a set of random graphs with the same prediction score as the original graph. Bipartite graphs can be viewed as **hyper graphs** with individuals at vertices and properties at hyper edges to detect overlapping communities. Alternatively, you can view properties as nodes and individuals as hyper edges. **Clustering** is the process of assembling a collection of related objects into structures called clusters. The social network graph can be clustered to reveal a wealth of information about the participants' relationships, qualities, and hidden attributes as well as how they interact with one another. The two most popular clustering strategies in literature are **hierarchical clustering** and the partitioning method. Communities are gradually built by eliminating the edges from the initial graph that the algorithm found to be most "between" the communities [Nicholas Monath, *et al*. (2019)]. In a genetic algorithm, chromosomes are the initial set of solutions for which fitness functions are calculated. If the best-fitting solution is discovered, one continues; if not, one employs crossover and mutation operators to generate the new set of solutions from the existing set with a specific probability. In the context of community detection, an objective function that matches the intuition of a community with better internal connectivity than outward connectivity may be chosen to be optimized. [O. E. David and I. Greental, (2018)]. The spreading of a label among various nodes in a network is known as **label propagation.** Each node acquires the label held by the greatest possible number of its neighboring nodes. Some label propagation-based community discovery strategies are covered in this section. Random walks are performed to detect communities in **dynamical method.** For example, the random walk determines node distances and the probability of community participation. The minimal-length encoding is used by Information Mapping (InfoMap). Label Propagation Algorithm (LPA) uses an information propagation method to identify diffusion communities by the approach to identify communities after first locating all of the network's cliques [H. Sun, *et al*. (2020)].

III. RELATED WORK

The results of community detection using conventional approaches for gathering links in large complicated networks may not be sufficient. As a result, a new growing branch called deep learning-based community detection has emerged. Its general framework converts high-dimensional data from complex structural relationships into lower-dimensional vectors. As a result, it facilitates knowledge discoveries better than using cutting-edge machine learning and data mining approaches. Further, by jointly recognizing sets of data from nodes, edges, neighborhoods, or multi-graphs, the deep learning technique can be applied to successfully identify communities. A more and more large-scale, highly sparse, complicated structural, dynamic network in real-world scenarios may be examined because of deep learning's capacity to use huge data [D. Jin, *et al*., (2021)]. To improve the understanding of community memberships, the framework with representations can further include non-structural elements, such as node properties.

Community detection becomes an even more difficult task when overlapping communities are taken into account since, in addition to characterizing the facts of the groupings, the organization of the overlapping regions must also be assumed. A module called "community" has a collection of nodes with a lot of commonalities, interactions, and key activities. If any of the nodes took part in the creation of various modules, overlapping communities may exist. Each node has a varied strength or membership value in various modules. Different indicators are used to assess the community's strength. There are several techniques in the literature that seek to identify overlapping community edifice, and it can be challenging to distinguish between the techniques for community detection and the authors' presumptions about how a community should be described. As a response, it can be claimed that when a certain criterion is set to be optimized as an objective function by a community detection technique, an implicit assumption is made relating to the definition of an overlapping community structure. In this section authors will discuss the overlapping community detection methods traditionally used and need for deep learning techniques.

The Clique Percolation Method (CPM) is predicated on the notion that a community is made up of overlapping clique sets. By looking for nearby cliques, CPM identifies all cliques of size k and finds community structures. In the worst scenario, CPM's temporal complexity is exponential because it must compute all of the network's maximal cliques, and as a result, frequently enormous social networks' overlapping communities go undetected by CPM based algorithms such as SCP, FCP and CFinder. Seed expansion-based techniques begin with a tiny cluster of nodes or a single node, and then they add neighboring nodes that have a local benefit function to identify a community. Popularity of this method lies in that this technique IS, LFM, SSE finds seed nodes based on specific criteria, then iteratively expands or merges the seed nodes until a local quality function cannot be further improved. The fitness function used in these methods has three stages: planting, screening, and growing the seed set. Algorithms' flaw is that they produce inconsistent results from run to run. By allowing a node to possess several labels, such as COPRA and SLPA, the label propagation technique has also been expanded to discover overlapping community structures. While label propagation techniques can find overlapping communities in minimal time, the label clustering still exhibits some no determinacy. By partitioning links rather than nodes, link partitioning-based approaches LC, CDAEO, MELC finds overlapping community structures. They transform the original network into a line graph and use disjoint community recognition techniques to find the non-overlapping link communities. The disadvantage of these algorithms is the resolution limit problem. Numerous dynamical algorithms that can also be used to find overlapping communities have been published, and they are based on methods including synchronization, label propagation, spin dynamics, and random walk. Nevertheless, a parameter must be provided to restrict the number of communities a node may join and thus community s becomes unstable.

