Jianke HU, Yin ZHANG, 2022. NGAT: attention in breadth and depth exploration for semi-supervised graph representation learning. *Frontiers of Information Technology & Electronic Engineering*, 23(3):409-421.
https://doi.org/10.1631/FITEE.2000657

NGAT: attention in breadth and depth exploration for semi-supervised graph representation learning

Key words: Graph learning; Semi-supervised learning; Node classification; Attention

Corresponding author: Yin ZHANG
E-mail: yinzh@zju.edu.cn
ORCID: https://orcid.org/0000-0001-6986-4227

Motivation

1. Graph convolutional networks (GCNs) with deeper layers might show worse performance due to the problem of oversmoothing, which becomes the major bottleneck for these models to be scalable to large graphs.

2. Oversmoothing causes each node in a connected component to converge to a similar embedding, which means that the features learned by GCNs become indistinguishable.

Main idea

1. We propose an end-to-end deep architecture for graph representation learning. It directly accepts graphs as inputs without any feature engineering, and provides high-quality node embedding as outputs.
2. We propose a novel neighborhood aggregation algorithm to alleviate the problem of oversmoothing. As far as we know, this is the first attempt to design a depth-wise feature aggregation scheme in self-attention fashion.
3. The experimental results on a number of semi-supervised node classification datasets show that our nested graph attention network (NGAT) is highly competitive compared to other novel models for node classification.

