Resource-efficient Graph Representation Learning

Chuxu Zhang

Computer Science Department

Brandeis University

A Little About Myself

❑ Chuxu Zhang, Ph.D.

- Assistant Professor, Computer Science, Brandeis University
- Ph.D. degree (2020), University of Notre Dame
- Personal webpage: https://chuxuzhang.github.io/

❑ Research

- Fundamental Research

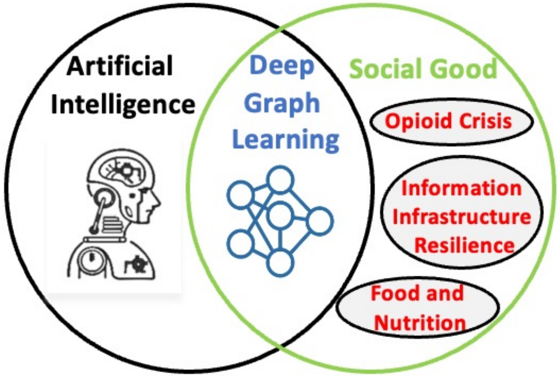
Graph learning with a focus on resource efficiency and trustworthiness

- Applied Research

AI and graph learning for social good and interdisciplinary studies

- More than 50 Papers in Major Conferences

NeurIPS, KDD, WWW, EMNLP, WSDM, AAAI, IJCAI, ICDM, CIKM, SDM, PKDD

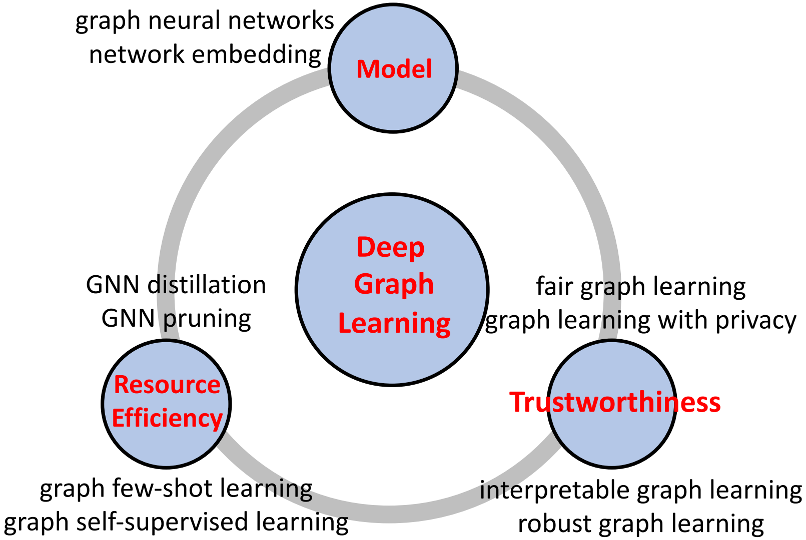


❑ Award/Honor

- Best Paper Award (or nomination): CIKM-21, WWW-19, APWeb/WAIM-16

Fundamental Research - Graph Learning

- **Model**
  - Network Embedding
  - Graph Neural Networks
- **Resource Efficiency**
  - Data Efficiency
  - Model Efficiency
  - Data-Model Co-efficiency
- **Trustworthiness**
  - Robustness
  - Explainability
  - Fairness
  - Privacy

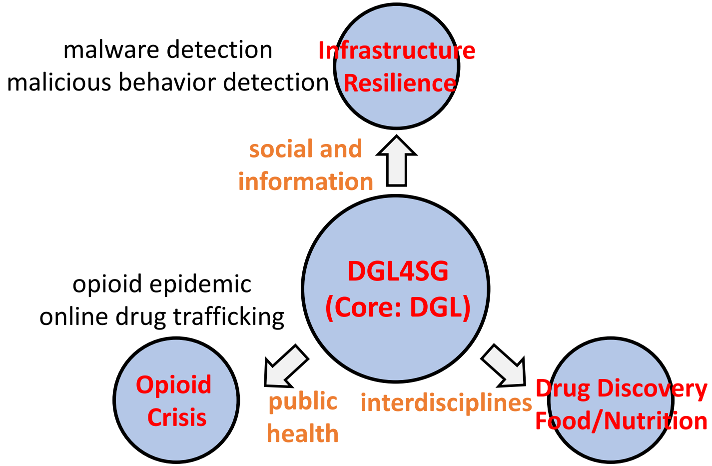


Applied Research – AI and Graph Learning for Social Good

- Public Health
  - Opioid Crisis

- Social/Information Infrastructure Resilience
  - Software/repository-based Applications
  - Anomaly Detection

- Interdisciplinary Studies
  - Molecular Analysis
  - Food and Nutrition Service



Resource-efficient Graph Representation Learning

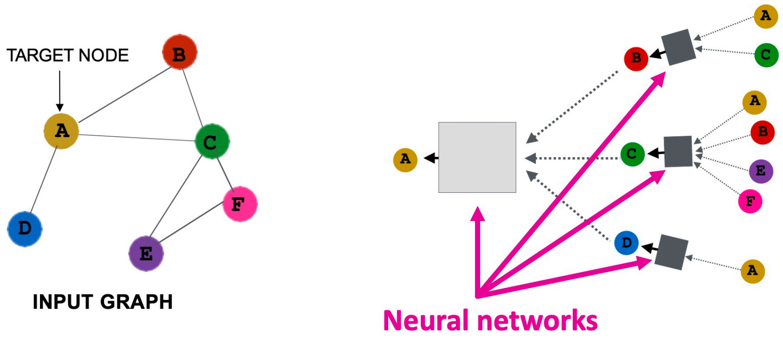
- ❑ Introduction
- ❑ Resource-efficient GRL
- ❑ Social Good Application
- ❑ Conclusion

Introduction - GRL

- Graph Representation Learning (GRL)
  - Graph Neural Networks: Learn node representations through message passing/neighbor aggregation

Node-level: $$h_v^{(l+1)} = \mathrm{COM}\left(h_v^{(l)}, \left[\mathrm{AGG}\left(\left\{h_u^{(l)} \mid \forall u \in \mathcal{N}_v\right\}\right)\right]\right)$$

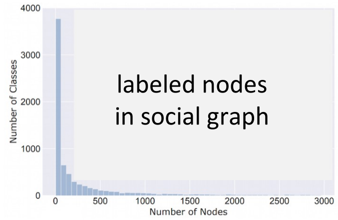
Graph-level: $$h_G^{(l)} = \mathrm{READOUT}\left\{h_v^{(l)} \mid \forall v \in V\right\}$$

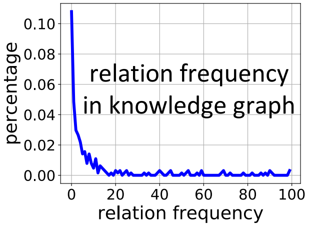


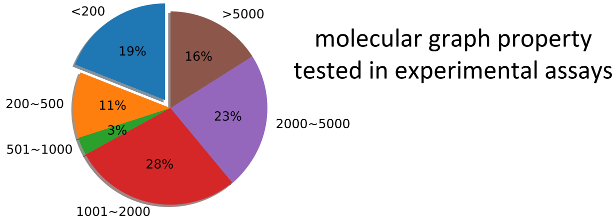
Introduction - Resource Constraint Challenges

❑ Data-level Challenge

- Small labeled data in many real applications







❑ Model-level Challenge

- Large graphs and large models but limited computing resources or small devices

OPEN GRAPH BENCHMARK

AMiner

❑ Other Challenges

- Noisy data, data imbalance, hardware constraint, etc.

