Hao WANG, Li-yan DONG, Tie-hu FAN, Ming-hui SUN, 2020. A local density optimization method based on a graph convolutional network. *Frontiers of Information Technology & Electronic Engineering*, 21(12):1795-1803. https://doi.org/10.1631/FITEE.1900663

Front Inform Technol Electron Eng

A local density optimization method based on a graph convolutional network

Key words: Semi-supervised learning; Graph convolutional network; Graph embedding; Local density

Corresponding author: Tie-hu FAN
E-mail: fth@jlu.edu.cn
ORCID: https://orcid.org/0000-0003-1496-9464

Motivation

1. GCN ignores some local information at each node in the graph, so data preprocessing is incomplete and the model generated is not accurate enough.

2. Unsupervised graph analysis methods pay more attention to the local adjacency of nodes in the graph, so it is a feasible idea that they can solve the problem and optimize data processing of GCN.

3. When processing GCN data, we do not hope to expand the scale of the original data.

Front Inform Technol Electron Eng

Main idea

1. Feature mapping in the spectral domain is required during training. The mapped matrix is the decomposition matrix of the Laplacian matrix, so we think that adding local features to the Laplacian matrix can achieve the goal.

2. Define the rules of local features to make the nearest local features more important.

3. When calculating the high-order matrix, we can adjust the calculation order of each node because the calculation is synchronous.

Front Inform Technol Electron Eng

Method

1. Unbalanced method: For the input original matrix, we directly calculate the high-order degree matrix of each node one by one, and then use the degree matrix to calculate the Laplacian matrix.

2. Balanced method: For the original input matrix, we first sort it according to the local density of each node, and then calculate their higher-order matrix.

Front Inform Technol Electron Eng

Major results

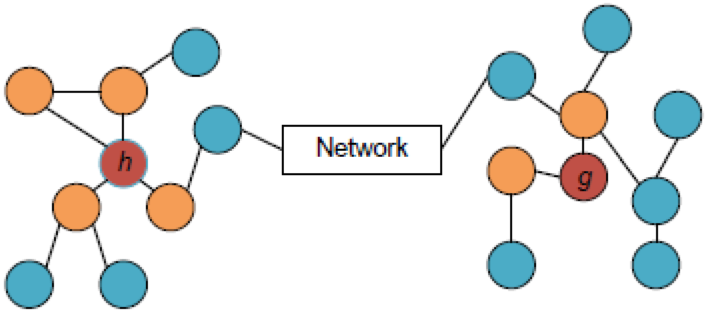


Fig. 1 Difference between the local characteristics of nodes *g* and *h* in the rules of local features

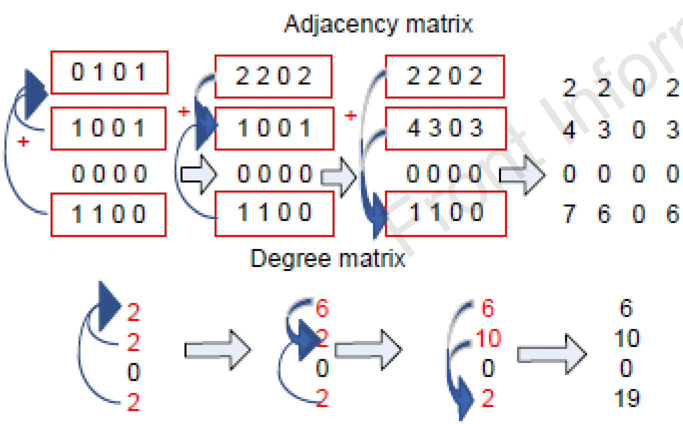


Fig. 2 Calculation process of the second degree matrix in a graph with four nodes