Method

1. We present the architecture of the NGAT, which deploys the node- and layer-wise attention mechanisms.

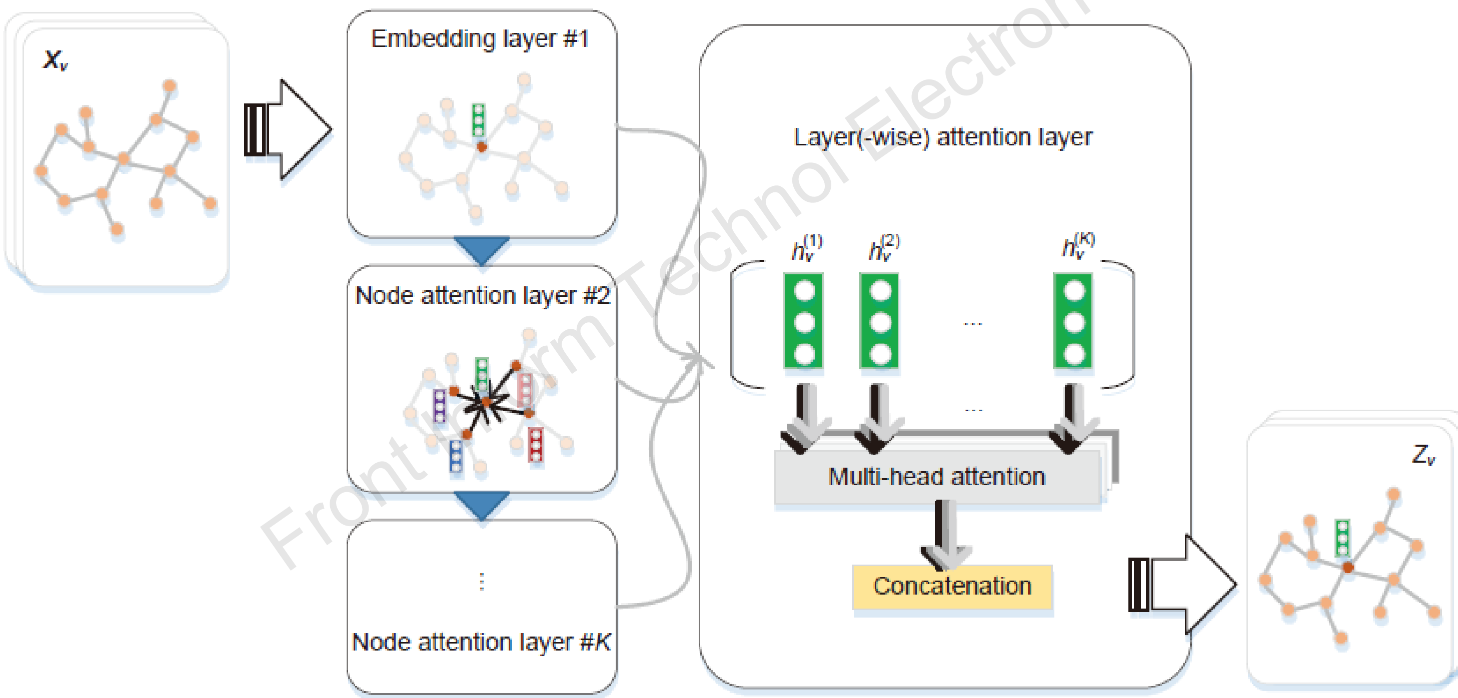
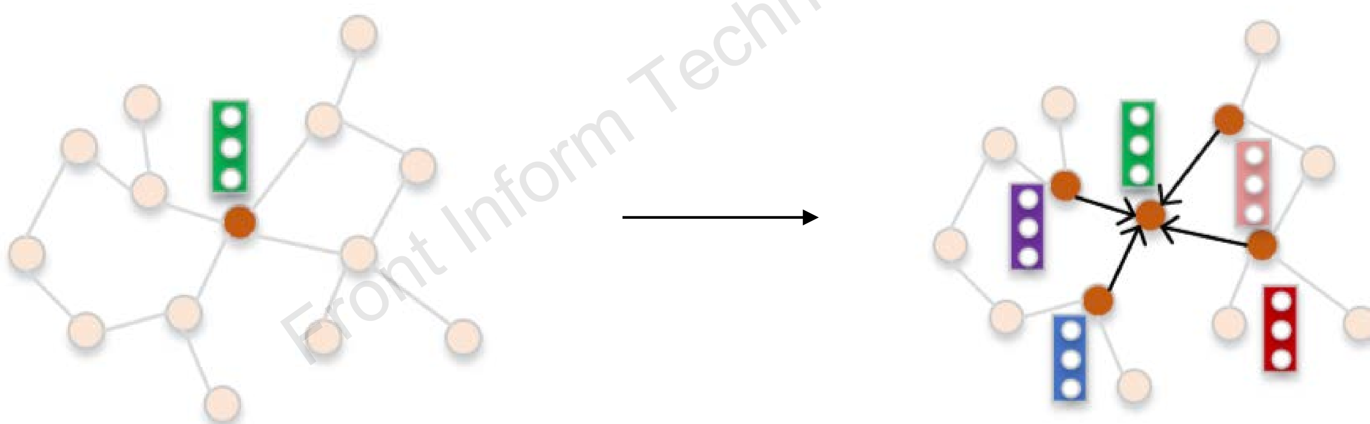


Fig. 2 Schematic layout of a K-layer nested graph attention network (NGAT) for graph representation learning. The architecture consists of three parts: (1) the embedding layer does a linear transformation with non-linear activation; (2) the node attention layer aggregates the feature information from neighbors in attention fashion; (3) the layer(-wise) attention layer selectively aggregates the feature information from different depths, and outputs the final embedding Z_v

Method (Cont’d)

2. Node-wise attention improves the neighborhood aggregation strategy by distinguishing the importance of each neighboring node. Cosine similarity is used to measure the correlation between different nodes.



Method (Cont'd)

3. Layer-wise attention is to selectively leverage these hidden embeddings of all layers to generate an informative embedding. We provide two attention mechanisms: parametric and non-parametric attention mechanisms.

