Demystify the Degree-related Bias in Recommender System

Yu Wang

Yuying Zhao

Yi Zhang

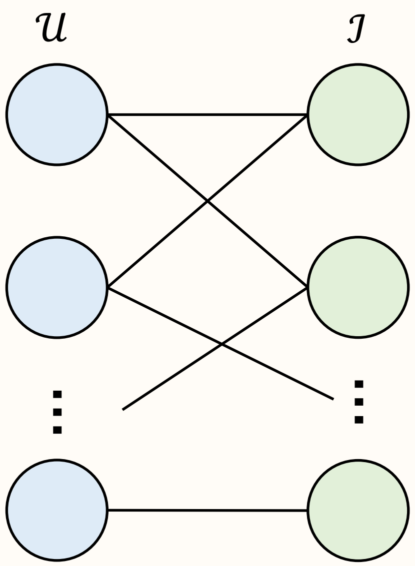
Tyler Derr

Network and Data Science Lab
Vanderbilt University

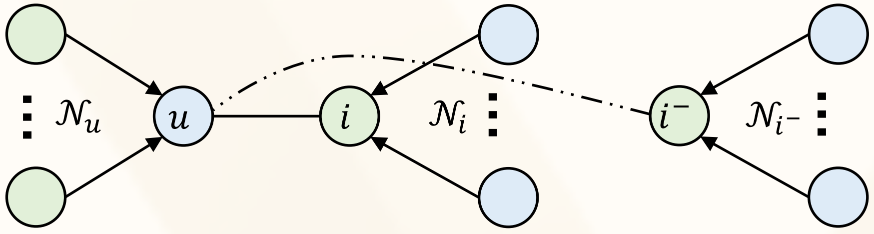
Background – Message-passing in graph-based method for recommendation

○ User node $\mathcal{N}_u$ Neighborhood set of u ⟶ Message passing

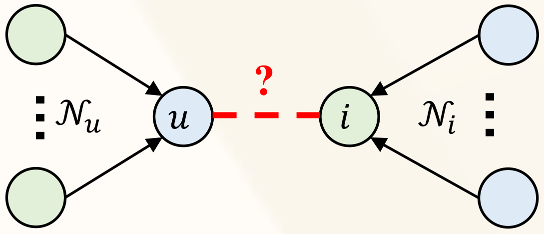
○ Item node i^- Negative sample of u



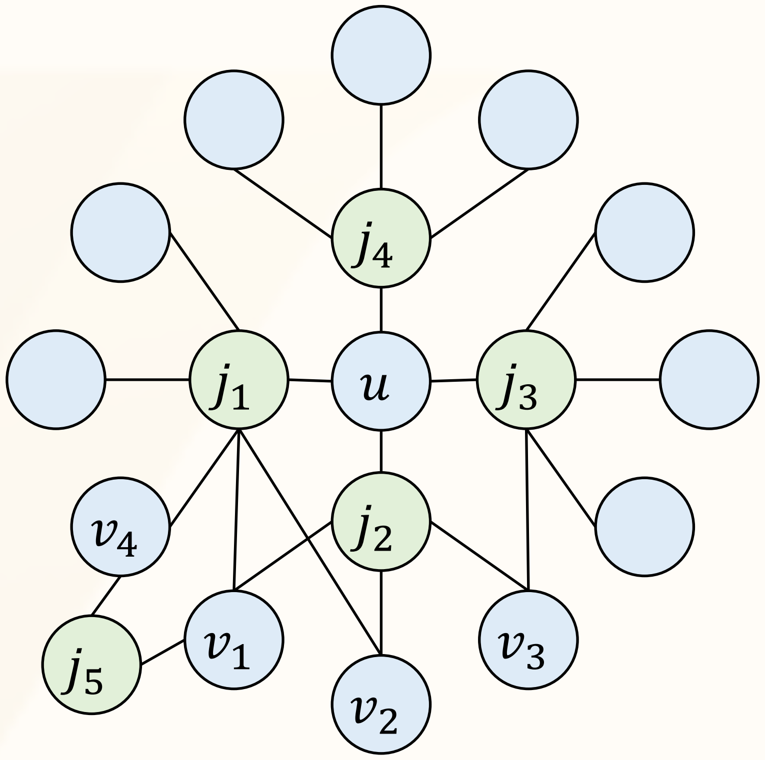
(a) User-item interacted bipartite graph