Resource-efficient Graph Representation Learning

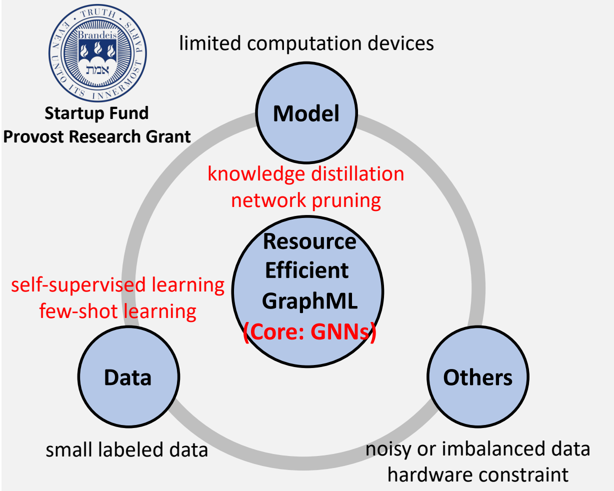
- ❑ Introduction
- ❑ Resource-efficient GRL
- ❑ Social Good Application
- ❑ Conclusion

Introduction - Tackling Resource Constraint Challenges

- ❑ Data-efficient GRL
  - Unlabeled Data Efficiency: Graph Self-supervised Learning
  - Label Efficiency: Graph Few-shot Learning

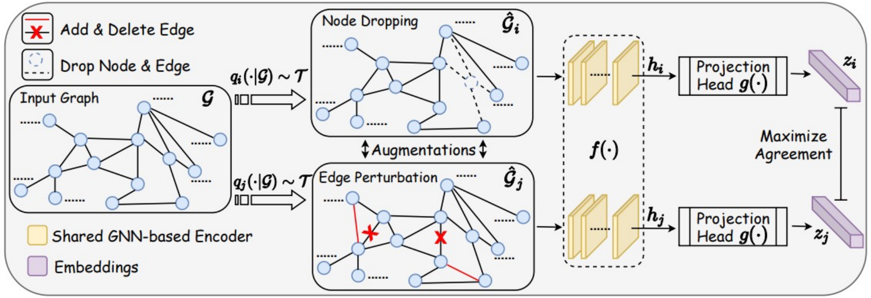
- ❑ Model-efficient GRL
  - Graph Knowledge Distillation
  - GNN Pruning

- ❑ Data-Model Co-efficient GRL
  - GNN Pruning + Graph Self-supervised Learning



Data-efficient GRL - Unlabeled Data Efficiency

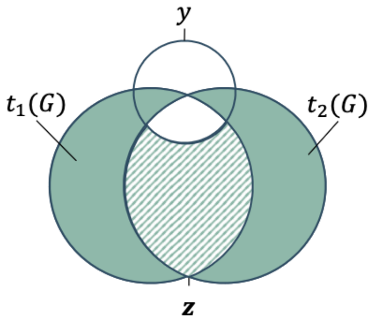
❑ Graph Self-supervised Learning: Graph Contrastive Learning (GCL)



Qiu et al., GCC, KDD-20
You et al., GraphCL, NeurIPS-20
Thakoor et al., BGRL, ICLR-22; etc.

Mutual Information Maximization Principle

Velickovic et al., DGI, ICLR-19

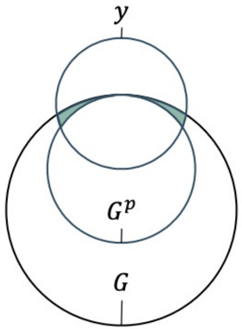


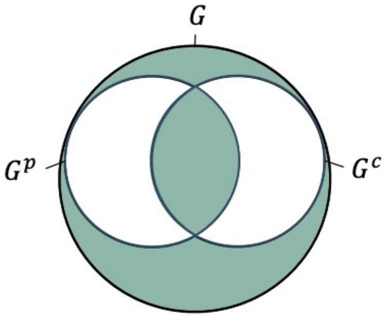
- The shared information: the learned graph representation **z**.
- **z:** both predictive information G^p and non-predictive G^c information. (Suresh et al., AD-GCL, NeurIPS-21)
- Ideal situation: G^p is sufficient for y, G^p and G^c are disentangled, $I(G^p; G^c) = 0$

Graph Contrastive Learning with Cross-view Reconstruction, arXiv, Under Review at ICLR-23

Data-efficient GRL - Unlabeled Data Efficiency

❑ Ideal Situation: Zero Shaded Region





Sufficiency: G^p covers the information which is sufficient to make correct graph label prediction, G^c is redundant.

Disentanglement: G^p and G^c are supposed to be mutually disentangled with each other, $I(G^p;G^c) = 0$. The union of them cover all the information of original data.

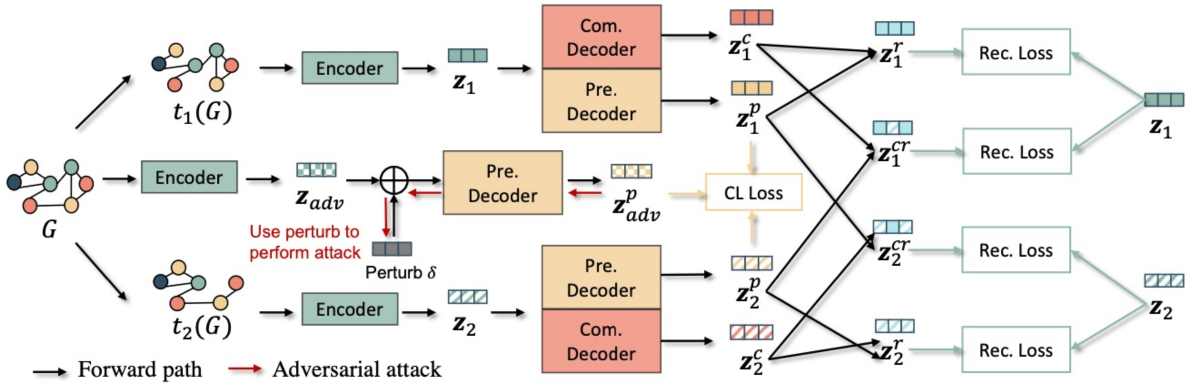
Information Bottleneck Principle: Minimal yet Sufficient Information

Tishby et al., The information bottleneck method, arXiv

Graph Contrastive Learning with Cross-view Reconstruction, arXiv, Under Review at ICLR-23

Data-efficient GRL - Unlabeled Data Efficiency

❑ Our Model: Graph Contrastive Learning with Cross-view Reconstruction (GraphCV)



- Two types of factors: $\mathbf{z^p}$ captures the predictive factors and $\mathbf{z^c}$ keeps other complementary non-predictive factors.

- Minimal: Two pairs of representations to reconstruct $\mathbf{z_1}$ and $\mathbf{z_2}$ in both the intra-view and inter-view.