The [Palla G et al] Cfinder have used Clique Percolation Method (CPM) and Nodes, threshold weight. The method is distinct in that it allows simultaneously examine the network at a higher level of organization and identify the communities that are important to the web of communities. It was unable to identify the overlapping characteristic and hierarchical structure.

The [Lancichinetti Algorithm] have used Fitness function with random Seed Normalized Mutual information and Function for Fitness. One can probe the network at different scales and look into any potential community hierarchies by varying the resolution parameter. An intriguing outcome of our approach is the capacity to quantify the engagement of overlapping nodes in their communities by the metrics of their (node's) fitness with relation to each group they belong to.

The [Shen et al] (EAGLE) have used Agglomerative Hierarchical Clustering and Similarity between two adjusting communities. The algorithm uses an agglomerative framework and works with the set of maximal cliques. Note that only un-weighted and undirected networks are considered in this paper.

The [Evans et al] have used Line graph, Clique graph and Links, Partition. After the line graph has been clustered, overlapping clusters can be produced by mapping the output backward to the input graph. Others have utilized this general strategy on citation graphs. However, because a line graph's size is significantly larger than its input graph's size, line graph approaches are not highly scalable.

The [Du et al (ComTector)] have used Kernels based clustering and Set of all kernels. The nuclei of any possible community are formed by the confluence of maximal cliques. Based on the relative degree matrices, iteratively add the left vertices to their nearest kernels using an agglomerative method. The final division of the network is made up of the finally achieved community structures and other elements. The connection between the community members in social networks and the knowledge dissemination mechanism to better understand network dynamics from both the micro and macro perspectives.

The [Lee et al (GCE)] have used Cliques based expansion and Fitness function. Greedy Clique Expansion is an algorithm that was developed that combines the greedy expansion strategy used in other algorithms with the methodology for detecting cliques based on graph structures. More advanced local algorithms that cleverly explore the most promising regions of the search space should be investigated in future research.

The [Gregory et al] CONGA, CONGO Peacock algorithms have used Split betweenness and Vertex, short paths ratio of max. edge and max. split betweenness. The algorithm can handle a greater overlap as the number of communities grows. Although the algorithm is slow, it is comparable in speed to the GN algorithm from which it was created.

The [Y. Pan et al] have used genetic algorithm approach and In degree, out degree, belongingness Coefficient. For directed graphs with overlapping communities, a modularity function extension was created. Regarding the goal and structure of the original modularity function, we use an enhanced null-model that takes into consideration nodes belonging to many communities simultaneously. The formal method used to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
***Article Received**: 01 October 2022 **Revised**: 12 November 2022 **Accepted**: 29 November 2022*

determine the generalized modularity is unaffected by the function used to quantify an edge's contribution to the modularity of a community, hence it has been left out for further study.

The [Pizzuti et a (GA-NET)] have used GA based and Community score. The approach uses the line graph that corresponds to the graph that models the network to optimize a fitness function that can spot densely connected clusters of nodes. Cluster density not automatically decided by the fitness function's ideal value.

The [S. Sharma and H. M. Pandey (GaoCD)] have used Objective function: Partition density and Size of population, ratio of crossover, mutation. The method utilizes a novel genetic algorithm to cluster on edges and the specific property of links. Phases of nodes are remained to be isolated.

The [Xing et al] (OCDLCE) have used Community Detection and Nodes, edges, neighbours of node. The network's local information to gather small-scale local communities before merging any that have significant overlap. It goes over every node that is not assigned to a community and splits them into communities based on node fitness. The node attributes are not considered for fitness function.

The [Bhat et al (OCMiner)] have used Density based and Threshold Ƞ.Uses a unique distance function that can find communities in unweighted networks and can also use the edge weights in weighted networks. The complexity reaches highest by weighted graphs.

IV. METHODOLOGY

The suggested model has two basic modules: a clustering algorithm component and a unsupervised neural module. The community detection method extracts community data from networks using modularity techniques. The unsupervised neural modules can combine regional and social data with learning vertex centrality from other communities n the social networks. As an outcome, the suggested model uses these two modules to generate the relevant result from the original input data.

1.1. *Modular centrality model*

This section outlines the number of fundamentals that form the foundation of the suggested method. The definition of Intersecting Modular Centrality and the formula to calculate its elements are provided.

