Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 24, Issues 4, pp. 985–997, DOI 10.12694/scpe.v24i4.2475

A DEEP COMMUNITY DETECTION APPROACH IN REAL TIME NETWORKS

DEEPJYOTI CHOUDHURY* AND TAPODHIR ACHARJEE†

Abstract. Community detection in real time networks is one of the important aspect of social network analysis. Deep learning has been applied successfully in a variety of research fields in recent years. Proximity matrix is frequently used as the representation of the network structure. However, there are issues with the proximity matrix's insufficient spatial contiguity information. As a result, this research provides a deep learning applied community identification approach that combines the reorganization of the matrices, spatial attribute uprooting, and community identification. For obtaining a spatial proximity matrix, the primary proximity matrices in a real time graph is recreated using the highest weight and adjacent users. The dimensional proximity matrix can obtain a subdomain of the network, allowing the convolutional neural network (CNN) to draw out dimensional localization more easily and fast. Ten different real time datasets of social networks are used in tests to examine our proposed approach. Our results show that the proposed community identification approach has higher compatibility than existing deep learning-based strategies. As a result, the proposed deep community identification approach is capable of detecting the excellent clusters in real time networks.

Key words: real-time network, deep learning, community detection, social network, proximity

1. Introduction. It is conventional that social networks have been extensively deliberated to analyze behaviors of human considering a number of layouts, including information extraction, domination analysis, community detection, individual profile details, social data privacy etc [1]. Community identification in real-time social networks is a well-known aspect of networked systems in biology, economics, politics, and computer science.

Deep learning (DL) has showed excellent performance in a wide number of research domains, including real-time networks for analysing user structural information [2]. DL-applied network embedding can be executed in both ways: with random walk [3] and without random walk [4][5]. As we know that proximity matrix is used to reserve the information of the connected nodes in the network. However, the adjacency matrix has inadequate spatial proximity information.Several Auto-Encoder (AE)-based network entrenching studies improved the input vector by performing divergent pre-processing on proximity matrices to improve the elicitation correctness of spatial characteristics extraction to overcome this complication and boost the correctness of feature uprooting [6][7][8]. Convolutional neural network (CNN) is used in network embedding because it is an effective technique for extracting spatial localisation [9][10]. The convolution operation can be simulated on the graph. As a result, the difficulty is how to enhance the proximity matrix such that it can store spatial closeness among vertices.

1.1. Contribution. Given the resilience and efficacy of AE and CNN-applied network embedding, this paper integrates 'AutoEnc+CNN' to increase the aspect of feature uprooting from the nodes. As a result, this research offers a deep community detection approach that combines (a) matrix reorganization, (b) 'AutoEnc+CNN'-applied spatial feature uprooting, and (c) community identification. Furthermore, this paper provides a spatial characteristics uprooting strategy based on AutoEnc and CNN to uproot the spatial features of the reorganised proximity matrix.

In recent decades, efforts have been considerably made to build efficient models to identify communities in social networks. The main objectives of this paper are mentioned below:

1. To obtain the dimensional proximity matrices in real-time networks, a matrix reorganisation strategy based on a unique structure reorganisation approach is proposed, which can aid CNN in quickly and

*Department of Computer Science and Engineering, Assam University, Silchar, Assam, India 788011 (deepjyotichoudhury05@gmail.com).

†Department of Computer Science and Engineering, Assam University, Silchar, Assam, India 788011 (tapacharjee@gmail.com).

easily determining geographical localization.

2. To obtain a dimensional features uprooting strategy based on autoencoder and CNN, which is proposed to derive spatial eigenvectors and successfully uproot the spatial properties of real-time networks.
3. To explore the topographic architecture of real-time networks based on the matrix reorganisation and dimensional characteristics uprooting approach in order to improve social community detection.

After a brief introduction in Section 1, Section 2 presents important related investigations performed in recent years. In Section 3, proposed approach is discussed in detailed manner. Execution & Results Analysis is elaborated in Section 4. We have compared some previous works with our proposed approach in this section. Section 5 highlights the conclusion, challenges and future work.

2. Related Works. It has been observed in recent years that the applications of Deep Neural Network (DNN) can embed networks using arbitrary walks while maintaining network properties. DeepWalk is one of the most well-known deep learning method along with the enhancement of arbitrary walk. Without arbitrary walk, the same method applies deep approaches over the entire network. AE and CNN are two prominent deep learning models used in network embedding that do not employ random walks. Resler et al. [11] presented in their paper that the use of a metric learning enhanced deep CNN on an archaeological dataset. The associations between sites were determined by them using a community detection algorithm based on the confusion matrix data. A deep learning-based weighted network community discovery technique was also presented [12] based on a deep sparse autoencoder. A solution to problem of overlapping community structure for large graph was presented using autoencoder [13]. Ferraro et al. [14] presented a different technique to resolve community detection problem based on DL and they proposed a hypergraph-based data model for representing all forms of user connections within an MSN, which were frequently mediated by multimedia data.

