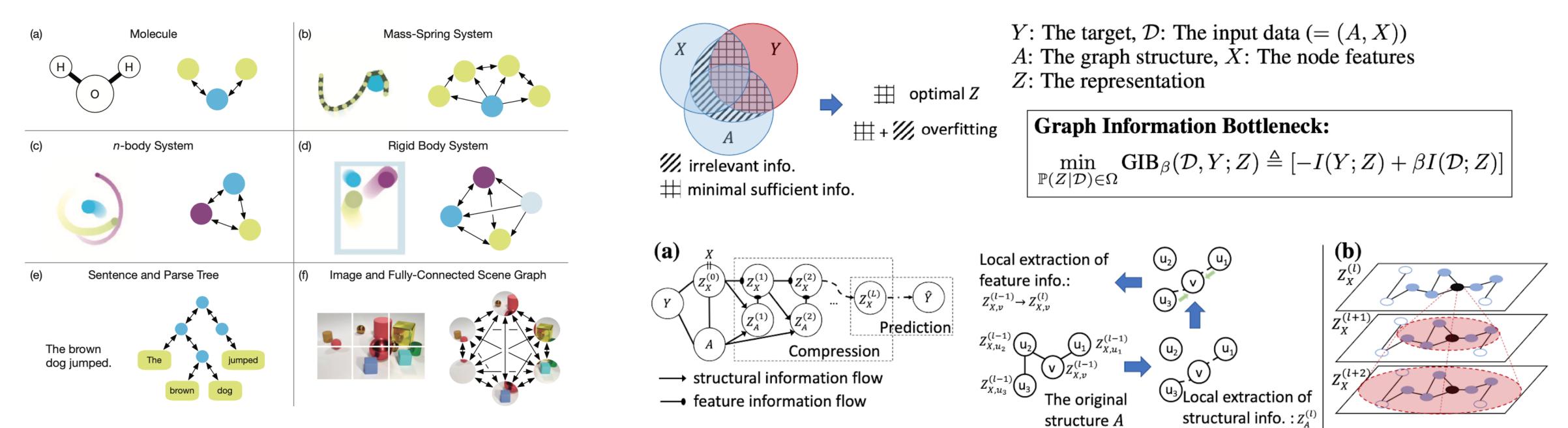
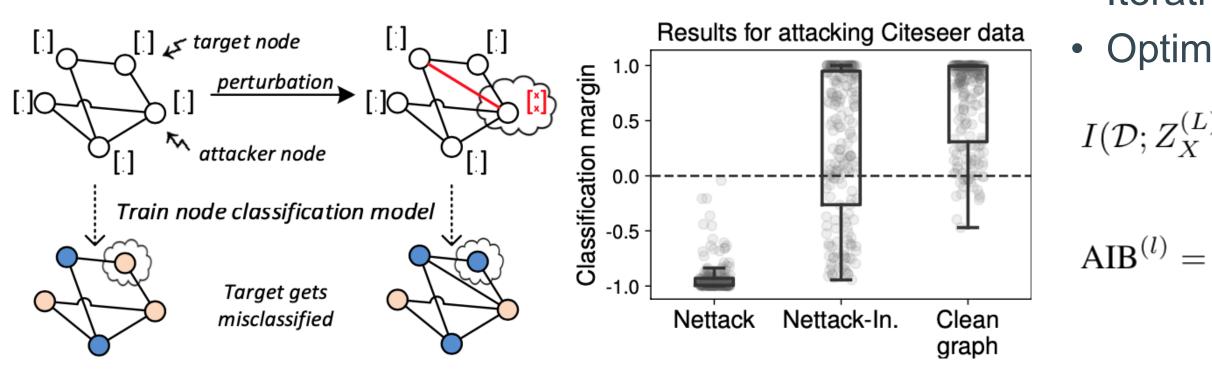
## **Graph Information Bottleneck** Tailin Wu\*, Hongyu Ren\*, Pan Li, Jure Leskovec Stanford University, Purdue University **Representation Learning on Graphs** Graph Information Bottleneck

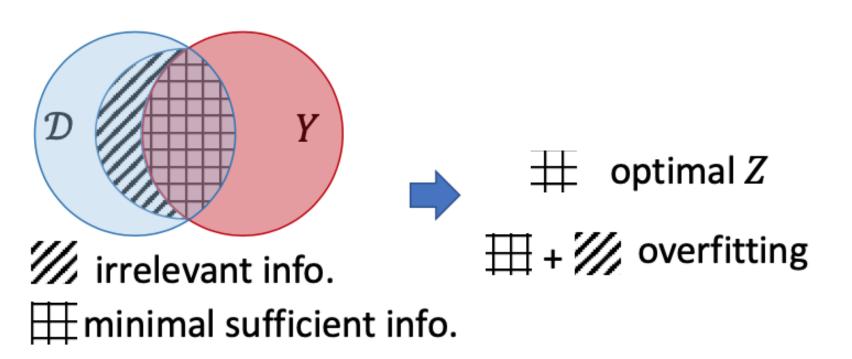


However, they are susceptible to adversarial attacks on node features or graph structure.



