Active Learning for Semi Supervised Node Classification with Selective Features

Dr M V Jagannatha Reddy
Assistant Professor, Department of CSE
Gandhi Institute of Technology and Management (GITAM Deemed to be University)
Bengaluru Campus, Doddaballapura,
Karnataka, INDIA-561203
mvjagannathareddy@gmail.com

Dr G.JeraldPrasath
Assistant Professor, Department of CSE
Gandhi Institute of Technology and Management (GITAM Deemed to be University)
Bengaluru Campus, Doddaballapura,
Karnataka, INDIA-561203
jgeorge@gitam.edu

Dr Nagaraju Mysore
Associate Professor, Department of CSE
Gandhi Institute of Technology and Management (GITAM Deemed to be University)
Bengaluru Campus, Doddaballapura,
Karnataka,INDIA-561203
mnraju14101969@gmail.com

Dr J.Jegan
Associate Professor, Department of CSE
Aditya College of Engineering,Madanamalle
Chittoor Dt. Andhrapradesh, INDIA-517325
jegan.deepa@gmail.com

Dr Justin Verghese
Professor, Department of CSE
Gandhi Institute of Technology and Management (GITAM Deemed to be University)
Bengaluru Campus, Doddaballapura,
Karnataka,INDIA-561203
jvarghes@gitam.edu

ArticleInfo
Page Number: 353-365
PublicationIssue:
Vol.72 No. 1 (2023)

Abstract
Node classification is a fundamental process for graph information which basically establishes node classifications based on their properties and structure of graph. Current learning algorithms from graphically represented information have increased interest in Graph Neural Networks (GNNs) for predictive problems including node categorization or edge forecasting. The genuine success of GNNs is considerably constrained by the difficulty of obtaining a large number of node labels. In semi-supervised classification, active learning entails adding new labels to previously unlabeled data to boost the underlying classifier's performance. This paper presents active learning for semi supervised node classification (AL-SSC) with selective features. This approach use feature learning methods for reduce the dimensionality of the data and use SOM cluster-based active learning model for node categorization that utilize labeled and unlabeled data. Extensive experiments were used to analyze the accuracy of the suggested work's performance. Experiments are conducted using the three publicly accessible reference datasets Citeseer, Pubmed, and Cora. According to the findings, the suggested technique performs well compared with other previous methods.

ArticleHistory
Article Received: 12 October 2022
Revised: 24 November 2022
Accepted: 18 December 2022

Keywords Active Learning, Semi-supervised learning, node classification, feature learning

Introduction

Nodes with similar features are clustered [35] (or grouped) in order to classify them, which is a common unsupervised learning approach used for node categorization in graphs. When the labels for a small portion of the nodes are revealed, the unsupervised node categorization problem transforms into a semi-supervised learning problem [1]. Using a limited sample of nodes' class labels, semi-supervised node classification attempts to classify the nodes in a (attributed) network [2].

Graph neural networks (GNNs) are frequently used in information retrieval applications, such as recommender systems [3] [4], query response [5][6], and online search [7]. By recursively aggregating the neighbourhood data, graph convolution-based algorithms [8] [9] specifically incorporate rich node properties as well as the structural data of graphs. Despite their effectiveness, GNNs' node classification performance suffers noticeably when just a few numbers of labelled nodes are provided because they have a tendency to overfit to those few nodes.This research work uses active learning algorithm for node categorization.

Only some GNN-based active learning techniques have recently been developed for attributed graphs [10] [11]. However, in terms of node categorization, their performance is still far from adequate. These techniques do not fully utilise the appropriate representation power of unlabeled data nodes and largely use labelled nodes for label extraction.The information provided in uncategorized nodes is not fully utilised by the existing techniques. It is simple to employ attributes extrapolated from a learned unsupervised form to decide which nodes to choose in order to make advantage of the information in unlabelled nodes. This research work uses dimensionality reduction concept to select important features.