Again, a parallel deep learning-applied community identification strategy using particle swarm optimization in massive composite networks was proposed [15]. The findings demonstrated how well the suggested deep learning with hybrid optimisation works for identifying communities in large networks. Essaid et al. [16] presented a method for detecting communities inside the Bitcoin network that uses a deep feature representation algorithm and Deep Feedforward Autoencoders. Their findings demonstrated that, compared to a random P2P network, the Bitcoin network has a stronger clustering coefficient and community structure. Sun et al. [17] developed a system that combined CNN and Transformer, used wavelet and inverse wavelet transforms for encoding and decoding, and employed wavelet transform and inverse wavelet transform for learning. A dual graph autoencoder (DGAE) was suggested by Zhang et al. [18] to develop discriminative representations for hyperspectral images. In order to characterise the geometric structures of hyperspectral pictures, DGAE first built the super pixel-based similarity graph with spatial information and band-based similarity graph using the relationships of pair-wise pixels within homogeneous regions and pair-wise spectral bands. More discriminative feature representations were learned from the hidden layer via the encoder-decoder structure of DGAE using the newly created dual graph convolution.

Self-Supervised Contrastive Graph Clustering (SCGC) was suggested by Kulatilleke et al. [19], which enforced graph structure using contrastive loss signals to acquire discriminative node representations and iteratively revised soft cluster labels. A thorough analysis of AE-based industrial applications was published by Qian et al. [20]. It was primarily divided into two sections: AE-based representation learning and monitoring techniques, which showed how AE-based monitoring methods are designed from start to finish. Second, a thorough analysis of AE-based representation learning from the viewpoints of industrial data characteristics was conducted. A study was carried out by Lim et al. [21] recently in order to offer a thorough overview and to investigate potential future possibilities for the best reinforcement learning-based virtual network embedding solutions.

Proximity matrix was well elaborated by Goel et al. [22] and by utilising the user's own behaviours as well as those of other users in their social network, they suggested a novel way to build a strong User Interest Profile (UIP) in community detection. The same authors proposed a methodology [23] that focused on UIP augmentation using multiple strategies, as well as a novel approach to handle outlier tags that caused ambiguity in the collective Resource Illustration Profile (RIP). The fuzzy satisfaction requirement-based novel mapping functions were designed to measure query relevance score and user interest relevance score for a web resource. A novel approach [24] was proposed that employed Cohen's k as a similarity measure for each pair of nodes;

the values were then clustered to discover communities. A new community detection approach was presented by Wu et al. [25] that proved the spatial proximity matrix could obtain a subdomain of the graph, allowing the convolution neural network to extract spatial localization more easily and fast. An auto-encoder based on a convolution neural network can extract the spatial eigenvector of the reconstructed adjacency matrix to improve modularity.

With the aid of some recent works [26][27][28][29][30], we also want to lay the groundwork for how researchers have implemented deep learning models in the network embedding technique:

- **AE-applied Graph Embedding Approach:** The AE algorithm is a powerful facts condense technology. AutoEnc-applied graph embedding approaches frequently modify the parameters of the input and converts it into a new depiction. To absorb the composition of the network, SDNE was proposed to design a mini-supervised model based on DNN that enhanced the inputs by combining 1^{st} and 2^{nd} order adjacency values.
- **CNN-applied Graph Embedding Approach:** CNN and its variations have found widespread application in network embedding. CNN-applied graph embedding employs the original CNN model, which was built for both Euclidean and non-Euclidean domains.

2.1. Motivation. The following are the primary observations that motivated us to present the proposed work provided in this paper:

1. 'Matrix reorganisation' is needed because it is a strategy technique used to improve efficiency, flexibility, and overall performance to identify communities in a large network. The network structure is restructured from a typical hierarchical model to a matrix form that contains features of both functional and project-based structures. Matrix reorganisation can convert network data into a more appropriate structure, such as an adjacency matrix or a modularity matrix, making it easier to discover communities or groups of nodes within the network. Matrix reorganisation can help increase the scalability of community detection algorithms in large-scale social networks. It decreases computing complexity and enables more effective analysis of massive datasets, both of which are critical in comprehending and managing online communities with millions of members. Matrix reorganization can assist in identifying influential nodes or key community members within social networks. This is why we have adopted this approach in our proposed method.
2. 'Spatial feature uprooting' method is required because of its ability to select the most discriminating features, where a deep learning based method can ingest more readily. Communities in many real-world circumstances are defined not only by social connections but also by geographical proximity. The inclusion of geographic information into spatial feature uprooting allows community detection algorithms to incorporate both social and spatial dimensions. Different regions within a network may have varying degrees of community structure. In order to measure spatial heterogeneity and pinpoint areas with distinct community boundaries where communities converge, spatial feature uprooting is used. When spatial linkages and geographic context are important, spatial feature uprooting is crucial for community detection. It makes it possible to analyse networks more thoroughly by taking into account both social and spatial dimensions.
3. 'Community identification' in real time networks is the final goal based on matrix reorganisation and spatial feature uprooting method. Communities frequently display recognisable behavioural patterns. Researchers can learn more about the habits, pursuits, and pursuits of various groups inside the network by identifying communities. In social networks, community identification is essential for a variety of purposes, from boosting user experiences and marketing tactics to upholding safety and comprehending the social dynamics of online communities. It offers insightful information that aids in decision-making and promotes a better comprehension of the intricate linkages and behaviours seen in social networks.

3. Proposed Approach. As mentioned in the Section 1, our proposed approach is the combination of matrix reorganization approach, spatial feature uprooting and finally community identification. Let's elaborate these approaches in the subsections below.

