Graph Deep Learning for Medical Applications

Improving Graph Neural Networks with Simple Architecture Design

Sunil Kumar Maurya, Xin Liu, Tsuyoshi Murata

Student: Berfin Elif Erdogan
Tutor: Franz Rieger
03.05.2022

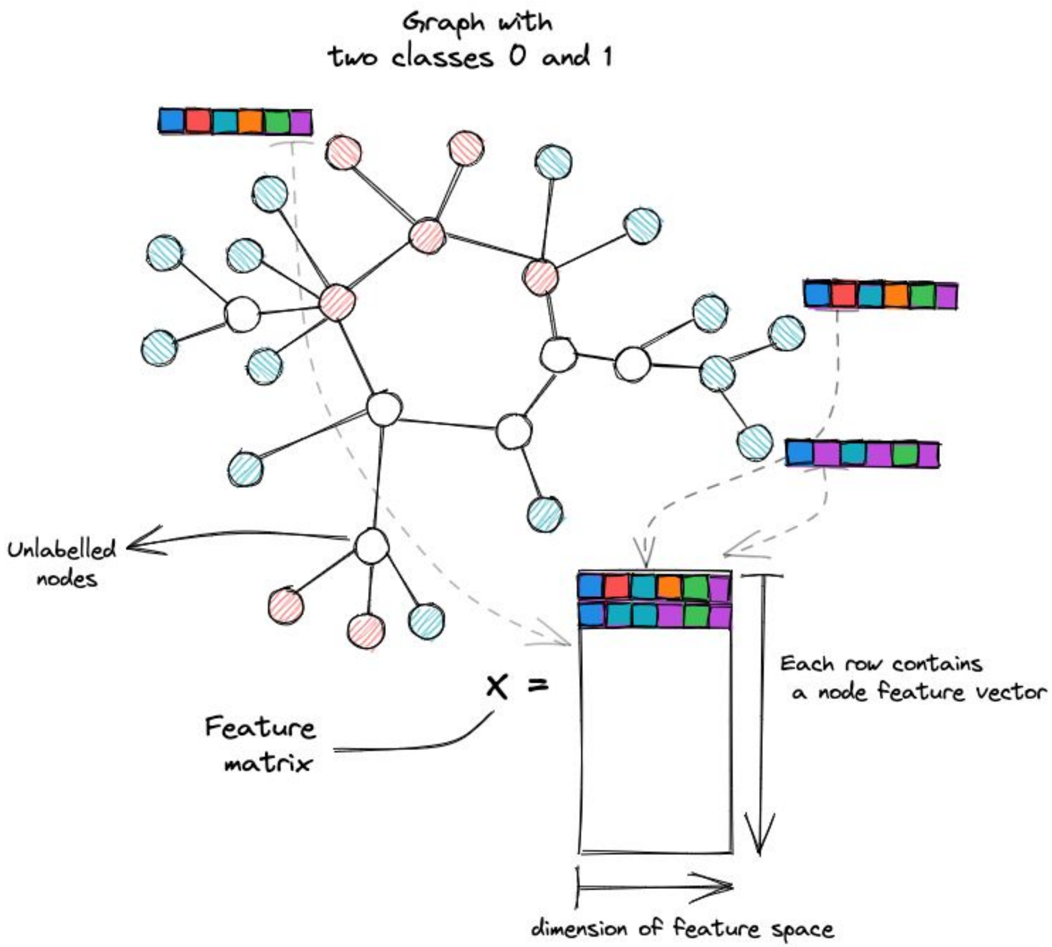
Technische Universität München + Johns Hopkins Whiting School of Engineering

Introduction

Technische Universität München

Johns Hopkins
Whiting School
of Engineering

Graph NN



Graph with
two classes 0 and 1
Unlabelled
nodes
X =
Feature
matrix
Each row contains
a node feature vector
dimension of feature space

