



TEXAS A&M UNIVERSITY

Engineering

When Does Self-Supervision Help Graph Convolutional Networks?

Yuning You*, Tianlong Chen*, Zhangyang Wang, Yang Shen

Texas A&M University

* Equal Contribution

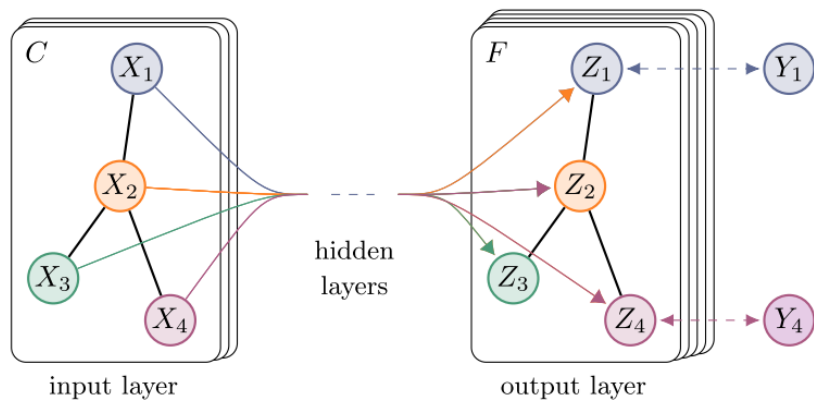


目录

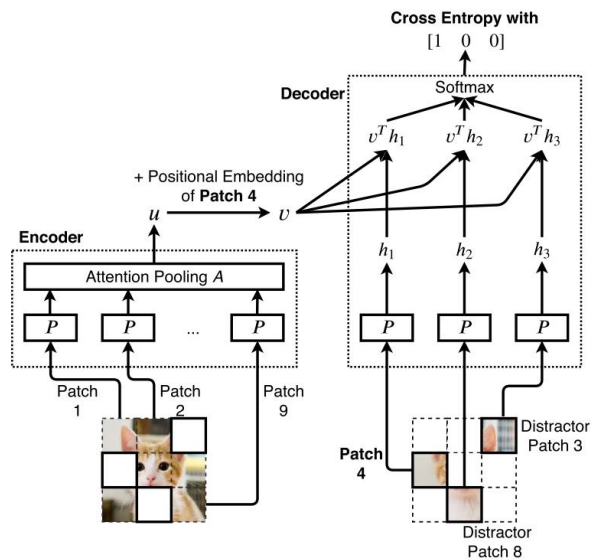
- 简介
- 研究问题
- 主要贡献
- 贡献1. 如何结合自监督学习和图卷积神经网络？
- 贡献2. 如何设计自监督学习任务以提升泛化性能？
- 贡献3. 自监督学习是否能提升鲁棒性？
- 结论

简介

- 图卷积神经网络
(GCNs, ICLR'17):



- 图像中的自监督学习
(e.g. Selfie, preprint'19):





目录

- 简介
- 研究问题
- 主要贡献
- 贡献1. 如何结合自监督学习和图卷积神经网络？
- 贡献2. 如何设计自监督学习任务以提升泛化性能？
- 贡献3. 自监督学习是否能提升鲁棒性？
- 结论

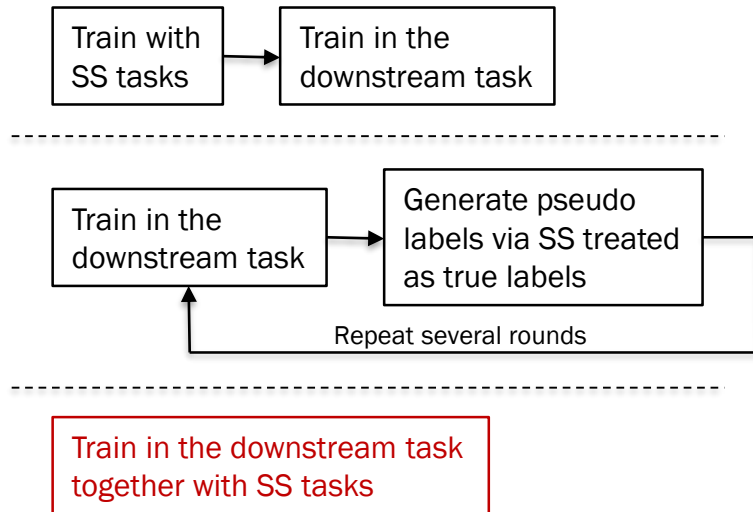
- 半监督学习在基于图数据结构的场景中有着重要的应用，而在这个场景下我们可利用大量的无标签数据；
- 自监督学习在少样本（半监督）的场景下是一种具有保证性的技术，其通过利用无标签数据来提升性能，近些年来其在图像上受到人们广泛的关注；
- 然而，自监督学习在图数据结构上的探索比较少，仅在少量的工作中被讨论 (M3S, AAAI'19).