Fact: $\mathbf{z^c}$ is easily influenced by different augmentations while **$\mathbf{z^p}$** is not (Suresh et al., AD-GCL, NeurIPS-21)
that is **$\mathbf{z^c}$** combines **$\mathbf{z^p}$** in the same view or other reviews should reconstruct **$\mathbf{z}$**, so as to learn minimal **$\mathbf{z^p}$**

- Sufficient: Third view - adversarial view $\mathbf{z^p_{adv}}$ (can be extracted from the original graph but the adversarial graph has better result).

avoid partial or even trivial information of **$\mathbf{z^p}$**

Graph Contrastive Learning with Cross-view Reconstruction, arXiv, Under Review at ICLR-23

Data-efficient GRL - Unlabeled Data Efficiency

❑ GraphCV: Loss Function

Theorem 1 *Suppose $f(\cdot)$ is a GNN encoder as powerful as 1-WL test. Let $\mathbf{z}_1^p$ and $\mathbf{z}_2^p$ be specific to the predictive information of G, meanwhile $\mathbf{z}_1^c$ and $\mathbf{z}_2^c$ account for the non-predictive factors of $t_1(G)$ and $t_2(G)$. Then we have:*

$$I\left(t_1(G);\mathbf{z}_2^p,\mathbf{z}_2^c\right) \geqslant I\left(\mathbf{z}_1^p;\mathbf{z}_2^p\right) \text{ where } G \in \mathcal{G} \text{ and } t_1(\cdot), t_2(\cdot) \in \mathcal{T}.$$

$$\mathcal{L}_{\text{pre}} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_{\text{CL}}(\mathbf{z}_{1,i}^p, \mathbf{z}_{2,i}^p)$$

Theorem 2 *Assume q is a Gaussian distribution, g_r is the parameterized reconstruction model which infers $\mathbf{z}_w$ from $(\mathbf{z}_{w'}^p, \mathbf{z}_w^c)$. Then we have:*

$$H\left(\mathbf{z}_w \mid \mathbf{z}_w^p, \mathbf{z}_{w'}^c\right) \leqslant \left\|\mathbf{z}_w - g_r\left(\mathbf{z}_w^p \odot \mathbf{z}_{w'}^c\right)\right\|_2^2 \text{ where } w = w' \text{ or } w \neq w'.$$

$$H\left(\mathbf{z}_w \mid \mathbf{z}_{w'}^p, \mathbf{z}_w^c\right) = -\mathbb{E}_{p\left(\mathbf{z}_w, \mathbf{z}_{w'}^p, \mathbf{z}_w^c\right)}\left[\log p\left(\mathbf{z}_w \mid \mathbf{z}_{w'}^p, \mathbf{z}_w^c\right)\right] = 0,$$

$$\mathcal{L}_{\text{recon}} = \frac{1}{2N}\sum_{i=1}^{N}\sum_{w=1}^{2}\left[\left\|\mathbf{z}_{w,i} - \mathbf{z}_{w,i}^r\right\|_2^2 + \left\|\mathbf{z}_{w,i} - \mathbf{z}_{w,i}^{cr}\right\|_2^2\right]$$

$$\mathcal{L}_{\text{adv}} = \frac{1}{N}\sum_{i=1}^{N}\max_{\delta*}\left[\mathcal{L}_{\text{CL}}\left(\mathbf{z}_{1,i}^p, G+\delta^*\right) + \mathcal{L}_{\text{CL}}\left(\mathbf{z}_{2,i}^p, G+\delta^*\right)\right]$$

$$\min_{f,g}\mathbb{E}_{G\in\mathbf{G}}\left[\mathcal{L}_{\text{pre}} + \lambda_r\mathcal{L}_{\text{recon}} + \lambda_a \max_{\|\delta\|_\infty \leqslant \epsilon}\mathcal{L}_{\text{adv}}\right]$$

Graph Contrastive Learning with Cross-view Reconstruction, arXiv, Under Review at ICLR-23

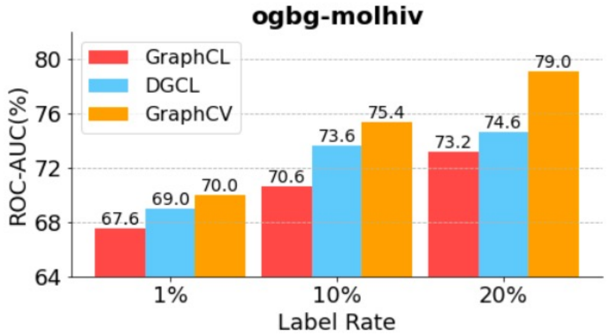
Data-efficient GRL - Unlabeled Data Efficiency

❑ GraphCV: Experiments

graph classification (unsupervised setting)

	MUTAG	PTC-MR	COLLAB	NCI1	PROTEINS	IMDB-B	RDT-B	IMDB-M	DD
node2vec	72.6±10.2	58.6±8.0	-	54.9±1.6	57.5±3.6	-	-	-	-
graph2vec	83.2±9.3	60.2±6.9	-	73.2±1.8	73.3±2.1	71.1±0.5	75.8±1.0	50.4±0.9	-
InfoGraph	89.0±1.1	61.7±1.4	70.7±1.1	76.2±1.1	74.4±0.3	73.0±0.9	82.5±1.4	49.7±0.5	72.9±1.8
VGAE	87.7±0.7	61.2±1.8	-	-	-	70.7±0.7	87.1±0.1	49.3±0.4	-
MVGRL	89.7±1.1	62.5±1.7	-	-	-	74.2±0.7	84.5±0.6	51.2±0.5	-
GraphCL	86.8±1.3	63.6±1.8	71.4±1.2	77.9±0.4	74.4±0.5	71.1±0.4	89.5±0.8	-	78.6±0.4
InfoGCL	91.2±1.3	63.5±1.5	80.0±1.3	80.2±0.6	-	75.1±0.9	-	51.4±0.8	-
DGCL	92.1±0.8	65.8±1.5	**81.2±0.3**	81.9±0.2	76.4±0.5	**75.9±0.7**	91.8±0.2	51.9±0.4	-
AD-GCL	89.7±1.0	-	73.3±0.6	69.7±0.5	73.8±0.5	72.3±0.6	85.5±0.8	49.9±0.7	75.1±0.4
RGCL	87.7±1.0	-	70.9±0.7	78.1±1.1	75.0±0.4	71.9±0.8	90.3±0.6	-	78.9±0.5
GASSL	90.9±7.9	64.6±6.1	78.0±2.0	80.2±1.9	-	74.2±0.5	-	51.7±2.5	-
GraphCV	**92.3±0.7**	**67.4±1.3**	80.5±0.5	**82.0±1.0**	**76.8±0.4**	75.6±0.4	**92.4±0.9**	**52.2±0.5**	**80.5±0.5**

graph classification (semi-supervised setting)



graph classification (transfer learning setting)

	BBBP	Tox21	ToxCast	SIDER	ClinTox	MUV	HIV	BACE	Avg
No Pre-Train	65.8±4.5	74.0±0.8	63.4 ±0.6	57.3±1.6	58.0±4.4	71.8±2.5	75.3±1.9	70.1±5.4	67.0
AttrMasking	64.3±2.8	**76.7±0.4**	**64.2±0.5**	61.0±0.7	71.8±4.1	74.7±1.4	77.2±1.1	79.3±1.6	71.1
ContextPred	68.0±2.0	75.7±0.7	63.9±0.6	60.9±0.6	65.9±3.8	75.8±1.7	77.3±1.0	79.6±1.2	70.9
GraphCL	69.5±0.5	75.4±0.9	63.8±0.4	60.8±0.7	70.1±1.9	74.5±1.3	77.6±0.9	78.2±1.2	70.8
GraphLoG	**72.5±0.8**	75.7±0.5	63.5±0.7	61.2±1.1	76.7±3.3	76.0±1.1	77.8±0.8	**83.5±1.2**	73.4
JOAO	70.2±1.0	75.0±0.3	62.9±0.5	60.0±0.8	81.3±2.5	71.7±1.4	76.7±1.2	51.5±0.4	71.9
RGCL	71.4±0.7	75.2±0.3	63.3±0.2	**61.4±0.6**	83.4 ±0.9	76.7 ±1.0	77.9±0.8	76.03±0.8	73.2
GraphCV	71.6±0.6	75.7±0.6	63.2±0.5	**62.2±0.7**	**83.6±1.5**	76.4±0.8	**77.9±1.0**	80.8±1.8	**73.9**

