Block Modeling-Guided Graph Convolutional Neural Networks

Dongxiao He, Chundong Liang, Huixin Liu, Mingxiang Wen, Pengfei Jiao, Zhiyong Feng

College of Intelligence and Computing, Tianjin University

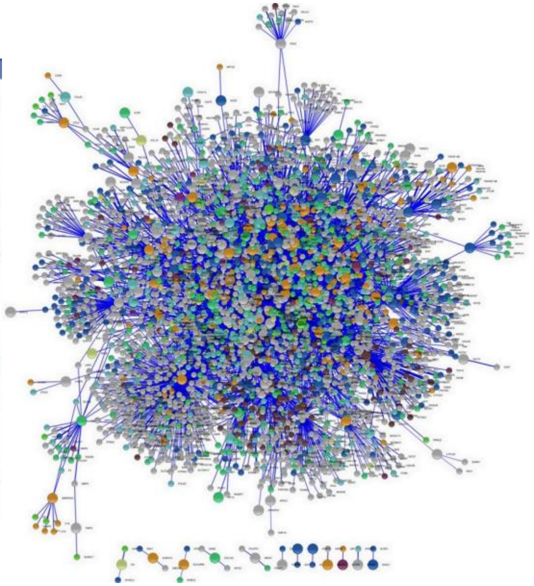
School of Cyberspace, Hangzhou Dianzi University

Outline

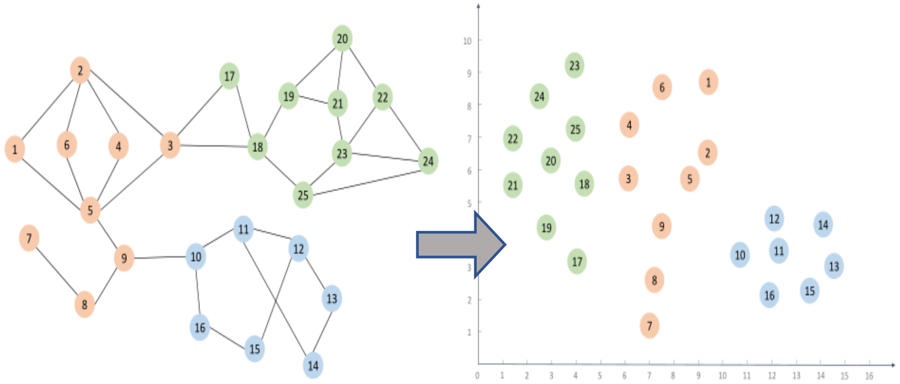
- Motivation
- Method
- Experiments
- Conclusion