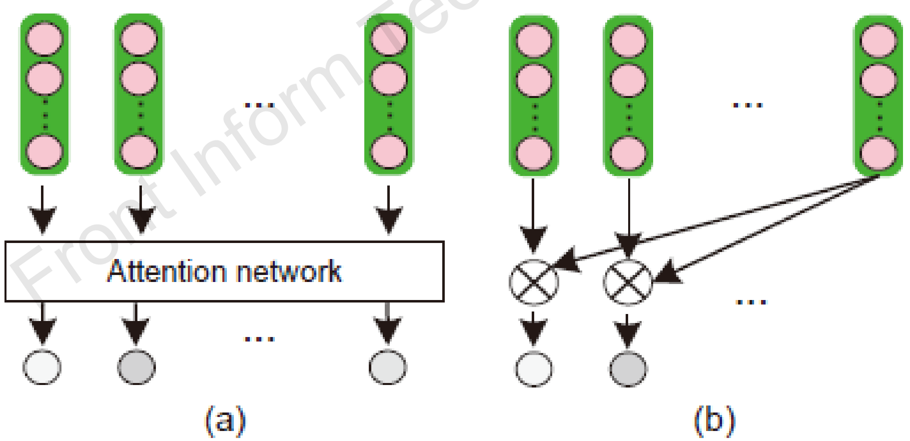


Fig. 3 Two versions of layer-wise attention: (a) parametric attention mechanism (self-attention); (b) non-parametric attention mechanism

Major results

Planetoid split

Table 2 Node classification accuracy on Planetoid split from Yang et al. (2016)

	Model	Node classification accuracy (%)				
		Cora	Citeseer	PubMed	Academic CS	Academic Physics
From literature*	GCN	81.5	70.3	79.0	91.1±0.5**	92.8±1.0**
	GAT	83.0±0.7	72.5±0.7	79.0±0.3	90.5±0.6**	92.5±0.9**
	AGNN	83.1±0.1	71.7±0.1	79.9±0.1	–	–
	DGI	82.3±0.6	71.8±0.7	76.8±0.6	–	–
	SGC	81.0±0.0	71.9±0.1	78.9±0.0	–	–
	LanczosNet	79.5±1.8	66.2±1.9	78.3±0.3	–	–
	AdaLNet	80.4±1.1	68.7±1.0	78.1±0.4	–	–
Our experiments	GCN	81.6±0.6	70.6±0.8	78.5±0.8	89.8±0.4	91.1±1.3
	GAT	83.2±0.4	72.3±0.7	79.2±0.5	90.1±0.8	92.1±0.8
	AGNN	83.8±0.2	71.0±0.4	79.8±0.4	90.9±0.3	91.5±0.3
	JK-Net	82.6±0.8	72.4±1.0	77.9±0.5	91.4±0.4	92.4±0.7
Our models	NP-NGAT	84.9±0.4	**72.7**±0.6	80.8±0.8	91.0±0.5	92.6±0.4
	P-NGAT	**85.1**±0.3	72.6±0.9	**81.1**±0.6	**91.7**±1.1	**93.2**±0.8

The best result in the corresponding dataset is in bold. * means that the results are cited from their original papers; ** means that the reported numbers are taken from Shchur et al. (2018); – denotes no result for the corresponding dataset. The reported numbers in "Our experiments" are the test accuracy values averaged over 10 runs. NP-NGAT and P-NGAT both outperform the compared models on all datasets based on Student's t-test ($p < 0.05$)

Major results (Cont'd)