目录

- 简介
- 研究问题
- **主要贡献**
- 贡献1. 如何结合自监督学习和图卷积神经网络？
- 贡献2. 如何设计自监督学习任务以提升泛化性能？
- 贡献3. 自监督学习是否能提升鲁棒性？
- 结论

- 我们做了系统的实验来研究自监督学习在图数据结构上的表现：
 - 1. 如何结合自监督学习和图卷积神经网络？
 - 预训练 & 微调；
 - 自训练 (M3S, AAAI'19);
 - 多任务学习.



- 我们做了系统的实验来研究自监督学习在图数据结构上的表现：
 - 2. 如何设计自监督学习任务以提升泛化性能？
 - 我们探索设计了三种自监督学习任务：节点聚类，子图分割，以及图特征补全；
 - 我们的观察是：不同的自监督学习任务对不同的场景（数据集，网络构架）的泛化性能的帮助不同。

- 我们做了系统的实验来研究自监督学习在图数据结构上的表现：
 - 3. 自监督学习是否能提升鲁棒性？
 - 我们首次将自监督学习引入对抗训练；
 - 我们的实验结果显示：在不需要额外的更大的模型与更多的数据的情况下，自监督学习可提升模型的鲁棒性。

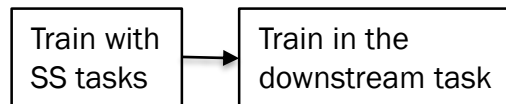


目录

- 简介
- 研究问题
- 主要贡献
- **贡献1. 如何结合自监督学习和图卷积神经网络？**
- **贡献2. 如何设计自监督学习任务以提升泛化性能？**
- **贡献3. 自监督学习是否能提升鲁棒性？**
- 结论

贡献1. 如何结合自监督学习和图卷积神经网络？

- 预训练 & 微调：



- 在我们的大数据集PubMed上提升不大；
- 猜想：这个提升表现是来自于损失函数从自监督任务转换到监督任务在较浅的图卷积神经网络导致的；
- 较浅的图神经网络在从自监督转换到监督这一过程很容易被“过度改写”。

Table 1: Comparing performances of GCN through pretraining & finetuning (P&F) and multi-task learning (MTL) with graph partitioning (see Section 3.3) on the PubMed dataset. Reported numbers correspond to classification accuracy in percent.

Pipeline	GCN	P&F	MTL
Accuracy	79.10 \pm 0.21	79.19 \pm 0.21	80.00 \pm 0.74

贡献1. 如何结合自监督学习和图卷积神经网络？

- 自训练 (M3S, AAAI'19):
 - 当标签率逐渐上升时，性能提升遭遇“**饱和**”；
 - 猜想：伪标签的分配是基于他们和真实标签的距离决定的，分配到的伪标签是最**邻近**的真实标签；
 - 在生成伪标签上受限于真实标签，与多任务学习相比**不太容易泛化**。

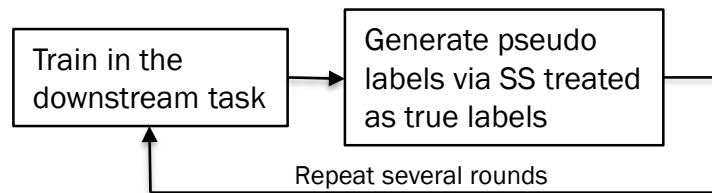


Table 2: Experiments for GCN through M3S. Gray numbers are from (Sun et al., 2019).