Network Representation Learning

social network



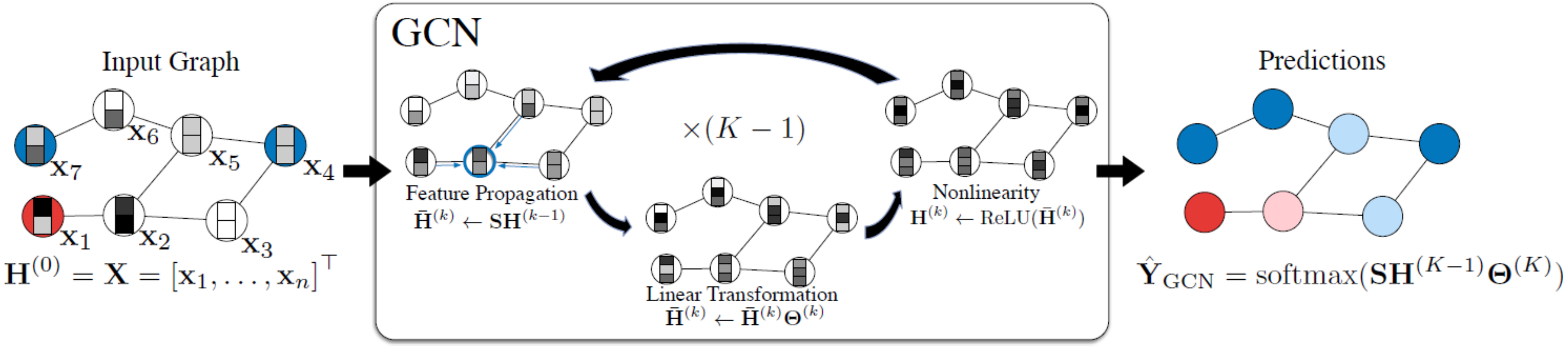
biology network



network representation learning

Graph Convolutional Networks

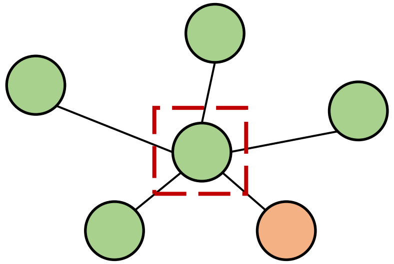
- **Essence**: propagate node attributes in neighbors guided by graph structure.
- **Goal**: encode nodes to embedding space by preserving network structure and properties.



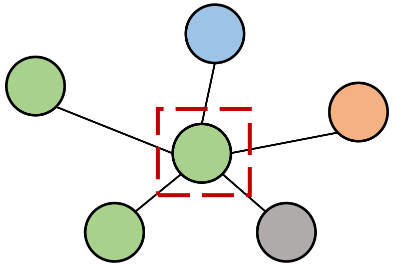
[1] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. ICLR, 2017.
[2] Wu F, Souza A, Zhang T, et al. Simplifying graph convolutional networks. PMLR, 2019.

Is GCN Universal?

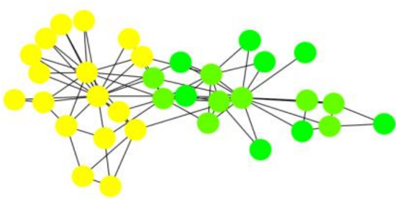
- **Question**: Can neighbor information represent a node?



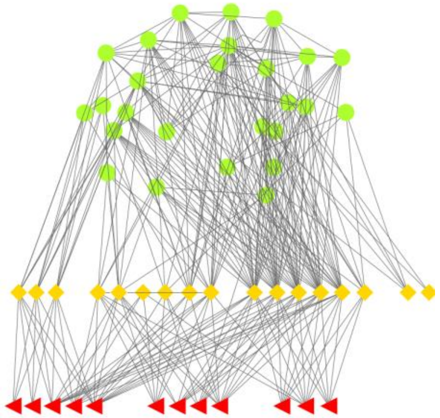
✓ GCN work



✗ GCN **NOT** work



homophily network



heterophily network

- **Homophily Assumption**: GCN only works on homophily networks.

most connections happen among
nodes in the same or similar classes

Existing Related Work

- **Aggregating higher-order neighbors**:
  - H_2GCN [Zhu *et al.*, 2020]
  - MixHop [Abu-El-Haija *et al.*, 2019]
  - …
- **Passing signed messages**:
  - GGCN [Yan *et al.*, 2021]
  - GPR-GNN [Chien *et al.*, 2019]
  - …

⚠ damage network topology

⚠ fail to define optimal aggregating mechanism

A BETTER WAY

automatically learn corresponding aggregation rules for neighbors of different classes

Homophily Ratio & Block Matrix

- **Homophily Ratio**: measure the overall homophily level in a network.

$$h = \frac{1}{|\mathcal{V}|} \sum_{v_i \in \mathcal{V}} \frac{|\{v_j | v_j \in \mathcal{N}_i, Y_j = Y_i\}|}{|\mathcal{N}_i|}$$

- **Block Matrix**: measure the connected possibility of nodes in any two classes.

$$H = (Y^T A \boxed{Y}) \oslash (Y^T A E)$$

Challenge 1: how to derive the block matrix in GCN without all known labels.