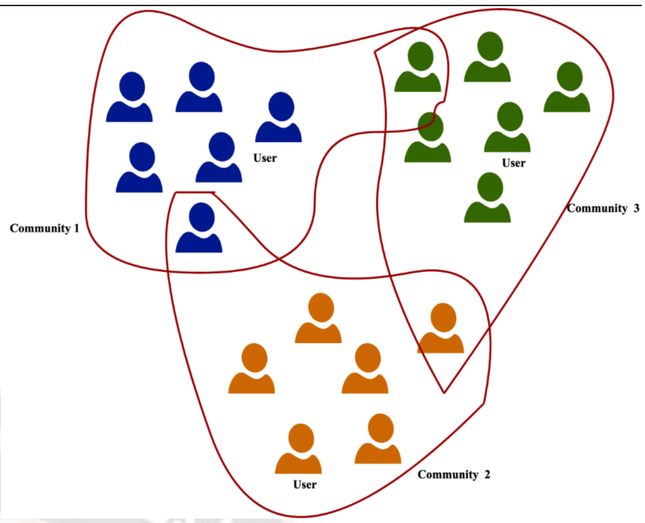

Fig. 1. Overlapping community in social networks

The overlapping community in a social network is shown in Fig. 1. The social media users are clustered based on their interests. The users may have different interests and cluster into different communities. This research suggests an Auto Encoder-based overlapping community detection in a social network. A deep learning model is used to find the interpretation between social network data and local information.

The group of community is denoted by $G(X,E)$, a straightforward undirected network. The formula X where denote the group of cells X_0 and X_{N0} correspondingly, stand for the sets of overlapped and nonoverlapping vertices, and E is the directed graph. N is the total number of villages and $K = \{k_1, k_2, \cdots, k_N\}$ is the collection of societies.

The effect of nodes in systems with an overlapped communal architecture can be divided into two categories.

- A local impact linked to connections with members of society.
  The initial network is divided into a home network that records connections within communities and an extensive network that allows for connections across communities to determine these two aspects. The type of component that is considered affects how the local networks are defined. An overlapped node contains all the groups it belongs to, a nonoverlapping entity shrinks to its specific community.
- A global influence linked to contacts with vertices from nearby communities.
  The community of all linked graph elements that remain after eliminating all intra-community ties and solitary nodes from the core graph is referred to as the global networks, or N_g. The collection of connected elements, $C = \{c_1, c_2, \cdots, c_N\}$, is represented by $N_g = V_x^y C_x$. The community is C_x and the vectex is denoted V_x^y. It

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
***Article Received**: 01 October 2022 **Revised**: 12 November 2022 **Accepted**: 29 November 2022*

defines $M_g(V_x^y)$ as the global element of the overlapped modular coherence of the cluster $N_x \in C_x$. It is calculated using all the C_x elements of the N_x Global graph.

The overlapped modular relevance of separated nodes' global element is zeroed out. The centrality of modules that overlap is computed. For a nodes $v_x \in$ V, the Overlapping Module extension of a centrality measurement M was created for non-modular systems. The overlapping modular extension is denoted in Equation (1)

$$E_{OM}(v_x) = M_L(V_x^y), M_z(V_x^z) \quad (1)$$

The local and global components of the social networks are denoted $\mathbf{M_L}$ and $\mathbf{M_Z}$. The center of the overlapping modularity of a community is denoted $\mathbf{v_x}$. Remember that the recovered local and global systems are not divisions of the initial modular system. These are collections of independently connected parts created by duplicating a few of the initial network' nodes and connections. Algorithm 1 illustrates how to specify the overlapping module center calculation process:

Algorithm Modular Centrality:

input a social network data, output an Overlapping Nodes Matrix
Select a centrality measure for non-modular networks V.
Create a set of the overlapping nodes as local network N_{OL} d as the local neighborhoods.
For each module $M \in M_x$
Calculate local element M for overlapping community
Delete the community edge from internal sets and replicate all the overlapping communities
For each community $K \in K_x$
Calculate the local element M_L for non-overlapping community
Delete the community edges from outlier sets to build the global network M_Z
For each module $C \in C_x$
Compute Overlapping Modular Centrality M_Z of isolated nodes.
Add M_L and M_Z to the Overlapping Nodes Matrix$E_{OM}(v_x)$.

1.2. Auto encoder based overlapping community detection

Information retrieval, image search, picture restoration, machine translation, and feature selection are just a few of the activities that autoencoders are employed. Autoencoders are an unsupervised, deep neural learning technique. Due to the autoencoder's ability to translate significant input data into an intermediate representation, certain applications are made possible. The neural models known as auto-encoders are straightforward but crucial since they transform high-dimensional (network) data into low-dimensional representations.

Auto-encoders specifically using the encoder and decoder components, a new data representation can be unsupportively learned. Their symmetrical architecture and multiple hidden layers are constants, and the output of one layer serves as the input for the layer after it. After comparing the output result with the input result and calculating the reconstruction errors, the automatic encoder's weight matrix is modified using the back-propagation technique. Until the numbers of iterations or the reconstruction errors are fewer than the set range, the reconstruction errors are calculated once more and iterated continually. The outcome is similar to or identical to the input. The minimizing of the reconstruction error describes the procedure of utilizing a back-propagation method to train a neural network. The output of the encoder, or encoding, is finally considered to be the output of the automatic encoder.

