Journal of Theoretical and Applied Information Technology
15th February 2023. Vol.101. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EVOLUTIONARY AUTO ENCODER TECHNIQUE FOR DETECTION OF OVERLAPPING COMMUNITY IN SOCIAL NETWORK WITH INTERPRETABILITY USING AEOCDSN ALGORITHM

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
[2]Professor, Department of CSE, Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP , 522302, India. Orcid: 0000000301398025

Email: Email: pranavati.jadhav@gmail.com, vijay_gemini@kluniversity.in

ABSTRACT

Social networks have increased and are gaining popularity because of the Internet's rapid technological advancement. Researchers' interest in a study on numerous social networks is growing. Community detection is crucial for social networks regarding issues with online security, such as user behavior analysis and anomalous community detection. The unique feature of a social network community is the multi-membership of a node, which leads to overlapping communities. However, solving the overlapping community's detection issue costs much computing. In social networks, communities develop as a result of nodes' self-interest. The social network's nodes play the role of self-interested people who want to get the most from interactions as communities. An Auto Encoder for Overlapping Community Detection in Social Network (AEOCDSN) is suggested in this paper to detect the overlapping community in the social network. This detection helps in recommending and understanding user interest in different matter. Low-dimensional vertices models are learned using Label Propagation Algorithm (LPA) algorithms and community input in the suggested model. The modularization architecture makes the system more adaptable with a lesser loss function. Overlapping community detection techniques are used to gather community data in the social network. In addition to effectively integrating and preserving data from regional and global viewpoints, LPA and Infomap approaches can improve the models' robustness.

Keywords – *Sparse Auto Encoder, Overlapping Community, Social Network, LPA, Infomap*

1. INTRODUCTION

The most famous websites on the Internet nowadays are social media platforms like Facebook, Snapchat, and Instagram, which typically include millions or billions of members [1]. These users communicate with one another, creating densely knit communities that go by the name of communities. A group of more connected users with similar characteristics or interests (such as birthplace, school, or city of residence) is called a tribe. A community is described as a group of tightly linked endpoints inside and outside. Because the framework of online communities is typically represented as a graph for whom the vertices represent the community members and edges portray the links among them, such as companionship, chat logs, etc.

The relationships between people in online communities shift with time in a lively way. Consider Facebook as an instance: individuals can add or remove friends, new users, and individuals who have left the site. The dynamics of networks suggest that communities change in composition over time and undergo continuous evolution. Recently, the scientific community has begun to pay attention to the complex problem of detecting developing groups in large systems [2]. Community detection is an effective method for understanding the internal organization of the system, the interactions among members, and how they engage, even for extensive systems, indicating its applicability. Its uses for community discovery exist, including connection inference, recommender, and friend suggestion.

Community detection is essential for comprehending

various network architectures in the real world [3]. It significantly impacts the research of complex processes and daily lives. According to the network's ecosystems, specific networks are linked tightly. Using the hierarchical architecture of graphs as its source of knowledge, community detection groups vertices into units in the graph such that the interior side counts of the component are higher than the line connecting components [4].

In the literature, a lot of population detection techniques were suggested. Most of them just highlighted the linear features. However, real-world networks' vertices have non-linear connections with one another [5]. A convolutional neural network-based classifier effectively produces non-linear, lower-dimensional annotations. Data point translation into lower-dimensional domains, such as clustering algorithms, is a job best adapted for auto-encoder [6]. Recent studies have demonstrated the viability and spectral clustering algorithms as alternatives to conventional neural network-based Sparse Auto

Encoders for representing lower-dimensional data [7]. While the current technique finds communities utilizing complex patterns, it only identifies disjoint groups without considering the graph's architecture. Additionally, the graph's structure is essential for massive graphs and includes a significant role in finding the dimension of the group.

The primary contribution of this paper is listed below

- Overlapping community detection in social media using Sparse Auto Encoder is suggested in this research, namely AEOCDSN.
- Label Propagation Algorithm (LPA) and info map methods are combined to find the overlapping community detection in social networks.
- A sparse autoencoder is designed to reduce the problem of overfitting and hence enhance the accuracy.
- An agent-based model evaluation is proposed to find the community's social power, social exclusion, and interaction diversity.

The remainder of the paper is organized as follows: section 2 discusses the background to the social networks and different overlapping community detection models. An Auto Encoder for Overlapping Community Detection in Social Network (AEOCDSN) is suggested to detect the overlapping community in social media using LPA-INFO and

Sparse Auto Encoder, as shown in section 3. The software analysis and simulation findings are enumerated in section 4. The conclusion and future study of the research are listed in section 5.

2. LITERATURE SURVEY

Two main issues were identified in the detection of overlapped and nonoverlapping groups. In other words, it was difficult to easily apply the techniques to find nonoverlapping communities to their overlapping equivalents. As a result, many different techniques have been created to display the architecture of overlapped groups.

The most critical subfields of analysis of social networks were community identification. The detection of overlapped groups is the main topic of this work suggested by Asmi et al. [8]. The initial community selection and expansion were the key two stages of the suggested methodology. A new weighted membership level was suggested to be used to increase accuracy. The findings were evaluated using the normalized similarity matrix and the overlapped version of modules. Research on many synthetic and actual networks demonstrated the high quality of outcome communities.

