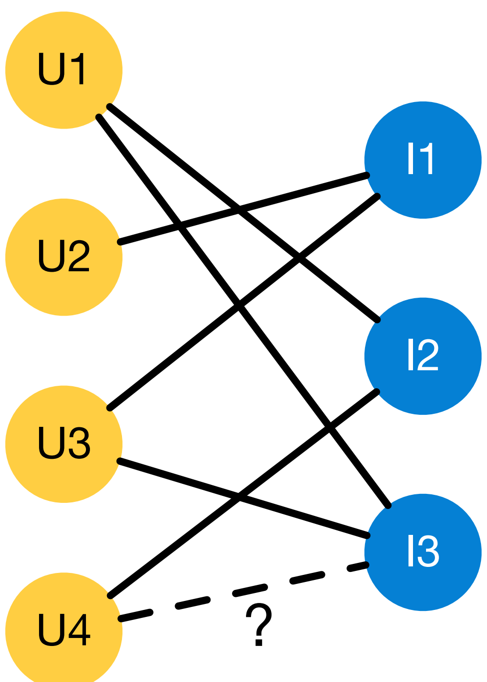
EFLEC: Efficient Feature-LEakage Correction in GNN based Recommendation Systems

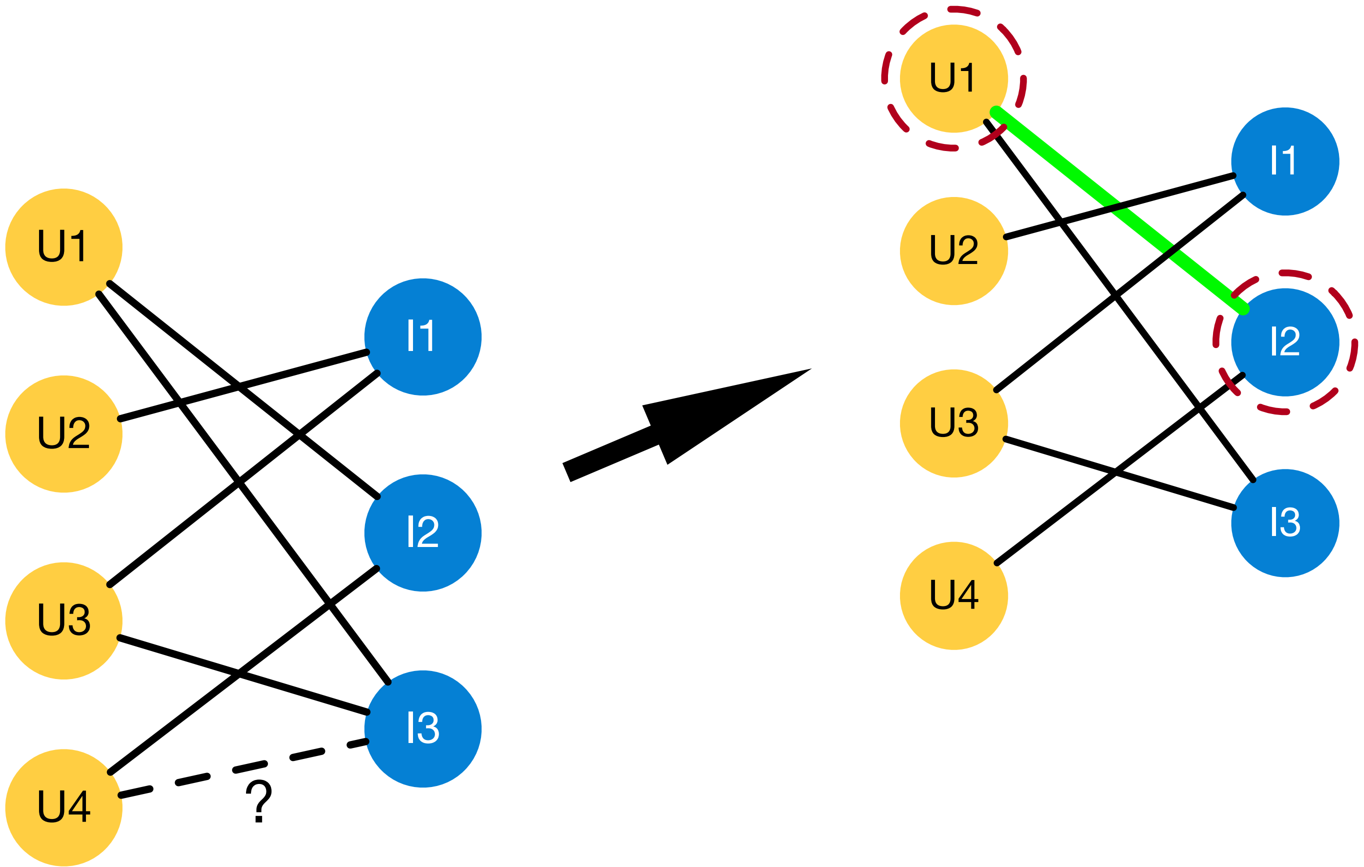
Isshan Kumar, Yaochen Hu (Presenter), Yingxue Zhang
Huawei Noah's Ark Lab