Motivation

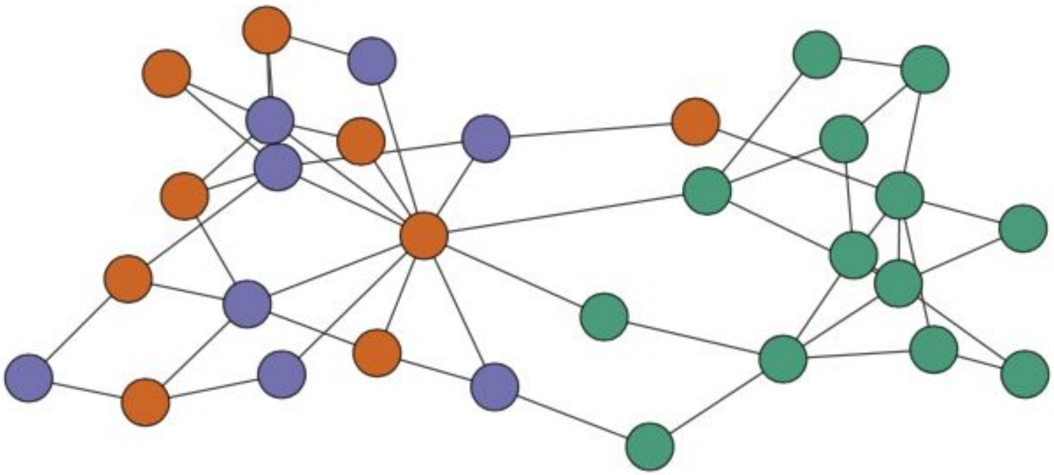
CHALLANGES

- Importance of features
- Expressiveness of Neural Network Layers

TO OVERCOME

- Unique mapping of features
- Feature selection
- Hop-Normalization

Homophily vs Heterophily



[2][2]J. Zhu, "Revisiting the problem of heterophily for GNNs", *Jiong Zhu*, 2022. [Online]. Available: https://www.jiongzhu.net/revisiting-heterophily-GNNs/. [Accessed: 03- May- 2022].

Current State of the Art and related work

Neighbor Sampling Approaches
- GraphSAGE [3]
- FastGCN [4]

Improve Feature Propagation
- APPNP [6]
- JK [7]
- Geom-GCN [8]

Attention mechanism
- GAT [5]

Deep GNN
- DropEdge [9]
- GCNII [10]

[3]W. Hamilton et al. Inductive representation learning on large graphs. In I. Guyon et al., editors, NIPS, pages 1024–1034. 2017
[4] J. Chen et al. FastGCN: fast learning with graph convolu- tional networks via importance sampling. In International Conference on Learning Representations, Feb. 15, 2018.
[5] P. Velickovic et al. Graph attention networks. ArXiv, abs/ 1710.10903, 2017.
[6] J. Klicpera et al. Predict then propagate: combining neural networks with personalized pagerank for classification on graphs. 2018.
[7] K. Xu et al. Representation learning on graphs with jumping knowledge networks. In International Conference on Machine Learning. International Conference on Machine Learning, pages 5453–5462. PMLR, July 3, 2018. ISSN: 2640-3498.
[8] H. Pei et al. Geom-GCN: geometric graph convolutional networks. ICLR, 2020.
[9]Y. Rong et al. DropEdge: towards deep graph convolutional networks on node classification. In ICLR, 2020.
[10]M. Chen et al. Simple and deep graph convolutional net- works. ICML, 2020.