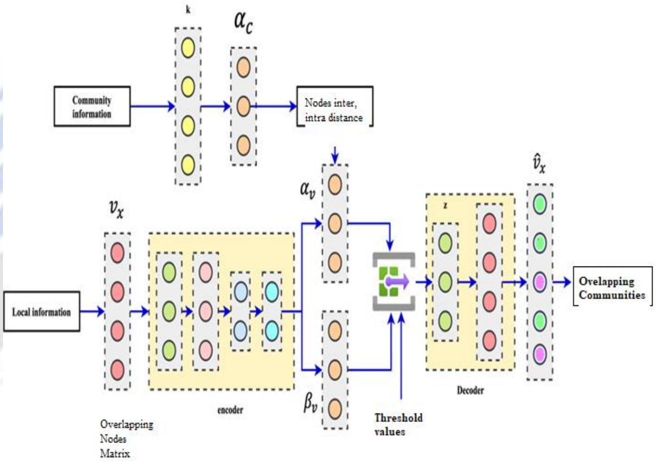

Fig 2: Proposed AEOCDSN Method Diagram

AEOCDSN Algorithm Steps:

As the input matrix, Let $\mathbf{V_x}$ be the social network graph G containing overlapping nodes of dimension n.
where $x_i \in (0, 1)$, $X \in M\ (n \times n)$. $x_i \in M\ (n \times 1)$
$P1 \in M\ (m \times n)$ is the overlapping nodes matrix gives $\boldsymbol{\beta_v}$
$P2 \in M\ (d \times m)$ is the inter and intra distance matrix gives $\boldsymbol{\alpha_v}$
h $\in$ V(d × 1) is the vector of the hidden layer .
q $\in$ V (n × 1) is the vector of the input layer.
The coding layer output is obtained by equation no. 1:

$$\hat{H} = \sum \beta_{V.} [(P1_i . v_x) . b_i]$$

(1)

The output of community information with all nodes s obtained by equation no. 2:

$$\hat{O} = \sum \alpha_v . [(P2_i . v_x) . c_i] \ / n_i$$

(2)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
***Article Received**: 01 October 2022 **Revised**: 12 November 2022 **Accepted**: 29 November 2022*

The matrix M' obtained at this time is a matrix after dimensionality reduction.

The output Ř of the decoding layer is obtained by equation no 3:

$$Ř = \varepsilon\,(\hat{H}\,.\,\hat{O}\,)\,.\,v_x \quad (3)$$

Where Ř ∈ V (n × d) is the decoded ith column vector. ε is the threshold value of centrality modularity for community detected.. The resulting matrix Ř is the same as the input matrix gives final overlapping communities.

The proposed algorithm completes the process with first finding overlapped nodes matrix by considering central modularity of each community. The matrix works as input to the sparse auto encoder along with each node inter and intra distance for community clustering coefficient. The encoder reduces the matrix with high dimensionality with lower dimensionality with threshold value for acquiring overlapping community. The reconstruction process of this matrix is completed by applying threshold vales of specific community locally. The loss function has same values as per input matrix with delta of overlapping threshold values.

1.3. *Experimental Process*

Vertex categorization and overlapping community identification tasks are used in experiments to test how healthy vertex and community models work on real-world datasets. On four frequently used networking datasets, the experiments were run and analyzed on following datasets:

- A network of research citations called Cora was created by research users and their publication. It has 2710 publications on computer domains with 7 labels.
- Another widely used dataset for research articles site is Citeseer. This dataset has 3265 articles and seven categories.
- Wiki is a linguistic network of 12760 edges connecting 2400 websites from 18 groups of users having different likings..
- The computer programming bibliography data in DBLP. Each paper has an abstract, authors, year, location, and title connected with it. The data set can be used for topic modeling analysis, determining the most influential publications, examining influence in the citation network, clustering with network and side information, etc.

Data set	No. of Nodes	No of Edges	Density of Nodes	Clustering Coefficient
Cora	6534	52312	0.03	0.081
Citeseer	1864	43545	0.08	0.092
Wiki	1432	36542	0.005	0.142
DBLP	2435	14354	0.07	0.131

Table 1. Dataset description

The dataset description is shown in Table 1. Four different datasets are considered for the analysis: Cora, Citeseer, Wiki and DBLP. The number of social media users in the dataset is considered the number of nodes; the number of the edges, density of the nodes and clustering co-efficient are analyzed. The lower density dataset results in loose interconnectivity between nodes, and lesser clustering co-efficient lead to high chances for getting more overlapping community.