Label Rate	0.03%	0.1%	0.3% (Conventional dataset split)
GCN	51.1	67.5	79.10 ± 0.21
M3S	59.2	70.6	79.28 ± 0.30

贡献1. 如何结合自监督学习和图卷积神经网络？

- 多任务学习：

Train in the downstream task
together with SS tasks

- 在经验上（实验结果）**超越**另外两种结合方案；
- 我们将在多任务自监督学习中的自监督任务看做是网络训练过程中的**正则项**；
- 不同于传统的正则项（例如拉普拉斯平滑），它是**数据驱动**的。

Table 1: Comparing performances of GCN through pretraining & finetuning (P&F) and multi-task learning (MTL) with graph partitioning (see Section 3.3) on the PubMed dataset. Reported numbers correspond to classification accuracy in percent.

Pipeline	GCN	P&F	MTL
Accuracy	79.10 \pm 0.21	79.19 \pm 0.21	80.00 \pm 0.74



目录

- 简介
- 研究问题
- 主要贡献
- 贡献1. 如何结合自监督学习和图卷积神经网络？
- 贡献2. 如何设计自监督学习任务以提升泛化性能？
- 贡献3. 自监督学习是否能提升鲁棒性？
- 结论

贡献2. 如何设计自监督学习任务以提升泛化性能？

Table 3: Overview of three self-supervised tasks.

Task	Relied Feature	Primary Assumption	Type
Clustering	Nodes	Feature Similarity	Classification
Partitioning	Edges	Connection Density	Classification
Completion	Nodes & Edges	Context based Representation	Regression

- 我们探索设计了三种自监督学习任务：

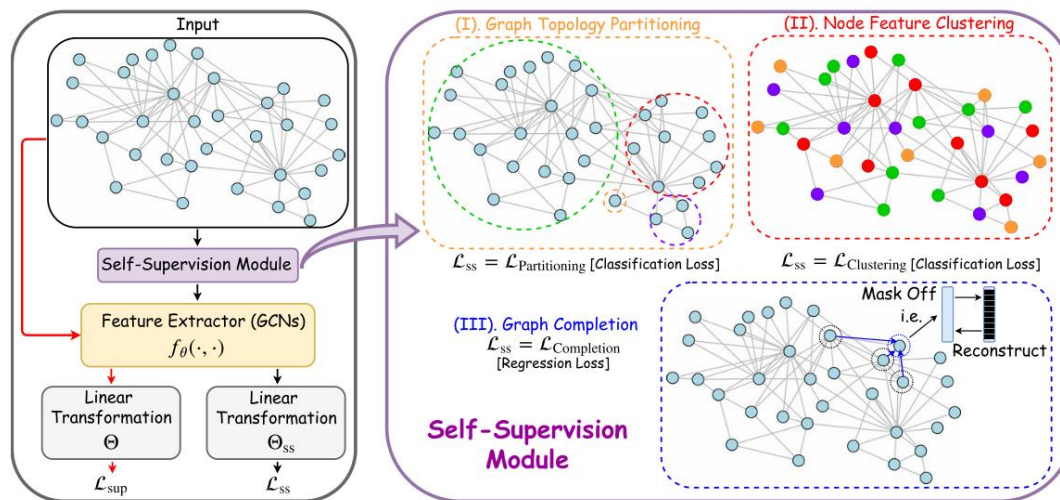


Figure 1: The overall framework for self-supervision on GCN through *multi-task learning*. The target task and auxiliary self-supervised tasks share the same feature extractor $f_{\theta}(\cdot, \cdot)$ with their individual linear transformation parameters Θ, Θ_{ss} .

贡献2. 如何设计自监督学习任务以提升泛化性能？

- 我们的观察是：不同的自监督学习任务对不同的场景（数据集，网络构架）的泛化性能的帮助不同：
 - 节点聚类 (Clu) 假设节点特征的相似性预示着下游标签的相似性；
 - 因此其表现在大数据集+低节点特征维度的情况下会受到挑战 (例如 PubMed).

Table 6: Experiments on SOTAs (GCN, GAT, GIN, GMNN, and GraphMix) with multi-task self-supervision. Red numbers indicate the best two performances for each SOTA.