Methodology

Technische Universität München + Johns Hopkins Whiting School of Engineering

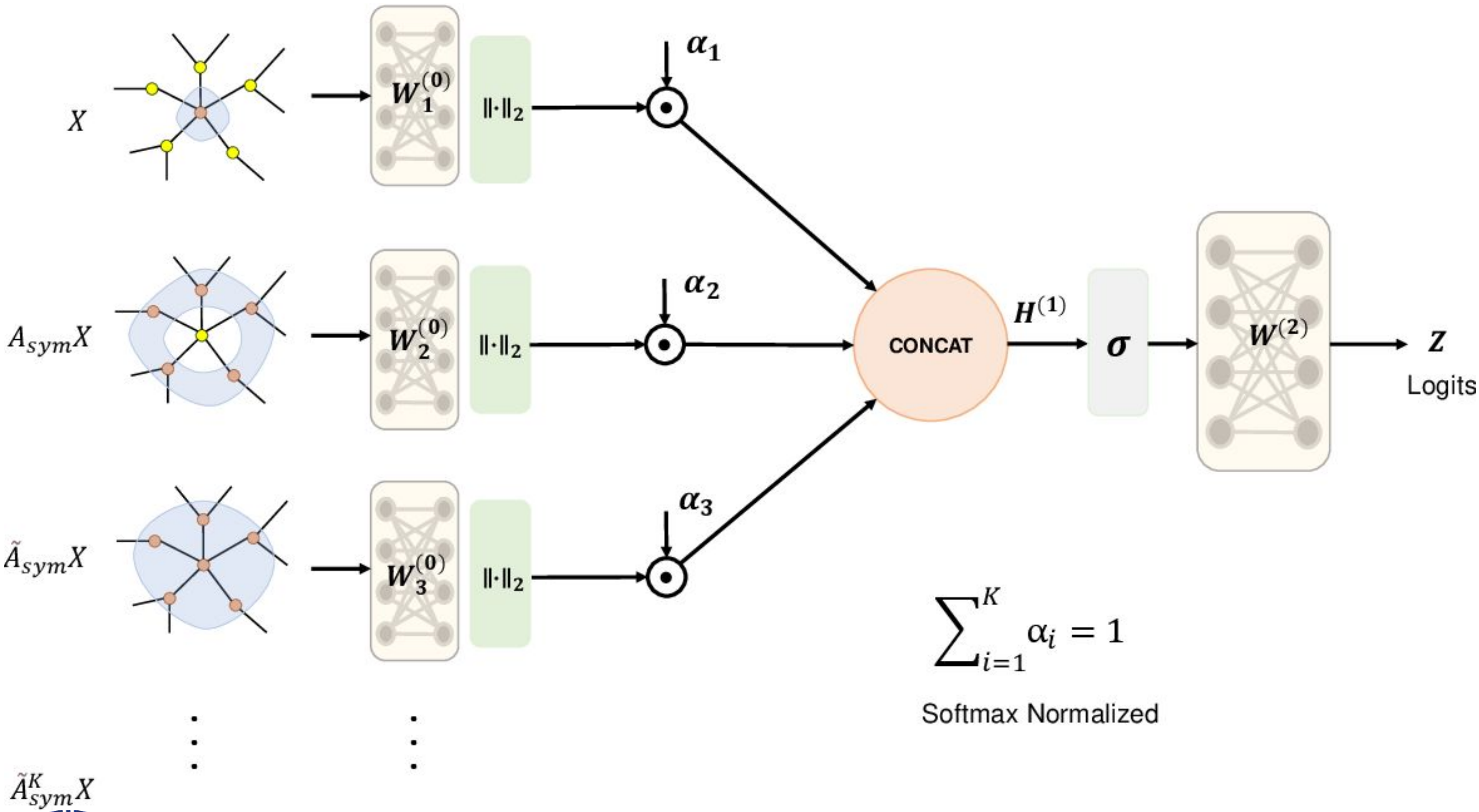
$$H^{(i+1)} = \sigma(\tilde{A}_{sym} H^{(i)} W^{(i)})$$

$$H^{(i+1)} = \sigma(A_{sym} H^{(i)} W^{(i)})$$

$$Z = \tilde{A}_{sym} \sigma(\tilde{A}_{sym} X W^{(0)}) W^{(1)}$$

CAMP

FSGNN



X
$A_{sym}X$
$\tilde{A}_{sym}X$
$\tilde{A}^K_{sym}X$
$W_1^{(0)}$
$W_2^{(0)}$
$W_3^{(0)}$
$\|\cdot\|_2$
α_1
α_2
α_3
CONCAT
$H^{(1)}$
σ
$W^{(2)}$
Z
Logits
$\sum_{i=1}^{K} \alpha_i = 1$
Softmax Normalized

Experimental Setup

Datasets

Datasets	Hom. Ratio	Nodes	Edges	Features	Classes
Cora	0.81	2,708	5,429	1,433	7
Citeseer	0.74	3,327	4,732	3,703	6
Pubmed	0.80	19,717	44,338	500	3
Chameleon	0.23	2,277	36,101	2,325	4
Wisconsin	0.21	251	499	1,703	5
Texas	0.11	183	309	1,703	5
Cornell	0.30	183	295	1,703	5
Squirrel	0.22	5,201	198,353	2,089	5
Actor	0.22	7,600	26,659	932	5
ogbn-papers100M		111,059,956	1,615,685,872	128	172

