Adversarial Graph Augmentation to Improve Graph Contrastive Learning

Susheel Suresh
Purdue University
suresh43@purdue.edu

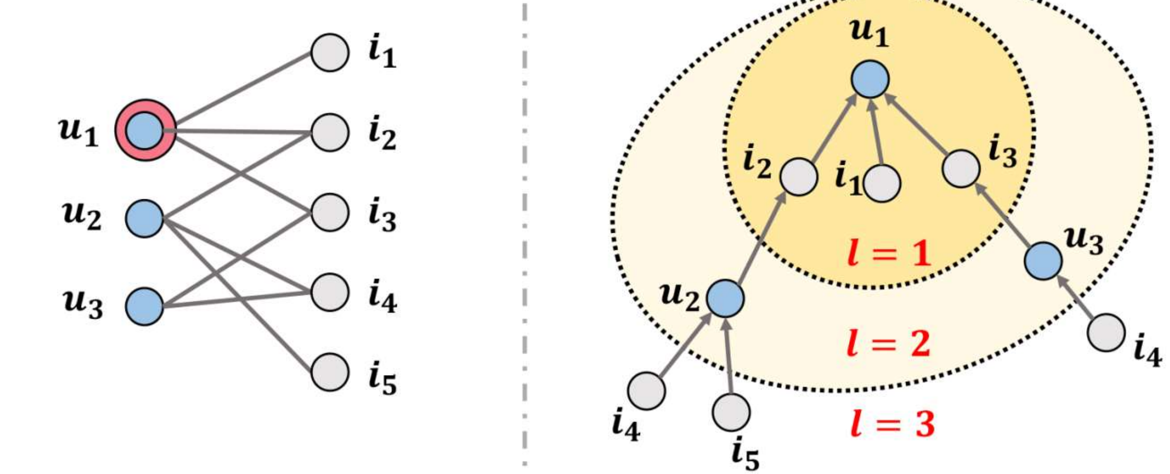
Pan Li*
Purdue University
panli@purdue.edu

Cong Hao
Georgia Tech
callie.hao@gatech.edu

Jennifer Neville
Purdue University and Microsoft Research
jenneville@microsoft.com

Background

Neural Graph Collaborative Filtering*



User-Item Interaction Graph   High-order Connectivity for u_1

对比学习

CL

compare negtive with positive samples ro differ features of different classes

→

图对比学习

GCL

using CL to capture key feature of Graph

图卷积神经网络

GNN

to capture infomations between nodes and graph

→

图表示学习

GRL

vectorize data to lower dimensions

classfication、prediction、structure analysis.....

Challenge

Problems:

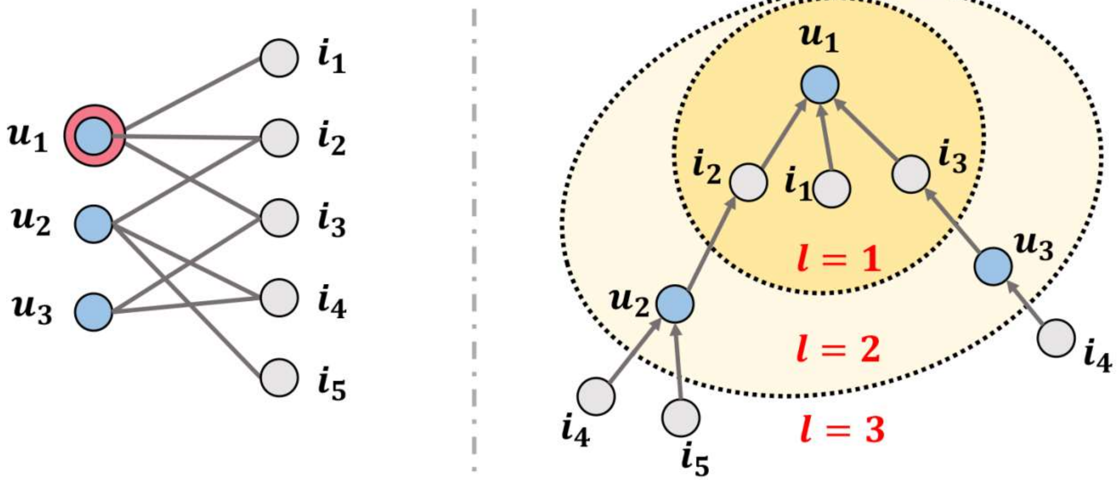
视图的缺乏，不能单纯修改颜色角度等

图结构的**不规则性**需设计有效的对比对

Way to enable contrasitive learning:

不同的图部分对比	Node - Node Node - Graph Node - Subgraph
图数据增强（GDA）	Edge Drop Node Drop Random walk

Neural Graph Collaborative Filtering*



User-Item Interaction Graph **High-order Connectivity for u_1**

Related Work

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

1·GraphCL：进行了广泛的GDA组合研究，需要大量评估来确定最佳组合，缺乏优化增强的考虑。

2·MVGRL and GCA：利用网络科学的领域知识进行GDA，但可能限制了方法的通用性。

以上方法均没有考虑到优化增强

3·JOAO：采用双层优化框架，但主要侧重于统一的扰动，缺乏对非统一扰动的支持。

扰动方法多样：JOAO方法设置为统一扰动的增强，而本文的AD-GCL允许非统一扰动的增强

4·InfoMin：需要下游任务的知识来指导增强，限制了其在没有明确下游任务的情况下的应用。

限制：InfoMin给出的最佳增强点需要下游任务知识的情况下实现

实证中用途不同：InfoMin应用于卷积神经网络CNN，而AD-GCL应用于图卷积神经网络GNN

Method

Learning Graph Representations

$G = (V, E)$

$\{X_v \in \mathbb{R}^F \mid v \in V\}$

$\{X_e \in \mathbb{R}^F \mid e \in E\}$

节点表示

h_v initialized as $h_v^{(0)} = X_v$

Graph Neural Networks (GNNs)

将节点表示和边属性 X_{uv}映射到一个聚合向量

在第k次迭代中：

$$h_v^{(k)} = \mathrm{UPDATE}^{(k)}\left(h_v^{(k-1)},\ \mathrm{AGGREGATE}^{(k)}\left(\{(h_u^{(k-1)}, X_{uv}) \mid u \in \mathcal{N}_v\}\right)\right)$$

将v的当前表示和聚合向量映射到v的更新表示

经过K次迭代后得到聚合的图的表示

$$f(G) :\triangleq h_G = \mathrm{POOL}\left(\{h_v^{(K)} \mid v \in V\}\right)$$

The Mutual Information Maximization Principle

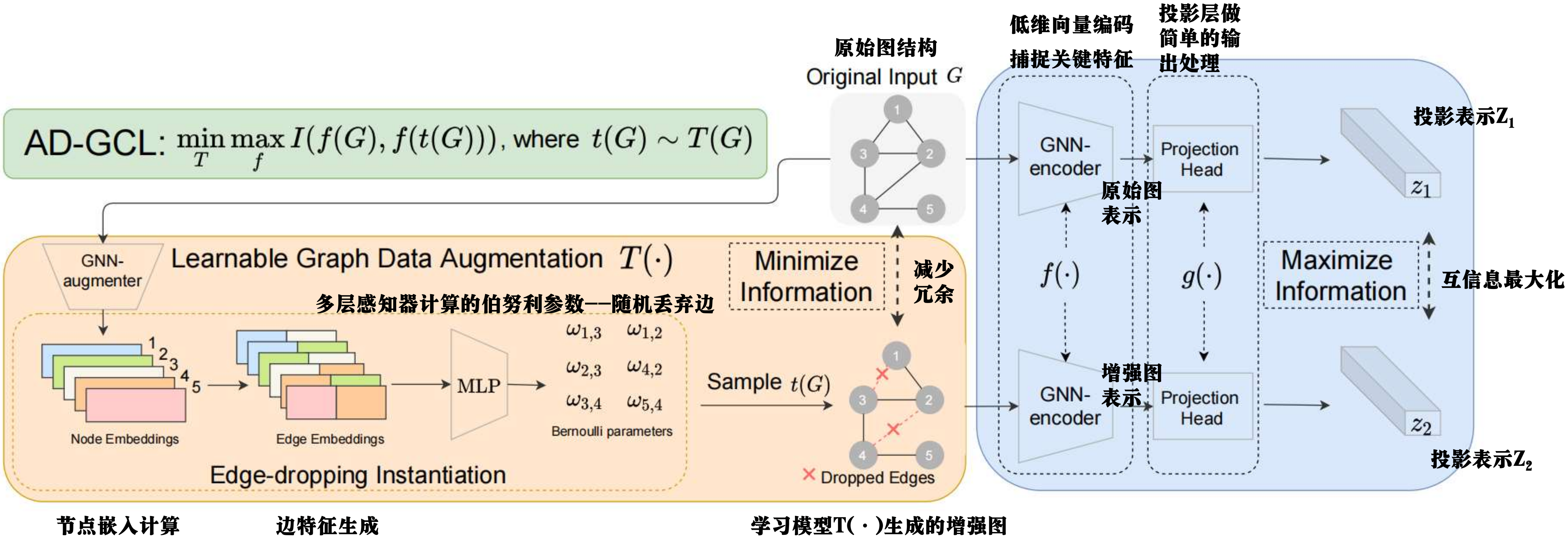
$$\text{InfoMax:} \quad \max_f I(G; f(G)), \quad \text{where } G \sim \mathbb{P}_{\mathcal{G}}$$