In many disciplines, including machine learning [34], pattern recognition, and image processing, graph-based feature extraction has drawn significant interest due to its effectiveness in reducing the dimensionality of data. However, traditional graph-based dimensionality reduction algorithms frequently rely on a similarity matrix with set weights, which has a significant impact on the subsequent feature extraction procedure [12]. There is an implicit or explicit assumption made by traditional attribute selection methods for networkinformation that there is a substantial amount of labelled training data. However, the labels of graph information are extremely expensive or challenging to collect in many real-world applications. It can be excessively expensive, prolonged, or yet impossible to create a large training dataset. This research combine unsupervised and supervised model for dimensionality reduction.

The key contribution of this research work is:

- A new active learning for semi-supervised node categorization called AL-SSC is suggested.
- Unsupervised and Supervised feature learning model is combined to select features from graph data.
- Based on the selected features, SOM based active learning algorithm is suggested for node classification.
- Comprehensive investigations of the reference datasets (Citeseer, Cora, Pubmed) on semi-supervised classification tasks reveal that AL-SSC outperforms baseline techniques in term of accuracy.

The remaining portion of this research paper is arranged as follows: Section 2 give explanationof the related work on active learning, semi supervised classification, node categorization and feature learning. The preliminary concepts associated to the suggested approach are described in Section 3. The proposed methodology explained in Section4 and Section 5 evaluates the performance of the suggested method. Lastly section 6 provides the conclusion.

Related Work

This section explains some related research work which includes node classification, feature learning and active learning.

Node Classification

Jia et al., [13] suggest a semi supervised node categorization approach which combines discriminable Squeeze and Excitation with GCN.To selectively enhance important characteristics and achieve the adaptive calibration of feature dimensions, graph convolutional networks are employed. For semi supervised node learning tasks, Liu et al. [14] offer a new higher-order GCN that can simultaneously gather data from different neighbourhoods by creating high-order convolution. By combining the neighbourhood information via multi-scale neighbourhood pooling, it minimizes parameter counts by

employing a weight distribution approach. The complexity of this method is low because it only requires a few hidden units and fits a few parameters.

For graph convolutional networks, creating an autonomous procedure to establish a problem-dependent architecture can significantly lessen the requirement for manual construction of the model's structure during training. A technique for autonomously creating compacted and task particular graph convolutional networks are presented by Heidari et al. [15]. It uses a learning mechanism that makes use of the data provided to each layer to expand the graph structure in both depth and width. For semi-supervised node classification, Hu et al. [16] offer a new deep Hierarchical Graph Convolutional Network (H-GCN). In order to restore the representation for each node, it continually combined structurally related nodes to hyper-nodes and then processes the changed network to the baseline. Most of the node classification approach use GCN. It needs a lengthy hyper-parameter search procedure in order to find good network architecture is one of their key disadvantages.

Feature Learning

Dimensionality reduction is essential in feature learning models because it makes it easier to solve issues like interpreting high dimension information and lessens the dimensionality curse. Liu et al [17] propose new structured optimum network based sparse feature extraction method. This approach simultaneously frames label propagation, sparse representation, and local structure learning to reduce the dimensionality of the input. A novel graph embedding-based dimensionality reduction strategy based on the hypothesis margin of heterogeneous near neighbours is proposed by Tian et al. in [18].

Gou et al. [19] offer a novel exclusionary dimension and locally maintaining network embedding graph-based feature reduction approach. Extreme learning machine-based graph embedding dimension reduction approach is proposed by Yang et al. [20]. All samples are rebuilt using the weights from a graph matrix containing the supervised data. An effective method for dimensionality reduction implementations is provided by many feature learning models. For large, complex problems, feature learning models may not perform as expected.

Active Learning

Using attributed graphs, Li et al. [21] propose a semi-supervised adversarial active learning framework that fully utilises the representational power of deep neural networks and develops a novel AL query approach for node categorization. A batch mode active learning method is suggested by Ping et al. in [22] that chooses which nodes to label by using information-theoretic methods and random walks. It doesn't need to know how many blocks there are in advance, merely has to know the network architecture.

For the purpose of classifying nodes, Liu et al. [23] suggest a clustering basedlearning approach that fully exploits the representational power of both labelled and unlabelled nodes. This framework specifically utilize the K-Medoids clustering on a latent space based on a dynamic mixture of both unsupervised and supervised features to choose nodes for labelling. A unique Model-Change active learning acquisition function and a broad framework that

unifies several graph-based semi-supervised models are presented by Miller et al. in [24]. It uses the graph Laplacian matrix's spectrum, which can be reduced to prevent unreasonably high complexity and costs.