Depth experiment

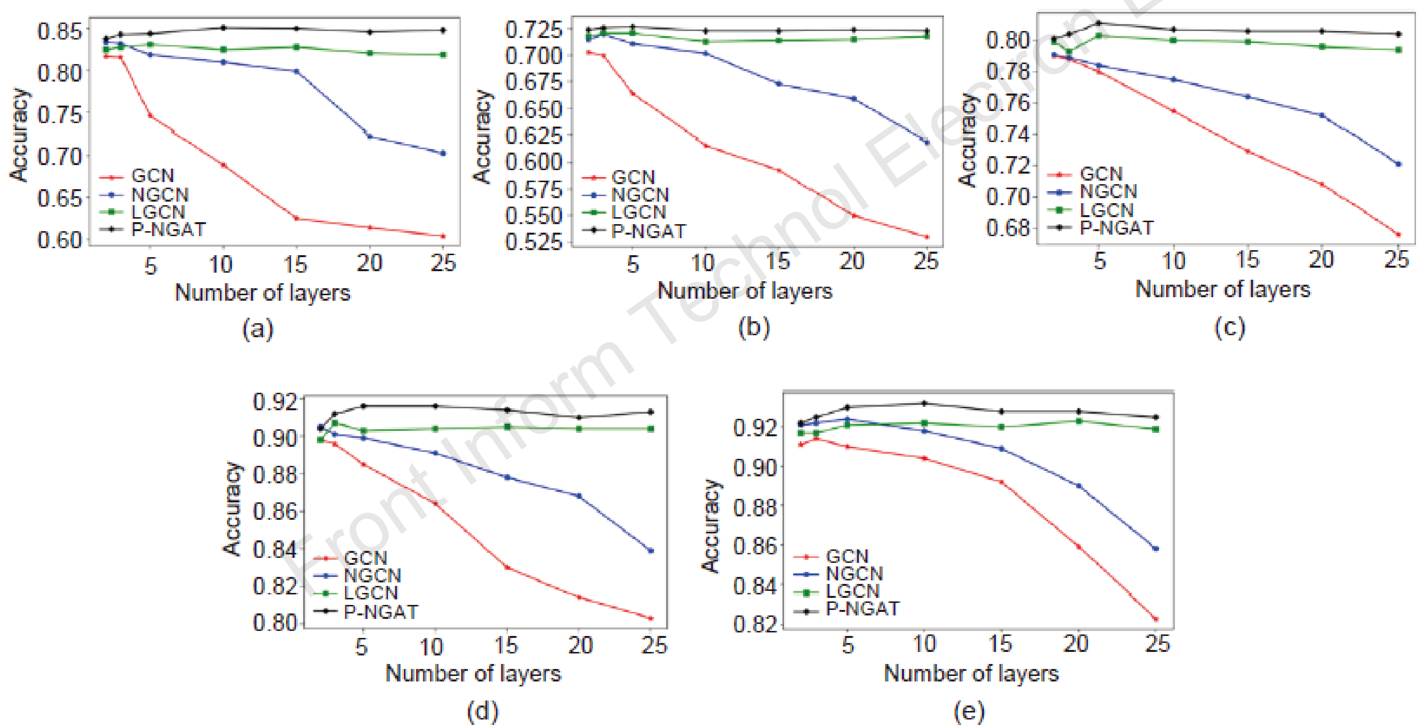


Fig. 5 Influence of model depth (number of layers) on node classification performance. The experimental results of P-NGAT (using self-attention layer-wise aggregation) are compared to those of a standard GCN model, an NGAT with only node-wise attention (NGCN), and an NGAT with only layer-wise attention (LGCN) on datasets Cora (a), Citeseer (b), PubMed (c), Academic CS (d), and Academic Physics (e). Markers denote test accuracy averaged over 10 runs

Major results (Cont'd)