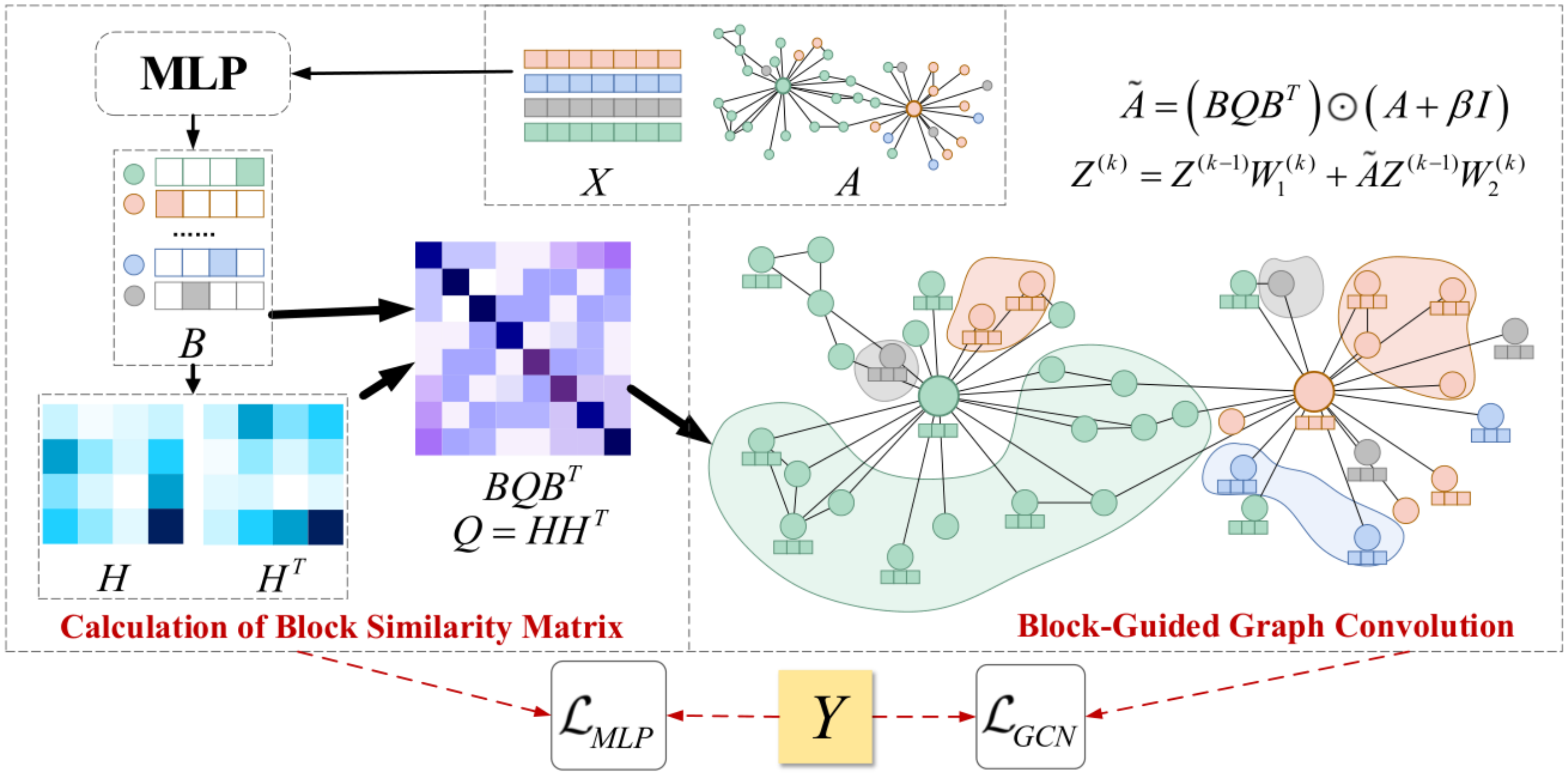
Challenge 2: how to design aggregation mechanism based on block matrix.

block matrix depicts the heterophilic property of the network in heterophilic situations.