编码器f最大化图及其表示之间的互信息或对应关系

Method

$$p_e \sim \text{Bernoulli}(\omega_e)$$
$$\delta \sim \text{Uniform}(0,1)$$

$$p_e = \text{Sigmoid}\left(\frac{\log(\delta) - \log(1-\delta) + \omega_e}{\tau}\right)$$



Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

对于图像分类，捕获与图像标签完全无关的信息也能够最大化互信息，但这样的表示对于图像分类显然是无用的。

信息最大化原则 InfoMax: $\max_{f} I(G; f(G)), \quad \text{where } G \sim \mathbb{P}_{\mathcal{G}}$

最大化来自原图的信息

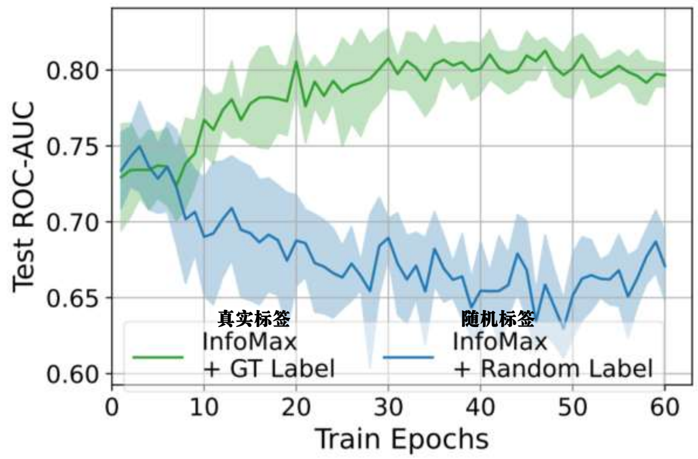
β **是一个正常数**

图信息瓶颈 GIB: $\max_{f} I(f(G); Y) - \beta I(G; f(G))$

应用于图表示学习

最大化与下游任务相关的信息
最小化来自原图的信息

数据集ogbg-molbace中的二元图分类
训练以保持图表示与输入图之间的互信息最大化



预测真实标签的测试性能

GIB需要从下游任务中获取类标签Y的知识，因此不适用于几乎没有或没有标签的GNN的自监督训练。

受GIB启发：不再简单的应用固定的数据增强技术。

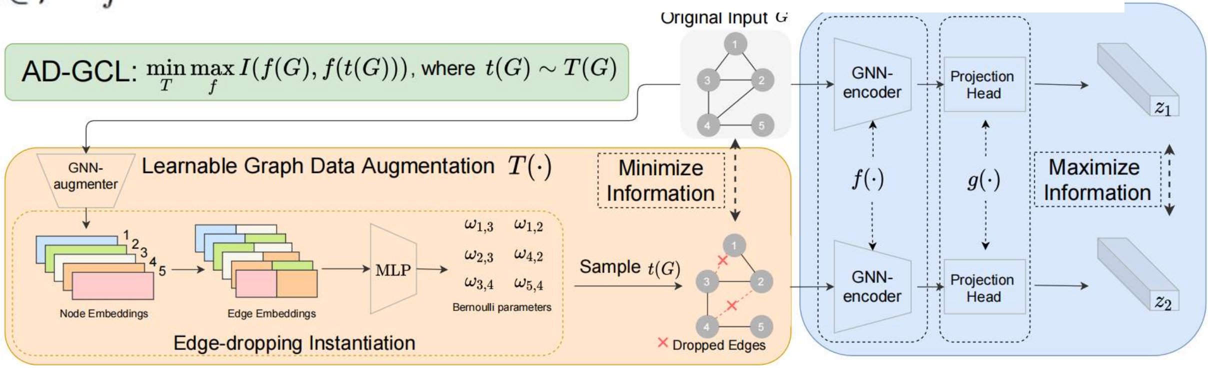
- **GDA使用图增强扰动减少编码信息量，在学习中确定那些增强最有效**
- **对扰动的图（不同的GDA）训练编码器f捕捉剩余信息使编码器捕捉到足够识别每个图的最小必要信息**