July, 2022

Feature Leakage Problem in GNN based Recommender Systems



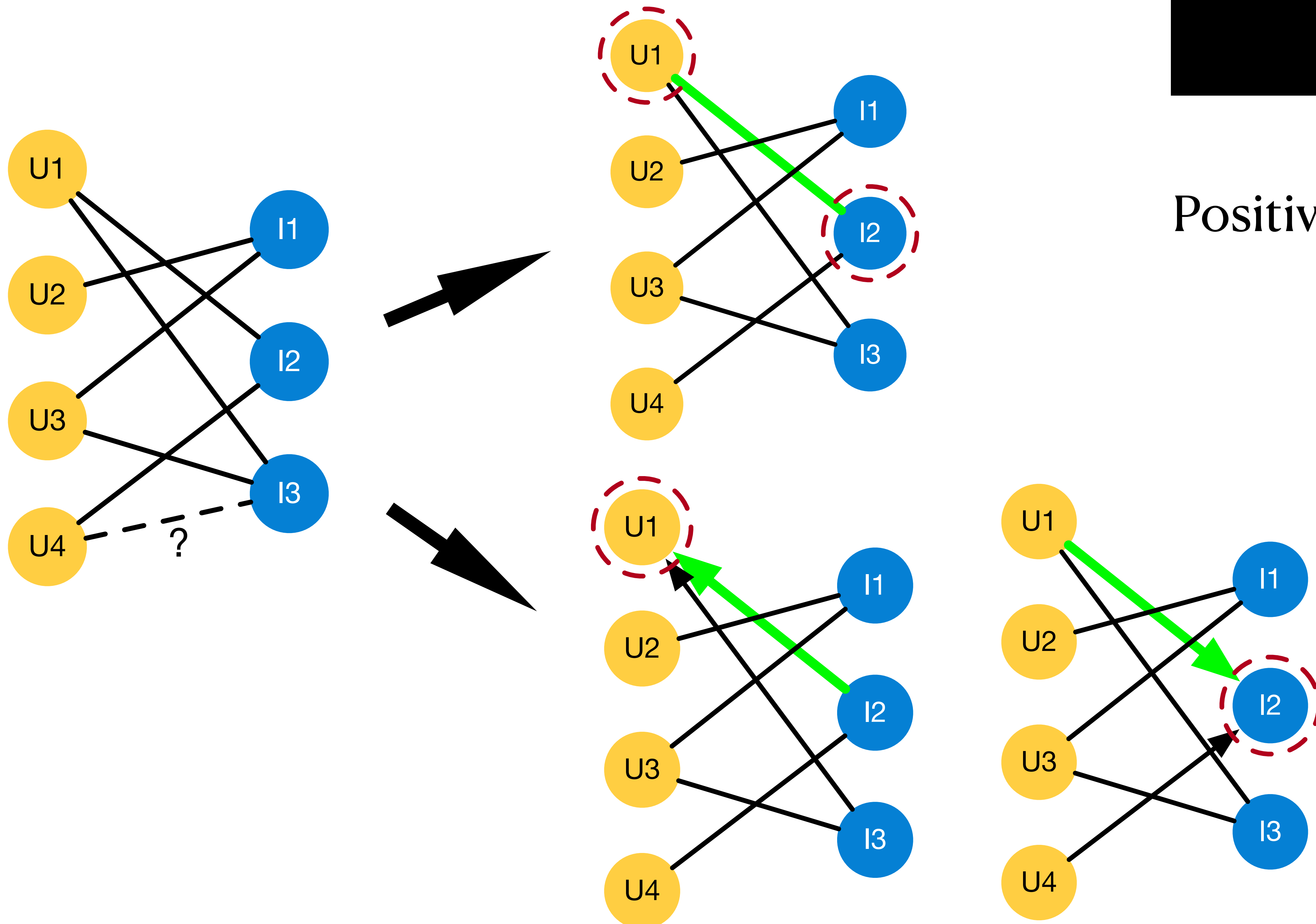
Feature Leakage Problem in GNN based Recommender Systems