Datasets	Cora	Citeseer	PubMed
GCN	81.00 ± 0.67	70.85 ± 0.70	79.10 ± 0.21
GCN+Clu	81.57 ± 0.59	70.73 ± 0.84	78.79 ± 0.36
GCN+Par	81.83 ± 0.65	71.34 ± 0.69	80.00 ± 0.74
GCN+Comp	81.03 ± 0.68	71.66 ± 0.48	79.14 ± 0.28
GAT	77.66 ± 1.08	68.90 ± 1.07	78.05 ± 0.46
GAT+Clu	79.40 ± 0.73	69.88 ± 1.13	77.80 ± 0.28
GAT+Par	80.11 ± 0.84	69.76 ± 0.81	80.11 ± 0.34
GAT+Comp	80.47 ± 1.22	70.62 ± 1.26	77.10 ± 0.67
GIN	77.27 ± 0.52	68.83 ± 0.40	77.38 ± 0.59
GIN+Clu	78.43 ± 0.80	68.86 ± 0.91	76.71 ± 0.36
GIN+Par	81.83 ± 0.58	71.50 ± 0.44	80.28 ± 1.34
GIN+Comp	76.62 ± 1.17	68.71 ± 1.01	78.70 ± 0.69
GMNN	83.28 ± 0.81	72.83 ± 0.72	81.34 ± 0.59
GMNN+Clu	83.49 ± 0.65	73.13 ± 0.72	79.45 ± 0.76
GMNN+Par	83.51 ± 0.50	73.62 ± 0.65	80.92 ± 0.77
GMNN+Comp	83.31 ± 0.81	72.93 ± 0.79	81.33 ± 0.59
GraphMix	83.91 ± 0.63	74.33 ± 0.65	80.68 ± 0.57
GraphMix+Clu	83.87 ± 0.56	75.16 ± 0.52	79.99 ± 0.82
GraphMix+Par	84.04 ± 0.57	74.93 ± 0.43	81.36 ± 0.33
GraphMix+Comp	83.76 ± 0.64	74.43 ± 0.72	80.82 ± 0.54

贡献2. 如何设计自监督学习任务以提升泛化性能？

- 我们的观察是：不同的自监督学习任务对不同的场景（数据集，网络构架）的泛化性能的帮助不同：

- 子图分割 (Par) 假设在图拓扑结构上的连接预示着下游标签的相似性；

- 这个假设在所研究的三个数据集中（引用网络）是大概率成立的。

Table 6: Experiments on SOTAs (GCN, GAT, GIN, GMNN, and GraphMix) with multi-task self-supervision. Red numbers indicate the best two performances for each SOTA.

Datasets	Cora	Citeseer	PubMed
GCN	81.00 ± 0.67	70.85 ± 0.70	79.10 ± 0.21
GCN+Clu	81.57 ± 0.59	70.73 ± 0.84	78.79 ± 0.36
GCN+Par	81.83 ± 0.65	71.34 ± 0.69	80.00 ± 0.74
GCN+Comp	81.03 ± 0.68	71.66 ± 0.48	79.14 ± 0.28
GAT	77.66 ± 1.08	68.90 ± 1.07	78.05 ± 0.46
GAT+Clu	79.40 ± 0.73	69.88 ± 1.13	77.80 ± 0.28
GAT+Par	80.11 ± 0.84	69.76 ± 0.81	80.11 ± 0.34
GAT+Comp	80.47 ± 1.22	70.62 ± 1.26	77.10 ± 0.67
GIN	77.27 ± 0.52	68.83 ± 0.40	77.38 ± 0.59
GIN+Clu	78.43 ± 0.80	68.86 ± 0.91	76.71 ± 0.36
GIN+Par	81.83 ± 0.58	71.50 ± 0.44	80.28 ± 1.34
GIN+Comp	76.62 ± 1.17	68.71 ± 1.01	78.70 ± 0.69
GMNN	83.28 ± 0.81	72.83 ± 0.72	81.34 ± 0.59
GMNN+Clu	83.49 ± 0.65	73.13 ± 0.72	79.45 ± 0.76
GMNN+Par	83.51 ± 0.50	73.62 ± 0.65	80.92 ± 0.77
GMNN+Comp	83.31 ± 0.81	72.93 ± 0.79	81.33 ± 0.59
GraphMix	83.91 ± 0.63	74.33 ± 0.65	80.68 ± 0.57
GraphMix+Clu	83.87 ± 0.56	75.16 ± 0.52	79.99 ± 0.82
GraphMix+Par	84.04 ± 0.57	74.93 ± 0.43	81.36 ± 0.33
GraphMix+Comp	83.76 ± 0.64	74.43 ± 0.72	80.82 ± 0.54