The prior setting of the community count was one of the computation's main drawbacks. The process employed a modularity approach to generate several overlapped communities was suggested by Hasan et al.to overcome the drawback of fuzzy grouping [9]. There were two main phases to the approach. The technique calculated the similarity measure to investigate the characteristics of the vertices in the initial stage utilizing a random walk. In the second stage, overlapping groups were found using a fuzzy clustering algorithm that utilizes non-linear feature reduction. The approach used the fuzzy classification approach as input for low-dimensional information. The usefulness of this approach has been demonstrated in experiments using both actual and synthetic networks.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Zhou et al. [10] introduced a game theory-based method to identify overlapped structures and functions. The process of social action is turned into a game when all participants (nodes) cannot increase their utility. The utility function comprised profit and transfer functions, and this research introduced a novel gaining experience. In addition, two new methods for every agent to refresh its name were devised throughout the tournament. Every agent assessed and contrasted the approaches to determine the best outcome. Thatwas distinct from selecting action at chance among joining, quitting, and swapping on every agent to acquire a new label.

A technique for identifying overlapping groups in complicated networks was presented by Ma et al. [11]. The method offered the overlapped society that the clique served as the center of the community. Each repetition of the method chose the component with the highest fitness score of the network neighbor. The process started with a population of random vertices. The community was expanded to encompass all the datatype k-cliques. Negatively fit vertices were eliminated during the procedure. The split of network communities was realized, and overlapped nodes were found.

Chu et al. proposed a fuzzy clustering-based social network community discovery methodology for social networks [12]. A strategy for mending imperfect fuzzy relations built on split communities were put forward. The research presented a two-stage agreement reaching technique to combine the number of rounds and alteration of the initial choice data, where the group proximity center of the society was employed to gauge its significance. It was the process for large-scale group planning with intuitionistic fuzzy relationships utilizing social networking community research.

Based on overlapping individuals, a community discovery algorithm for use across various social media platforms was suggested by Zhu et al. [13]. After defining the idea of overlapping individuals, an algorithm was created to find the stub groups among overlapped users according to their social importance. Following that, it added members with high levels of similarity to every stub group across various social media sites, and various communities were determined across networks. Finally, tests were run to confirm the community's rationality.

Gao et al. suggested a unique seeding strategy for a community detection discovery technique for large-scale networks that relies on local growth [14]. Additionally, it improved the conductivity of populations by (1) altering incorrect community connections through degrees of activity, (2) merging highly overlapping populations with a novel merging function, and (3) locating populations for outliers using the proposed theory. Optimization improved community reliability optimization, according to experimental findings in social networks.

Zhao et al.provided an incremental strategy to detect community by managing subgraphs to overcome the classification issue [15]. The research suggested various update tactics. It demonstrated the algorithms for progressively detecting groups in dynamic developing networks. The suggested method was efficient and performed better than numerous employed methods, according to test findings on actual data sets.

Without disclosing a lot of communities in advance, you et al. offered a three-stage approach to detect groups utilizing local and worldwide information [16]. The stages identified the core network, label dissemination, and community combination. The proposed clustering method labeled nodes with identical colors when they had reached the maximum resemblance. The community combination method merged two organizations into one of the improvements was positive and highest when the two communities were merged. The central locations were selected based on the distance within them.

The majority of existing densely connected identification algorithms create significant, larger-than-expected delays among groups and do not permit users to communicate with the technology to control the community size, except those that allow for the addition of problematic limitations.

3. PROPOSED METHODOLOGY

The architecture of the proposed AEOCDSN framework is depicted in Fig. 3. The AEOCDSN framework accesses social network graphs such as Facebook, Instagram, Snapchat, and citeseer. Then the graph is processed with LPA-Infomap and the maps are encoded using sparse autoencoder. That enhances the taxonomy and encoding accuracy. An LPA model is used to find the interpretation between social network data and local information. 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 and a global influence linked to contacts with vertices from nearby communities. 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, and a nonoverlapping entity shrinks to its specific community. Infomaps are used to decode the data on the encoding and decoding side. The final overlapping community is detected, and the loss function is computed to find the effectiveness of the proposed AEOCDSN framework. The proposed AEOCDSN is designed in this section with LPA-Infomap and a Sparse Auto Encoder model to detect the overlapping community in social networks. The agent-based model evaluates the proposed AEOCDSN and interprets the relationship between overlapping communities.

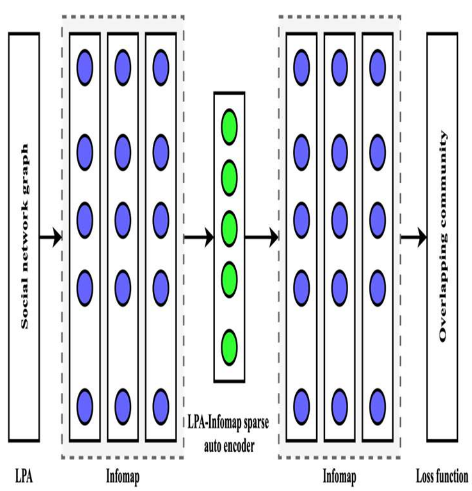

Fig. 1. Architecture Of The Proposed AEOCDSN Framework

3.1 Traditional Community Detection Module

The community data can expose some network relationships concealed from a broad perspective. Community detection must be done to gather information about the community as possible community sensing devices; two well-known and efficient algorithms—Label Propagation Algorithmic (LPA) and Infomap used.