(b) Training stage

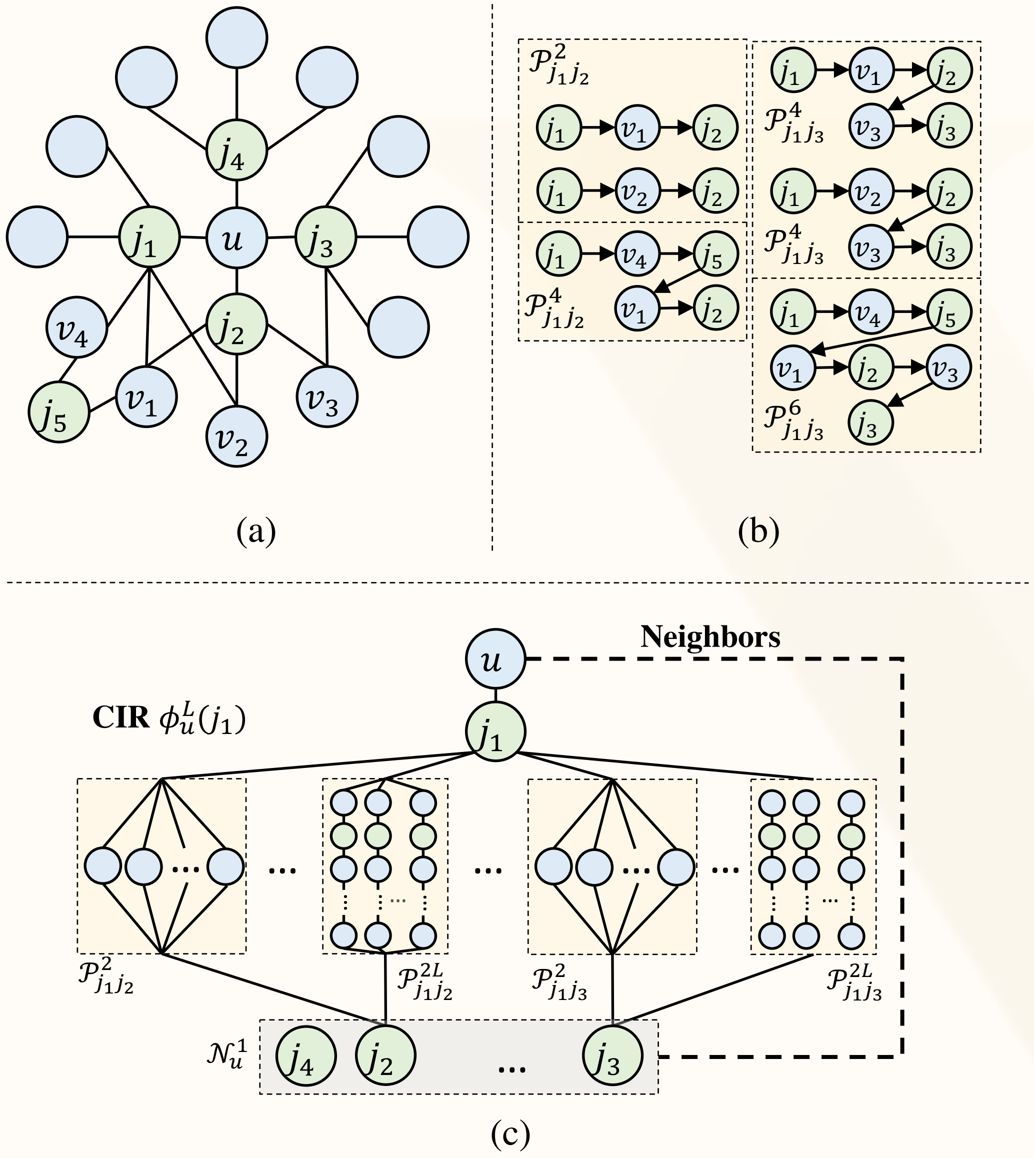


(c) Inference stage

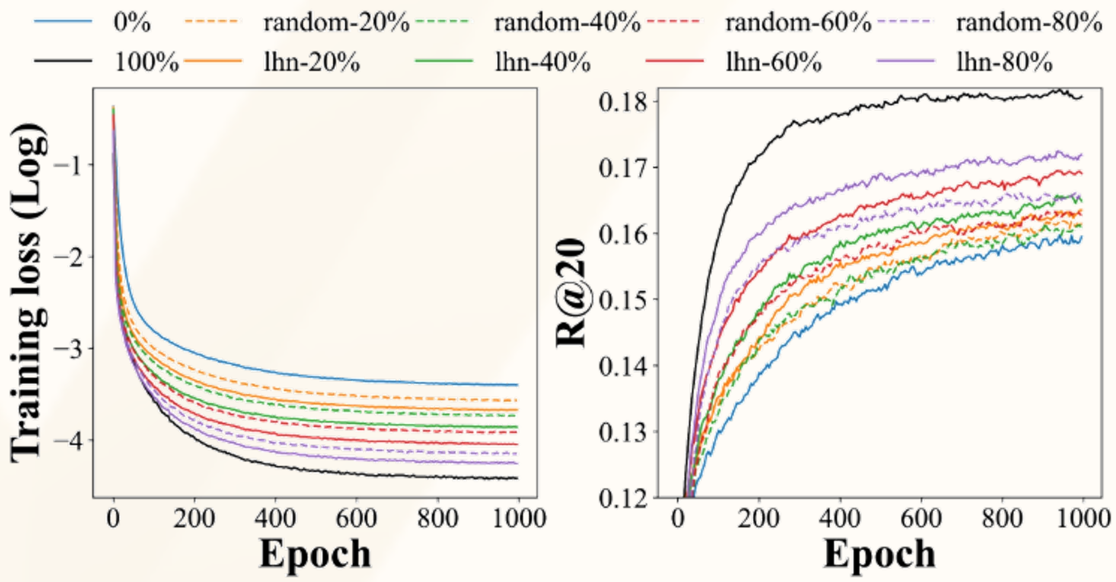


j_1 is more connected to u's neighborhood than j_4

Method – Common Interacted Ratio (CIR)



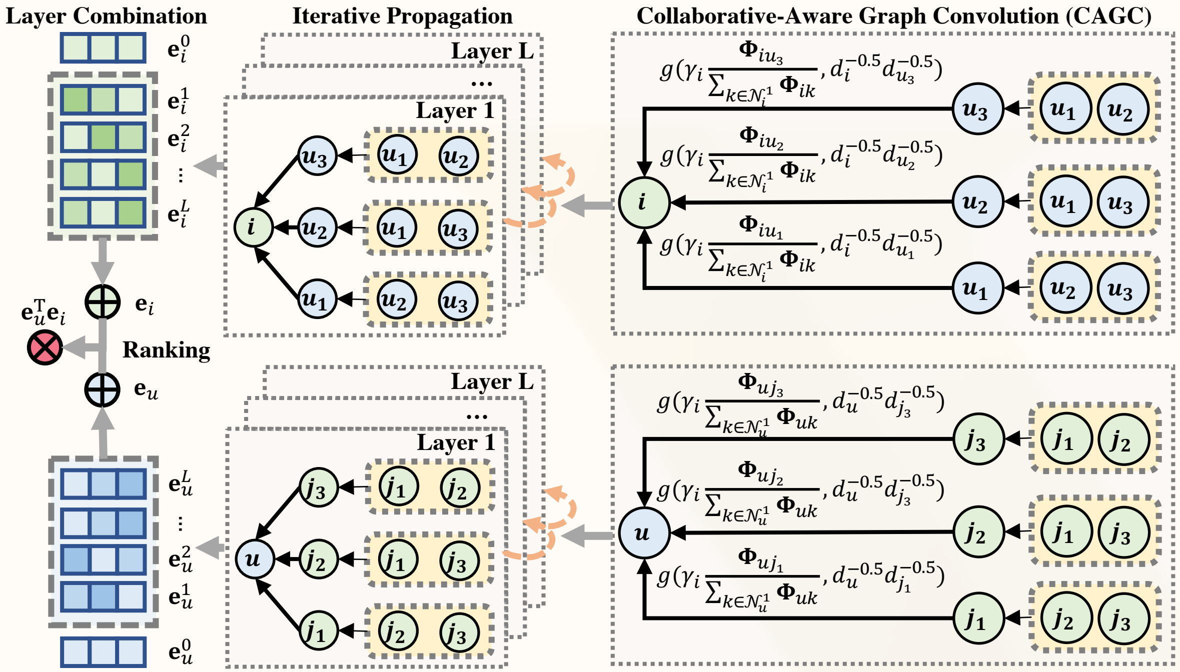
$$\phi_u^L(j) = \frac{1}{|\mathcal{N}_u^1|}\sum_{i\in\mathcal{N}_u^1}\sum_{l=1}^{L}\beta^{2l}\sum_{P_{ji}^{2l}\in\mathscr{P}_{ji}^{2l}}\frac{1}{f(\{\mathcal{N}_k^1|k\in P_{ji}^{2l}\})}, \quad \forall j\in\mathcal{N}_u^1, \forall u\in\mathcal{U}$$