Preliminary Concept

Node Classification

When the information is represented as a network, node categorization plays a significant part in learning issues. V nodes and E edges join pairs of nodes to form a graph, G. A graph's edges can also have directions. Node categorization is frequently employed in real-world contexts, including social network analysis [25], and recommender systems [26].For the given attribute G = (V, E) with N nodes and M features, each node has set of valued feature vector $x_i \in \mathbb{R}^{N\times F}$ with one class label C. The adjacency matrix and node feature matrix of graph G is represented by $AdM \in \mathbb{R}^{N\times N}$ and $AtM \in \mathbb{R}^{N\times F}$. The element in AdM is represented as w_{ij}. If $w_{ij} = 1$ when there is a connection between nodes v_i and v_j,Otherwise$w_{ij} = 0$.AdM is symmetric when the graph G is undirected. A degree matrix is the diagonal matrix represented by $DiM = \{d_1, d_2, .., d_N\}$where$d_i = \sum_{j=1}^{N} w_{ij}$.

Active Learning

The pool based active learning (AL) is depicted in figure 1. By choosing currently unlabeled points (U) to label via an oracle (i.e., human in the loop) to add to the labelled data and hence enhance the underlying semi-supervised classifier, active learning takes the next logical step from the semi-supervised learning problem.

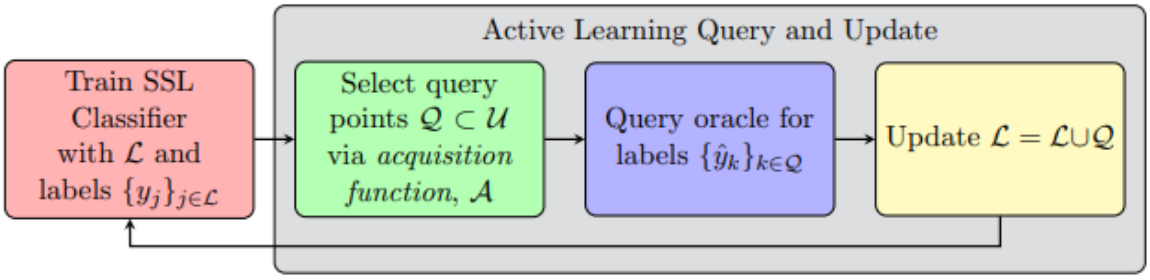

Figure 1 Pool based AL [24]

The active learning method iterates between (1) determining a graph-based semi-supervised classifier given the labelled data (L)available and (2) locating a query set (Q □ U) of unlabeled points to label and then update the labelled data. A real-valued acquisition function A: U →R that measures the value of labelling a single unlabeled point x_i, i ϵ U yields the query set Q.

The majority of active learning acquisition functions fall into one of three categories: uncertainty [27], margin [28] and clustering [29]. Unlabeled points whose classification is "uncertain" are preferred by uncertainty-based acquisition functions. Margin-based acquisition functions, which favour unlabeled points close to the decision border of the

current SSL classifier, are closely connected to uncertainty-based acquisition functions. Clustering-based approaches leverage their active learning query selection on the geometric clustering structure of the input data.

Proposed Methodology

This section explains proposed active learning for semi supervised node classification with selective features. The general steps for the proposed approach as follows:

Step1: The unsupervised and supervised combination of feature selection algorithm is applied to the given graph structured data.

Step2: SOM (Self Organizing Map) clustering algorithm is applied

Step3: Based on the Clustered data, the unlabelled data is classified through active learning.

The problem definition of semi supervised node categorization is described as:

Consider the Graph (G) having n nodes (V) and m edges (E). The node attribute and label vector of the G is represented as, $D \in \mathbb{R}^{n*f}$ and C. Each node $v \in V$ is specifically connected to a length-f feature vector d_v and one-hot label vector c_v. Node classification seeks to identify a model (M) that determines the label for each node in G such that the accuracy (performance) rateP(M | G, D, C) over the inputs (G, D, C) is maximized for a given a graph and its associated feature matrix.

