Measuring and Relieving the Over-smoothing Problem for Graph Neural Networks from the Topological View

Deli Chen, Yankai Lin, Li Wei, Peng Li, Jie Zhou, Xu Sun

Accepted by AAAI 2020

The Over-smoothing Problem

Background

➢ 问题描述

- ✓ smoothing 指图上节点表示变得相似
- ✓ over-smoothing 指图上节点表示变得过于相似，以致于无法区分不同类别节点

➢ 现有工作

- ✓ 很多工作提到在叠加多层图网络时出现了 Over-smoothing问题
- ✓ Yang et al. (2019) 提出了节点属性的smoothing直接导致了节点被误分类
- ✓ Li, Han, and Wu (2018) 提出并证明了GCN本质是一种Laplace平滑，在高层图网络中难以避免
- ✓ Li et al.,（2019）和 Rong et al（2019）等工作则尝试克服smoothing问题来叠加多层图网络

Motivation & Proposal

- **定量研究**
  - ✓ 我们设计了两个定量指标，分别用于定量的衡量图的smoothing以及over-smoothing程度
  - ✓ 基于图的拓扑结构自动计算这两个指标，而不需要依赖节点的gold label

- **系统研究**
  - ✓ 我们复现了 10 种不同的，具有代表性的GNN模型并展开实验
  - ✓ 我们在论文引用，商品购买，合作作者三个领域选取了 7 个常用的图数据集开展实验

- **解决方案**
  - 我们通过定量和系统的分析，得到了对于over-smoothing深入理解
  - 基于我们的分析结果，我们提出了两种通用的，模型无关的缓解over-smoothing的方法

MAD: Metric for Graph Smoothing

- Mean Average Distance

$$D_{ij} = 1 - \frac{\boldsymbol{H}_{i,:} \cdot \boldsymbol{H}_{j,:}}{|\boldsymbol{H}_{i,:}| \cdot |\boldsymbol{H}_{j,:}|} \tag{1}$$

$$\boldsymbol{D}^{tgt} = \boldsymbol{D} \circ \boldsymbol{M}^{tgt}, \tag{2}$$

$$\bar{\boldsymbol{D}}_i^{tgt} = \frac{\sum_{j=0}^{n} \boldsymbol{D}_{ij}^{tgt}}{\sum_{j=0}^{n} \mathbb{1}\left(\boldsymbol{D}_{ij}^{tgt}\right)}, \tag{3}$$

$$\mathrm{MAD}^{\mathrm{tgt}} = \frac{\sum_{i=0}^{n} \bar{\boldsymbol{D}}_i^{tgt}}{\sum_{i=0}^{n} \mathbb{1}\left(\bar{\boldsymbol{D}}_i^{tgt}\right)}. \tag{4}$$