- LPA

After initializing every vertex with a distinct community label, LPA combines each vertex's communal label by allowing neighbors to vote until convergence. As a result, label polling and community merging allow the global propagation of community data from one-hop relationships to higher hop relationships. In addition, LPA is a community detection technique that runs almost linearly. LPA is a community detection method due to its highly efficient data propagation mode.

- Infomap

In contrast to LPA, Infomap employs a deterministic greedy selection approach to find groups and focuses on recording vertex patterns with the most petite length possible based on data theory. A random variable is employed to gather higher-order data and obtain vertex patterns. Additionally, the greedy search method connects communities and interactions worldwide. Because greedy searching and randomized walking are two standard methods for gathering information about a community, Infomap is used as another community detector, and a model named Infomap is created.

3.2 Deep Learning Methodology

This section explains the Sparse Auto Encoder's ability to use local and regional data to learn vertex models and the Sparse Auto Encoder's LPA design. The overlapping community detection in social networks using Sparse Auto Encoder is shown in Fig.1. The community information is given as the input, and the local information is fetched with community data. The closer nodes are selected, and the Sparse Auto Encoder merges local information with community data. The encoded data is combined with the probability distribution function and decoded at the receiver end. The final predicted result of overlapping communities in social networks is displayed.

The component's majority of levels are connected to get the maximum information out of the input vectors. Hence, given the input vectors I_vThe actual output vector O_vAll those fully-connected levels can be determined. The output vector is denoted in Equation (1).

$$O_v = A(K_1 I_v + K_2) + (K_1 - 1)I_v + (K_2 + 1)I_v \quad (1)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The weight matrix and the biasing vector are denoted K_1 and K_2. The input vector is denoted $I_v$and the activation function is denoted $A(.)$.Non-linear functions must be used for the perceptron in encoding and decoding to acquire non-linear translations from the source vertex field to the subspace. The encoder design added a layer after each fully-connected layer to prevent overfitting. In the learning phase, the dropout level randomly turns off a subset of neural, which can improve the aggregation platform performance.

The code h follows quantization as the outcome of the preceding level as the input of the next layer. The sizes are each decreased by one. An autoencoder with unlabeled data called a sparse autoencoders trained to find the most effective way to convey the eigenvectors. The sparse autoencoder is often used to minimize dimensions.

The encoder (encoding) and the decoding are the two components of the automated encoder (decoding). From the intake level to the concealed level, encoding takes place. The incoming information is now treated to density reductions to create a coding encrypted as the encoder's response. The command is employed as an input to the decoder for deciphering, and the decrypted product is the same dimensions as the original information, utilized as the decoder's outcome. The back-propagation technique is used to modify the weights of the automated encoding after comparing the outcome with the intake outcome and computing the reconstructing faults. Until the number of repetitions or the rebuilding mistakes is lower than the set range, the rebuilding mistakes are computed and executed continually. The outcome is similar to identical to the input. Minimizing the reconstructing error describes the procedure of utilizing a back-propagation method to train a computational model. The product of the converter, or processing, is possible input to be the result of the automated encoder.

The following actions are detailed: Let N be the k-dimensional resemblance matrix for the network structure G. Using $n_x \in (0,1)$ as the input matrix, where (0, 1). The kth vertical vector in N is represented by $N \in R(kxk)$, the intake layer's parameter is $B_1 \in R(kxl)$, and the concealed layer's parameter is $B_2 \in R(kx1)$.The offsetting column matrix of the concealed level is represented by $w \in R(lx1)$. The displacement column matrix of the input neurons is denoted by $v \in R(kx1)$. Equation (2) yields the outcome h of the encoding matrix:

$$c_x = f(B_1 p_x + w) \tag{2}$$

The column matrix represented is $c_x \in R(lx1)$. The weight is denoted w, and the biasing condition is indicated $B_1$and the input is denoted p_1. The sigmoid activation function (f) is chosen as the activating rate, as indicated by Equation (3).

$$f(k) = \frac{1}{1+\exp(-k)} \tag{3}$$

At this point, the generated vector k has undergone size reductions. Equation (4) is used to calculate the deciphering layer's result.

$$O_x = f(B_2 c_x + t) \tag{4}$$

The weight of this layer is denoted t, and the biasing condition is indicated $B_2$and the concealed layer output is indicated $c_x.O_x \in R(lx1)$ is the xth column matrix that has been deciphered. The activating factor is called f. The original vector corresponds to the resultant vector O. Applying Equations (2) and (4) yields the reconstructing error shown in Equation (5).

$$Er = \sum_{x=0}^{k} f(B_2(B_1 p_x + w) + t) \tag{5}$$

The different eights are denoted w, t, and the biasing condition is denoted B_1 and $B_2$and the input is denoted p_x. The neurons' mapping spectrum is exponential when the activating parameter is (0, 1). The output is referred to as active when it is near 1, while the result is inactive when it equals 0. Sparseness limitations are introduced to the buried level in a sparse autoencoder. Since neurons are largely muted under a sparse constraint, the outcome is nearly zero. Many systems, including target identification, voice identification, and behavioral identification, have been effectively implemented using sparse expressions. The following is the method for calculating sparseness:

Initially, the average price of the encoding layer's outputs, cl_y, is determined, and $c_y(k)$ stands for the concealed layer's the neuron's performance value (c_y), and the inputs are p. In the concealed layer, the neuron outputs have an estimated value

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

using Equation (6).