Algorithm-1 explains unsupervised and supervised feature learning

Algorithm-1 Feature learning

Input: Graph G, Features F, ClassLable CL, Parameter $p_1 \epsilon [0,1]$

Output: Selected Features SF

1. For i = 1: f

2. Compute Mean and Standard deviation

$$MS_i = \frac{\sum_{c=1}^{CL}\left(\mu_{i,c} - \widehat{\mu_i}\right)^2}{\sum_{c=1}^{CL}(\sigma_{i,c})^2}$$

3. Compute Mutual Information

$$MI_i = \sum_{y \in YL} \sum_{z \in f_i} p(z,y) \log\left(\frac{p(z,y)}{p(z)p(y)}\right)$$

4. Compute $a_i = p_1 * (MS_i + MI_i)$

5. For j = 1: f

6. Compute maximum standard deviation

$$SD_{ij} = \max(std(f_i), std(f_j))$$

7. Compute Correlation

$$CC = \frac{\sum(f_i - \bar{f})(f_j - \bar{f})}{\sqrt{\sum(f_i - \bar{f})^2 \sum(f_j - \bar{f})^2}}$$

8. Compute $UF_{ij} = p_1 * (SD_{ij} + (1 - CC))$

9. Compute $SF_{ij} = a_i * a_j$

10. End For

11. End For

12. Compute $s_1 = \max_{e_i \in \{e_0, .. e_{n-1}\}}(|e_i|)$ where e_i= eigenvalues of UF_{ij}

13. Compute $s_2 = \max_{e_i \in \{e_0, .. e_{n-1}\}}(|e_i|)$ where e_i= eigenvalues of SF_{ij}

14. Compute $r_1 = \frac{1}{s_1} \quad and\ r_2 = \frac{1}{s_2}$

15. Find Identity Matrix for UF and SF

16. Find $W_1 = (I(UF) - r_1 UF)^{-1}$ and $W_2 = (I(SF) - r_2 SF)^{-1}$

17. Select $SF = Top_k(W_1) \cap Top_k(W_2)$

Algorithm-2 explains active learning for semi supervised node classification.

Algorithm-2 AL-SSC

Input: Graph G, Features F, Query K, Class Label CL

1. For i = 1 to K
2. Select SF using feature learning algorithm-1
3. Apply SOM Clustering
4. For each cls in Cluster
5. Select the best unlabelled data (uld) based on the cluster center
6. Get label of Luld
7. Update label set CL = CL U Luld
8. Update the model Mnew= P(M(G), CL)
9. End For

Experimental Results

This section evaluates the effectiveness of the suggested strategy to different number of relevant methods on three citation datasets. Three publicly accessible datasets, baseline comparison techniques, and outcomes comparison are all explained in this section.

Datasets

To test the effectiveness of the suggested technique, this research work use three citation data including CiteSeer, Cora, and PubMed [30].The term "citation network" refers to a network made up of papers and the connections between them, such as co-authors and citation relationships. The citation network has simple structure and article classification and link prediction are the challenges of this data.Each node in a citation network consists of a class label describing the article's subject and a binary Bag-of-Words (BoW) attribute vector extracted from the description. With the help of the list of undirected edges connecting the nodes, the symmetric binary adjacency matrix is constructed, and the goal at hand is to forecast the subject of the articles using the BoW features and the citations made by those articles to other works. Table1 shows the summary of citation data.

Table 1Citation Data Summary

Dataset	Nodes	Edges	Features	Classes
Citeseer	3327	4732	3703	6
Cora	2708	5429	1433	7
Pubmed	19717	44338	500	3

CiteSeer

The majority of the scientific articles in the CiteSeer dataset are computer-related, including topics like machine language, artificial intelligence, databases, and information retrieval.It contains 3327 scientific articles which are separated into the following categories: Agents, AI, DB, IR, ML, and HCI. Each article is characterized by a 0/1-valued vector whichspecifies whether the related term is included in the dictionary or not. The dictionary has 3703 unique terms.