$$\text{AD-GCL:}\quad \min_{T\in\mathcal{T}}\max_{f} I(f(G); f(t(G))),\quad \text{where } G\sim\mathbb{P}_{\mathcal{G}}, t(G)\sim T(G)$$

$$\{G_i\}_{i=1}^{m}$$

$$z_{i,1} = g(f_{\Theta}(G_i))$$

$$z_{i,2} = g(f_{\Theta}(t(G_i)))$$



$$I(f_{\Theta}(G); f_{\Theta}(t(G))) \rightarrow \hat{I} = \frac{1}{m}\sum_{i=1}^{m}\log\frac{\exp(sim(z_{i,1}, z_{i,2}))}{\sum_{i'=1, i'\neq i}^{m}\exp(sim(z_{i,1}, z_{i',2}))}$$

给定$z_{i,1}$和$z_{i,2}$增强图的表示之间的相似性的指数↑

互信息项

对所有m图的相似性指数的总和

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

生化与社交网络分类任务

FIX–固定参数
OPT–优化参数
NAD–非对抗
AD–对抗

	Dataset	NCI1	PROTEINS	MUTAG	DD	COLLAB	RDT-B	RDT-M5K	IMDB-B	IMDB-M
	F-GIN	78.27 ± 1.35	72.39 ± 2.76	90.41 ± 4.61	74.87 ± 3.56	74.82 ± 0.92	86.79 ± 2.04	53.28 ± 3.17	71.83 ± 1.93	48.46 ± 2.31
Baselines	RU-GIN [72]	62.98 ± 0.10	69.03 ± 0.33	87.61 ± 0.39	74.22 ± 0.30	63.08 ± 0.10	58.97 ± 0.13	27.52 ± 0.61	51.86 ± 0.33	32.81 ± 0.57
Baselines	InfoGraph [18]	68.13 ± 0.59	72.57 ± 0.65	87.71 ± 1.77	75.23 ± 0.39	70.35 ± 0.64	78.79 ± 2.14	51.11 ± 0.55	71.11 ± 0.88	48.66 ± 0.67
Baselines	GraphCL [24]	68.54 ± 0.55	72.86 ± 1.01	88.29 ± 1.31	74.70 ± 0.70	71.26 ± 0.55	82.63 ± 0.99	53.05 ± 0.40	70.80 ± 0.77	48.49 ± 0.63
AB-S	NAD-GCL-FIX	69.23 ± 0.60	72.81 ± 0.71	88.58 ± 1.58	74.55 ± 0.55	71.56 ± 0.58	83.41 ± 0.66	52.72 ± 0.71	70.94 ± 0.77	48.33 ± 0.47
AB-S	NAD-GCL-OPT	69.30 ± 0.32	73.18 ± 0.71	89.05 ± 1.06	74.55 ± 0.55	72.04 ± 0.67	83.74 ± 0.76	53.43 ± 0.26	71.94 ± 0.59	49.01 ± 0.93
Ours	AD-GCL-FIX	**69.67 ± 0.51***	**73.59 ± 0.65**	**89.25 ± 1.45**	74.49 ± 0.52	**73.32 ± 0.61***	**85.52 ± 0.79***	53.00 ± 0.82	**71.57 ± 1.01**	**49.04 ± 0.53**
Ours	AD-GCL-OPT	**69.67 ± 0.51***	**73.81 ± 0.46***	**89.70 ± 1.03**	75.10 ± 0.39	**73.32 ± 0.61***	**85.52 ± 0.79***	**54.93 ± 0.43***	**72.33 ± 0.56***	**49.89 ± 0.66***