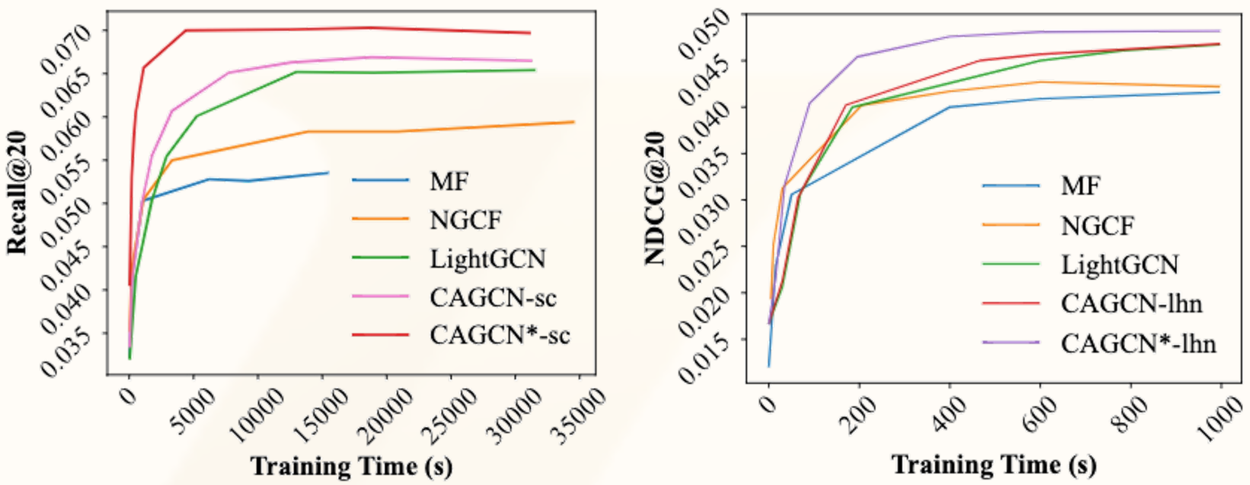


(a) Gowalla-Loss (b) Gowalla-Recall

Leveraging collaborations from u's neighboring node j with higher CIR would cause more benefits to u's ranking

Model – Collaboration-aware GNN (CAGCN)