Cora

The Cora dataset is primarily made up of publications on machine learning that use techniques like reinforcement learning, genetic algorithms, and neural networks. The 2708 scholarly articles in the Cora dataset are categorized into the following categories: networks (818), probabilistic approaches (426), evolutionary algorithms (418), hypothesis (351), case-based (298), fortification (217), and rule mining (180). This dataset has 5429 connections.

Every article is represented by a 0/1-valued vector that indicates whether the related term is included in the dictionary or not. There are 1433 unique keywords in the dictionary.

PubMed

Scientific articles in the field of biomedicine comprise the majority of the PubMed dataset. Nodes represent texts, while edges represent citation relationships. The Diabetes dataset from the PubMed database includes 19717 scholarly articles about diabetic that are broken down into three types (type 1, type 2, and experimental diabetes mellitus). The citation network contains 44,338 links in total. Word vectors with TF/IDF weighting from a 500 different words are used to characterize each publication.

Baseline Methods

The following baseline methods are utilized to compare the effectiveness of the suggested approach: GCN [31], HGCN [16], NAGCN [32], D-SEGCN [13], HCNP [14], PGCN [15], DLM-SSC [33]. The description of base line methods is shown in Table 2.

Table 2 Baseline methods for comparison

Method	**Author and Year**	**Description**
GCN [31]	Kipf et al., 2017	Works on graph directly based on CNN
HGCN [16]	Hu et al., 2019	To enhance the receptive field, an unique deep hierarchical GCN is used for semi-supervised categorization.
NAGCN [32]	Gong et al., 2019	A convolutional network for neighbourhood adaptive graphs that is utilized for semi-supervised categorization tasks.
D-SEGCN [13]	Jia et al., 2020	Feature based GCN for semi-supervised node classification.
HCNP [14]	Liu et al., 2021	GCN for semi-supervised approach concurrentlycombined data from several neighbourhoods by building a high-order convolution.
PGCN [15]	Heidari et al., 2021	It automatically creates convolutional graph networks that are small and task-specific.
DLM-SSC [33]	Mandapati et al., 2022	It combines feature learning and high-order convolution for semi-supervised classification.

Result Comparison

The efficiency of the suggestedmethod is evaluated on the basis of node classification accuracy and compares with the base line methods.Table 3 shows the accuracy comparison.

Table 3Result Comparison of Baseline Vs AL-SSC

Method	**Citeseer**	**Cora**	**Pubmed**
GCN [ex2]	70.3	81.5	79
HGCN [16]	72.8	84.5	79.8
NAGCN [ex3]	72	82.8	79.7
D-SEGCN [13]	74	84.8	81.2
HCNP [14]	71.7	82.5	79.3
PGCN [15]	74.3	83.1	80.2
DLM-SSC [ex4]	76.5	**85.3**	78.7
AL-SSC	**79.3**	84.2	**84.9**

Figure 2 compares the accuracy of several baseline approaches for three sets of citation data. For the citeseer and pubmed datasets, the suggested technique has improved classification accuracy of 79.3 and 84.9. For cora data, the D-SEGCN [33] provides more accuracy.