The accuracy and modularity of the AEOCDSN system are represented in Fig. 3(a) and 3(b). The social network data is collected from different datasets and clustered into several groups. The community is varied from 1 to 8 for the analysis with the different overlapping communities. The accuracy in detecting overlapping communities and modularity to find the social media data in the overlapping community are computed. The accuracy is computed as the ratio of correctly classified social network data to the total data. The modularity is computed as the ratio of the deviation in one community to the other community in social networks.

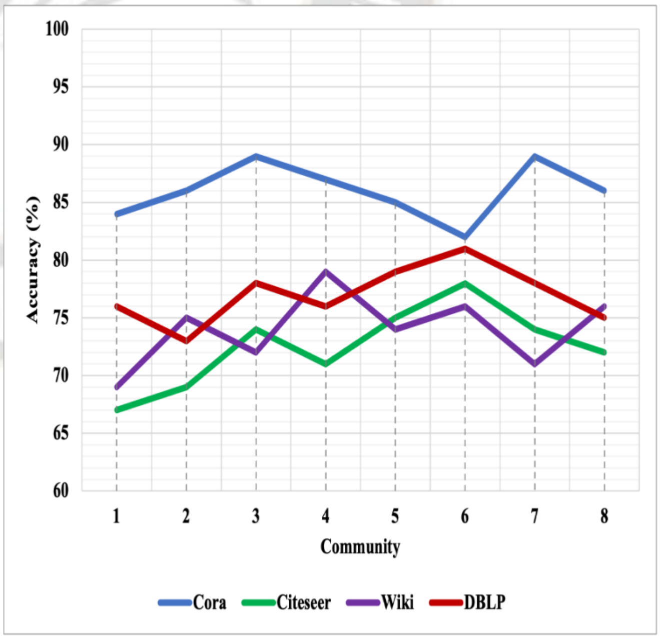

Fig. 3(a). Accuracy analysis of AEOCDSN system

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
***Article Received**: 01 October 2022 **Revised**: 12 November 2022 **Accepted**: 29 November 2022*

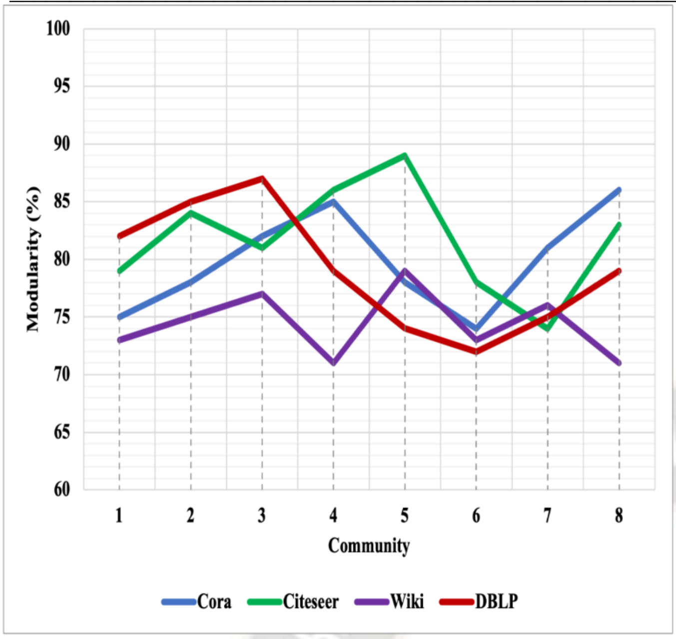

Fig. 3(b). Modularity analysis of the AEOCDSN system

V. Result and Discussion:

The simulation outcomes of the proposed AEOCDSN in terms of F score, modularity, normalized mutual information, and conductance are computed and compared with existing deep and machine learning models in Figure 4a and 4b. Results from the AEOCDSN are associated with those from currently used systems, including Support Vector Machine (SVM), Fuzzy Inference System (FIS), Convolutional Neural Network (CNN), Deep Neural Network (DNN), Decision Tree (DT), and Random Forest (RF).

The accuracy of a test is measured by the F-measure, also referred to as the F-score, which is employed in statistical analyses of binary categorization. The formula for F-Measure is (2 * Precision * Recall) / (Precision + Recall). The value for F-score is improved by use of threshold values of distances of overlapped nodes.

Modularity counts the number of edges within each community and the number of edges that leave the community for each community. Even in the extreme instance when they are only connected to the rest of the network through a single edge, modules less than the minimal size may not be resolved through modularity optimization. This has the practical effect of allowing smaller clusters to "hide" inside bigger clusters, making it challenging to locate them. The proposed system has given centrality modularity with average values for each community nodes.