(a) Yelp (b) Loseit

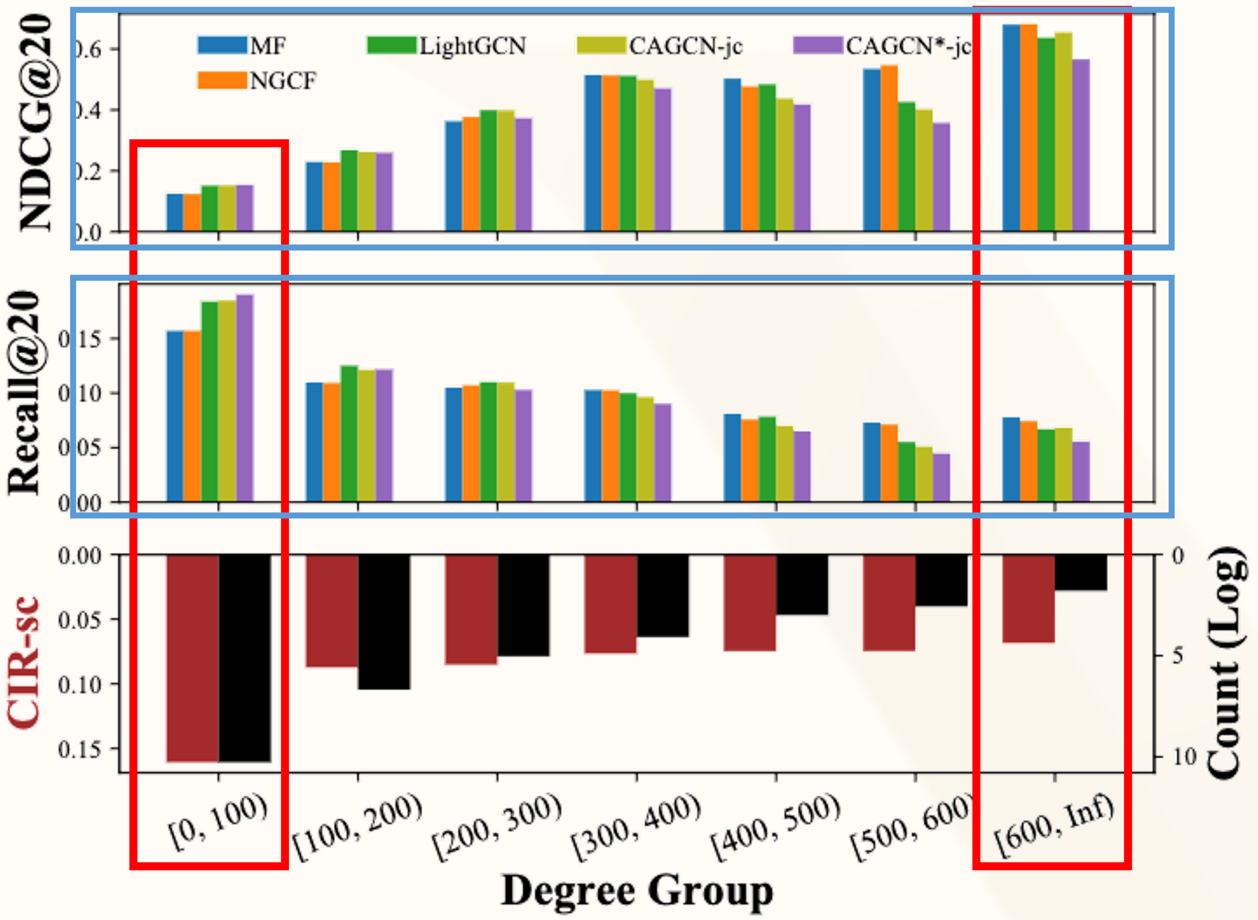
$$\Phi_{ij} = \begin{cases} \phi_i(j), & \text{if } \mathbf{A}_{ij} > 0 \\ 0, & \text{if } \mathbf{A}_{ij} = 0 \end{cases}, \forall i, j \in \mathcal{V}$$

$$\mathbf{e}_i^{l+1} = \sum_{j \in \mathcal{N}_i^1} g\left(\gamma_i \frac{\Phi_{ij}}{\sum_{k \in \mathcal{N}_i^1} \Phi_{ik}}, d_i^{-0.5} d_j^{-0.5}\right) \mathbf{e}_j^l, \forall i \in \mathcal{V}$$

Table 4: Efficiency comparison of CAGCN* with LightGCN.

Model	Stage	Gowalla	Yelp	Amazon	Ml-1M	Loseit	News
LightGCN	Training	16432.0	28788.0	81976.5	18872.3	39031.0	13860.8
CAGCN*	Preprocess	167.4	281.6	1035.8	33.8	31.4	169.0
	Training	2963.2	1904.4	1983.9	11304.7	10417.7	1088.4
	Total	3130.6	2186.0	3019.7	11338.5	10449.1	1157.4
Improve	Training	82.0%	93.4%	97.6%	40.1%	73.3%	92.1%
	Total	80.9%	92.4%	96.3%	39.9%	73.2%	91.6%

Analysis – Performance differs per degree and metric



(1) CAGCN performs better on low-degree users while worse on high-degree users, which also aligns with CIR

(2) Different trend when using NDCG and Recall

Which one really performs well? High-degree or low-degree nodes?