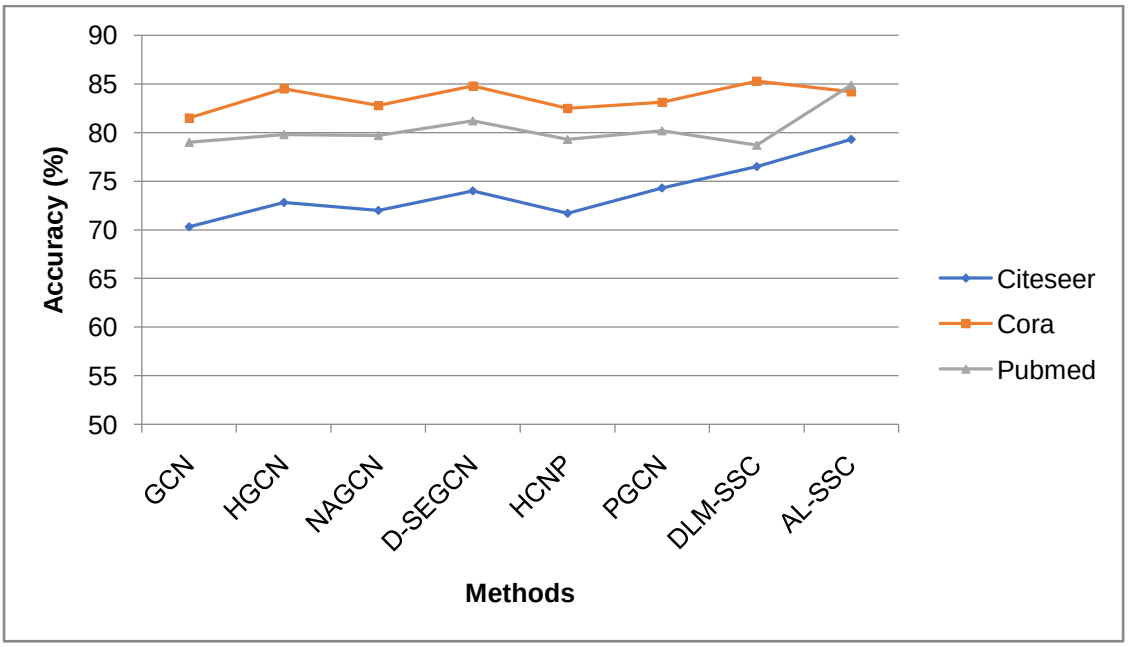

Figure 2 Accuracy Comparison

Conclusion

This research work presents active learning for semi-supervised node classification with selective features. Unsupervised and supervised feature learning is used to select the features from graph data. This work highlights a novel approach for active learning on graphs thatconstruct unsupervised and supervised feature learning with SOM clustering that is more appropriate for active learning.Three publicly accessible citation datasets (Citeseer, Cora, and Pubmed) are utilizedfor experiments. From results, the suggested approachattainsaccuracy 79.3% for Citeseer, 84.2% for Cora, and 84.9% for the Pubmed citation dataset. The experiments demonstrate that thesuggested method performs better than baseline approaches across three datasets.

References

1. Esmaeili, M., &Nosratinia, A. (2020). Semi-supervised node classification by graph convolutional networks and extracted side information. arXiv preprint arXiv:2009.13734.
2. London, B., &Getoor, L. (2014). Collective Classification of Network Data. Data Classification: Algorithms and Applications, 399.
3. Jin, B., Gao, C., He, X., Jin, D., & Li, Y. (2020, July). Multi-behavior recommendation with graph convolutional networks. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 659-668).
4. Yu, J., Yin, H., Li, J., Gao, M., Huang, Z., & Cui, L. (2020). Enhance social recommendation with adversarial graph convolutional networks. IEEE Transactions on Knowledge and Data Engineering.
5. Yusuf, A. A., Chong, F., &Xianling, M. (2022). An analysis of graph convolutional networks and recent datasets for visual question answering. Artificial Intelligence Review, 1-24.
6. Xu, X., Wang, T., Yang, Y., Hanjalic, A., & Shen, H. T. (2020). Radial graph convolutional network for visual question generation. IEEE transactions on neural networks and learning systems, 32(4), 1654-1667.
7. Mao, K., Xiao, X., Zhu, J., Lu, B., Tang, R., & He, X. (2020, July). Item tagging for information retrieval: A tripartite graph neural network based approach. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2327-2336).
8. Rong, Y., Huang, W., Xu, T., & Huang, J. (2019). Dropedge: Towards deep graph convolutional networks on node classification. arXiv preprint arXiv:1907.10903.
9. Zhang, H., Lu, G., Zhan, M., & Zhang, B. (2022). Semi-supervised classification of graph convolutional networks with Laplacian rank constraints. Neural Processing Letters, 54(4), 2645-2656.
10. Wu, Y., Xu, Y., Singh, A., Yang, Y., Dubrawski, A. (2019). Active learning for graph neural networks via node feature propagation. In: Proceedings of NeurIPS 2019 Graph Representation Learning Workshop (GRL)