3.1. Matrix Reorganization Approach. Workflow of the proposed approach has been shown in Fig. 3.1. The proposed approach comprises of three different sub-approaches: Highest Weight User Selection, Adjacent

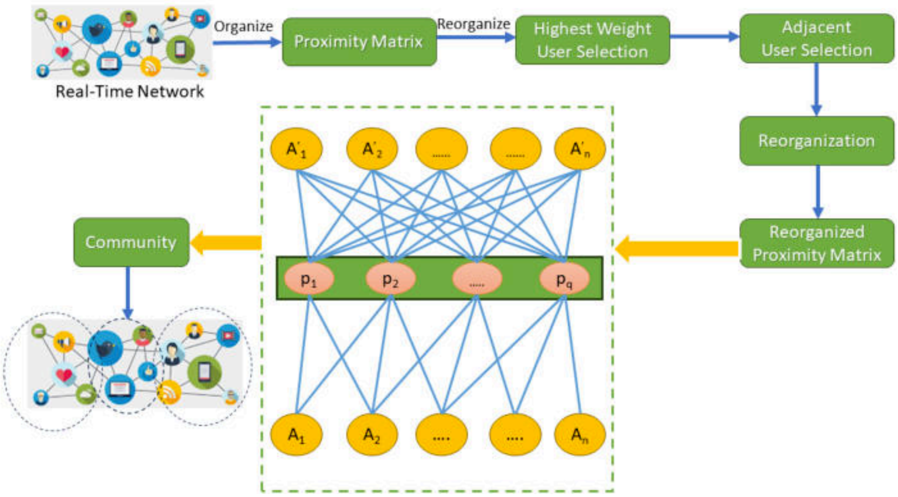

Fig. 3.1: Workflow of the Proposed Matrix Reorganization Approach

User Selection and Reorganization of Proximity Matrix. Let's discuss these sub-approaches below.

3.1.1. Highest Weight User Selection. The user is a key member of a group who may influence the views of other members using this strategy. In real-time social networks, numerous users can follow and make friend with the most connected person in a community. As a result, the Highest Weight User Selection technique proposed here would assess each node's influence on the next most significant node in the proximity matrix to identify a starting node for matrix reconstruction. The pseudo codes of this method is presented in Algorithm 1.

Algorithm 1 Proposed Highest Weight User Selection Algorithm

Require: a proximity matrix with nodes and the number of repetitions
Ensure: the user with the highest weight

Organize the proximity matrix A_{nxn}
while $p = 1, ..., n$, $q = 1, ..., n$ **do**
    Calculate the weight of the connections from user a_p to a_q based on the number of interconnections of a_p
end while
for $p = 1, ..., n$ **do**
    Initialize the weight of user a_p in the starting moment as 1.
end for
while (the amount of repetitions is less than the user having highest weight) **do**
    for $p = 1, ..., n$ **do**
        Initialize the weight of user a_p in the next repetition as 0.
        for $q = 1, ..., n$ **do**
            Compute the weight of user a_p grown by the connection weight from user a_q to user a_p as the merit of substitute.
            Initialize the weight of user a_p in the next repetition along with the merit of substitute.
        end for
    end for
end while
Return the user has the maximum weight.

3.1.2. Adjacent User Selection. After identifying the person with the highest weight (superior), an adjacent user selection method is presented to identify the adjacent user who is most relevant to the superior user.

Based on Equation 3.2, the superior user having the highest weight is represented as user a_p. Here, we have applied Euclidean Distance (ED), $z(p, q)$ to compute the distance between user a_p and user a_q. According to Equation 3.3, the user a_q with the least distance based on the nearest neighbor is initiated.

$$z(p, q) = \sqrt{\sum_{l=1}^{n} d(p, q, l)^2} \tag{3.1}$$

where

$$d(p, q, l) = a_{p,l} - a_{q,l} \tag{3.2}$$

$$AdjacentNeighbor = least_{z(p,q)} \tag{3.3}$$

where $q \neq p, 1 \leq q \leq n$

3.1.3. Reorganization of Proximity Matrix. The proposed highest weight user selection and adjacent user selection methods may be used to determine the order of users in the proximity matrix A. The superior user may be selected as the starting user, and the adjacent of the superior user can be selected as the next user; then, using Equations 3.2 and 3.3, the adjacent neighbor of the second user can be selected, and so on. Algorithm 2 can rebuild the proximity matrix A as matrix Z.

Algorithm 2 Reorganization of Proximity Matrix

```
Require: a proximity matrix A with n nodes
Ensure: the reorganized proximity matrix Z
Organize the proximity matrix A_nxn
Construct the directory R_n
Construct the directory S_n
for p = 1, ...., n do
    Enter the user a_p into the directory R_n.
end for
Initialize the superior leader from the highest weight user selection approach as user a_p.
while Length of R_n > 0 do
    Extract user a_p from R_n
    for q = 1, ...., n do
        Determine the Euclidean distance from user a_p to user a_q
    end for
    Detect as well as initialize adjacent user as user a_p
    Extract user a_p from S_n
end while
Build the proximity y matrix Z_nxn based on S_n
Return the reorganized proximity matrix Z.
```

3.2. Spatial Feature Uprooting. In a basic instance, both the input as well as output layers have 4 number of neurons, that means 4 users present in the network. The first hidden layer (the convolutional layer) has a filter size of 1 by 3, therefore two neurons are instructed in the concealed layer. Figure 3.2 depicts the architecture of a basic CNN-based auto-encoder.

In a normal scenario, the rebuilt proximity matrix may be divided into n records, with each record having a dimension of 1 by n. Both the input and output layers have n number of neurons. The first concealed layer has q neurons.

Figure 3.3 depicts the general case structure of a CNN-based auto-encoder. The loss function considers mean squared deviation (MSD). During the execution and performance stages, spatial information may be retrieved based on neuron values.