Positive training label

[redacted]

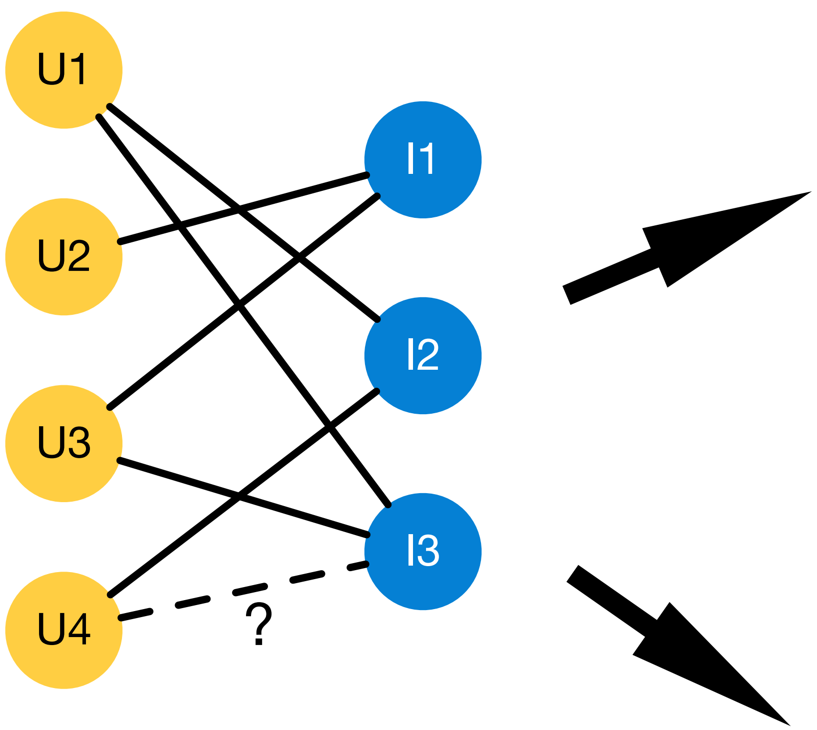
Feature Leakage Problem in GNN based Recommender Systems