Graph Contrastive Learning with Cross-view Reconstruction, arXiv, Under Review at ICLR-23

Data-efficient GRL - Label Efficiency

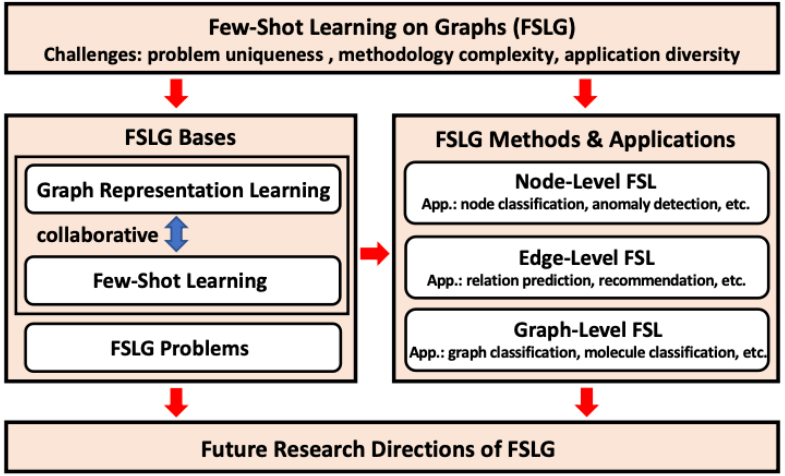
❑ Graph Few-shot Learning

- ○ Problem: GRL with few labeled samples on classes as supervision signals
- ○ Method: Combine strengths of GNNs and few-shot learning

○ Node-level GFL: few-shot node classification
GFL (AAAI-20), MetaHG (NeurIPS-21), TENT (KDD-22), CrossHG-Meta (KDD-22), etc.

○ Edge-level GFL: few-shot relation prediction
FSRL (AAAI-20), FIRE (EMNLP-20), etc.

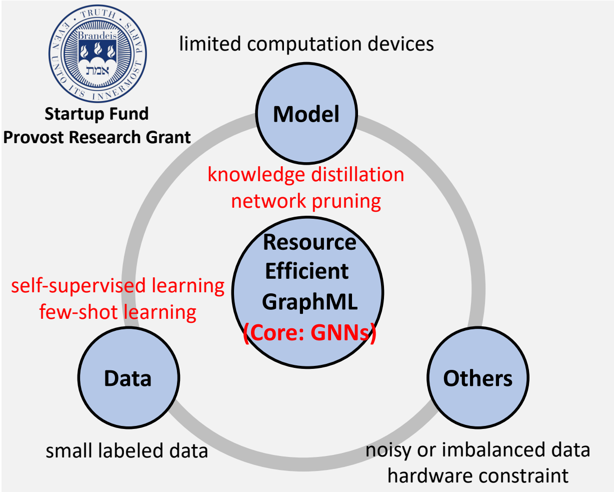
○ Graph-level GFL: few-shot graph classification
MGNN (WWW-21), etc.



Towards Graph Minimally-Supervised Learning, Tutorial at KDD-22
Few-Shot Learning on Graphs, IJCAI-22

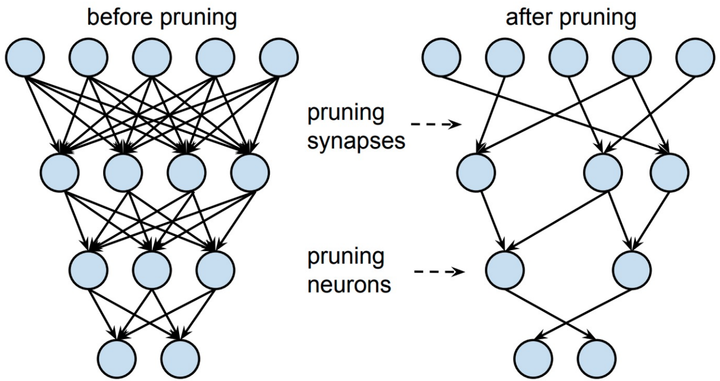
Introduction - Tackling Resource Constraint Challenges

- Data-efficient GRL
  - Unlabeled Data Efficiency: Graph Self-supervised Learning
  - Label Efficiency: Graph Few-shot Learning
- Model-efficient GRL
  - Graph Knowledge Distillation
  - GNN Pruning
- Data-Model Co-efficient GRL
  - GNN Pruning + Graph Self-supervised Learning



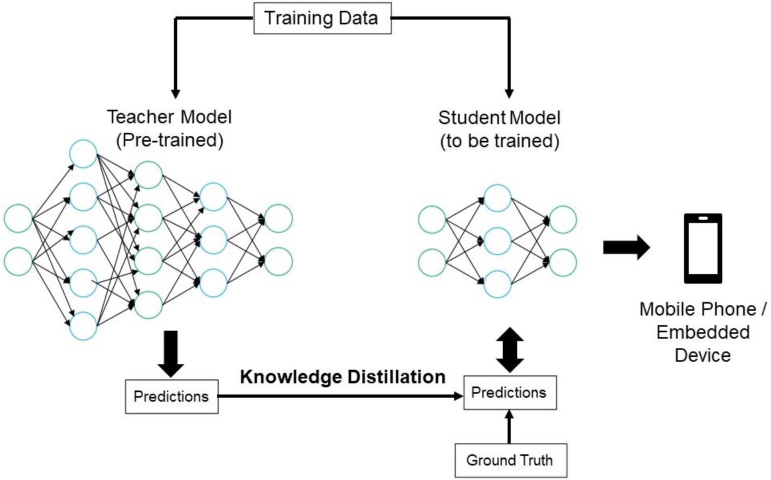
Model-efficient GRL

- Model-efficient GRL
  - Network Pruning
  - Knowledge Distillation



extract subnetworks (winning tickets)
to approximate full network

Frankle et al., Lottery Ticket, ICLR-19



$$\mathcal{L} = \sum_{v \in \mathcal{V}^L} \mathcal{L}_{GT}(\hat{\boldsymbol{y}}_v, \boldsymbol{y}_v) + \lambda \sum_{v \in \mathcal{V}} \mathcal{L}_{SL}(\hat{\boldsymbol{y}}_v, \boldsymbol{z}_v)$$

Hinton et al., Distilling the Knowledge in a Neural Network, arXiv

Learning MLPS On Graphs: A Unified View of Effectiveness, Robustness, and Efficiency, arXiv, Under Review at ICLR-23