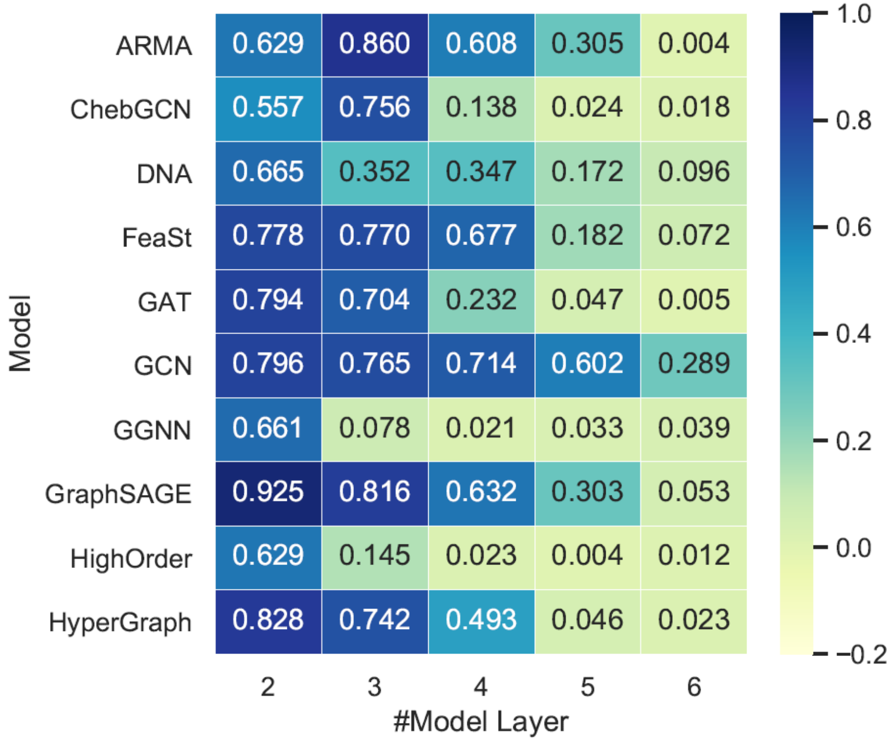


Figure 2: The MAD values of various GNNs with different layers on the *CORA* dataset. Darker color means larger MAD value. We can find that the smoothness of graph representation rises as the model layer increases.

MADGap: Metric for Graph Over-Smoothing

Metric

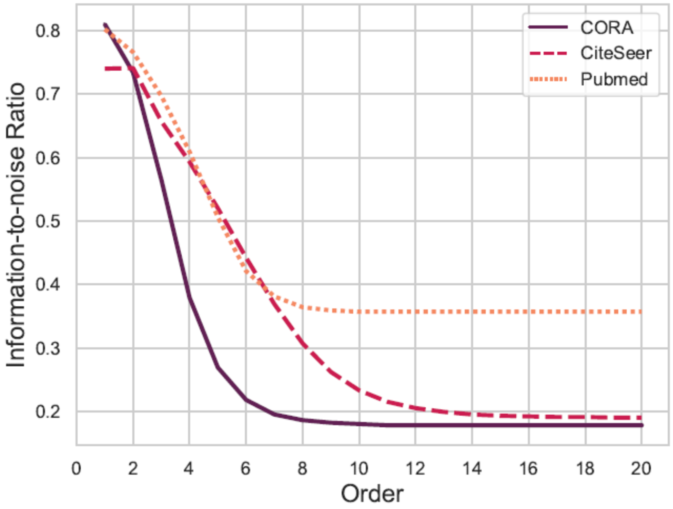


Figure 3: The information-to-noise ratio at different neighbor orders (accumulated) for the *CORA/CiteSeer/PubMed* datasets. We can find that the information-to-noise ratio declines as the orders increases in all these three datasets.

$$\mathrm{MADGap} = \mathrm{MAD}^{\mathrm{rmt}} - \mathrm{MAD}^{\mathrm{neb}}$$

Model	CORA	CiteSeer	PubMed
GCN	0.986**	0.948**	0.971**
ChebGCN	0.945**	0.984**	0.969**
HyperGraph	0.990**	0.965**	0.932**
FeaSt	0.993**	0.986**	0.906*
GraphSAGE	0.965**	0.995*	0.883
GAT	0.960**	0.998**	0.965**
ARMA	0.909*	0.780	0.787
HighOrder	0.986**	0.800	0.999**
DNA	0.945**	0.884*	0.887*
GGNN	0.940*	0.900*	0.998**

Table 2: The Pearson coefficient between accuracy and MADGap for various models on *CORA/CiteSeer/PubMed* datasets. Pearson coefficient is calculated based on the results of models with different layers (1-6). * means statistically significant with $p < 0.05$ and ** means $p < 0.01$.