贡献2. 如何设计自监督学习任务以提升泛化性能？

- 我们的观察是：不同的自监督学习任务对不同的场景（数据集，网络构架）的泛化性能的帮助不同：
 - 图特征补全 (Comp) 假设节点特征在小范围邻域内的平滑性；
 - 其提升了邻域范围较小的数据集 (例如平均度最小的 Citeseer).

Table 6: Experiments on SOTAs (GCN, GAT, GIN, GMNN, and GraphMix) with multi-task self-supervision. Red numbers indicate the best two performances for each SOTA.

Datasets	Cora	Citeseer	PubMed
GCN	81.00 \pm 0.67	70.85 \pm 0.70	79.10 \pm 0.21
GCN+Clu	81.57 \pm 0.59	70.73 \pm 0.84	78.79 \pm 0.36
GCN+Par	81.83 \pm 0.65	71.34 \pm 0.69	80.00 \pm 0.74
GCN+Comp	81.03 \pm 0.68	71.66 \pm 0.48	79.14 \pm 0.28
GAT	77.66 \pm 1.08	68.90 \pm 1.07	78.05 \pm 0.46
GAT+Clu	79.40 \pm 0.73	69.88 \pm 1.13	77.80 \pm 0.28
GAT+Par	80.11 \pm 0.84	69.76 \pm 0.81	80.11 \pm 0.34
GAT+Comp	80.47 \pm 1.22	70.62 \pm 1.26	77.10 \pm 0.67
GIN	77.27 \pm 0.52	68.83 \pm 0.40	77.38 \pm 0.59
GIN+Clu	78.43 \pm 0.80	68.86 \pm 0.91	76.71 \pm 0.36
GIN+Par	81.83 \pm 0.58	71.50 \pm 0.44	80.28 \pm 1.34
GIN+Comp	76.62 \pm 1.17	68.71 \pm 1.01	78.70 \pm 0.69
GMNN	83.28 \pm 0.81	72.83 \pm 0.72	81.34 \pm 0.59
GMNN+Clu	83.49 \pm 0.65	73.13 \pm 0.72	79.45 \pm 0.76
GMNN+Par	83.51 \pm 0.50	73.62 \pm 0.65	80.92 \pm 0.77
GMNN+Comp	83.31 \pm 0.81	72.93 \pm 0.79	81.33 \pm 0.59
GraphMix	83.91 \pm 0.63	74.33 \pm 0.65	80.68 \pm 0.57
GraphMix+Clu	83.87 \pm 0.56	75.16 \pm 0.52	79.99 \pm 0.82
GraphMix+Par	84.04 \pm 0.57	74.93 \pm 0.43	81.36 \pm 0.33
GraphMix+Comp	83.76 \pm 0.64	74.43 \pm 0.72	80.82 \pm 0.54

贡献2. 如何设计自监督学习任务以提升泛化性能？

- 我们的观察是：不同的自监督学习任务对不同的场景（数据集，网络构架）的泛化性能的帮助不同：

— 网络结构也会影响自监督学习的表现；

— 嵌入较少的先验知识的网络结构会从自监督学习中得到更大的提升.

Table 6: Experiments on SOTAs (GCN, GAT, GIN, GMNN, and GraphMix) with multi-task self-supervision. Red numbers indicate the best two performances for each SOTA.