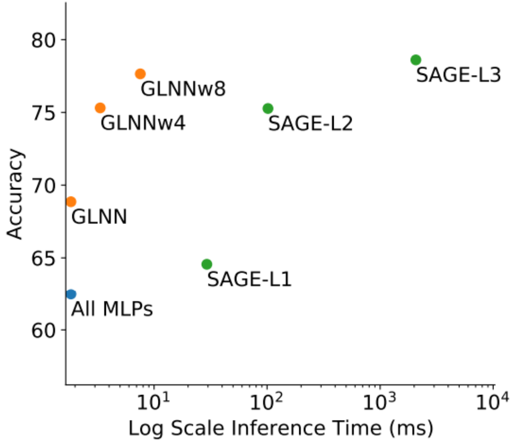
Model-efficient GRL

❑ GNN Model Distillation (Large Teacher to Small Student)

- What we need is just MLP during inference (Zhang et al., GLNN, ICLR-22)

Limitation: sacrifice in effectiveness, especially for large graphs

- RQ: Can MLP outperforms GNN in both effectiveness and efficiency during inference?



much faster in inference

Datasets	SAGE	MLP	GLNN	Δ_{MLP}	Δ_{GNN}
Cora	80.52 ± 1.77	59.22 ± 1.31	$\mathbf{80.54 \pm 1.35}$	21.32 (36.00%)	0.02 (0.02%)
Citeseer	70.33 ± 1.97	59.61 ± 2.88	$\mathbf{71.77 \pm 2.01}$	12.16 (20.40%)	1.44 (2.05%)
Pubmed	75.39 ± 2.09	67.55 ± 2.31	$\mathbf{75.42 \pm 2.31}$	7.87 (11.65%)	0.03 (0.04%)
A-computer	82.97 ± 2.16	67.80 ± 1.06	$\mathbf{83.03 \pm 1.87}$	15.23 (22.46%)	0.06 (0.07%)
A-photo	90.90 ± 0.84	78.77 ± 1.74	$\mathbf{92.11 \pm 1.08}$	13.34 (16.94%)	1.21 (1.33%)
Arxiv	$\mathbf{70.92 \pm 0.17}$	56.05 ± 0.46	63.46 ± 0.45	7.41 (13.24%)	-7.46 (-10.52%)
Products	$\mathbf{78.61 \pm 0.49}$	62.47 ± 0.10	68.86 ± 0.46	6.39 (10.23%)	-9.75 (-12.4%)

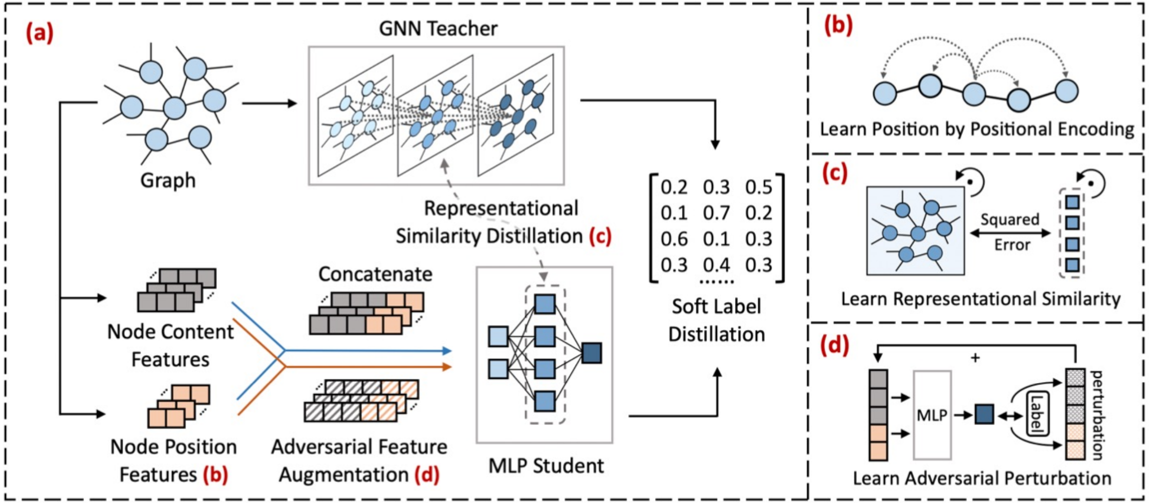
close or even worse performance

- lack rich (global) structure information
- strict hard matching to teacher's output (response-based knowledge)

Learning MLPS On Graphs: A Unified View of Effectiveness, Robustness, and Efficiency, arXiv, Under Review at ICLR-22

Model-efficient GRL

❑ Our Model: NOise-robust Structure-aware MLPs On Graphs (NOSMOG)



- Richer feature: augment node content feature with position (structure) feature to add structure information
- Representational similarity distillation (RSD): force embedding similarity in hidden state (feature-based knowledge)
- Robust feature augmentation: add adversarial feature to ensure robustness

$$\mathcal{L} = \mathcal{L}_{GT} + \lambda\mathcal{L}_{SL} + \mu\mathcal{L}_{RSD} + \eta\mathcal{L}_{ADV},$$

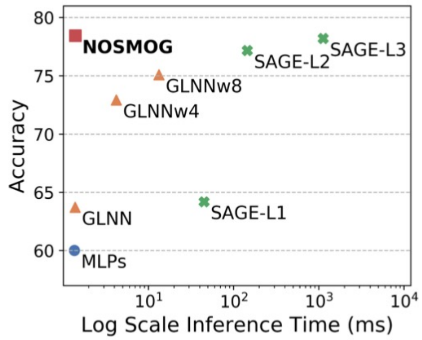
Learning MLPS On Graphs: A Unified View of Effectiveness, Robustness, and Efficiency, arXiv, Under Review at ICLR-23

Model-efficient GRL

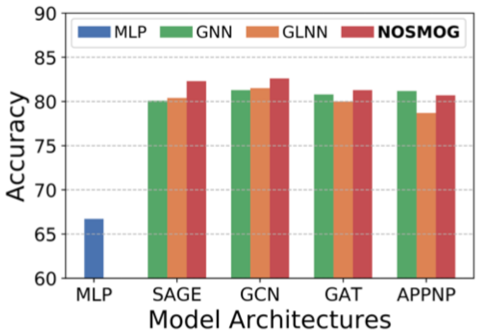
❑ NOSMOG: Experiments

node classification

Datasets	SAGE	MLP	GLNN	NOSMOG	Δ_{GNN}	Δ_{MLP}	Δ_{GLNN}
Cora	80.64 ± 1.57	59.18 ± 1.60	80.26 ± 1.66	**83.04 ± 1.26**	↑ 2.98%	↑ 40.32%	↑ 3.46%
Citeseer	70.49 ± 1.53	58.50 ± 1.86	71.22 ± 1.50	**73.78 ± 1.54**	↑ 4.67%	↑ 26.12%	↑ 3.59%
Pubmed	75.56 ± 2.06	68.39 ± 3.09	75.59 ± 2.46	**77.34 ± 2.36**	↑ 2.36%	↑ 13.09%	↑ 2.32%
A-computer	82.82 ± 1.37	67.62 ± 2.21	82.71 ± 1.18	**84.04 ± 1.01**	↑ 1.47%	↑ 24.28%	↑ 1.61%
A-photo	90.85 ± 0.87	77.29 ± 1.79	91.95 ± 1.04	**93.36 ± 0.69**	↑ 2.76%	↑ 20.79%	↑ 1.53%
Arxiv	70.73 ± 0.35	55.67 ± 0.24	63.75 ± 0.48	**71.65 ± 0.29**	↑ 1.30%	↑ 28.70%	↑ 12.39%
Products	77.17 ± 0.32	60.02 ± 0.10	63.71 ± 0.31	**78.45 ± 0.38**	↑ 1.66%	↑ 30.71%	↑ 23.14%