Information2noise Ratio Decides Over-smoothing

Conclusion

➢ **对于over-smoothing的观点**

✓ 图网络通过邻接节点接触来传递信息

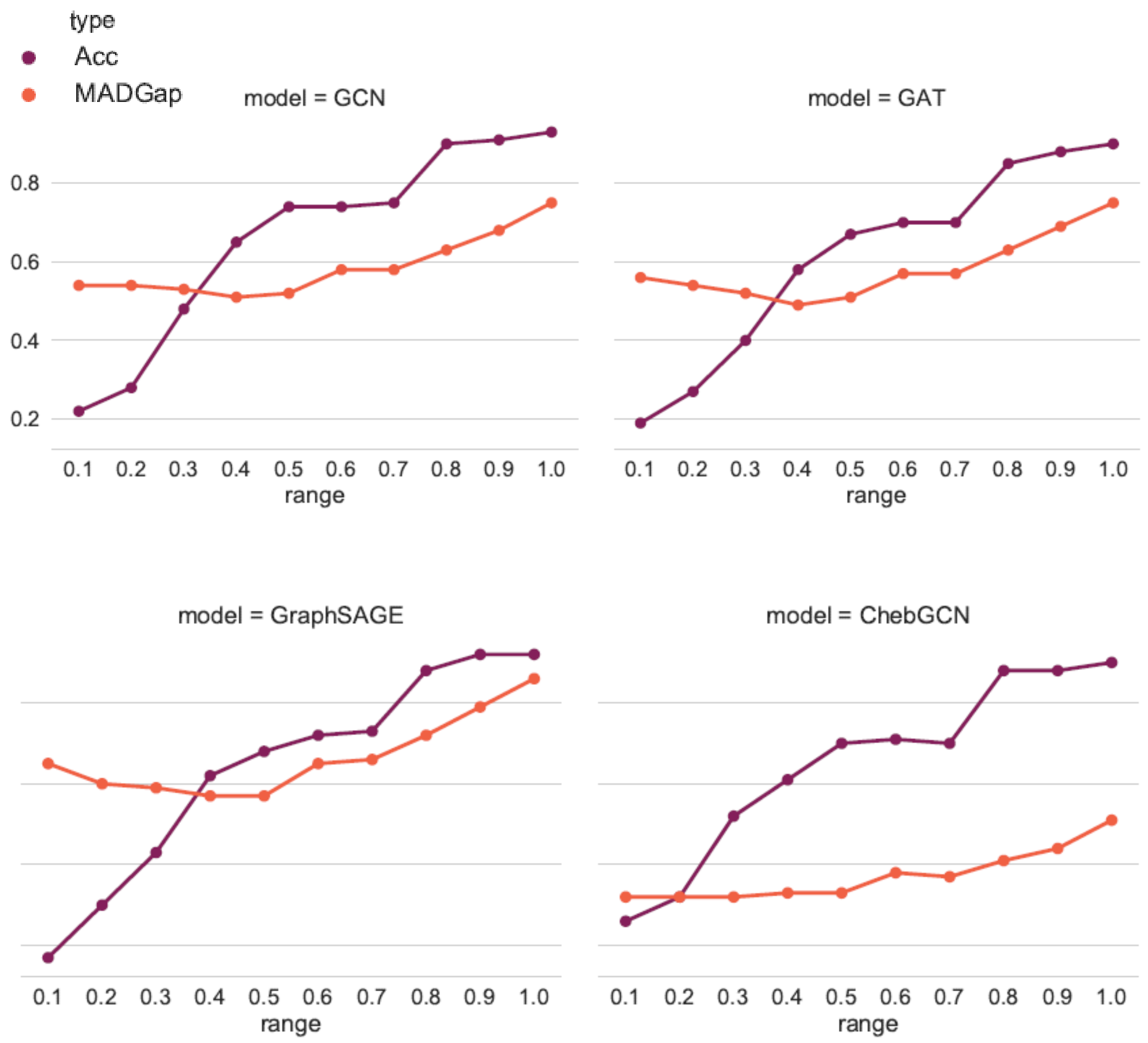
✓ 而节点间信息传递会带来节点表示的smoothing，这是GNN的本质特征

✓ 而smoothing可以分为两种

- Helpful smoothing : 同类节点接触
- Harmful smoothing：不同类节点接触

➢ **Information-to-noise Ratio**