Analysis – Bias of different evaluation metrics

$$\mathrm{R@}K_i = \frac{|\hat{\mathcal{N}}_i^1 \cap \widetilde{\mathcal{N}}_i^1|}{|\hat{\mathcal{N}}_i^1|} \quad \longrightarrow \quad E(\mathrm{R@K}|d) = \frac{K}{n}, \quad \frac{\partial E(\mathrm{R@K}|d)}{\partial d} = 0,$$

$$\mathrm{P@}K_i = \frac{|\hat{\mathcal{N}}_i^1 \cap \widetilde{\mathcal{N}}_i^1|}{K} \quad \longrightarrow \quad E(\mathrm{P@K}|d) = \frac{d}{n}, \quad \frac{\partial E(\mathrm{P@K}|d)}{\partial d} = 1,$$

$$\mathrm{F1@}K_i = 2\frac{\mathrm{P@}K \cdot \mathrm{R@}K}{\mathrm{R@}K + \mathrm{P@}K} = \frac{2|\hat{\mathcal{N}}_i^1 \cap \widetilde{\mathcal{N}}_i^1|}{K + |\hat{\mathcal{N}}_i^1|} \quad \longrightarrow \quad E(\mathrm{F1@K}|d) = \frac{2K}{n}\frac{d}{K+d}, \quad \frac{\partial E(\mathrm{F1@K}|d)}{\partial d} = \frac{2K^2}{n}\frac{1}{(K+d)^2}$$

$$\mathrm{N@}K_i = \frac{\sum_{k=1}^{K} \frac{1[v_{\phi_i^k} \in (\hat{\mathcal{N}}_i^1 \cap \widetilde{\mathcal{N}}_i^1)]}{\log_2(k+1)}}{\sum_{k=1}^{K} \frac{1}{\log_2(k+1)}} \quad \longrightarrow \quad E(\mathrm{N@K}|d) = \frac{d}{n}, \quad \frac{\partial E(\mathrm{N@K}|d)}{\partial d} = 1.$$

$\widehat{\mathcal{N}}_i^1 \cap \widetilde{\mathcal{N}}_i^1$ follows hyper-geometric distribution if $\widetilde{\mathcal{N}}_i^1$ is given by an unbiased recommender system

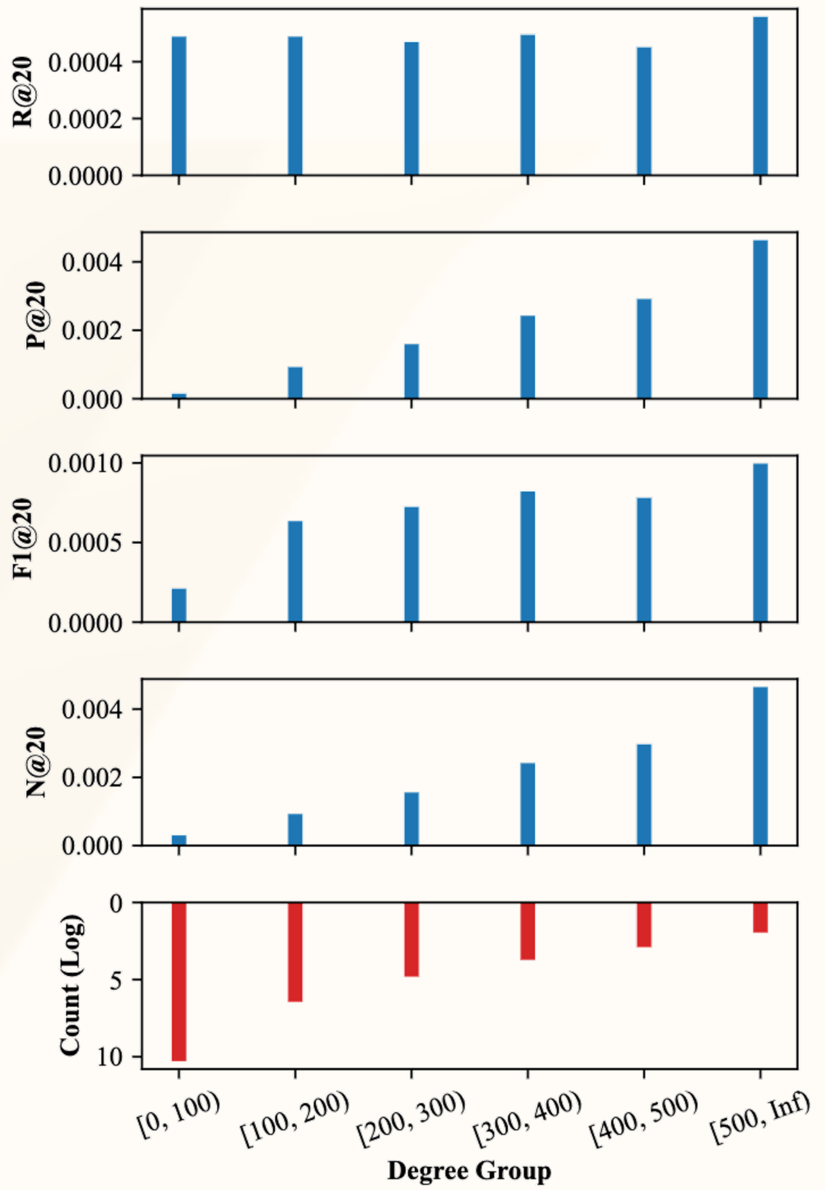
Analysis – Bias of different evaluation metrics