inference time



teacher model

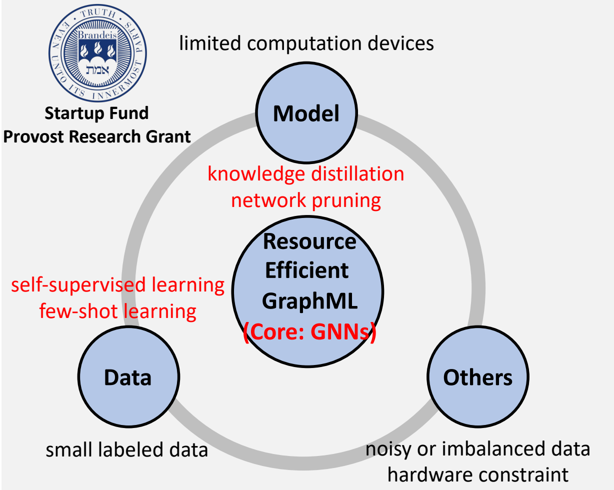
Learning MLPS On Graphs: A Unified View Of Effectiveness, Robustness, and Efficiency, arXiv, Under Review at ICLR-23

Introduction - Tackling Resource Constraint Challenges

- ❑ Data-efficient GRL
  - ○ Unlabeled Data Efficiency: Graph Self-supervised Learning
  - ○ Label Efficiency: Graph Few-shot Learning

- ❑ Model-efficient GRL
  - ○ Graph Knowledge Distillation
  - ○ GNN Pruning

- ❑ Data-Model Co-efficient GRL
  - ○ GNN Pruning + Graph Self-supervised Learning



Data-Model Co-Efficient GRL

❑ Data-Model Co-efficient GRL for Class-Imbalanced Graph Data

○ Massive data: more data, larger time cost

RQ: Can we find a subset of data to approximate the full data, so as to reduce the time cost

Theorem 1 *For a data selection algorithm, we assume the model is optimized via full gradient descent. At epoch t where $t \in [1, T]$, denote the model's parameters as $\theta^{(t)}$ where $\left\|\theta^{(t)}\right\|^2 \leqslant d^2$ and d is constant, the optimal model's parameters as θ^*, subset data as $\mathcal{D}_S^{(t)}$, and learning rate as α. Define gradient error $Err(\mathcal{D}_S^{(t)}, \mathcal{L}, \mathcal{L}_{train}, \theta^{(t)}) = \left\|\sum_{i \in \mathcal{D}_S^{(t)}} \nabla_\theta \mathcal{L}^i_{train}(\theta^{(t)}) - \nabla_\theta \mathcal{L}(\theta^{(t)})\right\|$, where $\mathcal{L}$ denotes training loss $\mathcal{L}_{train}$ over the full training data or validation loss $\mathcal{L}_{val}$ over the full validation data. $\mathcal{L}$ is a convex function. Then we have the following guarantee:*

If $\mathcal{L}_{train}$ is Lipschitz continuous with parameter σ_T and $\alpha = \frac{d}{\sigma_T\sqrt{T}}$, then $\min_{t=1:T} \mathcal{L}(\theta^{(t)}) - \mathcal{L}(\theta^*) \leqslant \frac{d\sigma_T}{\sqrt{T}} + \frac{d}{T}\sum_{t=1}^{T-1} Err(\mathcal{D}_S^{(t)}, \mathcal{L}, \mathcal{L}_{train}, \theta^{(t)})$.

→ reducing the distance between gradients of the subset and the full set: minimize the performance difference of the model trained with the subset and the optimal model

extract the most important samples based on gradient values

○ Class-imbalanced data: cannot be alleviated with standard contrastive learning

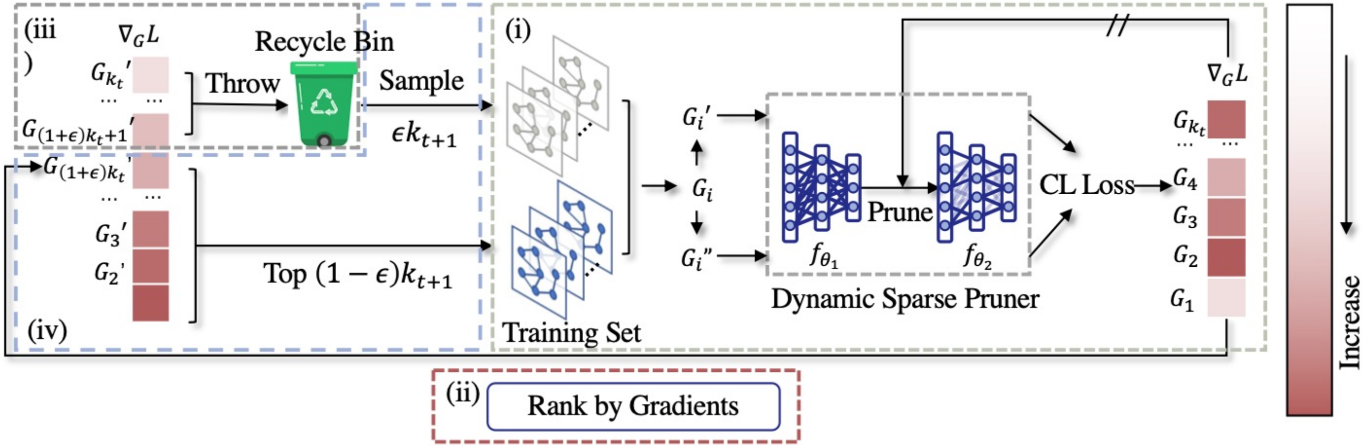
Fact: pruned network easily "forgets" the minority samples (DNN memorization)

Imbalanced contrastive learning: pruned vs. non-pruned - amplify the prediction differences between the pruned and non-pruned models, hence minority samples' weights will be implicitly increased in the overall loss (Jiang et al., Self-Damaging Contrastive Learning, ICML-21)

Diving Into Unified Data-Model Sparsity for Class-Imbalanced Graph Representation Learning, arXiv, Under Review at ICLR-23

Data-Model Co-Efficient GRL

❑ Our Model: A Unified Data-Model Dynamic Sparsity Framework (GraphDec)



- Imbalanced issue: the dynamic sparse graph contrastive learning model computes gradients for graph/node samples at each epoch

pruned vs. non-pruned branches for exploring minority samples information

- Massive data issue: extract informative subset samples according to gradients at each epoch, add some low-scored samples to recycle bin

Low-scored samples at current epoch still have the potential to be highly-scored in following epochs

Diving Into Unified Data-Model Sparsity for Class-Imbalanced Graph Representation Learning, arXiv, Under Review at ICLR-23