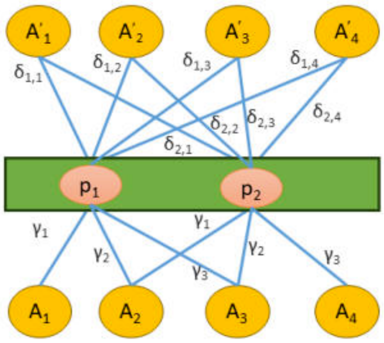

Fig. 3.2: A basic example of a CNN-applied auto-encoder

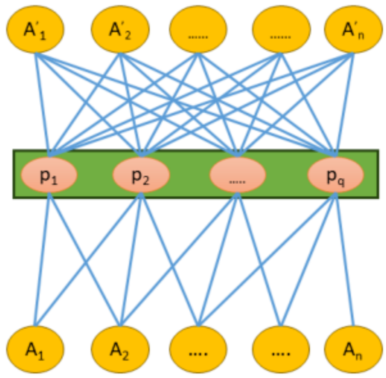

Fig. 3.3: A general instance of a CNN-applied auto-encoder

3.3. Community Identification. After extracting spatial information, the dimension of each record is represented as 1xn and used in the K-means method.

$$P_q = [p_{q,1}, p_{q,2}, ..., p_{q,n}] \tag{3.4}$$

The proposed community detection approach consists of three phases, which are as follows:

1. The k records are chosen at random from the n records to serve as k cluster centers.

$$R_p = [r_{p,1}, r_{p,2}, ..., r_{p,n}] \tag{3.5}$$

2. Based on Equation 3.1, ED is used to calculate the separation space from the q^{th} position to the p^{th} position of the cluster. Data n are divided into k number of clusters depending on their distance, also the core of every community is re-evaluated established on the data in the community.
3. If no modifications are made to any of the cluster centers, the community identification process is completed. Otherwise, Point 1 & 2 must be repeated.

4. Execution & Results Analysis. This section comprises of dataset description, evaluation matrices, followed by results analysis.

4.1. Dataset. We have considered here 10 different real time datasets to experiment using our proposed approach. The details is shown in Table 4.1.

4.2. Evaluation Matrices. We have used here Q-modularity, Normalized Mutual Information, Mean Reciprocal Rank, and Mean Average Precision as the evaluation matrices.

Table 4.1: Dataset [31]

Sl No.	Dataset	Nodes	Edges
1	Karate	34	78
2	Football	115	613
3	Dolphins	62	159
4	Polbooks	105	441
5	Cora	2,708	5,429
6	Facebook	4,039	81,800
7	Artists	50,515	819,306
8	CiteSeer	3,312	4,732
9	Polblogs	1,490	16,718
10	School	68	220

- Q-Modularity [32]: This measure is calculated as:

$$Q = \frac{1}{2m} \sum_{pq} (A_{pq} - \frac{l_p l_q}{2m}) \delta(t_p, t_q) \tag{4.1}$$

where, A represents an adjacency matrix, m represents the quantity of edge, l_p represents the degrees of the p-th node.

- Normalized Mutual Information [33]: NMI is defined as:

$$NMI(P,Q) = \frac{-2 \sum_{l=1}^{c_P} \sum_{m=1}^{c_Q} R_{lm} \log(\frac{R_{lm}.N}{R_{l.}R_{.m}})}{\sum_{l=1}^{c_P} R_{l.} \log(\frac{R_{l.}}{N}) + \sum_{m=1}^{c_Q} R_{.m} \log(\frac{R_{.m}}{N})}, \tag{4.2}$$

where, c_P & c_Q are considered as the numeral of the clusters in the partition, $P(Q)$. The total number of the users in the error matrix are depicted as $R_{l.}(R_{.m})$. N is considered here as the entire users in the network.

- Mean Reciprocal Rank [34]: MRR is calculated as:

$$\frac{1}{|p|} \sum_{a=1}^{|p|} \frac{1}{q_a} \tag{4.3}$$

where $|p|$ is a total number of queries, q_a is the rank position of a first relevant community among all the communities retrieved for the a^{th} query. A value of MRR ranges between 0 and 1.

- Mean Average Precision: MAP can be defined as:

$$\frac{1}{|p|} \sum_{a=1}^{p} AvgP_a \tag{4.4}$$

where $|p|$ is a total number of queries, MAP is the mean of Average Precision (AvgP) of each query in a query. A value of MAP ranges between 0 and 1.

4.3. Results Analysis. Four scenarios were created to assess the modularity of different approach combinations in order to evaluate the proposed approach.

1. To find communities in a social network, the auto-encoder approach is used to extract features from the original proximity matrix. Case (1)'s label is written as 'AutoEnc'.