Optimal representation based on the information bottleneck principle:

- Maximally informative of the prediction Y.
- Leverage minimally sufficient information from input graph D (node features X, graph structure A)







We design operators for GNN

• Iteratively compress the information of X and A.

Optimize variational bounds of structure IB and feature IB.

$$\begin{split} & (\mathcal{D}) \leq I(\mathcal{D}; \{Z_X^{(l)}\}_{l \in S_X} \cup \{Z_A^{(l)}\}_{l \in S_A}) \leq \sum_{l \in S_A} \operatorname{AIB}^{(l)} + \sum_{l \in S_X} \operatorname{XIB}^{(l)}, \text{where} \\ & = \mathbb{E}\left[\log \frac{\mathbb{P}(Z_A^{(l)} | A, Z_X^{(l-1)})}{\mathbb{Q}(Z_A^{(l)})}\right], \operatorname{XIB}^{(l)} = \mathbb{E}\left[\log \frac{\mathbb{P}(Z_X^{(l)} | Z_X^{(l-1)}, Z_A^{(l)})}{\mathbb{Q}(Z_X^{(l)})}\right], \end{split}$$

Apply GIB to GAT

• Model  $P(Z_A^{(l)}|A, Z_X^{(l-1)})$  as Bernoulli distribution: GIB-Bern. • Model  $P(Z_A^{(l)}|A, Z_X^{(l-1)})$  as categorical distribution: GIB-Cat. • Model  $P(Z_X^{(l)}|Z_X^{(l-1)}, Z_A^{(l)})$  as Gaussian distribution

Algorithm 2: NeighborSample (categorical)	Algorithm 3: NeighborSa
<b>Input:</b> $Z_X^l, \mathcal{T}, V_{vt}, a$ , as defined in Alg. 1;	<b>Input:</b> $Z_X^l, \mathcal{T}, V_{vt}, a$ , as de
<b>Output:</b> $Z_{A,v}^{(l+1)}$	Output: $Z_{A,v}^{(l+1)}$ 1.For $t \in [\mathcal{T}]$ , do:
1.For $t \in [\mathcal{T}]$ , do:	1.For $t \in [\mathcal{T}]$ , do:
2. $ \downarrow \phi_{vt}^{(l)} \leftarrow \operatorname{softmax}(\{(\tilde{Z}_{X,v}^{(l-1)} \oplus \tilde{Z}_{X,u}^{(l-1)})a^T\}_{u \in V_{vt}}) $	2. $ \lfloor \phi_{vt}^{(l)} \leftarrow \text{sigmoid}(\{(\tilde{Z}$
3. $Z_{A,v}^{(l+1)} \leftarrow \bigcup_{t=1}^{\mathcal{T}} \{ u \in V_{vt}   u \stackrel{\text{iid}}{\sim} \operatorname{Cat}(\phi_{vt}^{(l)}), k \text{ times} \}$	3. $Z_{A,v}^{(l+1)} \leftarrow \bigcup_{t=1}^{\mathcal{T}} \{ u \in V_{vt} \}$