Datasets	Cora	Citeseer	PubMed
GCN	81.00 ± 0.67	70.85 ± 0.70	79.10 ± 0.21
GCN+Clu	81.57 ± 0.59	70.73 ± 0.84	78.79 ± 0.36
GCN+Par	81.83 ± 0.65	71.34 ± 0.69	80.00 ± 0.74
GCN+Comp	81.03 ± 0.68	71.66 ± 0.48	79.14 ± 0.28
GAT	77.66 ± 1.08	68.90 ± 1.07	78.05 ± 0.46
GAT+Clu	79.40 ± 0.73	69.88 ± 1.13	77.80 ± 0.28
GAT+Par	80.11 ± 0.84	69.76 ± 0.81	80.11 ± 0.34
GAT+Comp	80.47 ± 1.22	70.62 ± 1.26	77.10 ± 0.67
GIN	77.27 ± 0.52	68.83 ± 0.40	77.38 ± 0.59
GIN+Clu	78.43 ± 0.80	68.86 ± 0.91	76.71 ± 0.36
GIN+Par	81.83 ± 0.58	71.50 ± 0.44	80.28 ± 1.34
GIN+Comp	76.62 ± 1.17	68.71 ± 1.01	78.70 ± 0.69
GMNN	83.28 ± 0.81	72.83 ± 0.72	81.34 ± 0.59
GMNN+Clu	83.49 ± 0.65	73.13 ± 0.72	79.45 ± 0.76
GMNN+Par	83.51 ± 0.50	73.62 ± 0.65	80.92 ± 0.77
GMNN+Comp	83.31 ± 0.81	72.93 ± 0.79	81.33 ± 0.59
GraphMix	83.91 ± 0.63	74.33 ± 0.65	80.68 ± 0.57
GraphMix+Clu	83.87 ± 0.56	75.16 ± 0.52	79.99 ± 0.82
GraphMix+Par	84.04 ± 0.57	74.93 ± 0.43	81.36 ± 0.33
GraphMix+Comp	83.76 ± 0.64	74.43 ± 0.72	80.82 ± 0.54



目录

- 简介
- 研究问题
- 主要贡献
- 贡献1. 如何结合自监督学习和图卷积神经网络？
- 贡献2. 如何设计自监督学习任务以提升泛化性能？
- 贡献3. 自监督学习是否能提升鲁棒性？
- 结论

贡献3. 自监督学习是否能提升鲁棒性？

- 我们首次将自监督学习引入对抗训练：

- 对抗训练：

$$\begin{aligned} Z &= f_{\theta}(X, \hat{A})\Theta, \quad Z' = f_{\theta}(X', A')\Theta, \\ \theta^*, \Theta^* &= \arg \min_{\theta, \Theta} (\mathcal{L}_{\text{sup}}(\theta, \Theta) + \alpha_3 \mathcal{L}_{\text{adv}}(\theta, \Theta)), \quad (6) \end{aligned}$$

- 自监督学习 + 对抗训练：

$$\begin{aligned} Z &= f_{\theta}(X, \hat{A})\Theta, \quad Z' = f_{\theta}(X', A')\Theta, \\ Z_{\text{ss}} &= f_{\theta}(X_{\text{ss}}, A_{\text{ss}}) \\ \theta^*, \Theta^*, \Theta_{\text{ss}}^* &= \arg \min_{\theta, \Theta, \Theta_{\text{ss}}} (\alpha_1 \mathcal{L}_{\text{sup}}(\theta, \Theta) \\ &\quad + \alpha_2 \mathcal{L}_{\text{ss}}(\theta, \Theta_{\text{ss}}) + \alpha_3 \mathcal{L}_{\text{adv}}(\theta, \Theta)), \quad (7) \end{aligned}$$

贡献3. 自监督学习是否能提升鲁棒性？

- 我们的实验结果显示：在不需要额外的更大的模型与更多的数据的情况下，自监督学习可提升模型的鲁棒性。

— 节点聚类 (Clu) 在对抗特征攻击的时候较为有效；

— 子图分割 (Par) 在对抗连接攻击的时候较为有效；

Table 7: Adversarial defense performances on Cora using adversarial training (abbr. AdvT) without or with graph self-supervision. Attacks include those on links, features (abbr. Feats), and both. Red numbers indicate the best two performances in each attack scenario (node classification accuracy; unit: %).