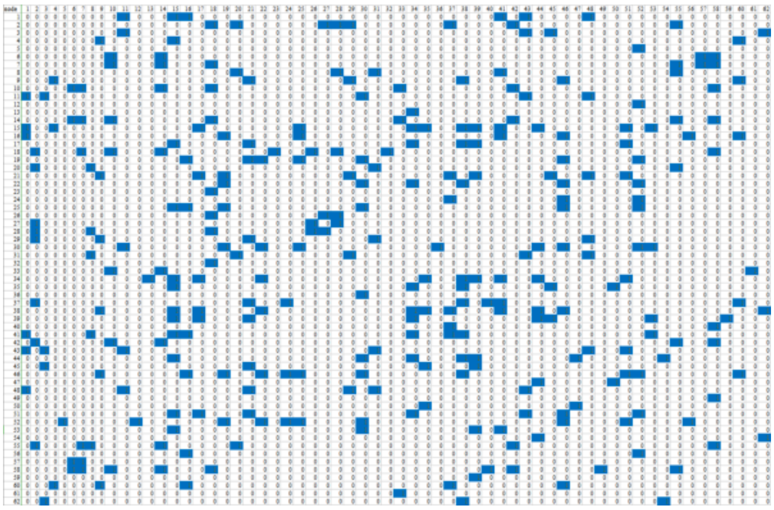

Fig. 4.1: Original Proximity Matrix of Dolphin Network

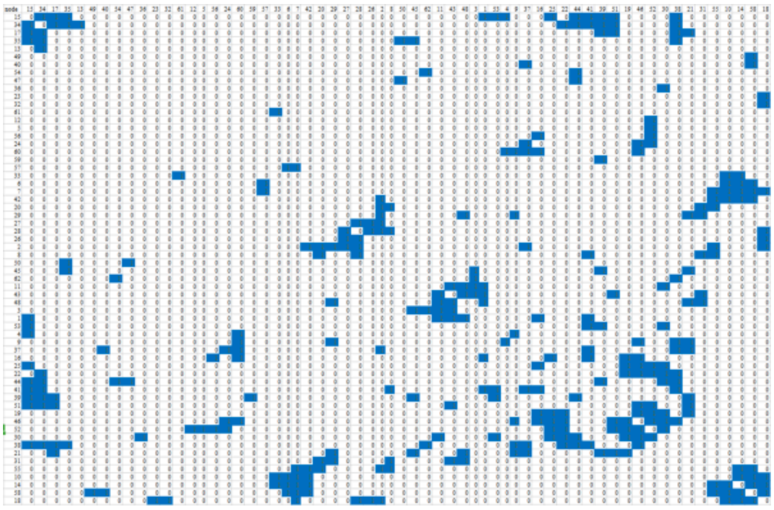

Fig. 4.2: Reorganized Proximity Matrix of Dolphin Network

2. To discover communities in a social network, the auto-encoder approach is used to extract characteristics from the rebuilt proximity matrix. Case (2)'s label is written as 'ReMat+AutoEnc'.
3. To bring out the characteristics of the primary proximity matrix in a real-time network to recognize community, the CNN-based auto-encoder approach is used. Case (3)'s label is written as 'CNN+AutoEnc'.
4. To extract the characteristics of a rebuilt proximity matrix in a real-time network to recognize community, an CNN-based auto-encoder approach is used. Case (4) is labeled as 'AutoEnc+ReMat+CNN'.

4.3.1. Execution of Reorganized Proximity Matrices. The process of the reorganization of proximity matrices is already elaborated in section 3. For the visualisation of the reorganized proximity matrices, we have considered two datasets as the samples. Figures 4.2 & 4.4 exhibit the reconstructed proximity matrices of dolphin and karate club network.