NMI, or normalized mutual information, is a commonly used metric to contrast various community discovery approaches. However, due to their propensity to select clustering solutions with more communities, information theoretic-based measures have recently been criticized for the necessity for adjustment. In order to fully explore this issue, an experimental evaluation of these metrics is conducted in this study, and a scaling modification is suggested. The more equitable behavior of scaled NMI is highlighted by experiments on artificially formed networks and a comparison of several well-known community detection approaches, particularly when the network topology does not clearly show a community structure.

The computed results of the proposed AEOCDSN outperform the existing deep and machine learning models with the autoencoder and deep learning model. The proposed algorithm corrects the NMI score as the auto encoder lowers the complexity of calculation. The overall comparable results with existing systems are given in Table 2.

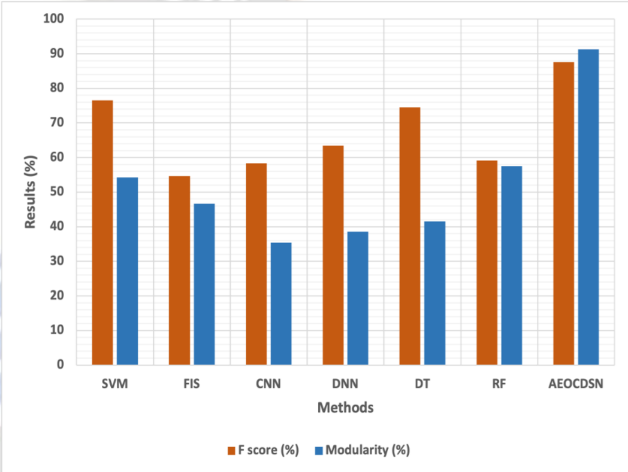

Fig. 4(a). F score and modularity analysis

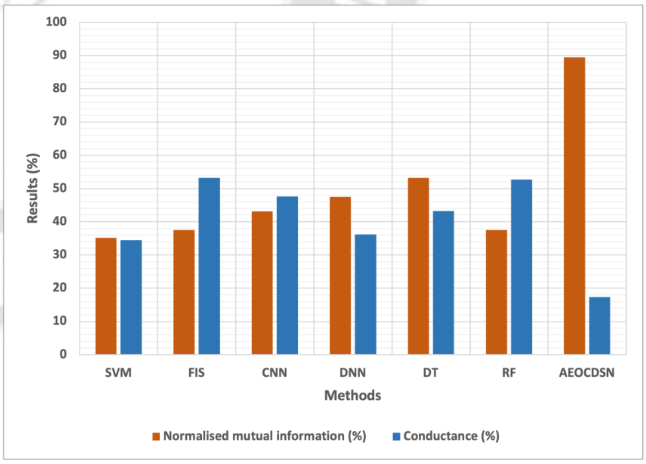

Fig. 4(b). Normalized mutual information and conductance analysis

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 12
DOI: https://doi.org/10.17762/ijritcc.v10i12.5839
***Article Received**: 01 October 2022 **Revised**: 12 November 2022 **Accepted**: 29 November 2022*

Methods	F score (%)	Modularity (%)	Normalized Mutual Information (%)	Conductance (%)
SVM	76.5	54.2	35.2	34.5
FIS	54.7	46.7	37.5	53.2
CNN	58.3	35.4	43.1	47.6
DNN	63.4	38.6	47.5	36.2
DT	74.5	41.5	53.2	43.2
RF	59.2	57.5	37.5	52.7
AEOCDSN	87.6	91.3	89.5	17.3

Table 2. Simulation outcome analysis

VI. CONCLUSION

The current existing approaches are restricted to overlapping community detection in social networks with limited scope for correctness of final output. They are unable to manage real-world networks where nodes frequently migrate between communities, join or leave communities, and change their activity levels at random or on a regular basis. To identify the overlapping community in the social network, the Auto Encode for Overlapping Community Detection in Social Network (AEOCDSN) is recommended. Clustering and an sparse autoencoder based deep learning combined model detect the overlapping community in social networks. The proposed system enhanced the centrality concept of community with deep learning autoencoder. It produces lesser loss than other existing deep learning and machine learning models. The proposed system checked several parameters to approve the expected results and the AEOCDSN proved with better and efficient algorithm to detect overlapping community in social networks. The system's performance is enhanced using artificial intelligence and big data analytics to optimize the results in the future.

VII. FUTURE WORK

The proposed system is working for dimensionality reduction concept with few parameters of communities on social networks. In real time the nodes behave differently or dynamically at various timestamps so detecting such overlapping community will be part of future work.

REFERENCES

[1] A. Rajagopal, G. P. Joshi, A. Ramachandran, R. T. Subhalakshmi,M. Khari, S. Jha, K. Shankar, and J. You, ``A deep learning model based on multi-objective particle swarm optimization for scene classification in unmanned aerial vehicles," IEEE Access, vol. 8, pp. 135383_135393, 2020, doi: 10.1109/ACCESS.2020.3011502.