$$E(\mathrm{R@K}|d) = \frac{K}{n}, \qquad \frac{\partial E(\mathrm{R@K}|d)}{\partial d} = 0,$$

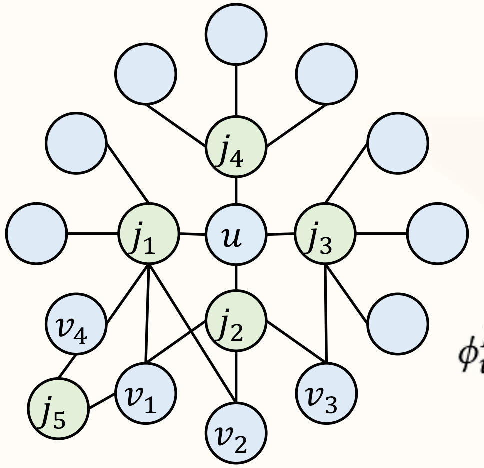
$$E(\mathrm{P@K}|d) = \frac{d}{n}, \qquad \frac{\partial E(\mathrm{P@K}|d)}{\partial d} = 1,$$

$$E(\mathrm{F1@K}|d) = \frac{2K}{n}\frac{d}{K+d}, \quad \frac{\partial E(\mathrm{F1@K}|d)}{\partial d} = \frac{2K^2}{n}\frac{1}{(K+d)^2}$$

$$E(\mathrm{N@K}|d) = \frac{d}{n}, \qquad \frac{\partial E(\mathrm{N@K}|d)}{\partial d} = 1.$$