Outline

- Motivation
- Method
- Experiments
- Conclusion

Overview – BM-GCN



Block Similarity Matrix

1) Pre-train MLP for soft labels

$$\bar{B} = \sigma\left(\mathrm{MLP}\left(X\right)\right)$$

$$B = \mathrm{softmax}\left(\bar{B}\right)$$

$$\mathcal{L}_{MLP} = \sum\nolimits_{v_i \in \mathcal{T}_{\mathcal{V}}} f\left(B_i, Y_i\right)$$

- obtain predicted soft labels via node attributes
- train MLP in a semi-supervised way

2) Calculate block matrix H

$$Y_s = \{Y_i, B_j | \forall v_i \in \mathcal{T}_{\mathcal{V}}, \forall v_j \notin \mathcal{T}_{\mathcal{V}}\}$$

$$H = \left(Y_s^T A Y_s\right) \oslash \left(Y_s^T A E\right)$$

- make full use of existing known labels
- H depicts the connecting pattern between classes

3) Calculate block similarity matrix Q

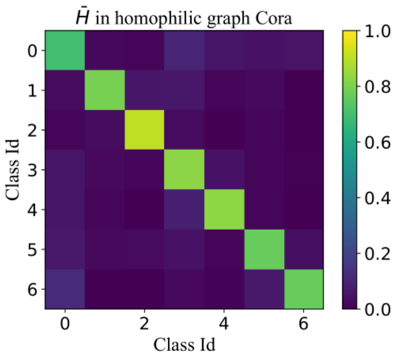
$$Q = HH^T$$

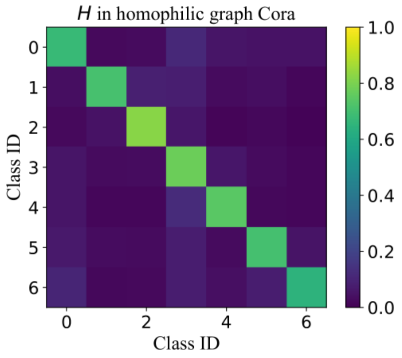
$$\mathrm{Diag}\left(Q\right) \leftarrow \alpha \cdot \mathrm{Diag}\left(Q\right)$$

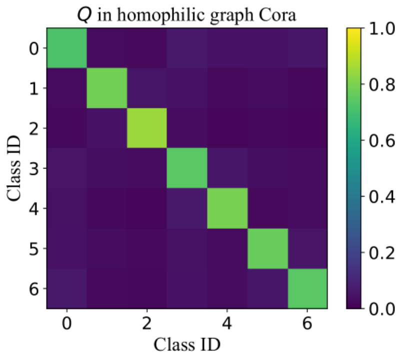
- Q depicts the similarity of connecting patterns between classes
- The more similar the two classes, the greater the value of the corresponding element in Q

Illustration on Why Block Modeling Effective

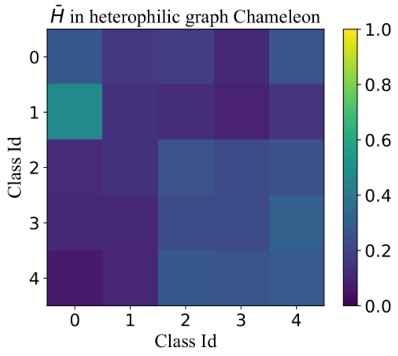
Homophilic Cora

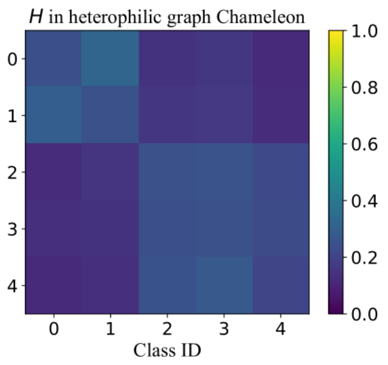


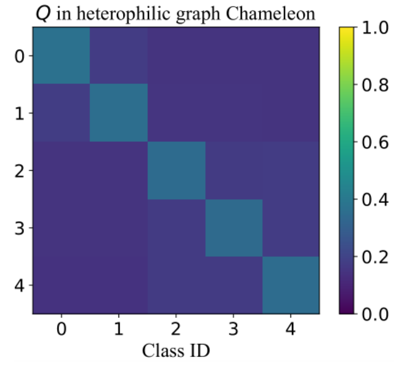




Heterophilic Chameleon







Block-Guided Graph Convolution Process

Soft label vector B_i: $\boxed{B_i^1 \mid B_i^2 \mid B_i^3 \mid B_i^4}$ The probability that v_i belongs to 2-th class