Experimental Results

	Cora	Citeseer	Pubmed	Chameleon	Wisconsin	Texas	Cornell	Squirrel	Actor
GCN	87.28±1.26	76.68±1.64	87.38±0.66	59.82±2.58	59.80±6.99	59.46±5.25	57.03±4.67	36.89±1.34	30.26±0.79
GAT	82.68±1.80	75.46±1.72	84.68±0.44	54.69±1.95	55.29±8.71	58.38±4.45	58.92±3.32	30.62±2.11	26.28±1.73
GraphSAGE	86.90±1.04	76.04±1.30	88.45±0.50	58.73±1.68	81.18±5.56	82.43±6.14	75.95±5.01	41.61±0.74	34.23±0.99
Cheby+JK	85.49±1.27	74.98±1.18	89.07±0.30	63.79±2.27	82.55±4.57	78.38±6.37	74.59±7.87	45.03±1.73	35.14±1.37
MixHop	87.61±0.85	76.26±1.33	85.31±0.61	60.50±2.53	75.88±4.90	77.84±7.73	73.51±6.34	43.80±1.48	32.22±2.34
GEOM-GCN	85.27	**77.99**	90.05	60.90	64.12	67.57	60.81	38.14	31.63
GCNII	**88.01±1.33**	77.13±1.38	**90.30±0.37**	62.48±2.74	81.57±4.98	77.84±5.64	76.49±4.37	N/A	N/A
H2GCN-1	86.92±1.37	77.07±1.64	89.40±0.34	57.11±1.58	86.67±4.69	<u>84.86±6.77</u>	82.16±4.80	36.42±1.89	**35.86±1.03**
Ours(3-hop)	87.73±1.36	77.19±1.35	89.73±0.39	<u>78.14±1.25</u>	**88.43±3.22**	**87.30±5.55**	<u>87.03±5.77</u>	<u>73.48±2.13</u>	35.67±0.69
Ours(8-hop)	<u>87.93±1.00</u>	<u>77.40±1.93</u>	<u>89.75±0.39</u>	**78.27±1.28**	<u>87.84±3.37</u>	**87.30±5.28**	**87.84±6.19**	**74.10±1.89**	<u>35.75±0.96</u>

Discussion

Technische Universität München

Johns Hopkins
Whiting School
of Engineering

Ablation Studies

	Cora	Citeseer	Pubmed	Chameleon	Wisconsin	Texas	Cornell	Squirrel	Actor
Proposed	83.68±2.22	74.48±1.44	**89.24±0.27**	**72.48±4.16**	81.48±5.62	**78.80±5.88**	**78.09±2.22**	**63.57±6.83**	33.54±1.21
Without soft-selection	**87.07±0.26**	**76.45±0.27**	89.09±0.39	72.27±1.34	78.03±6.55	76.28±6.72	74.32±6.54	61.73±4.15	34.15±0.64
Common weight ($W^{(0)}$)	83.19±1.41	72.15±1.02	88.96±0.28	68.24±6.03	70.56±10.94	68.45±7.65	68.18±9.13	56.63±8.54	32.73±1.48
Without Hop-normalization	77.12±3.49	71.40±10.01	87.72±0.77	53.06±6.18	**82.60±2.68**	76.33±3.87	76.18±3.43	32.60±6.38	**36.66±0.55**

Model Scalability

Method	Accuracy
SGC	63.29±0.19
Node2Vec	58.07±0.28
SIGN	65.11±0.14
FSGNN	**67.17±0.14**

Personal Review and Take-Home Message

Personal Review

Strengths

- Interesting novel concept of using homophily and heterophily datasets
- Outperforming the current state of the art GNN models on the node classification task
- Scalability
- Code is available on Github

Weaknesses

- Repeats itself
- Sometimes difficult to understand
- Insufficient description of the figures

Take-Home Message

- New architecture for node classification
  - Old methods in homophily datasets, proposed architecture in heterophily datasets performs well.

- Hyperparameter tuning is an important step

Thank you for your attention!
Questions?