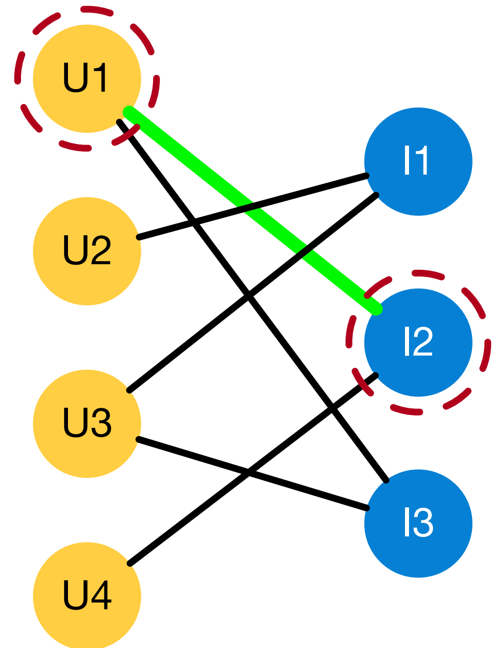
Positive training label

Signal for message passing

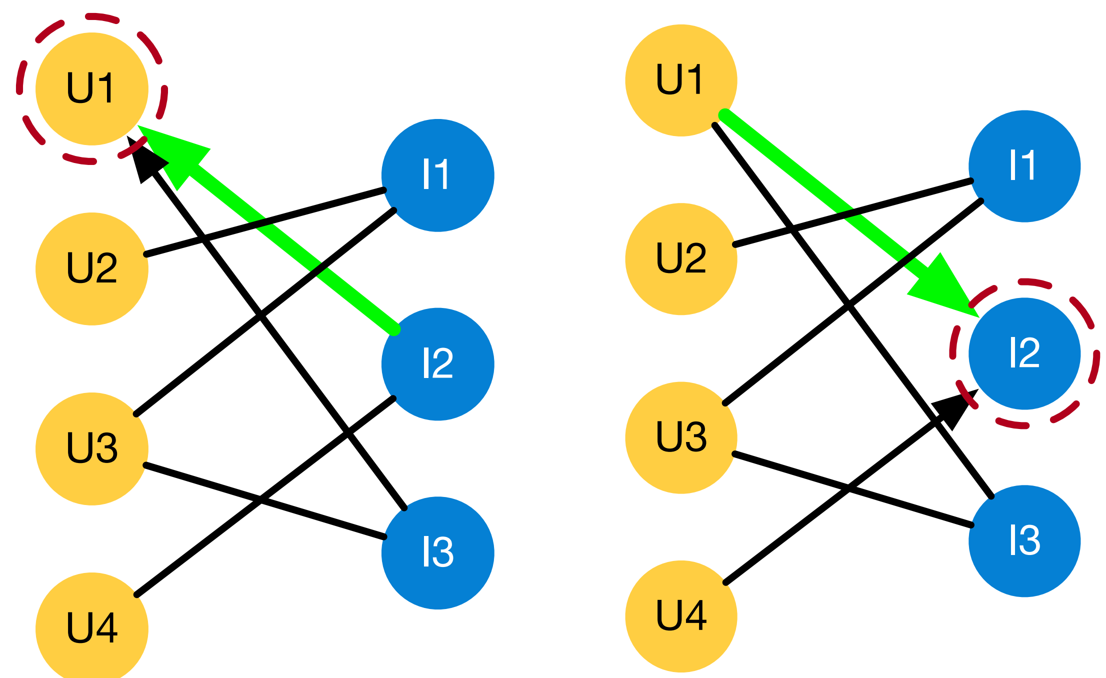
Feature Leakage Problem in GNN based Recommender Systems



In the prediction phase, every pair of nodes does not have an edge in the graph.



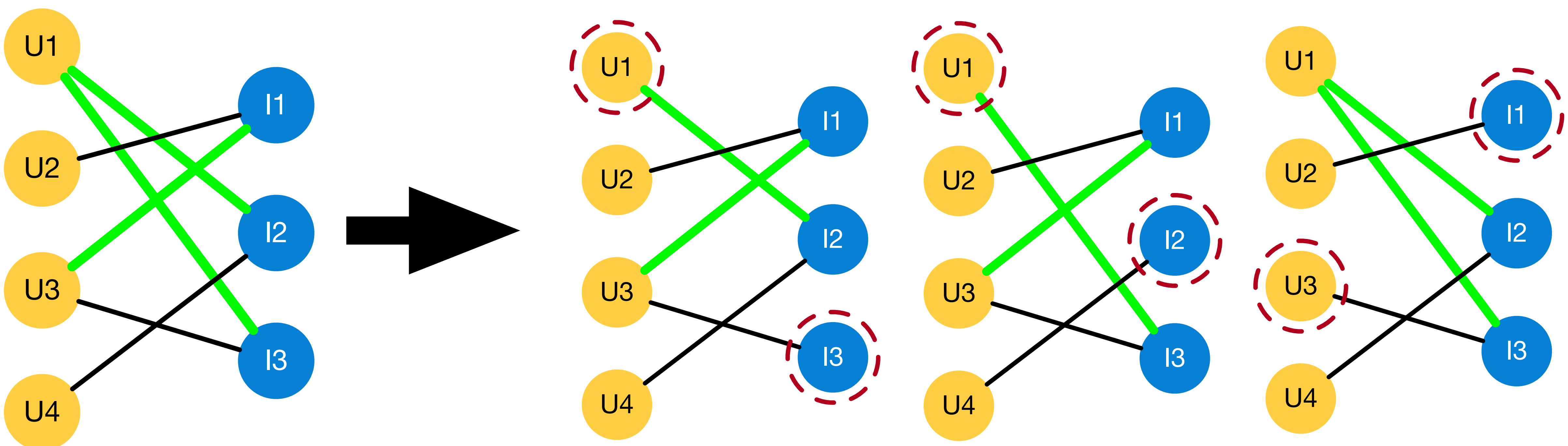
Positive training label



Signal for message passing

How to efficiently solve the feature leakage problem in GNN-based recommender systems?