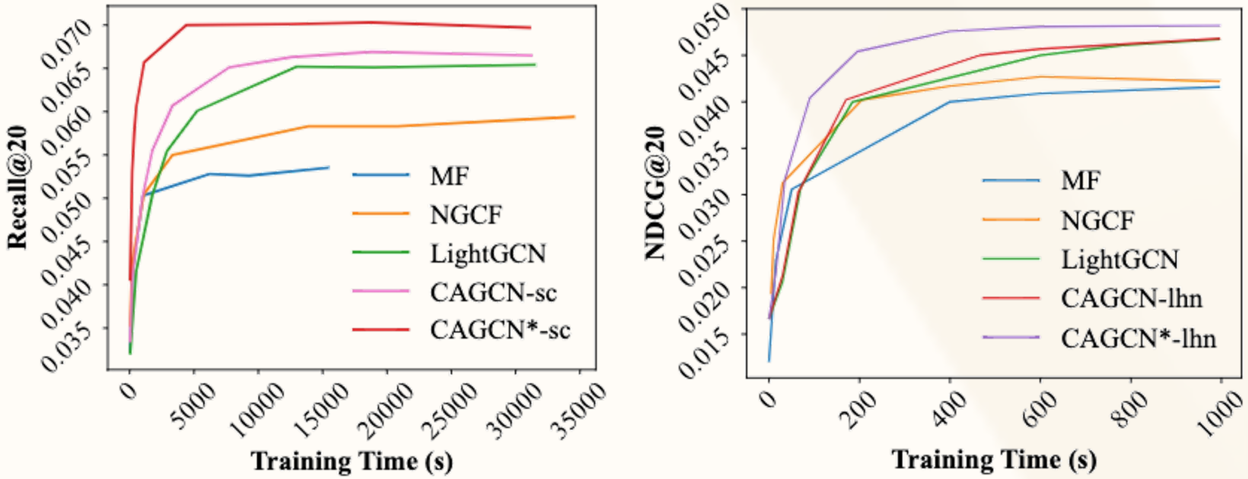
Conclusion



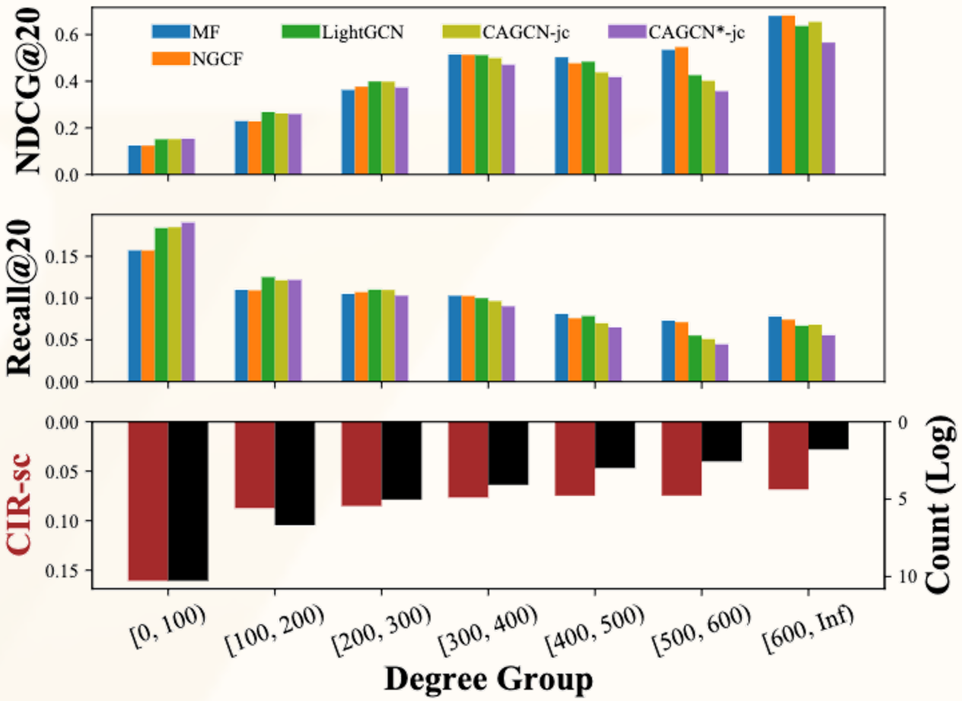
CIR-based message-passing

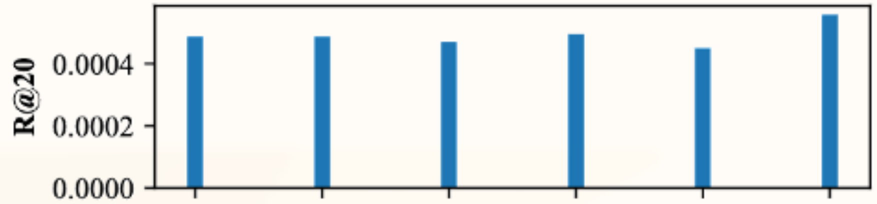
$$\phi_u^L(j) = \frac{1}{|\mathcal{N}_u^1|} \sum_{i \in \mathcal{N}_u^1} \sum_{l=1}^{L} \beta^{2l} \sum_{P_{ji}^{2l} \in \mathscr{P}_{ji}^{2l}} \frac{1}{f(\{\mathcal{N}_k^1 | k \in P_{ji}^{2l}\})},$$

CIR-based GNNs



(a) Yelp (b) Loseit





Recall is an unbiased estimator