Algorithm 1 Third-order degree matrix computation

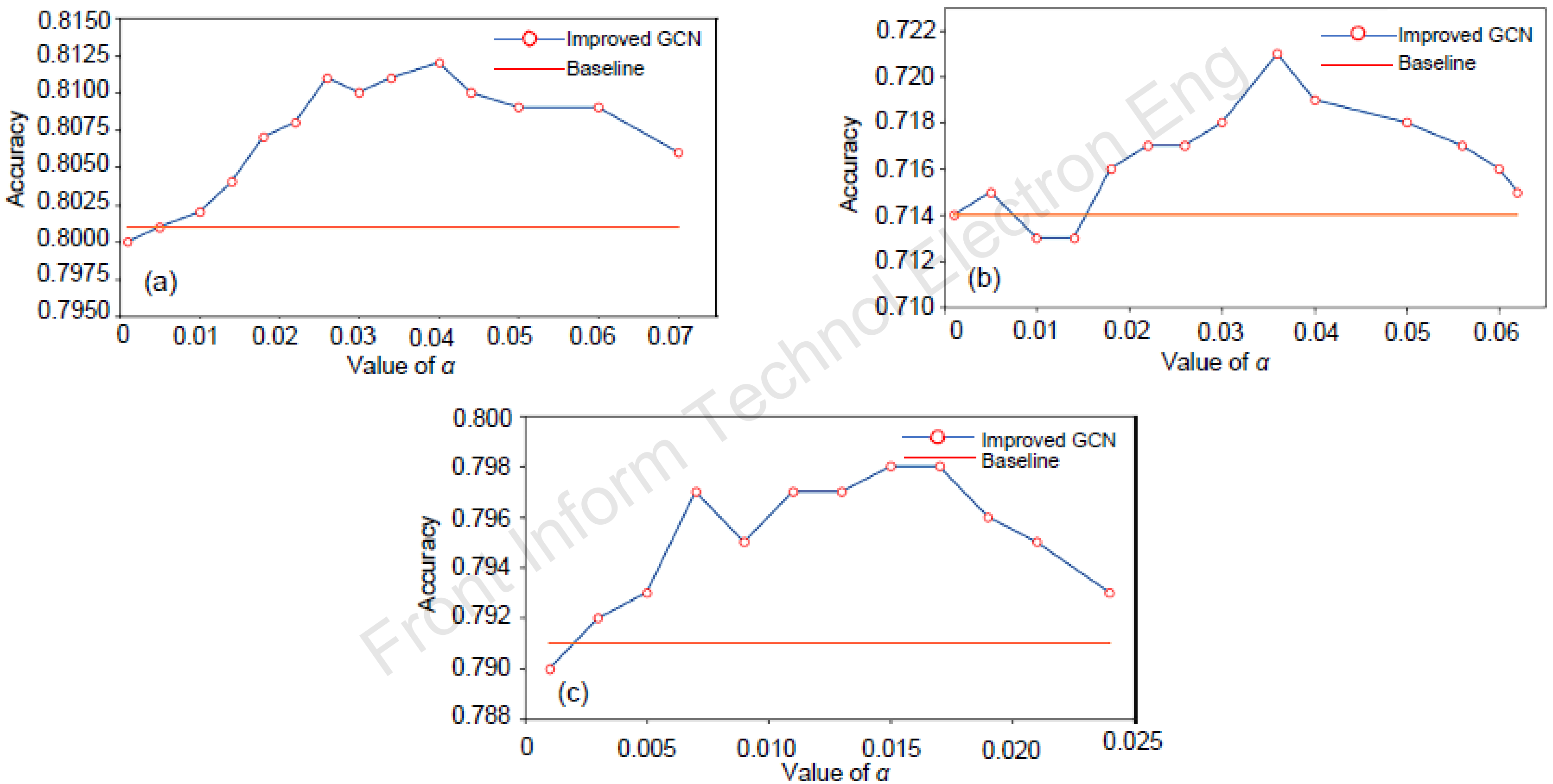
Input: adjacency matrix A

Output: Laplacian matrix L

D=sum(A.row)
define D_2 and D_3
D_2, D_3=D
for A_i in A
  for A_{ij} in A_i
    if nodes v_i and v_j are connected **do**
      calculate $D=D+\alpha_1 D_2$
      for A_{jk} in A_j
        if nodes v_j and v_k are connected **do**
          calculate $D=D+\alpha_2 D_3$
        end if
      end for
    end if
  end for
end for
calculate $L=D^{-1/2}AD^{-1/2}$
return matrix L