$$cl_y = \frac{1}{N}(c_1(p_1) + c_2(p_2) + \cdots + c_N(p_N)) \quad (6)$$

The concealed layer functions are denoted $c_1, c_2, \cdots, c_N$ and the inputs of the respective concealed layer are denoted $p, p_2, \cdots, p_N$. The total number of inputs is denoted N. A sparsity restriction must be added to attain sparsity, which is accomplished by Equation (7).

$$cl_y = \propto \quad (7)$$

where the sparsity variable is ∝, and ∝"1, is a value like 0.05. The buried level neurons' activity value is often near zero when Equation (7) is met. The reconstructing loss is given a sparsity limitation; in other words, a compensation term is appended to the reconstructing loss, and cl_yThe deviation ∝ is penalized. The compensation algorithm works using Equation (8).

$$\sum_{y-0}^{l} \propto \log\left(\frac{\propto}{cl_y}\right) + \sum_{y-0}^{l} 1-\propto \log\left(\frac{1-\propto}{1-cl_y}\right) \quad (8)$$

The sparsity variable is denoted ∝, and the encoding layer function is denoted cl_y.The number of neurons in the concealed level is denoted l. Kullback-Leibler(KL) divergence is the name of this Equation, also known as Equation (9).

$$\sum_{y=0}^{l} KL(\propto |cl_y) \quad (9)$$

The KL function is used to compute the output. The sparsity variable is expressed ∝, and the encoding layer function is expressed cl_y. In conclusion, Equations (8) and (9) are merged to provide the divergence, and it is expressed in Equation (10).

$$KL(\propto |cl_y) = \propto \log\left(\frac{\propto}{cl_y}\right) + 1-\propto \log\left(\frac{1-\propto}{1-cl_y}\right) \quad (10)$$

The sparsity variable is denoted ∝, and the encoding layer function is denoted cl_y. In light of this, the result that minimizes the sparse phrase penalty component shows that cl_yIs sealed ∝. The reconstructing inaccuracy has been modified using Equation (11).

$$Er = \sum_{x=0}^{k} f(B_2(B_1 p_x + w) + t) + \delta KL(\propto |cl_y) \quad (11)$$

The different eights are denoted w, t, and the biasing condition is denoted B_1 *and* $B_2$and the input is denoted p_x. The scant penalized factor's quantity is δ. The sparsity variable is denoted ∝, and the encoding layer function is denoted cl_y.

a. Agent-based modeling

The model's characteristics as an agent-based system are next examined. Any manually entered probabilities of W, b, and c seem excessively arbitrary and do not resemble any actual situation due to the great degree of flexibility of the current model. It analyzed the model's micro- and macro-level characteristics using the acquired endowments and variables as the input. The model demonstrates many complicated and well-known social worlds, indicating that these occurrences are brought about by the straightforward processes of gains and transaction costs among diverse actors.

An intriguing subject at the micro-scale is how social impacts their ego connections. It focuses on two factors, particularly for agents using the proposed approach. The first factor is social power, a numerical indicator of social standing. Social power is denoted in Equation (12).

$$SP(x) = \frac{b}{w_x} + \frac{b}{w_x - 1} + \frac{b}{w_x - 2} + \cdots \frac{b}{w_x - N} \quad (12)$$

The social benefit is denoted b, and the agent weight on the xth dimension is denoted w_x.Social influence, often known as the potential advantages one could provide another, is referred to as social power. The value of the xth agent on the yth dimension is w_{xy}. The likelihood that action helps another in the y-th dimension rises as $b_x \times w_{xy}$. It makes sense to use the dot products of b and w_xTo describe an agent's social standing. As a result, this variable's definition is congruent with the idea of social power. Given the usefulness of this public control for the social transaction, particularly within the homophile's community surrounding it.

"Social exclusion" is the second factor that gauges the degree to which an agent is marginalized. Social exclusion is denoted in Equation (13).

$$SE = \left|d \times \frac{w_x}{w_x+1}\right|_2 + \frac{1}{d-1} \quad (13)$$

The deviation of the data from the overlapping community in a social network is denoted d, and the agent weight on the xth dimension is indicated w_x. Remember that it sets a limit of 0 for the norms across all dimensions. Since a more significant cost is incurred when an agent connects to another random individual, it is assumed that someone is at the edge of that degree when she has a significant exact number on that dimension.
The relationship between data of their egocentric community and social dominance or social isolation (i.e., degree and grouping coefficient) is derived. It discovered that "social power" and degree have significant positive correlations while "social exclusion" and degree have strong negative correlations. This result is in line with what the suggested model implies: People on the margins of society have fewer possibilities, whereas those with significant (beneficial) assets can assist others more.

It investigated the relationships between public control or exclusions and the node grouping coefficients, which is more intriguing. An agent's neighbors are not diverse if there is a significant clustering coefficient, which implies that the neighbors are linked closely. Individuals with higher public control or less social isolation have lower grouping indices on the system; high status (power) individuals have more social networking sites, a well-known and significant feature of human connections.