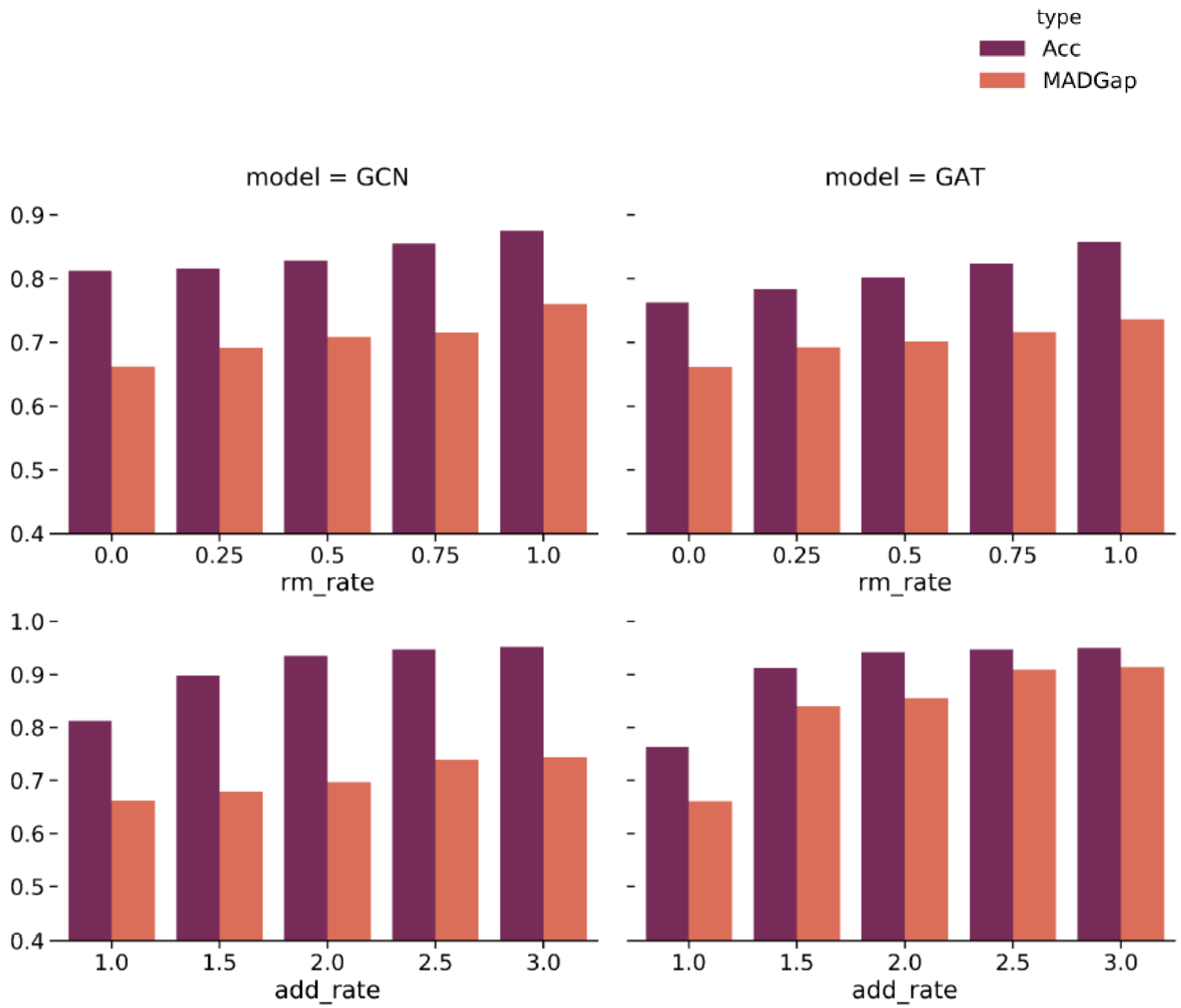
✓ 节点在N阶内接触到的同类节点的比例



Graph Topology Affects Information2noise Ratio

Conclusion

- **拓扑结构影响information2noise ratio**
  - ✓理想的拓扑结构是图中有尽可能多的类内边，从而只使得同类节点smoothing
  - ✓而当类间边很多的时候，就会降低节点的information2noise ratio，从而导致节点表示的over-smoothing