Soft label vector B_j: $\boxed{B_j^1 \mid B_j^2 \mid B_j^3 \mid B_j^4}$ The probability that v_j belongs to 3-th class

$B_i^2 B_j^3 \longrightarrow$ The probability that (v_i, v_j) belongs to (2-3) class-combination

$$Q = \begin{bmatrix} \; & \; & \; & \; \\ \; & \; & Q_{2,3} & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \\ \; & \; & \; & \; \end{bmatrix}$$

Information propagating ratio between 2-th class and 3-th class

$Q_{2,3} B_i^2 B_j^3 \longrightarrow$ Under the (2-3) class-combination,
How many message should be propagated between v_i and v_j

$$\omega_{ij} = \sum_{r=1}^{c} \sum_{t=1}^{c} q_{r,t} b_i^r b_j^t \longrightarrow$$

Consider all kinds of class-combinations,
information propagating ratio between v_i and v_j

Block-Guided Graph Convolution Process

Information propagating ratio for two nodes

$$\omega_{ij} = \sum_{r=1}^{c}\sum_{t=1}^{c} q_{r,t} b_i^r b_j^t$$

Information propagating ratio for all node pairs (in the form of matrix)

$$\Omega = BQB^T$$

A refined topology matrix based on Ω

$$A' = \Omega \odot (A + \beta I)$$

New graph convolutional layer

$$Z^{(k)} = Z^{(k-1)} W_1^{(k)} + \tilde{A} Z^{(k-1)} W_2^{(k)}$$

Semi-supervised Model Optimization (with fine-tuning MLP)

$$\mathcal{L}_{final} = \lambda \mathcal{L}_{GCN} + (1-\lambda)\,\mathcal{L}_{MLP} \quad , \quad \mathcal{L}_{GCN} = \sum_{v_i \in \mathcal{T}_{\mathcal{V}}} f\left(Z_i, Y_i\right)$$

Outline

- Motivation
- Method
- Experiments
- Conclusion

Experiments

- ## Datasets

Heterophilic Networks: texas, squirrel, chameleon

Homophilic Networks: cora, pubmed, citeseer

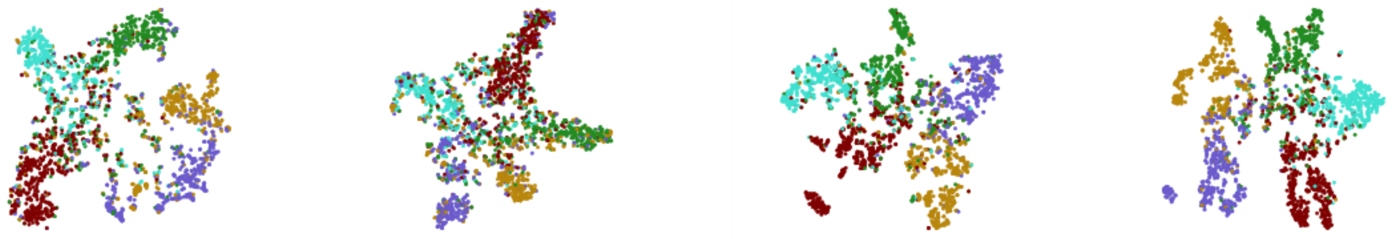
Datasets	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	c	d	h
texas	183	295	5	1703	0.11
squirrel	5,201	198,493	5	2,089	0.22
chameleon	2,277	31,421	5	2.325	0.23
cora	3,327	4,676	7	3,703	0.74
pubmed	19,717	44,327	6	500	0.80
citeseer	2,708	5,278	3	1,433	0.81