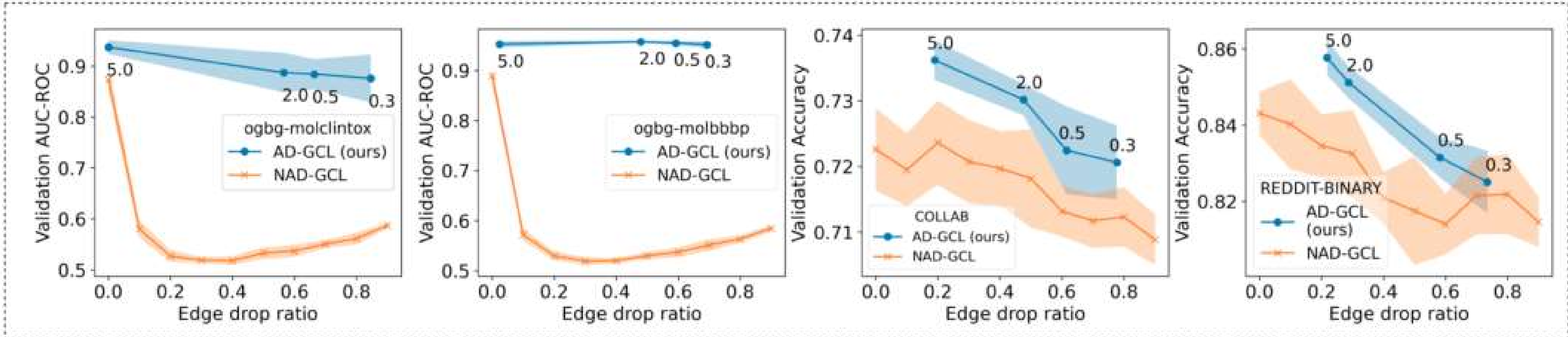
化学分子领域属性预测回归与分类性能

	Task	Regression (Downstream Classifier - Linear Regression + L2)				Classification (Downstream Classifier - Logistic Regression + L2)				
	Dataset	molesol	mollipo	molfreesolv	ZINC-10K	molbace	molbbbp	molclintox	moltox21	molsider
	Metric	RMSE (shared) (↓)			MAE (↓)	ROC-AUC % (shared) (↑)				
	F-GIN	1.173 ± 0.057	0.757 ± 0.018	2.755 ± 0.349	0.254 ± 0.005	72.97 ± 4.00	68.17 ± 1.48	88.14 ± 2.51	74.91 ± 0.51	57.60 ± 1.40
Baselines	RU-GIN [72]	1.706 ± 0.180	1.075 ± 0.022	7.526 ± 2.119	0.809 ± 0.022	75.07 ± 2.23	64.48 ± 2.46	72.29 ± 4.15	71.53 ± 0.74	62.29 ± 1.12
Baselines	InfoGraph [18]	1.344 ± 0.178	1.005 ± 0.023	10.005 ± 4.819	0.890 ± 0.017	74.74 ± 3.64	66.33 ± 2.79	64.50 ± 5.32	69.74 ± 0.57	60.54 ± 0.90
Baselines	GraphCL [24]	1.272 ± 0.089	0.910 ± 0.016	7.679 ± 2.748	0.627 ± 0.013	74.32 ± 2.70	68.22 ± 1.89	74.92 ± 4.42	72.40 ± 1.01	61.76 ± 1.11
AB-S	NAD-GCL-FIX	1.392 ± 0.065	0.952 ± 0.024	5.840 ± 0.877	0.609 ± 0.010	73.60 ± 2.73	66.12 ± 1.80	73.32 ± 3.66	71.65 ± 0.94	60.41 ± 1.48
AB-S	NAD-GCL-OPT	1.242 ± 0.096	0.897 ± 0.022	5.840 ± 0.877	0.609 ± 0.010	73.69 ± 3.67	67.70 ± 1.78	74.40 ± 4.92	71.65 ± 0.94	61.14 ± 1.43
Ours	AD-GCL-FIX	1.217 ± 0.087	0.842 ± 0.028*	**5.150 ± 0.624***	**0.578 ± 0.012***	**76.37 ± 2.03**	68.24 ± 1.47	80.77 ± 3.92	71.42 ± 0.73	**63.19 ± 0.95**
Ours	AD-GCL-OPT	1.136 ± 0.050*	0.812 ± 0.020*	**4.145 ± 0.369***	**0.544 ± 0.004***	**77.27 ± 2.56**	69.54 ± 1.92	80.77 ± 3.92	**72.92 ± 0.86**	**63.19 ± 0.95**