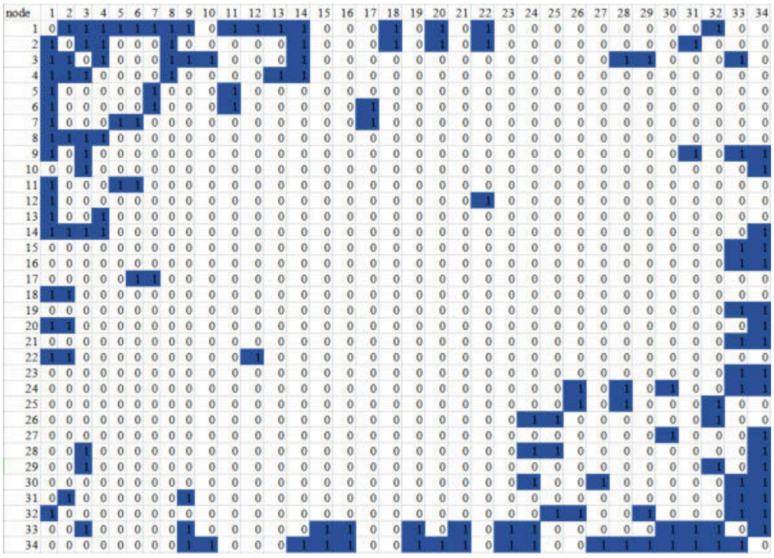

Fig. 4.3: Original Proximity Matrix of Karate Network

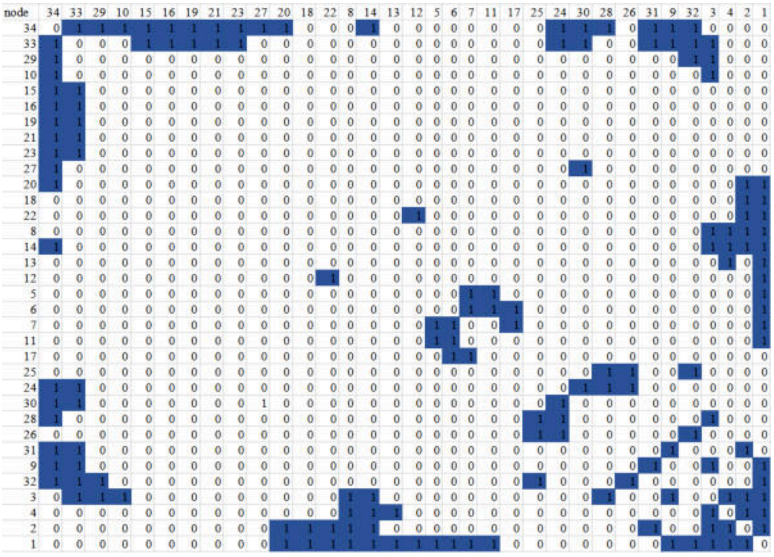

Fig. 4.4: Reorganized Proximity Matrix of Karate Network

4.3.2. Execution of Evaluation Matrices. This subsection provides the modularity score of all the four cases as shown in Table 4.2. Among all the four scenarios, 'AutoEnc+ReMat+CNN' has attained the highest modularity score. That is why we have considered the fourth case for comparing the modularity score with other existing algorithms. Here, we have considered other three popular algorithms for comparison, namely, 'Kmeans+NetRA','Kmeans+Node2Vec', and 'Kmeans+SDNE'. For all the 10 number of real time datasets, our proposed approach has gained the highest modularity score for community identification. The rebuilt proximity matrix is essentially a representation of the network's structure learned by the auto-encoder. The quality of the proximity matrix has a significant impact on how accurately communities are detected. A common technique for dimensionality reduction is auto-encoders. The reconstructed matrix's properties should

Table 4.2: Modularity Score of 4 Scenarios

Sl No.	Dataset	AutoEnc	ReMat+ AutoEnc	CNN+ AutoEnc	AutoEnc+ ReMat+ CNN
1	Karate	0.158	0.272	0.292	0.327
2	Football	0.539	0.629	0.730	0.768
3	Dolphins	0.484	0.539	0.549	0.625
4	Polbooks	0.544	0.593	0.637	0.683
5	Cora	0.358	0.397	0.478	0.483
6	Facebook	0.628	0.728	0.794	0.835
7	Artists	0.472	0.493	0.528	0.632
8	CiteSeer	0.469	0.528	0.573	0.624
9	Polblogs	0.372	0.448	0.472	0.527
10	School	0.573	0.576	0.638	0.735

Table 4.3: Modularity Score Compared with Existing Algorithms

Sl No.	Dataset	Kmeans+ NetRA	Kmeans+ Node2Vec	Kmeans+ SDNE	AutoEnc+ ReMat+ CNN
1	Karate	0.273	0.284	0.264	0.327
2	Football	0.528	0.618	0.593	0.768
3	Dolphins	0.528	0.492	0.519	0.625
4	Polbooks	0.492	0.439	0.542	0.683
5	Cora	0.293	0.346	0.274	0.483
6	Facebook	0.639	0.629	0.737	0.835
7	Artists	0.583	0.428	0.484	0.632
8	CiteSeer	0.529	0.553	0.514	0.624
9	Polblogs	0.384	0.418	0.474	0.527
10	School	0.618	0.531	0.683	0.735

preserve pertinent data while becoming less dimensional. Since the network is real-time, the characteristics of the rebuilt proximity matrix is generated quickly and efficiently. Real-time networks require low-latency processing, so the auto-encoder approach should be designed to produce the matrix in a timely manner. A rebuilt proximity matrix generated by a CNN-based auto-encoder is required for accurate, efficient, and flexible community detection in real-time networks. These qualities influence the approach's quality of community detection, scalability, tolerance to noise and changes, and overall efficacy in real-time applications. These are significance that our proposed approach (Case 4) has improved the performance among all the mentioned existing methods in essence.

The result is displayed in Table 4.3. Our proposed one, 'AutoEnc+ReMat+CNN' has attained the highest modularity score in Facebook network with 83.5% and achieved lowest in Karate club network with 32.7%. While 'Kmeans+NetRA' generates its better modularity score in Facebook and School network with 63.9% and 61.8% consecutively, 'Kmeans+Node2Vec' method provides its best modularity score in Football and Facebook network with 61.8% and 62.9%. 'Kmeans+SDNE' method attained its best modularity score in Facebook network with 73.7%.

Table 4.4 represents the NMI score comparison with the existing algorithms. In this experiment, our proposed 'AutoEnc+ReMat+CNN' approach has out-beats the other algorithms. Our method has achieved the highest NMI score in Karate network with 100% follwed by Dolphins network with 95%, Football network with 93% etc. 'Kmeans+SDNE' has also generated the better values of NMI than the other two existing

Table 4.4: NMI Score Compared with Existing Algorithms

Sl No.	Dataset	Kmeans+ NetRA	Kmeans+ Node2Vec	Kmeans+ SDNE	AutoEnc+ ReMat+ CNN
1	Karate	0.63	0.81	0.88	1
2	Football	0.59	0.72	0.83	0.93
3	Dolphins	0.65	0.75	0.87	0.95
4	Polbooks	0.62	0.68	0.72	0.78
5	Cora	0.53	0.59	0.60	0.62
6	Facebook	0.70	0.74	0.78	0.84
7	Artists	0.58	0.64	0.68	0.73
8	CiteSeer	0.61	0.67	0.68	0.69
9	Polblogs	0.52	0.63	0.66	0.71
10	School	0.71	0.83	0.85	0.91

Table 4.5: MRR Score Compared with Existing Algorithms

Sl No.	Dataset	Kmeans+ NetRA	Kmeans+ Node2Vec	Kmeans+ SDNE	AutoEnc+ ReMAT+ CNN
1	Karate	0.62	0.69	0.78	0.82
2	Football	0.66	0.73	0.81	0.84
3	Dolphins	0.71	0.79	0.83	0.88
4	Polbooks	0.68	0.74	0.79	0.83
5	Cora	0.63	0.71	0.75	0.81
6	Facebook	0.72	0.78	0.84	0.90
7	Artists	0.65	0.74	0.79	0.85
8	CiteSeer	0.63	0.69	0.74	0.80
9	Polblogs	0.59	0.66	0.71	0.78
10	School	0.75	0.80	0.84	0.92

algorithms. Karate club network has achieved 88% followed by Dolphins network with 87%, Footbal network with 83% NMI score in ‘Kmeans+SDNE’ method. ‘Kmeans+Node2Vec’ method provides its best NMI score in Karate club network with 81%, followed by Dolphins with 75% and Facebook with 74%. ‘Kmeans+NetRA’ generates its best NMI score in School network with 71%, followed by Facebook with 70%.