- **改进拓扑结构来提升相关任务**
  - ✓减少类间边
  - ✓增加类内边

Solutions: MADReg and AdaEdge

➢ MADGap as Regularizer

$$\mathcal{L} = \sum -l \log p(\hat{l} \,|\, \boldsymbol{X}, \boldsymbol{A}, \Theta) - \lambda \mathrm{MADGap}$$

➢ Adaptive Edge Optimization

- ✓ 使用self-training 的思路
- ✓ 基于预测结果，进行拓扑结构调整，然后在调整后的拓扑结构上进行新一轮的训练
- ✓ 过滤高置信度的节点来增删边
- ✓ 设计启发式规则保证增删边正确率

Relieving Over-smoothing in High-layer GNN

Result

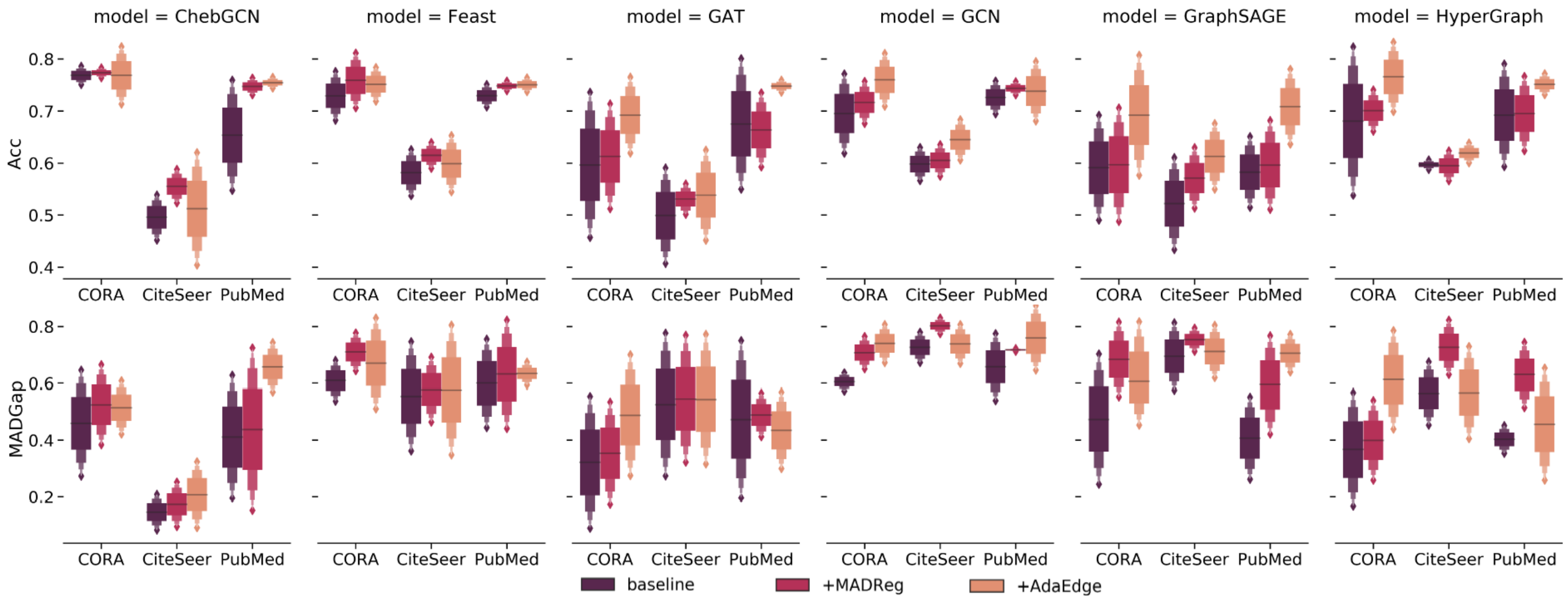


Figure 6: MADReg and AdaEdge results on the *CORA/CiteSeer/PubMed* datasets. The number of GNN layers is 4, where the over-smoothing issue is severe. The box plot shows the mean value and the standard deviation of the prediction accuracy and the MADGap values of 50 turns results (5 dataset splitting methods and 10 random seeds for each splitting following Shchur et al. (2018) and Sun, Koniusz, and Wang (2019). More details can be found in Appendix A). And we can find that the two proposed methods can effectively relieve the over-smoothing issue and improve model performance in most cases.