Data-Model Co-Efficient GRL

❑ GraphDec: Experiments

class-imbalanced graph classification

Rebalance Method	Basis	MUTAG (5:45) F1-ma.	MUTAG (5:45) F1-mi.	PROTEINS (30:270) F1-ma.	PROTEINS (30:270) F1-mi.	D&D (30:270) F1-ma.	D&D (30:270) F1-mi.	NCI1 (100:900) F1-ma.	NCI1 (100:900) F1-mi.	Sparsity (%) data	Sparsity (%) model
vanilla	GIN	52.50	56.77	25.33	28.50	9.99	11.88	18.24	18.94	100	100
	InfoGraph	69.11	69.68	35.91	36.81	21.41	27.68	33.09	34.03	100	100
	GraphCL	66.82	67.77	40.86	41.24	21.02	26.80	31.02	31.62	100	100
up-sampling	GIN	78.03	78.77	65.64	71.55	41.15	70.56	59.19	71.80	>100	100
	InfoGraph	78.62	79.09	62.68	66.02	41.55	71.34	53.38	62.20	>100	100
	GraphCL	80.06	80.45	64.21	65.76	38.96	64.23	49.92	58.29	>100	100
re-weight	GIN	77.00	77.68	54.54	55.77	28.49	40.79	36.84	39.19	100	100
	InfoGraph	80.85	81.68	65.73	69.60	41.92	72.43	53.05	62.45	100	100
	GraphCL	80.20	80.84	63.46	64.97	40.29	67.96	50.05	58.18	100	100
G^2GNN	remove edge	80.37	81.25	67.70	73.10	43.25	77.03	63.60	72.97	100	100
	mask node	83.01	83.59	67.39	73.30	43.93	**79.03**	64.78	74.91	100	100
GraphDec	dynamic sparsity	**85.71**	**85.71**	**76.92**	**76.89**	**77.97**	77.02	**76.30**	**76.29**	50	50

Rebalance Method	Basis	PTC-MR (9:81) F1-ma.	PTC-MR (9:81) F1-mi.	DHFR (12:108) F1-ma.	DHFR (12:108) F1-mi.	REDDIT-B (50:450) F1-ma.	REDDIT-B (50:450) F1-mi.	Avg. Rank F1-ma.	Avg. Rank F1-mi.	Sparsity (%) data	Sparsity (%) model
vanilla	GIN	17.74	20.30	35.96	49.46	33.19	36.02	12.00	12.00	100	100
	InfoGraph	25.85	26.71	50.62	56.28	57.67	67.10	11.00	11.14	100	100
	GraphCL	24.22	25.16	50.55	56.31	53.40	62.19	10.71	10.57	100	100
up-sampling	GIN	44.78	55.43	55.96	59.39	66.71	83.00	6.00	5.43	>100	100
	InfoGraph	44.29	48.91	59.49	61.62	67.01	78.68	6.00	6.00	>100	100
	GraphCL	45.12	53.50	60.29	61.71	62.01	75.84	6.29	6.43	>100	100
re-weight	GIN	36.96	43.09	55.16	57.78	45.17	51.92	9.86	9.86	100	100
	InfoGraph	44.09	49.17	58.67	60.24	65.79	77.35	5.43	5.29	100	100
	GraphCL	44.75	52.22	60.87	61.93	62.79	76.15	6.00	6.29	100	100
G^2GNN	remove edge	46.40	56.61	61.63	63.61	68.39	86.35	2.71	2.86	100	100
	mask node	46.61	56.70	59.72	61.27	67.52	85.43	2.71	2.71	100	100
GraphDec	dynamic sparsity	**54.03**	**61.17**	**64.25**	**67.91**	**69.70**	**87.00**	1.00	1.14	50	50

Diving Into Unified Data-Model Sparsity for Class-Imbalanced Graph Representation Learning, arXiv, Under Review at ICLR-23

Data-Model Co-Efficient GRL

❑ GraphDec: Experiments

class-imbalanced node classification

Method	Cora-LT			CiteSeer-LT			PubMed-LT			A.P. (ρ =82)		A.C. (ρ =244)		Sparsity (%)	
	Acc.	bAcc.	F1-ma.	Acc.	bAcc.	F1-ma.	Acc.	bAcc.	F1-ma.	(b)Acc.	F1-ma.	(b)Acc.	F1-ma.	data	model
vanilla	73.66	62.72	63.70	53.90	47.32	43.00	70.76	57.56	51.88	82.86	78.72	68.47	64.01	100	100
Re-Weight	75.20	68.79	69.27	62.56	55.80	53.74	77.44	72.80	73.66	92.94	92.95	90.04	90.11	100	100
Oversampling	77.44	70.73	72.40	62.78	56.01	53.99	76.70	68.49	69.50	92.46	92.47	89.79	89.85	>100	100
cRT	76.54	69.26	70.95	60.60	54.05	52.36	75.10	67.52	68.08	91.24	91.17	86.02	86.00	100	100
PC Softmax	76.42	71.30	71.24	65.70	**61.54**	61.49	76.92	75.82	74.19	93.32	93.32	86.59	86.62	100	100
DR-GCN	73.90	64.30	63.10	56.18	49.57	44.98	72.38	58.86	53.05	N/A	N/A	N/A	N/A	100	100
GraphSmote	76.76	69.31	70.21	62.58	55.94	54.09	75.98	70.96	71.85	92.65	92.61	89.31	89.39	>100	100
GraphENS	77.76	72.94	73.13	**66.92**	60.19	58.67	78.12	74.13	74.58	93.82	93.81	91.94	91.94	>100	100
GraphDec	**78.29**	**73.94**	**74.25**	**66.90**	**61.56**	**61.85**	**78.20**	**76.05**	**76.32**	**93.85**	**94.02**	**92.19**	**92.16**	50	50

training time (second)

Model	Method	PubMed-LT	Cora-LT	CiteSeer-LT	PROTEINS	PTC_MR	MUTAG
GCN	vanilla	2.436	2.154	2.129	12.798	4.295	2.989
	re-weight	2.330	2.282	2.150	12.903	4.410	3.125
	re(/over)-sample	3.241	2.860	2.794	15.996	5.734	4.022
	GraphCL	3.747	3.412	3.399	14.981	5.049	3.215
	GraphDec	2.243	1.995	1.952	10.614	4.212	2.090

Diving Into Unified Data-Model Sparsity For Class-Imbalanced Graph Representation Learning, arXiv, Under Review at ICLR-23

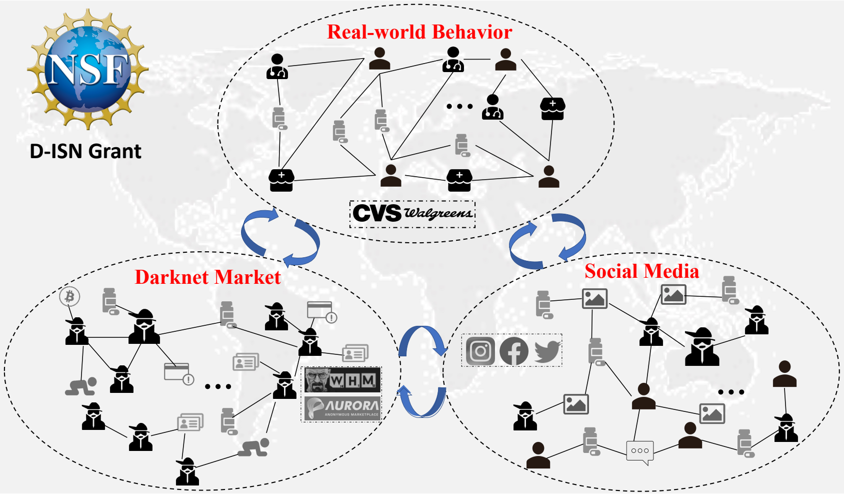
Resource-efficient Graph Representation Learning

- ❑ Introduction
- ❑ Resource-efficient GRL
- ❑ Social Good Application
- ❑ Conclusion

Deep Graph Learning for Social Good

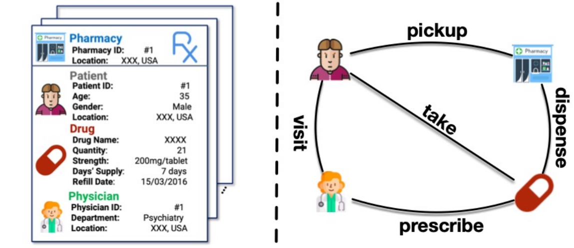
❑ Public Health: Combating the Opioid Crisis

the opioid crisis has co-evolved with modern systems



catastrophic consequences from public health to violent crimes in US

- Task-1: Opioid drug over-prescribing or abuse prediction.