Experiments

Nanjing University of Aeronautics and Astronautics

性能与丢边率对照



AD–GCL和非对抗性GCL（NAD–GCL）在不同边丢弃比率下的图分类任务的验证性能

激进的GDA情况下，即使丢边率很高AD–GCL也可以便显出强劲的性能

Experiments

南京航空航天大學 Nanjing University of Aeronautics and Astronautics

迁移学习性能（化学分子性质预测）

Pre-Train Dataset	ZINC 2M								PPI-306K
Fine-Tune Dataset	BBBP	Tox21	SIDER	ClinTox	BACE	HIV	MUV	ToxCast	PPI
No Pre-Train	65.8 ± 4.5	74.0 ± 0.8	57.3 ± 1.6	58.0 ± 4.4	70.1 ± 5.4	75.3 ± 1.9	71.8 ± 2.5	63.4 ± 0.6	64.8 ± 1.0
EdgePred [17]	67.3 ± 2.4	76.0 ± 0.6	60.4 ± 0.7	64.1 ± 3.7	79.9 ± 0.9	76.3 ± 1.0	74.1 ± 2.1	64.1± 0.6	65.7 ± 1.3
AttrMasking [17]	64.3 ± 2.8	76.7 ± 0.4	61.0 ± 0.7	71.8 ± 4.1	79.3 ± 1.6	77.2 ± 1.1	74.7 ± 1.4	64.2 ± 0.5	65.2 ± 1.6
ContextPred [17]	68.0 ± 2.0	75.7 ± 0.7	60.9 ± 0.6	65.9 ± 3.8	79.6 ± 1.2	77.3 ± 1.0	75.8 ± 1.7	63.9 ± 0.6	64.4 ± 1.3
InfoGraph [18]	68.8 ± 0.8	75.3 ± 0.5	58.4 ± 0.8	69.9 ±3.0	75.9 ± 1.6	76.0 ± 0.7	75.3 ± 2.5	62.7 ± 0.4	64.1 ± 1.5
GraphCL [24]	69.68 ± 0.67	73.87 ± 0.66	60.53 ± 0.88	75.99 ± 2.65	75.38 ± 1.44	78.47 ± 1.22	69.8 ± 2.66	62.40 ± 0.57	67.88 ± 0.85
AD-GCL-FIX	70.01 ±1.07	76.54 ± 0.82	**63.28 ± 0.79**	**79.78 ± 3.52**	78.51 ± 0.80	78.28 ± 0.97	72.30 ± 1.61	63.07 ± 0.72	**68.83 ± 1.26**
Our Ranks	1	2	1	1	4	2	5	5	1

半监督学习性能

Dataset	NCI1	PROTEINS	DD	COLLAB	RDT-B	RDT-M5K
No Pre-Train	73.72 ± 0.24	70.40 ± 1.54	73.56 ± 0.41	73.71± 0.27	86.63 ± 0.27	51.33 ± 0.44
SS-GCN-A	73.59 ± 0.32	70.29 ± 0.64	74.30 ± 0.81	74.19 ± 0.13	87.74 ± 0.39	52.01 ± 0.20
GAE [20]	74.36 ± 0.24	70.51 ± 0.17	74.54 ± 0.68	75.09 ± 0.19	87.69 ± 0.40	53.58 ± 0.13
InfoGraph [18]	74.86 ± 0.26	72.27 ± 0.40	75.78 ± 0.34	73.76 ± 0.29	88.66 ± 0.95	53.61 ± 0.31
GraphCL [24]	74.63 ± 0.25	74.17 ± 0.34	76.17 ± 1.37	74.23 ± 0.21	89.11 ± 0.19	52.55 ± 0.45
AD-GCL-FIX	**75.18 ± 0.31**	73.96 ± 0.47	**77.91 ± 0.73***	**75.82 ± 0.26***	**90.10 ± 0.15***	53.49 ± 0.28
Our Ranks	1	2	1	1	1	3

提高信息效率
对抗性训练
适应多种学习环境
扩展性和灵活性

Thanks