MRR score is also considered as one of the evaluation measures which displayed in Table 4.5. After the experiments, our proposed approach has attained the best MRR score in School network with 92%, followed by Facebook with 90%. Rest of the datasets have also performed well in our proposed approach. On the other hand, ‘Kmeans+NetRA’ generates its best MRR score in School network with 75%, followed by Facebook with 72% and Dolphins with 71%. ‘Kmeans+Node2Vec’ method provides its best MRR score in School network with 80%, followed by Dolphins network with 79% and Facebook with 78%. ‘Kmeans+SDNE’ has also generated the better values of MRR than the other two existing algorithms. It has achieved 84% in both Facebook and School network, followed by Dolphins with 83% and Football with 81%.

Table 4.6 represents the MAP score comparison with the existing algorithms. Our proposed method has outperformed among all the existing methods. It has attained the highest MAP score in School network with 88%, followed by Facebook with 87% and Artists with 83%. ‘Kmeans+NetRA’ generates its best MAP score in School network with 73%. School network has performed well in ‘Kmeans+Node2Vec’ also with 78% MAP score. But ‘Kmeans+SDNE’ has outperformed the other two methods and generate its best MAP score in Facebook network with 82%, followed by School with 81%.

Table 4.6: MAP Score Compared with Existing Algorithms

Sl No.	Dataset	Kmeans+ NetRA	Kmeans+ Node2Vec	Kmeans+ SDNE	AutoEnc+ ReMAT+ CNN
1	Karate	0.58	0.66	0.76	0.80
2	Football	0.63	0.71	0.78	0.82
3	Dolphins	0.65	0.72	0.75	0.81
4	Polbooks	0.64	0.70	0.76	0.80
5	Cora	0.58	0.68	0.72	0.78
6	Facebook	0.69	0.75	0.82	0.87
7	Artists	0.61	0.71	0.77	0.83
8	CiteSeer	0.60	0.66	0.71	0.78
9	Polblogs	0.57	0.64	0.69	0.75
10	School	0.73	0.78	0.81	0.88

5. Conclusion and Future Work. This research paper proposes a combined auto-encoder and CNN-based deep community identification approach for real time networks. During our experiments, we have first evaluated the modularity score on selected datasets using four different cases. Our proposed combined CNN and auto-encoder based method provides the prominent results on all the datasets. That is why we have considered our combined approach for evaluation and compare it with other existing approaches. To gather spatial adjacency matrices and reorganized proximity matrices, a novel matrix reorganization approach is proposed here. The matrix extends the standard proximity matrix with spatial closeness, obtaining obvious subspace features, and making convolutional processes simple and rapid to extract network spatial localisation. In this paper, the 'AutoEnc+ReMat+CNN' based approach is designed to obtain spatial eigenvectors, which spontaneously bring out the graph spatial properties and improve the modularity score. The combined model of the 'AutoEnc+ReMat+CNN'based community identification approach serves as the basis for community identification in a dynamic environment of the network.

The total amount of neurons in the input and output layers remains constant once the DL-based method is applied in spite of our approach 'AutoEnc+ReMat+CNN', which can be a useful investigation of the network embedding method. The interactions among the users in a real time community may change dynamically. Therefore, enhanced time-sequence approaches are necessary to draw out spatio-temporal properties for arbitrary real-time networks as future approach.

REFERENCES

[1] Yang, D., Liao, X., Wei, J., Chen, G. & Cheng, X. Modeling information diffusion with the external environment in social networks. *Journal Of Internet Technology.* **20**, 369-377 (2019)

[2] Perozzi, B., Al-Rfou, R. & Skiena, S. Deepwalk: Online learning of social representations. *Proceedings Of The 20th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining.* pp. 701-710 (2014)

[3] Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781.* (2013)

[4] Tao, Z., Liu, H., Li, J., Wang, Z. & Fu, Y. Adversarial graph embedding for ensemble clustering. *International Joint Conferences On Artificial Intelligence Organization.* (2019)

[5] Wang, C., Wang, C., Wang, Z., Ye, X., Yu, J. & Wang, B. DeepDirect: Learning directions of social ties with edge-based network embedding. *IEEE Transactions On Knowledge And Data Engineering.* **31**, 2277-2291 (2018)

[6] Cao, J., Jin, D., Yang, L. & Dang, J. Incorporating network structure with node contents for community detection on large networks using deep learning. *Neurocomputing.* **297** pp. 71-81 (2018)

[7] Cavallari, S., Zheng, V., Cai, H., Chang, K. & Cambria, E. Learning community embedding with community detection and node embedding on graphs. *Proceedings Of The 2017 ACM On Conference On Information And Knowledge Management.* pp. 377-386 (2017)

[8] Yu, W., Zheng, C., Cheng, W., Aggarwal, C., Song, D., Zong, B., Chen, H. & Wang, W. Learning deep network representations with adversarially regularized autoencoders. *Proceedings Of The 24th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining.* pp. 2663-2671 (2018)