11. Regol, F., Pal, S., Zhang, Y., Coates, M. (2020). Active learning on attributed graphs via graph cognizant logistic regression and preemptive query generation. In: ICML. pp. 8041–8050
12. Lai, J., Chen, H., Li, T., & Yang, X. (2022). Adaptive graph learning for semi-supervised feature selection with redundancy minimization. Information Sciences, 609, 465-488.
13. Jia, N., Tian, X., Zhang, Y., & Wang, F. (2020). Semi-supervised node classification with discriminable squeeze excitation graph convolutional networks. IEEE Access, 8, 148226-148236.
14. Liu X, Xia G, Lei F, Zhang Y, Chang S (2021) Higher-order graph convolutional networks with multi-scale neighborhood pooling for semi-supervised node classification. IEEE Access 9:31268–31275
15. Heidari, N., &Iosifidis, A. (2021). Progressive graph convolutional networks for semi-supervised node classification. IEEE Access, 9, 81957-81968.
16. F. Hu, Y. Zhu, S. Wu, L. Wang, and T. Tan, ‘‘Hierarchical graph convolutional networks for semi-supervised node classification,’’ in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 4532–4539
17. Liu Z, Lai Z, Ou W, Zhang K, Zheng R (2020) Structured optimal graph-based sparse feature extraction for semi-supervised learning. Signal Process 170:107456
18. Tian, Y., & Feng, X. (2021). Large Margin Graph Embedding-Based Discriminant Dimensionality Reduction. Scientific Programming, 2021.
19. Gou, J., Yang, Y., Yi, Z., Lv, J., Mao, Q., & Zhan, Y. (2020). Discriminative globality and locality preserving graph embedding for dimensionality reduction. Expert Systems with Applications, 144, 113079.
20. Yang, L., Song, S., Li, S., Chen, Y., & Huang, G. (2019). Graph embedding-based dimension reduction with extreme learning machine. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51(7), 4262-4273.
21. Li, Y., Yin, J., & Chen, L. (2020). Seal: Semisupervised adversarial active learning on attributed graphs. IEEE Transactions on Neural Networks and Learning Systems, 32(7), 3136-3147
22. Ping, S., Liu, D., Yang, B., Zhu, Y., Chen, H., & Wang, Z. (2017). Batch mode active learning for node classification in assortative and disassortative networks. IEEE Access, 6, 4750-4758.
23. Liu, J., Wang, Y., Hooi, B., Yang, R., & Xiao, X. (2020). Active learning for node classification: The additional learning ability from unlabelled nodes. arXiv preprint arXiv:2012.07065.
24. Miller, K., &Bertozzi, A. L. (2021). Model-change active learning in graph-based semi-supervised learning. arXiv preprint arXiv:2110.07739.
25. Chunaev, P. (2020). Community detection in node-attributed social networks: a survey. Computer Science Review, 37, 100286.
26. Salamat, A., Luo, X., & Jafari, A. (2021). HeteroGraphRec: A heterogeneous graph-based neural networks for social recommendations. Knowledge-Based Systems, 217, 106817.

27. Gal, Y., Islam, R., &Ghahramani, Z. (2017, July). Deep bayesian active learning with image data. In International Conference on Machine Learning (pp. 1183-1192). PMLR.
28. Jiang, H., & Gupta, M. R. (2021, August). Bootstrapping for batch active sampling. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (pp. 3086-3096).
29. Maggioni, M., & Murphy, J. M. (2019). Learning by active nonlinear diffusion. arXiv preprint arXiv:1905.12989.
30. Sen P, Namata G, Bilgic M, Getoor L, Galligher B, Eliassi-Rad T (2008) Collective classification in network data". AI Mag 29:93–93
31. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: Proc. int. conf. learn. represent., pp 1–13
32. Gong P, Ai L (2019) Neighborhood adaptive graph convolutional network for node classification. IEEE Access 7:170578–170588
33. Mandapati, S., Kadry, S., Kumar, R. L., Sutham, K., &Thinnukool, O. (2022). Deep learning model construction for a semi-supervised classification with feature learning. Complex & Intelligent Systems, 1-11.
34. Sirisha, P., &M V Jagannatha Reddy (2015). An Optimal Machine learning Approach for Large-Scale Data. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4(9).
35. M V Jagannatha Reddy. & Kavitha B. (2016). Clustering Incomplete Mixed Numerical and Categorical Datasets using Modified Squeezer Algorithm, International Journal of Computer Sciences and Engineering,4(5)