

## **ICPR2020**

Yuhang Zhang, Hongshuai Ren, Jiexia Ye,

Xitong Gao, Yang Wang, Kejiang Ye, Cheng-Zhong Xu

University of Chinese Academy of Sciences

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

State Key Lab of IOTSC, University of Macau

The first two authors contributed equally to this work. Kejiang Ye is the corresponding author.





#### Motivation

- Existing works on GCN have made some progress, however, one of their limitations is that the design of Adjacency Matrix (AM) as GCN input requires domain knowledge and such process is cumbersome and error-prone.
- 2 A fixed Adjacency Matrix is generally designed as binary values (i.e., ones and zeros) which can not reflect the real relationship between nodes.



# Part 02

# Methodology







adjacency matrix by using a binary value (If two nodes are adjacent, the value is denoted as 1, otherwise 0.). However, each node contributes differently to its neighbors, so, we propose an automatic optimized matrix based on node entropy to improve the performance of GCN.





Illustration of the optimized matrix. Fig. 1(a) is the sample distribution from Cora dataset. The blue rectangular block in Fig. 1(b) represents that all 200 nodes are initialized to 1. Fig. 1(c) is the optimized matrix by Section III-C. The horizontal and vertical coordinates represent nodes' positions, and different colors represent the new weights after network optimization.

博学驾志 格物明德



Algorithm	1	Graph	Convolutional	Networks	with	Adjacency
Modeling						

**Input**: The nodes' features X with dimension in  $N \times F$ . Pretrained GCN-64 model.

**Output**: Optimized matrix  $\boldsymbol{A}$  and the classification results.

- 1: Calculate the entropy  $\psi$  by the pre-trained model.
- 2: Initialize the adjacency matrix **A**.
- 3: for epoch of RL do
- 4: Initialize a random process N for action exploration.
  5: t = 1
- 6: while  $\epsilon \ll \epsilon'$  (Designed error threshold) do
- 7: Select action  $\alpha_t = \mu(s_t | \theta^{\mu}) + N_t$  according to the current policy and exploration noise  $N_t$ ;
- 8: Calculate the information entropy  $\psi$  by the input meta-model;
- 9: Use the equation (8) to update entropy  $\psi$ ;
- 10: Do  $a_{ij} = \alpha_t \psi$ ,
- 11: Update adjacency matrix A by  $a_{ij} = Max(a_{ij}, a_{ji})$ ;
- 12: Calculate  $Acc_{cur}$  with X, A;
  - t = t + 1
- 14: end while
- 15:  $R = Acc_{cur} Acc_{meta-model}$
- 16: Update agent;
- 17: end for

13:

18: return Best adjacency matrix A.

Our goal is to exploit the modeling method to optimize the value of the adjacency matrix in GCN. The Adjacency Matrix in most of existing works utilizes a binary value to represent the edge weight, which cannot fully reflect the real relationships between the adjacent nodes. Therefore, we first propose a method to find a better Adjacency Matrix and it does improve the performance of GCN. In this section, we do:

What do we optimize in GCN ?

#### Adjacency Matrix: A

How do we search the value of matrix in continuous space ?
 Deep Deterministic Policy Gradient

$$\begin{cases} \theta^{\mathcal{Q}} = \tau \theta^{\mathcal{Q}} + (1 - \tau) \theta^{\mathcal{Q}'} \\ \theta^{\mu} = \tau \theta^{\mu} + (1 - \tau) \theta^{\mu'}, \end{cases} \quad y_t = r_t + Q(s_{t+1}, \mu(s_{t+1}) | \theta^{\mathcal{Q}}) * \gamma - b$$

 How to update the policy generating matrix value ? Node entropy

Demo

**AOAM** 

Updating

Framework









AOAM

Updating



The illustration of calculating the node information entropy between nodes. *n* represents the number of nodes around node *A*.

# Part 03

Experiments





Attributes	PPI	Cora	Citeseer	Pubmed
#Nodes	56966	2708	3327	19717
#Edges	818716	5429 🛚	4732	44338
#Features	50	1433	3703	500
#Classes	121	7	6	3

Table I

500

Introduction for datasets: Protein-Protein Interaction (PPI) dataset consists 24 graphs corresponding to different human tissues, which is used for inductive learning. The rest three datasets in the Table I are mainly utilized in transductive task.

**Datasets** 

Model

Comparison



Datasets

Model

Comparison

Discussion

#### TABLE II: Results of transductive learning.

	Cora	Citeseer	Pubmed
ICA [29]	75.10%	69.10%	73.90%
LP [30]	68.00%	45.30%	63.00%
ManiReg [31]	59.50%	60.10%	70.70%
SemiEmb [32]	59.00%	59.60%	71.70%
DeepWalk [33]	67.20%	43.20%	65.30%
Planetoid [28]	75.70%	64.70%	77.20%
Chebyshev [14]	81.20%	69.80%	74.40%
MoNet [34]	81.70%	69.90%	78.80%
GAT [5]	83.30%	72.50%	79.10%
Auto-GNN [8]	83.60%	73.80%	79.70%
TOGCN [25]	83.10%	72.70%	79.50%
GCN-64	81.50 %	69.05 %	79.30 %
<b>Ours-AOAM</b>	84.60%	72.10%	81.60%

In this experiment, we use *Adam* as the optimizer, and the learning rate is set as 0.01. Elements of the adjacency matrix are initialized as 1 if the respective two nodes have an edge in the original graph otherwise set as 0, which remain fixed during training process.

The proposed algorithm applies L2 regularization with l = 0.0005. Using GCN-64 (2 layers for GCN) as the backbone to train a meta- model.



Datasets

Model Comparison

Discussion

- Our method in this paper performs better in the transductive learning (semisupervised) task.
- In the long run, the proposed approach is superior to the hand-crafted adjacency matrix. Our method is an end- to-end algorithm, and without human intervention in the process of matrix generation, and the size of the input matrix can be in any dimension.
- The effect of the proposed method partly depends on the accuracy of the metamodel since we initialize the weights by that. The meta-model from Citeseer dataset starts with a low accuracy and that is the reason why the Citeseer dataset performs worse compared with the other two datasets.



### **Contributions**



•In this paper, we develop an end-to-end optimal algorithm to define the input Adjacent Matrix without human intervention.

- •We find the relationship between entropy value and matrix. To that end, we are the first to propose a calculation approach (node information entropy) to update the matrix during the training process.
- •Our method achieves or matches state-of-the-art performance on two popular datasets: Cora and Pubmed.



# Thanks for Listening