Improving GNN Performance by AdaEdge

Result

Acc(%)	CORA		CiteSeer		PubMed		Amazon Photo		Amazon Comp.		Coauthor CS		Coauthor Phy.	
Model	baseline	+AE	baseline	+AE	baseline	+AE	baseline	+AE	baseline	+AE	baseline	+AE	baseline	+AE
GCN	$81.2_{\pm0.8}$	$82.3^{**}_{\pm0.8}$	$69.3_{\pm0.7}$	$69.7^{**}_{\pm0.9}$	$76.3_{\pm0.5}$	$77.4^{**}_{\pm0.5}$	$90.6_{\pm0.7}$	$91.5^{**}_{\pm0.5}$	$81.7_{\pm0.7}$	$82.4^{**}_{\pm1.1}$	$89.8_{\pm0.3}$	$90.3^{**}_{\pm0.4}$	$92.8_{\pm1.6}$	$93.0^{**}_{\pm1.1}$
ChebGCN	$78.6_{\pm0.6}$	$80.1^{**}_{\pm0.5}$	$67.4_{\pm1.0}$	$67.8^{*}_{\pm1.2}$	$76.7_{\pm0.1}$	$77.5^{**}_{\pm0.6}$	$89.6_{\pm1.6}$	$89.4^{*}_{\pm1.2}$	$80.8_{\pm2.4}$	$81.3^{**}_{\pm1.1}$	$90.5_{\pm0.4}$	$90.7^{*}_{\pm0.3}$	\	\
HyperGraph	$80.5_{\pm0.6}$	$81.4^{**}_{\pm0.8}$	$67.9_{\pm0.5}$	$68.5^{**}_{\pm0.5}$	$77.4_{\pm0.2}$	$77.3_{\pm0.7}$	$87.5_{\pm0.7}$	$88.6^{**}_{\pm0.3}$	$58.7_{\pm22.1}$	$61.4^{**}_{\pm26.8}$	$86.9_{\pm0.5}$	$87.3^{**}_{\pm0.4}$	$91.9_{\pm2.0}$	$92.2^{**}_{\pm1.4}$
FeaSt	$80.4_{\pm0.7}$	$81.6^{**}_{\pm0.7}$	$69.3_{\pm1.1}$	$69.4^{*}_{\pm1.0}$	$76.6_{\pm0.6}$	$77.2^{*}_{\pm0.4}$	$90.5_{\pm0.6}$	$90.8^{**}_{\pm0.6}$	$80.8_{\pm1.3}$	$81.7^{**}_{\pm0.9}$	$88.4_{\pm0.2}$	$88.9^{**}_{\pm0.2}$	\	\
GraphSAGE	$78.5_{\pm1.7}$	$80.2^{**}_{\pm1.2}$	$68.4_{\pm0.9}$	$69.4^{**}_{\pm0.8}$	$75.2_{\pm1.1}$	$77.2^{**}_{\pm0.8}$	$90.1_{\pm1.4}$	$90.6^{**}_{\pm0.5}$	$80.2_{\pm1.0}$	$81.1^{**}_{\pm1.0}$	$90.1_{\pm0.4}$	$90.3^{**}_{\pm0.4}$	$93.0_{\pm0.4}$	$92.7_{\pm0.2}$
GAT	$76.3_{\pm3.1}$	$77.9^{**}_{\pm2.0}$	$68.9_{\pm0.6}$	$69.1^{*}_{\pm0.8}$	$75.9_{\pm0.5}$	$76.6^{**}_{\pm0.2}$	$89.7_{\pm1.7}$	$90.8^{**}_{\pm0.9}$	$81.4_{\pm1.5}$	$81.1^{*}_{\pm1.6}$	$85.5_{\pm1.9}$	$86.6^{**}_{\pm1.6}$	$91.1_{\pm1.0}$	$91.4^{*}_{\pm1.0}$