[2] Anisha, P.R., Vijaya Babu, B., CEBPS: Cluster based effective breast cancer prediction system, International Journal of Recent and Technology Engineering,2019

[3] Bhat SY, Abulaish M. OCMiner: A Density-Based Overlapping Community Detection Method for Social Networks. Intelligent Data Analysis,IOS Press, 2015, 19(4),1-31. doi:10.3233/IDA-150751.

[4] C. Pizzuti, "Evolutionary computation for community detection in networks: A review," IEEE Trans. Evol. Comput., vol. 22, no. 3,

[5] D. Jin, B. Li, P. Jiao, D. He, and H. Shan, Community detection via joint graph convolutional network embedding in attribute network, in Proc. Int. Conf. Artif. Neural Netw. Cham, Switzerland: Springer, Sep. 2019, pp. 594_606, doi: 10.1007/978-3-030-30493-5_55.

[6] D. Jin, B. Li, P. Jiao, D. He, and W. Zhang, "Network-specific variational auto-encoder for embedding in attribute networks,"

[7] D. Jin, B. Li, P. Jiao, D. He, H. Shan, and W. Zhang, "Modelling with node popularities for autonomous overlapping community

[8] D. Jin, X. Wang, D. He, J. Dang, and W. Zhang, "Robust detection of link communities with summary description in social networks," IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2737– 2749, 2021.

[9] Du N, Wu B, Pei X, Wang B, Xu L. Community detection in large-scale social networks. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis: ACM; 2007:16-25.doi:10.1145/1348549.1348552.

[10] Evans T, Lambiotte R. Line graphs of weighted networks for overlapping communities. The European Physical Journal B 2010, 77 (2):265-272. doi:10.1140/epjb/e2010-00261-8.

[11] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," Data Min.

[12] Fortunato, S. and Hric, D., Community detection in networks: A user guide, Physics Reports, Vol. 659, (2016), 1-44.

[13] Genet. Evol. Comput., Jul. 2014, pp. 1451_1452, doi: 10.1145/2598394.2602287

[14] Goyal, P. and Ferrara, E., "Graph embedding techniques, applications, and performance: A survey", Knowledge-Based Systems, Vol. 151, (2018), 78-94.

[15] Gregory S. A fast algorithm to find overlapping communities in networks. In: ECML PKDD : European Conference on Machine Learning and Knowledge Discovery in Databases - Part I: Springer; 2008 : 408-423

[16] H. Sun, F. He, J. Huang, Y. Sun, Y. Li, C. Wang, L. He, Z. Sun,and X. Jia, "Network embedding for community detection in

[17] H. Sun, F. He, J. Huang, Y. Sun, Y. Li, C. Wang, L. He, Z. Sun,and X. Jia, Network embedding for community detection in attributed networks, ACM Trans. Knowl. Discov. Data, vol. 14,no. 3, pp. 1–25, 2020.

[18] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, Accurate efficient and scalable graph embedding,

in Proc. IEEE Int. Parallel Distrib. Process. Symp. (IPDPS), May 2019, pp. 462_471, doi:10.1109/IPDPS.2019.00056.

[19] H. Zhang, B. Qiu, C. L. Giles, H. C. Foley, and J. Yen, An lda based community structure discovery approach for large-scale Social networks in Proceedings of IJCAI, pp. 2663–2669, 2019.

[20] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure in complex networks. New Journal of Physics 2009, 11 (3):033015. doi:10.1088/1367-2630/11/3/033015.

[21] Lee C, Reid F, McDaid A, Hurley N. Detecting highly overlapping community structure by greedy clique expansion. arXiv preprint arXiv:1002.1827 2010.

[22] M. Al-Andoli, W. P. Cheah, and S. C. Tan, Deep autoencoder-based community detection in complex networks with particle swarm optimization and continuation algorithms, J. Intell. Fuzzy Syst., vol. 40, pp. 1_17, Mar. 2021, doi: 10.3233/JIFS-201342.

[23] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proc. Natl. Acad. Sci., vol. 99, no. 12,

[24] M. Gong, Q. Cai, X. Chen, and L. Ma, ``Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition," IEEE Trans. Evol. Comput., vol. 18, no. 1, pp. 82_97, Feb. 2014,doi: 10.1109/TEVC.2013.2260862

[25] M. K. Goldberg, M. Magdon-Ismail, S. Nambirajan, and J. Thompson, Tracking and predicting evolution of social communities.," in Proc. IEEE Int. Conf. Social Computing, (Boston, MA, USA), pp. 780{783, Oct. 2011.