- ## Node Classification

Method/ Accuracy (%)	**heterophilic graphs**			**homophilic graphs**		
	Texas $h = 0.09$	**Squirrel** $h = 0.23$	**Chameleon** $h = 0.22$	**Cora** $h = 0.81$	**Citeseer** $h = 0.74$	**Pubmed** $h = 0.8$
MLP	82.70±6.19	33.35±1.24	48.20±2.63	74.14±1.40	69.58±2.31	86.38±0.61
GCN	55.41±3.47	44.07±1.95	67.04±2.23	86.48±1.12	72.67±1.99	87.39±0.68
H2GCN	82.16±8.21	28.91±1.78	51.58±1.51	87.69±1.37	75.95±2.18	88.78±0.53
GPR-GNN	84.59±4.37	29.45±1.27	**69.78±1.97**	86.70±1.03	75.12±1.98	87.38±0.63
CPGNN-MLP	77.09±4.22	28.65±1.50	52.63±1.79	85.23±1.71	74.29±2.41	86.83±0.78
CPGNN-Cheby	77.03±5.83	30.95±1.24	54.05±4.67	86.82±1.11	75.42±1.85	89.08±0.67
BM-GCN(Ours)	**85.13±4.64**	**51.41±1.10**	69.58±2.90	**87.99±1.29**	**76.13±1.92**	**90.25±0.75**

Experiments

Visualization



(a) H2GCN (b) CPGNN (c) GPR-GNN (d) BM-GCN

Figure 3: Visualization results on Chameleon dataset. Different colors correspond to different ground truth classes.

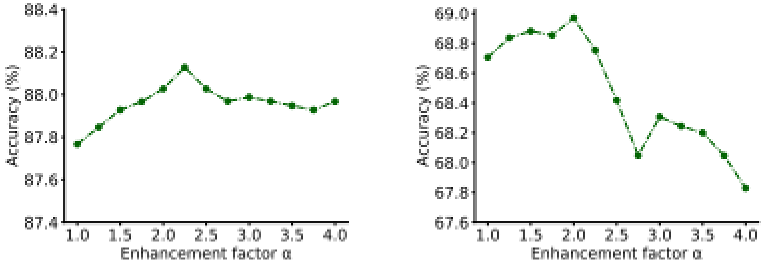
Parameter Analysis

Datasets/ Accuracy (%)	The number of graph convolutional layers k					
	1	2	3	4	5	6
Cora	81.27	86.92	87.99	87.30	60.10	40.16
Chameleon	58.71	67.74	69.58	65.00	49.28	34.69

Table 4: Node classification accuracy of BM-GCN with graph convolutional layers k varying from 1 to 6.

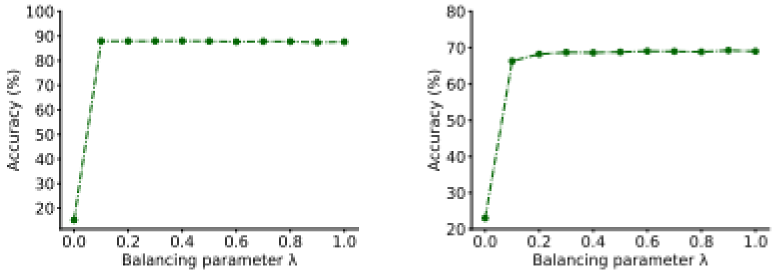
Experiments

◆ **Parameter Analysis**



(a) Cora (b) Chameleon

Figure 4: Parameter analysis of enhancement factor α in Q on Cora and Chameleon datasets. We report the average node classification accuracy over 10 random splits.



(a) Cora (b) Chameleon

Figure 5: Parameter analysis of loss balancing parameter λ on Cora and Chameleon datasets. We report the average node classification accuracy over 10 random splits.

Outline

- Motivation
- Method
- Experiments
- Conclusion

Conclusion

- Propose a new framework to make GCN applicable to both homophilic and heterophilic networks.
- Introduce block modeling technology to solve the problem of Homophily Assumption.
- Propose a novel design of block similarity matrix to enable block modeling technology to guide GCN to achieve classified aggregation.

Thanks

Data & Code: https://github.com/hedongxiao-tju/BM-GCN