Additionally, the suggested model is capable of forecasting network macrodynamic. It concentrates on the macro dynamics of the social system because of the systematic modification of the cost development of appropriate k i.e., changing $k^* = \frac{1-c}{k}; c \in (0,1)$. Advances in data technologies often reduce coordination expenses (e.g., a new railway). It uses agent-based modeling to rebuild the actual social

networking sites using the learned endowment vectors and functionalities. E is the network's input collection, and c is the amount after being reduced—it computes densities, average grouping coefficient, mean short distance in the huge component, and contact diversity. It includes the k in Equation (8) after it has been normalized by $|k|_2$.The social interaction diversity of a node is denoted in Equation (14).

$$ID = \frac{1}{N}\sum_{x=0}^{N} \frac{\left|d \times (w_x - w_y)\right|_2}{|k|_2} \quad (14)$$

The deviation of a node from the community is denoted d, and the agent weight in x and y dimensions are denoted w_x *and* w_y. The cost scaling variable is denoted k, and the total number of overlapping communities is denoted N. As the expense scalability parameters c drop, the density rises noticeably, but the grouping coefficient barely changes. It suggests that reducing agency costs increases linkages, community stability or balance, and the connectedness between neighbors. The declining trend in the adjacent nodes among pairs indicates that reducing the cost of coordinating might weaken the influence of social hierarchies. According to the tendency in contact diversity, linkages between more diverse persons increase as coordination costs go down. These results suggest that the decrease in expenses, typically brought about by technological advancements, leads to a society with much less hierarchy and more chances for interpersonal connection, particularly for persons from different social networks.

4. SIMULATION ANALYSIS AND PERFORMANCE EVALUATIONS

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.

- A network of research citations called Cora was created by [20]. It has 2387 publications on computer vision with 8 labels.
- Another widely used dataset for research articles is Citeseer [21], which has 3388 articles and six categories.
- Wiki [22] is a linguistic network of 11989 edges connecting 2434 websites from 18 groups.

Journal of Theoretical and Applied Information Technology
15th February 2023. Vol.101. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- The computer programming bibliography data in DBLP were created by [23].

Table 1. Dataset Description

Dataset	Number of nodes	Number of edges	Density	Clustering coefficient
Cora	7934	50312	0.062	0.087
Citeseer	1674	44505	0.075	0.093
Wiki	1432	36542	0.005	0.147
DBLP	2435	14354	0.089	0.134

The dataset description is shown in Table 1. Four 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 edges, density of nodes, and clustering coefficient are analyzed and plotted above. The higher density dataset results in higher interconnectivity between nodes, and higher clustering co-efficient lead to lesser overlapping community.

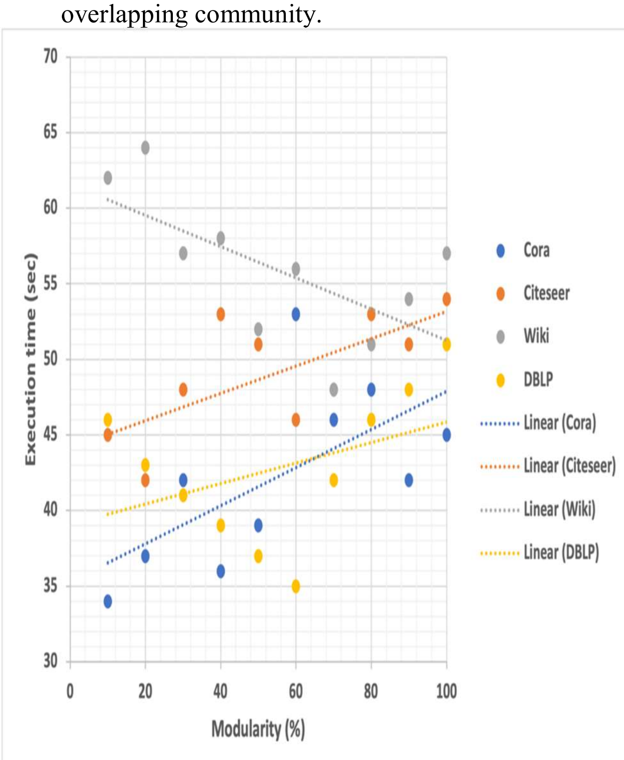

Fig. 2(A). Execution Time Analysis Of AEOCDSN

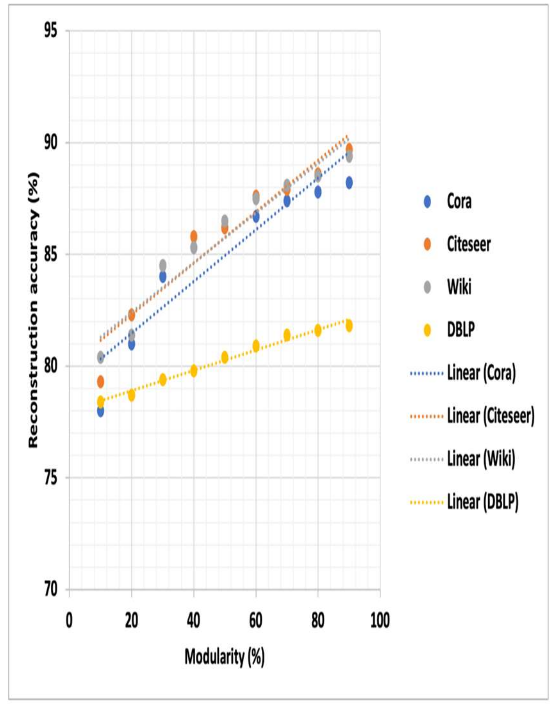

Fig. 2(B). Reconstruction Accuracy Analysis Of AEOCDSN

The execution and reconstruction time analysis of AEOCDSN is computed and plotted in Fig. 2(a) and 2(b), respectively. The execution time is analyzed as the total time required to analyze the social network data from the given datasets and detect the overlapping community in the social networks using the proposed AEOCDSN with Sparse Auto Encoder and LPA model. The proposed AEOCDSN with LPA and Sparse Auto Encoder produces results at a lesser time. In the same way, the higher reconstruction accuracy as the modularity varies from minimum to maximum. The respective outcomes of the AEOCDSN increase because of the higher possibility of overlapping communities in the social media data.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2. Simulation Outcome Analysis