[26] M. K. Rahman and A. Azad, "Evaluating the community structures from network images using neural networks," in Proceedings of Complex Networks and Their Applications, vol. 881, pp. 866–878, 2019.

[27] Newman M. Fast algorithm for detecting community structure in networks. Physical review E 2004, 69 (6):066133.doi:10.1103/PhysRevE.69.066133.

[28] Nicholas Monath, Manzil Zaheer, Daniel Silva, Andrew McCallum, and Amr Ahmed. 2019. Gradient-based Hierarchical Clustering using Continuous Representations of Trees in Hyperbolic Space. In KDD

[29] O. E. David and I. Greental, Genetic algorithms for evolving deep neural networks, in Proc. Companion Publication Annu. Conf.ic

[30] Palla G, Derenyi I, Farhas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. Nature 2005, 435:814–818. doi:10.1038/nature03607.

[31] Pizzuti C. Overlapped community detection in complex networks. In: Proceedings of the 11th Annual conference on Genetic and evolutionary computation: ACM; 2009,859-866.doi:10.1145/1569901.1570019. pp. 464–483, 2018.

[32] Pranavati J.and Dr. Vijaya Babu B. Detection of Community within Social Networks with Diverse Features of Network Analysis Jour of Adv Research in Dynamical & Control Systems, Vol. 11, 12-Special Issue, 2019

[33] Q. Li, J. Zhong, Q. Li, C. Wang, Z. Cao, A community merger of optimization algorithm to extract overlapping communities in networks. IEEE Access. 7, 3994–4005 (2019)

[34] R. Fei, J. Sha, Q. Xu, B. Hu, K. Wang, and S. Li, ``A new deep sparse autoencoder for community detection in complex networks, EURASIP J. Wireless Community. Netway., vol. 2020, no. 1, pp. 1_25, Dec. 2020, doi:10.1186/s13638-020-01706-4

[35] S. Sharma and H. M. Pandey, Genetic algorithm, particle swarm optimization and harmony search: A quick comparison, in Proc. 6th Int.Conf.-Cloud Syst. Big Data Eng. (Con_uence), Jan. 2016, pp. 40_44, doi:10.1109/CONFLUENCE.2016.7508044.

[36] Sandhya S Waghere , PothuRaju Raja Rajeswari, Design and Implementation of System which efficiently retrieve useful data for Detection of Dementia Disease SPRINGER 2021

[37] Sara E Garza and Satu Elisa Schaeer. 2019. Community detection with the label propagation algorithm: a survey. Physica A: Statistical Mechanics and its Applications (2019), 122058.

[38] Shen H, Cheng X, Cai K, Hu M-B. Detect overlapping and hierarchical community structure in networks. Physica A: Statistical Mechanics and its Applications 2009, 388 (8):1706-1712. doi:10.1016/j.physa.2008.12.021.

[39] Shi C, Cai Y, Fu D, Dong Y, Wu B. A link clustering based overlapping community detection algorithm. Data & Knowledge Engineering 2013, 87:394-404. doi:10.1016/j.datak.2013.05.004.

[40] Tianxi Li, Lihua Lei, Sharmodeep Bhattacharyya, Koen Van den Berge, Purnamrita Sarkar, Peter J Bicke, and Elizaveta Levina. Hierarchical community detection by recursive partitioning. J. Amer. Statist. Assoc

[41] X. Cheng, D. Ciuonzo, and P. S. Rossi, Multibit decentralized detection through fusing smart and dumb sensors based on rao test,' IEEE Trans. Aerosp. Electron. Syst., vol. 56, no. 2, pp. 1391_1405, Apr. 2020, doi:10.1109/TAES.2019.2936777

[42] Xing Y, Meng F, Zhou Y, Zhou R. Overlapping Community Detection by Local Community Expansion. Journal of Information Science And Engineering 2015, 31 (4):1213-1232.

[43] Y. Pan, Y. Yang, and W. Li, ``A deep learning trained by genetic algorithm to improve the efficiency of path planning for data collection with multi-UAV," IEEE Access, vol. 9, pp. 7994_8005, 2021, doi:10.1109/ACCESS.2021.3049892.

[44] Y. Xie, X. Wang, D. Jiang, and R. Xu, ``High-performance community detection in social networks using a deep transitive autoencoder," Inf. Sci., vol. 493, pp. 75_90, Aug. 2019, doi: 10.1016/j.ins.2019.04.018.

[45] Z. He, J. Liu, Y. Zeng, L. Wei, and Y. Huang, "Content to node: Self-translation network embedding," IEEE Trans. Knowl. Data Eng., vol. 33, no. 2, pp. 431–443, 2021.

[46] Zhou Z, Wang W, Wang L. Community Detection Based on an Improved Modularity. Pattern Recognition 2012:638-645. doi:10.1007/978-3-642-33506-8_78.