Visualization

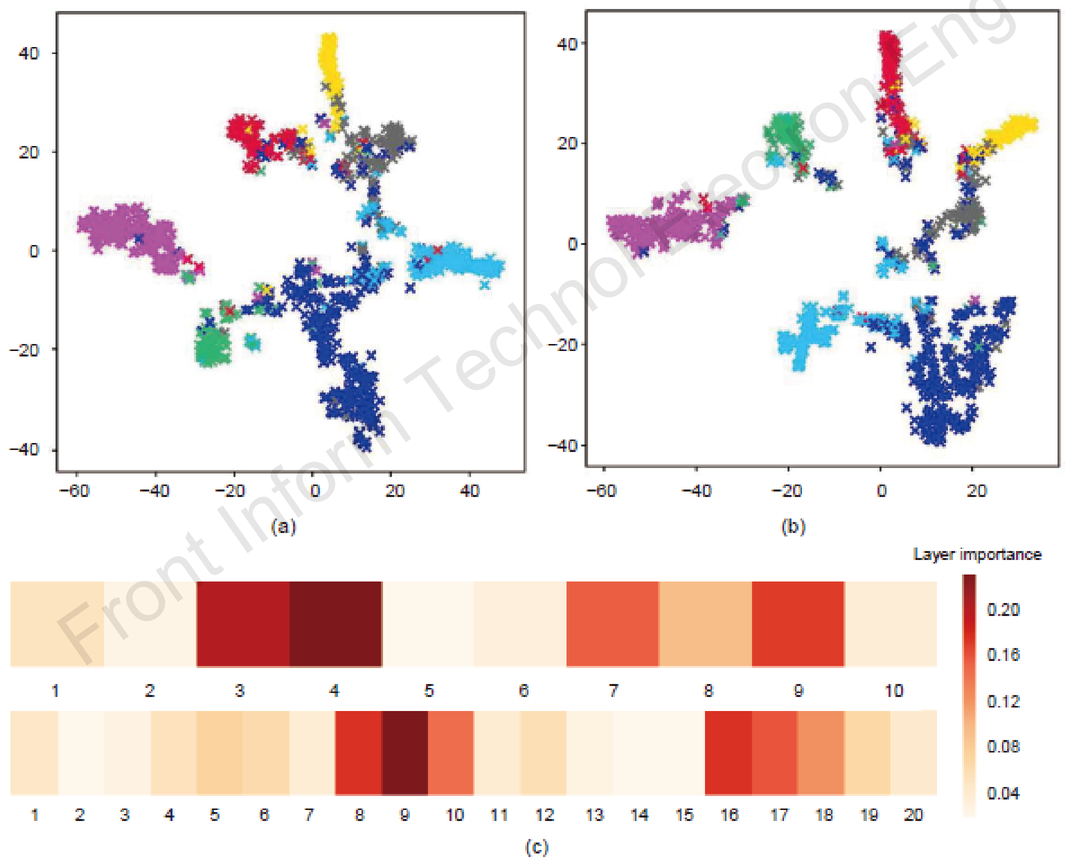


Fig. 6 Insight into NGAT performance on the Cora dataset. (a) and (b) are the t-SNE visualization of the classification of 1000 test nodes. Different colors mean different classes. Panel (a) is a 10-layer P-NGAT and panel (b) shows a 20-layer model. (c) depicts the layer importance of each hidden layer for representation. The numbers below the panel represent the index of the layer, and darker color means greater importance score

Conclusions

1. NGAT improves both breadth and depth exploration, and allows the stacking of more layers within a feasible scope.

2. Our work can be standardized as an extension of JK-Net. We have adopted a simple but powerful architecture for both neighborhood and layer aggregations.

3. On many node classification benchmarks, NGAT achieves state-of-the-art performance compared to some novel GNNs.

References

[1] Li QM, Han ZC, Wu XM, 2018. Deeper insights into graph convolutional networks for semi-supervised learning. Proc 32nd AAAI Conf on Artificial Intelligence, p.3538-3545.

[2] Niepert M, Ahmed M, Kutzkov K, 2016. Learning convolutional neural networks for graphs. Proc 33rd Int Conf on Machine Learning, p.2014-2023.

[3] Yang ZL, Cohen W, Salakhudinov R, 2016. Revisiting semi-supervised learning with graph embeddings. Proc 33rd Int Conf on Machine Learning, p.40-48.

[4] Zou DF, Hu ZN, Wang YW, et al., 2019. Layer-dependent importance sampling for training deep and large graph convolutional networks. Proc 33rd Int Conf on Neural Information Processing Systems, p.11247-11256.

[5] Knyazev B, Taylor GW, Amer MR, 2019. Understanding attention and generalization in graph neural networks. Proc 33rd Conf on Neural Information Processing Systems, p.4204-4214.