ARMA	$74.9_{\pm10.6}$	$76.4^{**}_{\pm5.6}$	$65.3_{\pm4.1}$	$66.1^{**}_{\pm4.3}$	$68.5_{\pm11.4}$	$68.9^{*}_{\pm12.2}$	$86.4_{\pm3.0}$	$87.0^{**}_{\pm1.9}$	$63.8_{\pm18.9}$	$71.7^{**}_{\pm8.1}$	$90.6_{\pm1.1}$	$90.9^{**}_{\pm0.6}$	$92.2_{\pm1.8}$	$92.6^{*}_{\pm1.0}$
HighOrder	$76.6_{\pm1.2}$	$72.5^{**}_{\pm4.1}$	$64.2_{\pm1.0}$	$63.3^{**}_{\pm1.0}$	$75.0_{\pm2.6}$	$76.9^{**}_{\pm1.3}$	$26.1_{\pm12.4}$	$30.3^{**}_{\pm10.2}$	$26.3_{\pm12.7}$	$23.9^{*}_{\pm13.4}$	$84.2_{\pm1.0}$	$85.6^{**}_{\pm0.7}$	$90.8_{\pm0.8}$	$90.9_{\pm0.6}$
DNA	$58.2_{\pm14.4}$	$60.1^{*}_{\pm10.8}$	$60.9_{\pm2.7}$	$61.3^{**}_{\pm2.2}$	$65.8_{\pm7.8}$	$66.8^{*}_{\pm9.6}$	$89.1_{\pm1.3}$	$89.8^{**}_{\pm0.6}$	$78.2_{\pm2.9}$	$79.8^{**}_{\pm2.0}$	$88.2_{\pm0.9}$	$90.0^{**}_{\pm0.6}$	$93.0_{\pm0.5}$	$93.3^{*}_{\pm0.4}$
GGNN	$47.3_{\pm6.1}$	$44.7^{**}_{\pm3.5}$	$55.5_{\pm2.8}$	$47.9^{**}_{\pm3.4}$	$66.1_{\pm4.4}$	$69.5^{**}_{\pm1.2}$	$74.1_{\pm12.3}$	$80.6^{**}_{\pm7.2}$	$42.4_{\pm26.7}$	$61.5^{**}_{\pm20.8}$	$86.6_{\pm1.4}$	$88.2^{**}_{\pm0.8}$	$91.2_{\pm1.2}$	$91.6^{**}_{\pm0.7}$

Table 3: Controlled experiments of AdaEdge (+AE) on all the 7 datasets. We show the mean value, the standard deviation and the t-test significance of 50 turns results. * means statistically significance with $p < 0.05$ and ** means $p < 0.01$. Darker color means larger improvement. The missing results are due to the huge consumption of GPU memory of large graphs.

Conclusion & Contribution

- 我们在众多图网络和图数据集中，开展了对于over-smoothing问题的系统而定量的研究。我们指出并证明了影响over-smoothing的关键在于节点的information2noise ratio，而图的拓扑结构会影响information2noise ratio.
- 我们从图的拓扑关系出发，设计了两个不依赖于gold lable的指标MAD和MADGap，分别用来衡量图网络学习到的节点表示的smoothing以及over-smoothing程度。统计检验结果表明，我们的MADGap和模型的准确率呈现显著的强正相关关系，说明我们的指标能够很好的衡量over-smoothing程度。
- 从我们的结论出发，我们设计了两种模型无关的缓解over-smoothing问题的方法：MADReg以及AdaEdge. 在众多图网络和图数据集的实验中，我们的方法能够显著缓解模型的over-smoothing程度，也能明显提升模型效果。

感谢聆听，欢迎交流！

论文链接

https://arxiv.org/abs/1909.03211