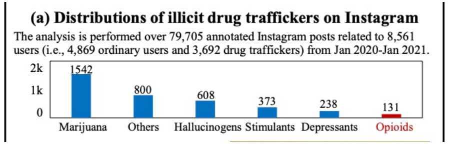


(a) Prescriptions (b) P&D Graph Schema

RxNet (CIKM-21)
(The Best Paper Award)

D^{2}HGNN (KDD-22)

- Task-2: illicit online drug trafficker detection.

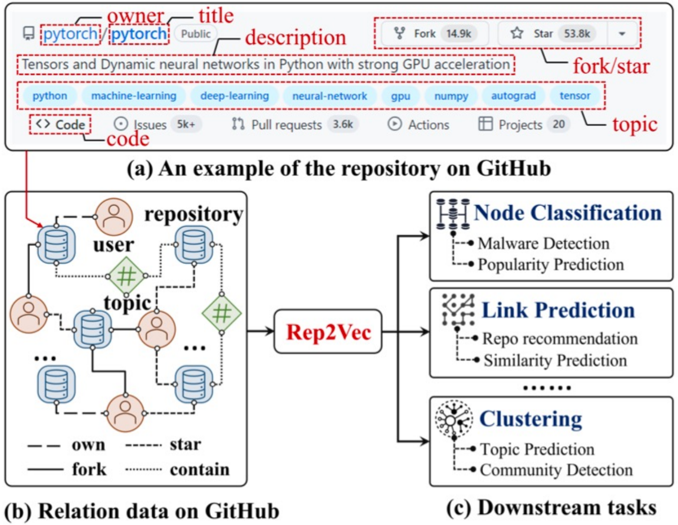


MetaHG (NeurIPS-21)

Deep Graph Learning for Social Good

❑ Cybersecurity: Enhancing Social and Information System Resilience

- Task-1: Software/repository-based applications
- Task-2: Anomaly detection



(a) An example of the repository on GitHub

(b) Relation data on GitHub

(c) Downstream tasks

Meta-AHIN (IJCAI-21)

Repo2Vec (KDD-22)

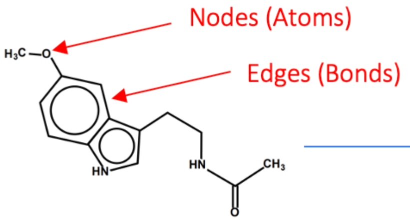
GraphBERT (ICDM-22)

CLA-HG (CIKM-22)

Deep Graph Learning for Social Good

❑ Interdisciplinary Applications

- Task-1: Molecular Analysis

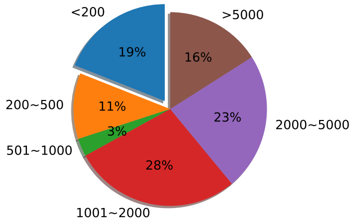


0 / 1 ?

Solubility

Toxicity

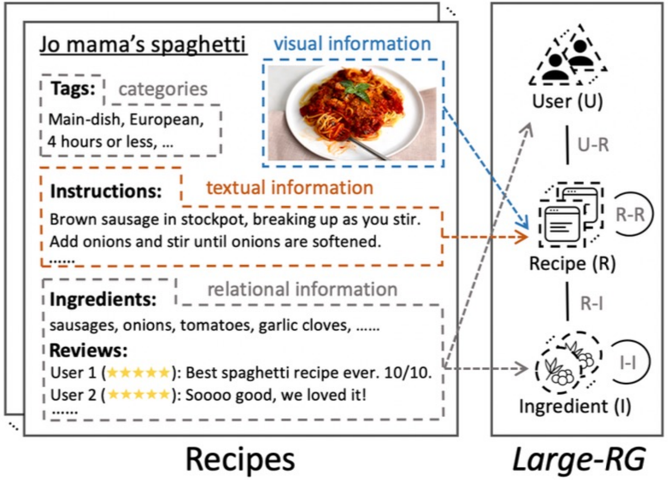
Activity



MGNN (WWW-21)

GraphSeq (CIKM-20)

- Task-2: Food Service



RecipeRec (IJCAI-22)

Recipe2Vec (IJCAI-22)

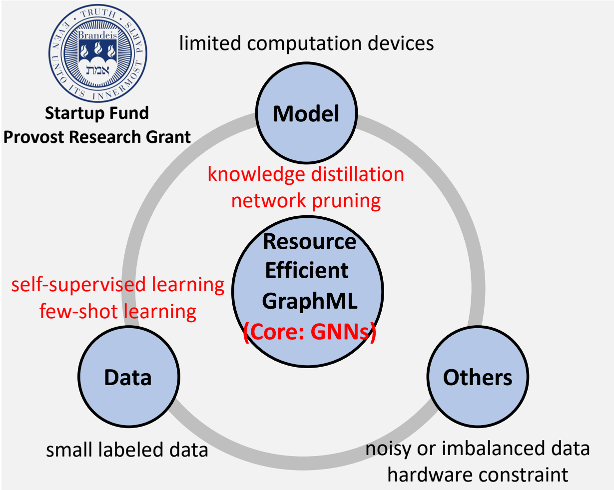
rn2vec (CIKM-21)

Resource-efficient Graph Representation Learning

- ❑ Introduction
- ❑ Resource-efficient GRL
- ❑ Social Good Application
- ❑ Conclusion

Conclusion and Future Work

- Tackling Resource Constraint Challenges of GRL
  - Data Efficiency: Graph self-supervised learning, Graph few-shot learning
  - Model Efficiency: Graph knowledge distillation, GNN pruning
  - Data-Model Co-efficiency: Data and model sparsity



- Future Direction A: Graph Models for RE-GRL
  - Heterogeneous graphs, Multiplex graphs
  - Graphs with different properties: heterophily graphs

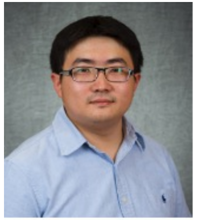
- Future Direction B: Explainability of RE-GRL Models
  - Improve model reliability and end-user trust
  - Information-based method to quantify the information loss

- Future Direction C: Broader Applications (Social Good)

Acknowledgement

- ❑ Faculty


Fanny Ye


Jundong Li


Xiangliang Zhang


Huan Liu


Nitesh Chawla

- ❑ Ph.D. Student


Chunhui Zhang


Qianlong Wen


Yiyue Qian


Mingxuan Ju


Yijun Tian


Zhichun Guo

- ❑ Funding

NSF

Brandeis UNIVERSITY

Thank you!

Q & A