Methods	Accuracy (%)	Modularity (%)	Timestamps(sec)	Loss function (%)
K-means	76.5	54.2	35.2	34.5
PSO	54.7	46.7	37.5	53.2
CNN	58.3	35.4	43.1	47.6
AEOCDSN	87.6	91.3	89.5	17.3

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 Table 2. The AEOCDSN results are compared with existing systems like K-means [17], Particle Swarm Optimization [18], and Convolutional Neural Networks (CNN) [19]. The computed results of the proposed AEOCDSN outperform the existing deep and machine learning models with the Sparse Auto Encoder and LPA model.

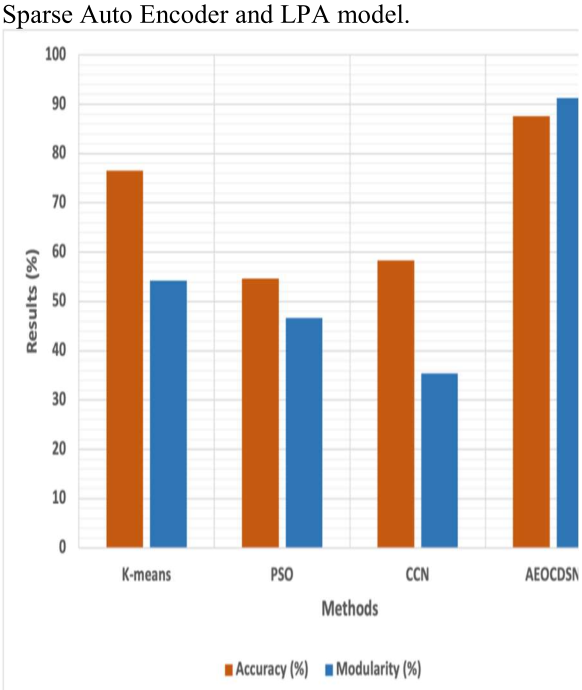

Fig. 3(A). F Score And Modularity Analysis

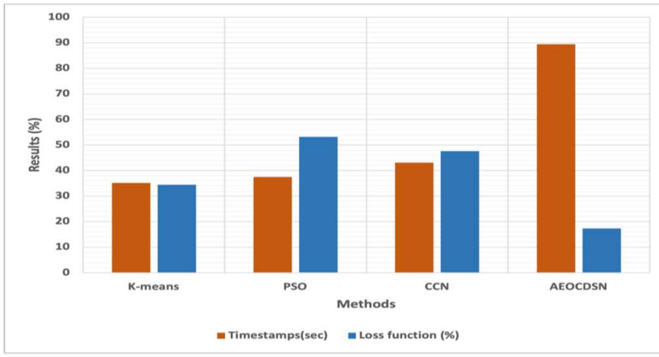

Fig. 3(B). Normalized Mutual Information And Conductance Analysis

The F score, modularity, normalized mutual information, and conductance are computed for the proposed AEOCDSN. The results are compared with the existing LPA and machine learning models such as K-means, PSO, and CNN. The proposed AEOCDSN with Sparse Auto Encoder and LPA model detects the overlapping community in social networks more accurately. The higher results lead to better identification of overlapping community in social network and hence finds the interpretation between the social network data.