[9] Xu, Y., Chi, Y. & Tian, Y. Deep convolutional neural networks for feature extraction of images generated from complex

networks topologies. *Wireless Personal Communications.* **103** pp. 327-338 (2018)
[10] Xu, J., Li, M., Jiang, J., Ge, B. & Cai, M. Early prediction of scientific impact based on multi-bibliographic features and convolutional neural network. *IEEE Access.* **7** pp. 92248-92258 (2019)
[11] Resler, A., Yeshurun, R., Natalio, F. & Giryes, R. A deep-learning model for predictive archaeology and archaeological community detection. *Humanities And Social Sciences Communications.* **8** (2021)
[12] Li, S., Jiang, L., Wu, X., Han, W., Zhao, D. & Wang, Z. A weighted network community detection algorithm based on deep learning. *Applied Mathematics And Computation.* **401** pp. 126012 (2021)
[13] Bhatia, V. & Rani, R. A distributed overlapping community detection model for large graphs using autoencoder. *Future Generation Computer Systems.* **94** pp. 16-26 (2019)
[14] Ferraro, A., Moscato, V. & Sperli, G. Deep learning-based community detection approach on multimedia social networks. *Applied Sciences.* **11**, 11447 (2021)
[15] Al-Andoli, M., Tan, S. & Cheah, W. Distributed parallel deep learning with a hybrid backpropagation-particle swarm optimization for community detection in large complex networks. *Information Sciences.* **600** pp. 94-117 (2022)
[16] Essaid, M. & Ju, H. Deep Learning-Based Community Detection Approach on Bitcoin Network. *Systems.* **10**, 203 (2022)
[17] Sun, K., Meng, F. & Tian, Y. Multi-level wavelet-based network embedded with edge enhancement information for underwater image enhancement. *Journal Of Marine Science And Engineering.* **10**, 884 (2022)
[18] Zhang, Y., Wang, Y., Chen, X., Jiang, X. & Zhou, Y. Spectral–spatial feature extraction with dual graph autoencoder for hyperspectral image clustering. *IEEE Transactions On Circuits And Systems For Video Technology.* **32**, 8500-8511 (2022)
[19] Kulatilleke, G., Portmann, M. & Chandra, S. SCGC: Self-supervised contrastive graph clustering. *ArXiv Preprint ArXiv:2204.12656.* (2022)
[20] Qian, J., Song, Z., Yao, Y., Zhu, Z. & Zhang, X. A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes. *Chemometrics And Intelligent Laboratory Systems.* pp. 104711 (2022)
[21] Lim, H., Ullah, I., Han, Y. & Kim, S. Reinforcement learning-based virtual network embedding: A comprehensive survey. *ICT Express.* (2023)
[22] Goel, S. & Kumar, R. Folksonomy-based user profile enrichment using clustering and community recommended tags in multiple levels. *Neurocomputing.* **315** pp. 425-438 (2018)
[23] Goel, S. & Kumar, R. Collaboratively augmented UIP–Filtered RIP with relevancy mapping for personalization of web search. *Information Sciences.* **547** pp. 163-186 (2021)
[24] Hoffman, M., Steinley, D., Gates, K., Prinstein, M. & Brusco, M. Detecting clusters/communities in social networks. *Multivariate Behavioral Research.* **53**, 57-73 (2018)
[25] Wu, L., Zhang, Q., Chen, C., Guo, K. & Wang, D. Deep learning techniques for community detection in social networks. *IEEE Access.* **8** pp. 96016-96026 (2020)
[26] Zhao, S., Peng, R., Hu, P. & Tan, L. Heterogeneous Network Embedding: A Survey. *CMES-Computer Modeling In Engineering & Sciences.* **137** (2023)
[27] Pham, P., Nguyen, L., Nguyen, N., Kozma, R. & Vo, B. A hierarchical fused fuzzy deep neural network with heterogeneous network embedding for recommendation. *Information Sciences.* **620** pp. 105-124 (2023)
[28] Abbas, K., Abbasi, A., Dong, S., Niu, L., Chen, L. & Chen, B. A Novel Temporal Network-Embedding Algorithm for Link Prediction in Dynamic Networks. *Entropy.* **25**, 257 (2023)
[29] Ding, Y., Zhai, Y., Hu, M. & Zhao, J. Deep Forest Auto-Encoder for Resource-Centric Attributes Graph Embedding. *Pattern Recognition.* pp. 109747 (2023)
[30] Vo, T. An integrated topic modeling and auto-encoder for semantic-rich network embedding and news recommendation. *Neural Computing And Applications.* pp. 1-16 (2023)
[31] Al-Andoli, M., Tan, S., Cheah, W. & Tan, S. A review on community detection in large complex networks from conventional to deep learning methods: A call for the use of parallel meta-heuristic algorithms. *IEEE Access.* **9** pp. 96501-96527 (2021)
[32] Lázár, A., Abel, D. & Vicsek, T. Modularity measure of networks with overlapping communities. *Europhysics Letters.* **90**, 18001 (2010)
[33] Kv, T. On normalized mutual information: measure derivations and properties. *Entropy.* **19**, 631 (2017)
[34] Goel, S., Kumar, R., Kumar, M. & Chopra, V. An efficient page ranking approach based on vector norms using sNorm (p) algorithm. *Information Processing & Management.* **56**, 1053-1066 (2019)

Edited by: Rajni Mohana
Special issue on: Sentiment Analysis and Affective computing in Multimedia Data on Social Network
Received: Jul 31, 2023
Accepted: Oct 5, 2023