Attacks	None	Links	Feats	Links & Feats
GCN	80.61 \pm 0.21	28.72 \pm 0.63	44.06 \pm 1.23	8.18 \pm 0.27
AdvT	80.24 \pm 0.74	54.58 \pm 2.57	75.25 \pm 1.26	39.08 \pm 3.05
AdvT+Clu	80.26 \pm 0.99	55.54 \pm 3.19	76.24 \pm 0.99	41.84 \pm 3.48
AdvT+Par	80.42 \pm 0.76	56.36 \pm 2.57	75.88 \pm 0.72	41.57 \pm 3.47
AdvT+Comp	79.64 \pm 0.99	59.05 \pm 3.29	76.04 \pm 0.68	47.14 \pm 3.01

Table 8: Adversarial defense performances on Citeseer using adversarial training without or with graph self-supervision.

Attacks	None	Links	Feats	Links & Feats
GCN	71.05 \pm 0.56	13.68 \pm 1.09	22.08 \pm 0.73	3.08 \pm 0.17
AdvT	69.98 \pm 1.03	39.32 \pm 2.39	63.12 \pm 0.62	26.20 \pm 2.09
AdvT+Clu	70.13 \pm 0.81	40.32 \pm 1.73	63.67 \pm 0.45	27.02 \pm 1.29
AdvT+Par	69.96 \pm 0.77	41.05 \pm 1.91	64.06 \pm 0.24	28.70 \pm 1.60
AdvT+Comp	69.98 \pm 0.82	40.42 \pm 2.09	63.50 \pm 0.31	27.16 \pm 1.69

贡献3. 自监督学习是否能提升鲁棒性？

- 我们的实验结果显示：在不需要额外的更大的模型与更多的数据的情况下，自监督学习可提升模型的鲁棒性。

- 令人惊奇的是，图特征补全 (Comp) 显著地提升 Cora 数据集上在面对特征攻击与连接攻击的鲁棒性。

Table 7: Adversarial defense performances on Cora using adversarial training (abbr. AdvT) without or with graph self-supervision. Attacks include those on links, features (abbr. Feats), and both. Red numbers indicate the best two performances in each attack scenario (node classification accuracy; unit: %).

Attacks	None	Links	Feats	Links & Feats
GCN	80.61 \pm 0.21	28.72 \pm 0.63	44.06 \pm 1.23	8.18 \pm 0.27
AdvT	80.24 \pm 0.74	54.58 \pm 2.57	75.25 \pm 1.26	39.08 \pm 3.05
AdvT+Clu	80.26 \pm 0.99	55.54 \pm 3.19	76.24 \pm 0.99	41.84 \pm 3.48
AdvT+Par	80.42 \pm 0.76	56.36 \pm 2.57	75.88 \pm 0.72	41.57 \pm 3.47
AdvT+Comp	79.64 \pm 0.99	59.05 \pm 3.29	76.04 \pm 0.68	47.14 \pm 3.01

Table 8: Adversarial defense performances on Citeseer using adversarial training without or with graph self-supervision.

Attacks	None	Links	Feats	Links & Feats
GCN	71.05 \pm 0.56	13.68 \pm 1.09	22.08 \pm 0.73	3.08 \pm 0.17
AdvT	69.98 \pm 1.03	39.32 \pm 2.39	63.12 \pm 0.62	26.20 \pm 2.09
AdvT+Clu	70.13 \pm 0.81	40.32 \pm 1.73	63.67 \pm 0.45	27.02 \pm 1.29
AdvT+Par	69.96 \pm 0.77	41.05 \pm 1.91	64.06 \pm 0.24	28.70 \pm 1.60
AdvT+Comp	69.98 \pm 0.82	40.42 \pm 2.09	63.50 \pm 0.31	27.16 \pm 1.69



目录

- 简介
- 研究问题
- 主要贡献
- 贡献1. 如何结合自监督学习和图卷积神经网络？
- 贡献2. 如何设计自监督学习任务以提升泛化性能？
- 贡献3. 自监督学习是否能提升鲁棒性？
- 结论

- 我们的研究观察到了通过**多任务自监督学习**来将自监督学习引入图卷积神经网络的有效性；
- 我们通过设计合适的多任务自监督学习任务来在不同情况下（数据集，网络结构）提升图卷积神经网络的**泛化性能**；
- 我们的实验结果显示：在不需要额外的更大的模型与更多的数据的情况下，自监督学习可提升模型的**鲁棒性**。



TEXAS A&M UNIVERSITY

Engineering

Thank you for listening. Q&A