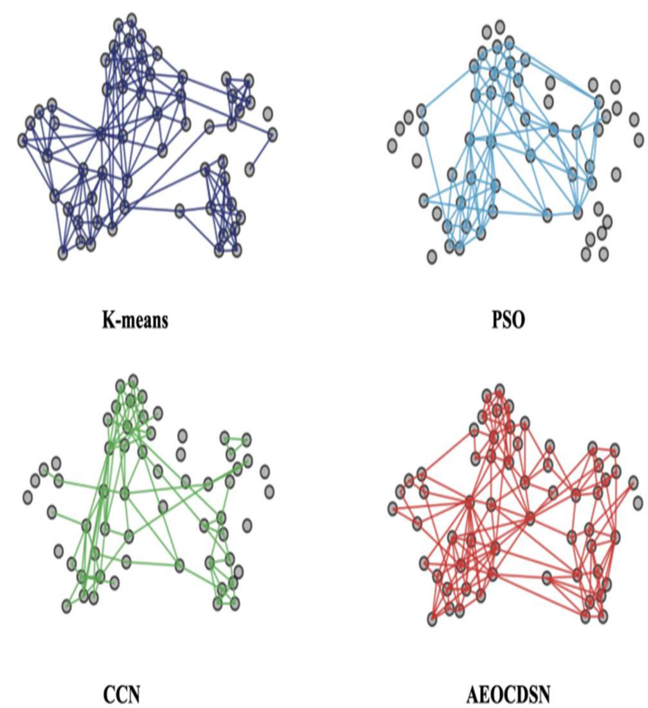

Fig. 4(A). Facebook Overlapping Diagram

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

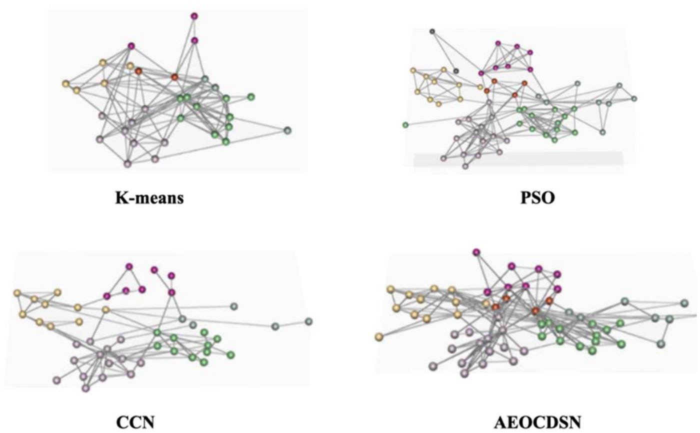

Fig. 4(B). Instagram Overlapping Diagram

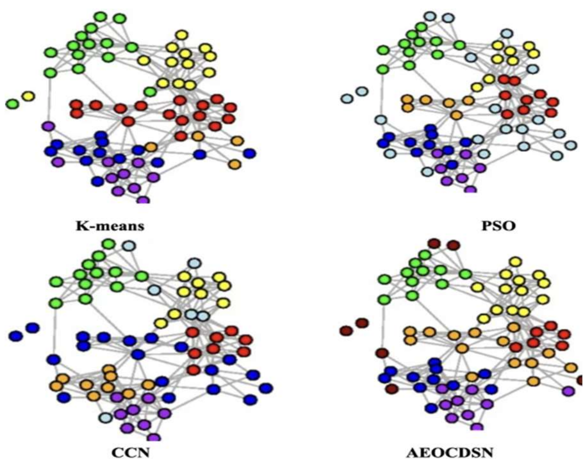

Fig. 4(C). Snapchat Overlapping Diagram

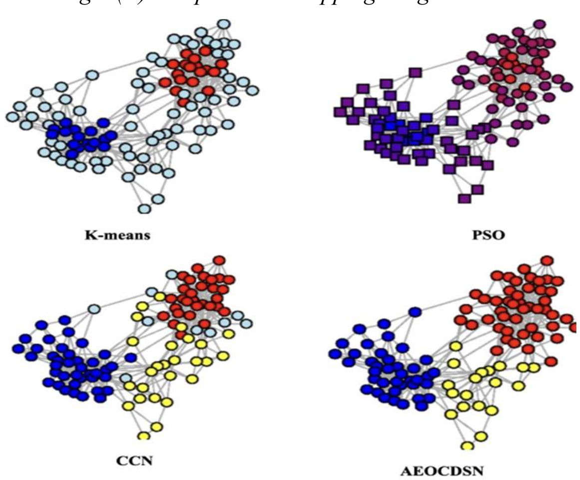

Fig. 4(D). Citeseer Overlapping Diagram

The overlapping diagram of different social network data such as Facebook, Instagram, Snapchat, and citeseer are analyzed and plotted in Fig. 4(a), Fig. 4(b), Fig. 4(c), and Fig. 4(d), respectively. The proposed AEOCDSN framework outcomes are compared with the existing models like K-means, PSO, and CNN. The proposed AEOCDSN framework with sparse autoencoder and LPA enhances the overall simulation and overlapping classification accuracy. The outcomes are analyzed using different datasets, and the proposed AEOCDSN framework outperforms the existing models in all the datasets.

The proposed AEOCDSN is evaluated in this section, and the results are compared with the existing machine and LPA models. The proposed AEOCDSN with Sparse Auto Encoder and LPA model efficiently detects the overlapping community in social networks.

5. CONCLUSION AND FUTURE WORK

The current approaches are restricted to overlapping community detection in social networks. They cannot handle real-world networks where nodes randomly or regularly change their activity levels, join or leave communities, and migrate between communities. An Auto Encoder for Overlapping Community Detection in Social Network (AEOCDSN) is suggested to detect the overlapping community in the social network. A Sparse Auto Encoder and LPA model detect the overlapping community in social networks. The agent-based model evaluates the system and interprets the social data in the overlapping community. The KL-based loss function is used to find the loss in the social data, and the proposed system with LPA and Sparse Auto Encoder produces lesser loss than other existing LPA and machine learning models. The Sparse Auto Encoder data limit the research, and the efficiency and accuracy are poor. The system's performance is enhanced using artificial intelligence and big data analytics to optimize the results in the future.

REFERENCE

[1]. Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. Computer Science Review, 37, 100286.

[2]. Azaouzi, M., Rhouma, D., & Ben Romdhane, L. (2019). Community detection in large-scale social networks: state-of-the-art and future directions. Social Network Analysis and Mining, 9(1), 1-32.