## Experiments

## Adversarial attack

	Model	Clean (%)	Evasive (%)			
	WIOUEI		1	2	3	4
Cora	GCN	<b>80.0</b> ±7.87	$51.5 \pm 4.87$	$38.0 \pm 6.22$	$31.0 \pm 2.24$	$26.0 \pm 3.79$
	GCNJaccard	$75.0 \pm 5.00$	$48.5 \pm 6.75$	$36.0 \pm 6.51$	$32.0 \pm 3.25$	$30.0 \pm 3.95$
	RGCN	<b>80.0</b> ±4.67	$49.5 \pm 6.47$	$36.0 \pm 5.18$	$30.5{\scriptstyle\pm3.25}$	$25.5 \pm 2.09$
	GAT	$77.8 \pm 3.97$	48.0±8.73	$39.5 \pm 5.70$	$36.5{\scriptstyle\pm 5.48}$	$32.5{\pm}5.30$
	GIB-Cat	$77.6 \pm 2.84$	<b>63.0</b> ±4.81	<b>52.5</b> ±3.54	<b>44.5</b> ±5.70	<b>36.5</b> ±6.75
	<b>GIB-Bern</b>	$78.4 \pm 4.07$	<b>64.0</b> ±5.18	<b>51.5</b> ±4.54	<b>43.0</b> ±3.26	<b>37.5</b> ±3.95
Pubmed	GCN	82.6±6.98	$39.5 \pm 4.81$	32.0±4.81	31.0±5.76	31.0±5.76
	GCNJaccard	82.0±7.15	$37.5 \pm 5.30$	$31.5 \pm 5.18$	30.0±3.95	30.0±3.95
	RGCN	$79.0 \pm 5.18$	$39.5 \pm 5.70$	$33.0 \pm 4.80$	$31.5 \pm 4.18$	$30.0 \pm 5.00$
	GAT	$78.6 \pm 6.70$	$41.0 \pm 8.40$	$33.5{\pm}4.18$	$30.5 {\pm} 4.47$	$31.0 \pm 4.18$
	GIB-Cat	<b>85.1</b> ±6.90	<b>72.0</b> ±3.26	<b>51.0</b> ±5.18	<b>37.5</b> ±5.30	<b>31.5</b> ±4.18
	<b>GIB-Bern</b>	<b>86.2</b> ±6.54	<b>76.0</b> ±3.79	<b>50.5</b> ±4.11	$\textbf{37.5}{\scriptstyle \pm 3.06}$	<b>31.5</b> ±1.37
	GCN	$71.8 \pm 6.94$	$42.5 \pm 7.07$	$27.5 \pm 6.37$	$18.0 \pm 3.26$	$15.0 \pm 2.50$
Citeseer	GCNJaccard	72.5±9.35	41.0±6.75	$32.5 \pm 3.95$	$20.5 \pm 3.70$	$13.0 \pm 1.11$
	RGCN	<b>73.5</b> ±8.40	$41.5 \pm 7.42$	$24.5 \pm 6.47$	$18.5 {\pm} 6.52$	$13.0 \pm 1.11$
	GAT	$72.3 \pm 8.38$	<b>49.0</b> ±9.12	33.0±5.97	$22.0 {\pm} 4.81$	$18.0 \pm 3.26$
	GIB-Cat	68.6±4.90	<b>51.0</b> ±4.54	<b>39.0</b> ±4.18	<b>32.0</b> ±4.81	<b>26.5</b> ±4.54
	<b>GIB-Bern</b>	$71.8 \pm 5.03$	<b>49.0</b> ±7.42	<b>37.5</b> ±7.71	$32.5{\scriptstyle\pm4.68}$	<b>23.5</b> ±7.42

## Feature attack

Dataset	Model	<b>Feature noise ratio</b> $(\lambda)$			
Dataset		0.5	1	1.5	
Cora	GCN	$64.0 \pm 2.05$	$41.3 \pm 2.05$	$31.4 \pm 2.81$	
	GCNJaccard	61.1±2.18	$41.2 \pm 2.28$	$31.8 \pm 2.63$	
	RGCN	57.7±2.27	$39.1 \pm 1.58$	$29.6 \pm 2.47$	
Cora	GAT	62.5±1.97	$41.7 \pm 2.32$	$29.8{\scriptstyle\pm2.98}$	
	AIB-Cat	$67.9 \pm 2.65$	<b>49.6</b> ±5.35	<b>38.4</b> ±5.06	
	AIB-Bern	<b>68.8</b> ±1.85	49.0±2.87	37.1±4.47	
	GIB-Cat	67.1±2.21	49.1±3.67	$37.5 \pm 4.76$	
	<b>GIB-Bern</b>	<b>69.0</b> ±1.91	<b>51.3</b> ±2.62	<b>38.9</b> ±3.38	
	GCN	$61.3 \pm 1.52$	$50.2 \pm 2.08$	44.3±1.43	
	GCNJaccard	$62.7 \pm 1.25$	51.9±1.53	$45.1 \pm 2.04$	
Pubmed	RGCN	58.4±1.74	$49.0 \pm 1.65$	43.9±1.29	
rubilleu	GAT	$62.7 \pm 1.68$	$50.2 \pm 2.35$	43.7±2.43	
	AIB-Cat	$64.5 \pm 2.13$	50.9±3.83	43.0±3.73	
	AIB-Bern	61.1±2.70	$47.8 \pm 3.65$	42.0±4.21	
	GIB-Cat	<b>67.1</b> ±4.33	<b>57.2</b> ±5.27	<b>51.5</b> ±4.84	
	<b>GIB-Bern</b>	<b>64.9</b> ±2.52	<b>54.7</b> ±1.83	<b>48.2</b> ±2.10	

ample (Bernoulli) lefined in Alg. 1;

 $\tilde{Z}_{X,v}^{(l-1)} \oplus \tilde{Z}_{X,u}^{(l-1)} a^T \}_{u \in V_{vt}})$  $\int_{vt} |u \stackrel{\text{iid}}{\sim} \text{Bernoulli}(\phi_{ut}^{(l)})\}$ 