Front Inform Technol Electron Eng

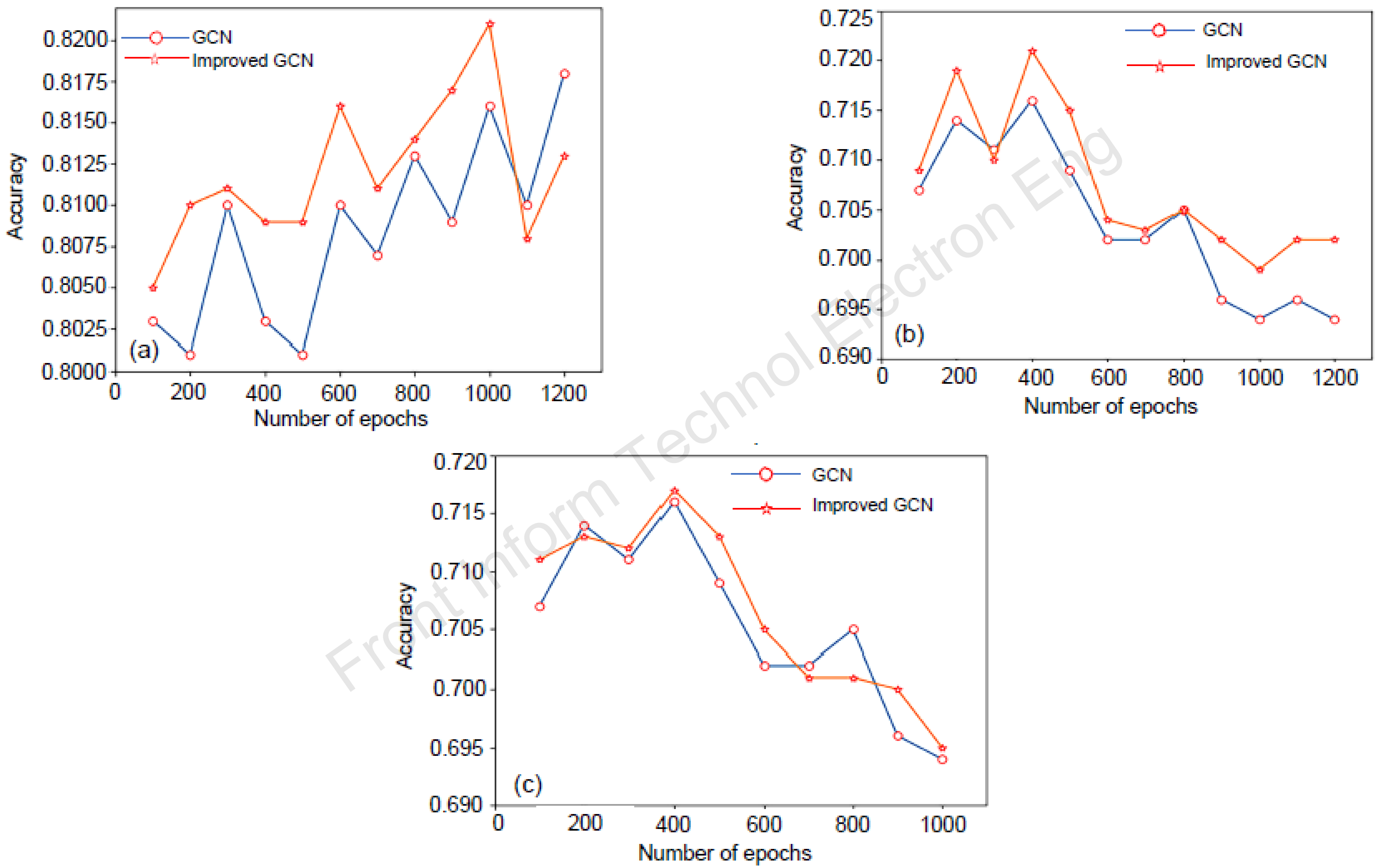
Major results (Cont’d)



Node classification prediction accuracy on the Cora (a), Citeseer (b), and Pubmed (c) datasets with changing α

Front Inform Technol Electron Eng

Major results (Cont'd)



Front Inform Technol Electron Eng

Node classification prediction accuracy on the Cora (a), Citeseer (b), and Pubmed (c) datasets with different numbers of epochs

Major results (Cont'd)

Table 8 Comparison of our LDGCN algorithm and five mainstream algorithms in terms of classification accuracies

Method	Accuracy		
	Cora	Citeseer	Pubmed
SemiEmb	0.590	0.596	0.717
DeepWalk	0.672	0.432	0.653
Planetoid	0.757	0.647	0.619
GCN	0.815	0.703	0.790
Graphite	**0.821**	0.710	0.793
LDGCN	0.819	**0.721**	**0.798**

Best results are in bold

Front Inform Technol Electron Eng

Conclusions

1. The existence of the extreme values indicates that the optimal coordination parameters are related to the dataset.

2. This optimization of input data is not affected by the epoch size.

3. When using different orders, the higher the order, the smaller the improvement of the algorithm, and the more time it takes to calculate the variable.

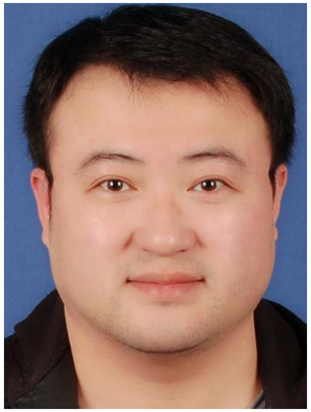
4. This improvement means that the focus of a GCN can move slightly into the graph structure like the unsupervised model.

Front Inform Technol Electron Eng

Hao WANG is currently a postgraduate at the College of Computer Science and Technology, Jilin University, China. His major research interests include graph data mining, data analysis, and machine learning. He is a CCF member.

Li-yan DONG received his PhD degree from Jilin University, China, in 2007. He is now a professor at the College of Computer Science and Technology, Jilin University. His research interests include database theory, data mining, distributed database and application, embedded database, and information retrieval.

Front Inform Technol Electron Eng



Tie-hu FAN received his BS and MS degrees from the School of Computer Science and Technology of Jilin University in computer science and technology in 2013 and 2006, respectively, and his PhD degree in computer intelligent control from Jilin University in 2010. He is now an associate professor. His major research interests include intelligent control, computational intelligence, and embedded operating systems.

Front Inform Technol Electron Eng

Ming-hui SUN received his PhD degree in computer science from the Kochi University of Technology, Japan, in 2011. He is currently an associate professor in the College of Computer Science and Technology in Jilin University, China. He is interested in using HCI methods to solve challenging real-world computing problems in many areas, including tactile interface, pen-based interface, and tangible interface.