[3]. Jokar, E., &Mosleh, M. (2019). Community detection in social networks is based on an improved label propagation algorithm and balanced link density. Physics Letters A, 383(8), 718-727.
[4]. Qin, M., Jin, D., Lei, K., Gabrys, B., & Musial-Gabrys, K. (2018). Adaptive community detection incorporating topology and content in social networks☆. Knowledge-based systems, 161, 342-356.
[5]. Singh, A., Garg, S., Batra, S., & Kumar, N. (2019). Probabilistic data structure-based community detection and storage scheme in online social networks. Future Generation Computer Systems, 94, 173-184.
[6]. Sun, K., Zhang, J., Zhang, C., Hu, J., 2016. Generalized extreme learning machine autoencoder and a new deep neural network. Neurocomputing 184, 232–242.
[7]. Tian, F., Gao, B., Cui, Q., Chen, E., Liu, T.-Y., 2014. Learning deep representations for graph clustering. In: Proceedings of 28th AAAI
[8]. Conference on Artificial Intelligence. pp. 1293–1299.
[9]. Asmi, K., &Abarda, A. (2022). An efficient local algorithm for overlapping community detection in social networks. Procedia Computer Science, 201, 789-794.
[10]. Hasan, A., & Kamal, A. (2022). LapEFCM: overlapping community detection using laplacianeigenmaps and fuzzy C-means clustering. International Journal of Information Technology, 1-12.
[11]. Zhou, X., Zhao, X., Liu, Y., & Sun, G. (2018). A game theoretic algorithm to detect overlapping community structures in networks. Physics Letters A, 382(13), 872-879.
[12]. Ma, J., & Fan, J. (2019). A Complex Network Overlapping Community Detection Algorithm based on K-Cliques and Fitness Function. International Journal of Performability Engineering, 15(12), 3161.
[13]. Chu, J., Wang, Y., Liu, X., & Liu, Y. (2020). Social network community analysis based large-scale group decision-making approach with incomplete fuzzy preference relations. Information Fusion, 60, 98-120.
[14]. Zhu, Z., Zhou, T., Jia, C., Liu, W., Liu, B., & Cao, J. (2022). Community detection across multiple social networks based on overlapping users. Transactions on Emerging Telecommunications Technologies, 33(6), e3928.
[15]. Gao, Y., Zhang, H., & Zhang, Y. (2019). Overlapping community detection based on conductance optimization in large-scale networks. Physica A: Statistical Mechanics and its Applications, 522, 69-79.
[16]. Zhao, Z., Li, C., Zhang, X., Chiclana, F., &Viedma, E. H. (2019). An incremental method to detect communities in dynamic evolving social networks. Knowledge-Based Systems, 163, 404-415.
[17]. You, X., Ma, Y., & Liu, Z. (2020). A three-stage algorithm on community detection in social networks. Knowledge-Based Systems, 187, 104822.
[18]. Sorensen, M., Sidiropoulos, N. D., & Swami, A. (2022). Overlapping Community Detection via Semi-binary Matrix Factorization: Identifiability and Algorithms. IEEE Transactions on Signal Processing.
[19]. Nasser Al-Andoli, M., Chiang Tan, S., & Ping Cheah, W. (2022). Distributed deep parallel learning with a hybrid backpropagation-particle swarm optimization for community detection in large complex networks.
[20]. Gupta, S. K., & Singh, D. P. (2022). Seed Community Identification Framework for Community Detection over Social Media. Arabian Journal for Science and Engineering, 1-15.
[21]. https://graphsandnetworks.com/the-cora-dataset/
[22]. https://relational.fit.cvut.cz/dataset/CiteSeer
[23]. https://snap.stanford.edu/data/wiki-topcats.html
[24]. https://dblp.org/db/conf/nlp/nlp2000.html
[25]. Kipf, T. N., and Welling, M. (2016). "Variational graph auto-encoders," in Proceedings of NeurIPS Bayesian Deep Learning Workshop (Barcelona).
[26]. Li, J., Dani, H., Hu, X., Tang, J., Chang, Y., and Liu, H. (2017). "Attributed network embedding for learning in a dynamic environment," in ACM Conference on Information and Knowledge Management (Singapore), 387–396.
[27]. Montavon, G., Samek, W., and Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Process. 17, 1–15. doi: 10.1016/j.dsp.2017.10.011

[28]. Sricharan, K., and Das, K. (2014). "Localizing anomalous changes in time-evolving graphs," in ACM SIGMOD International Conference on Management of Data (Snowbird, UT), 1347–1358.
[29]. Kumar, A., Abhishek, K., Vivekananda, K., Upadhyay, C.,Experimental study and optimization of process parameters during wedm taper cutting,(2019), Advances in Intelligent Systems and Computing Vol 817Issue10.1007/978-981-13-1595-4_57
[30]. Sreekar Reddy, M.B.S., Ratnam, C.H., Sharma, I.V., Manupati, V.K.,Social network analysis based evolutionary algorithimic approach to identify the influence of hubs on flexible scheduling problems,(2016), Proceedings of the International Conference on Industrial Engineering and Operations Management Vol 8-10 March 2016
[31]. Atmakur, V.K., Siva Kumar, P.,A prototype analysis of machine learning methodologies for sentiment analysis of social networks,(2018)International Journal of Engineering and Technology(UAE) Vol 7,Issue3
[32]. Anisha, P.R., Vijaya Babu, B.,BPS: Breast cancer prediction system using wisconsin breast cancer dataset,(2018),Journal of Advanced Research in Dynamical and Control Systems Vol no. 10,Issue4
[33]. Divya, T., Vijaya Babu, B.,An efficient feature subset algorithm for high dimensional data,(2016),Asian Journal of Information Technology Vol no. 15,Issue19,10.3923/ajit.2016.3730.3733
[34]. Kishor Kumar Reddy, C., Vijaya Babu, B., Rupa, C.H.,SLEAS: Supervised learning using entropy as attribute selection measure,(2014),International Journal of Engineering and Technology Vol no. 6,Issue5
[35]. Rajeswara Rao, D., Raghuveer, A., Murthy, J.V.R., Apparao Naidu, G.,Common user activities behind social networks using link based clustering,(2014),International Journal of Applied Engineering Research Vol no. 9,Issue15.