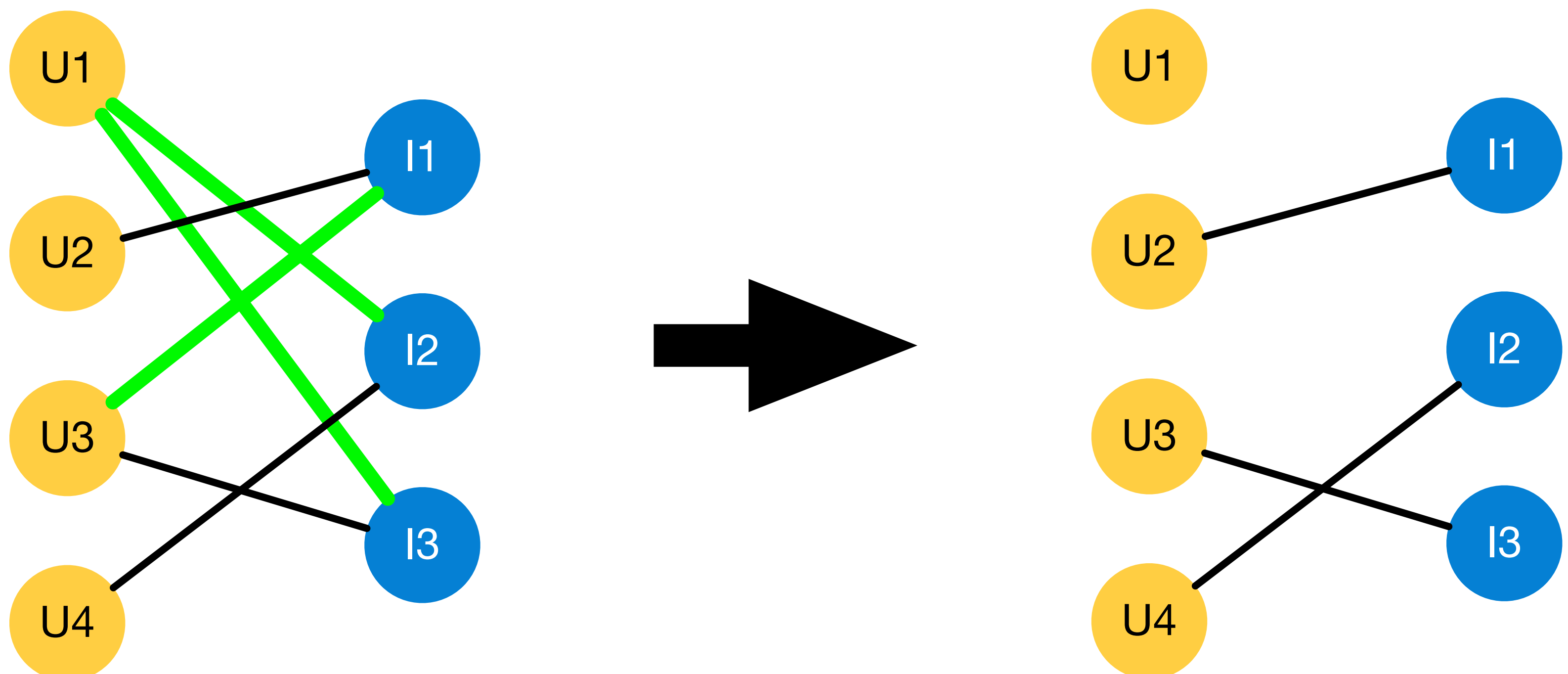
A Naive Solution: Accurate Removal



Cons: computationally infeasible

Sample and Removal Method

[redacted]



Jiani Zhang, Xingjian Shi, Shenglin Zhao, and Irwin King. 2019. STAR-GCN: Stacked and reconstructed graph convolutional networks for recommender sys- tems.

Cons: potential loss of information

Is it possible to use the graph information as an **accurate removal** method while keeping an acceptable computation complexity as the **sample and removal** method?

An Algebraic Trick

[redacted]

- Seek the relation of the node embeddings from the original graph and those after the accurate removal method.

$$\hat{\mathbf{A}}^{\boxed{k}}_{z,*}\mathbf{E}^{(0)} = |\mathcal{N}(z)|/|\hat{\mathcal{N}}(z)| \left(\mathbf{A}^{k}_{z,*}\mathbf{E}^{(0)} - \left(\mathrm{A}_{z\bar{z}}\mathbf{A}^{k-1}_{z,*}\mathbf{E}^{(0)} + \tilde{\Delta}^{k}_{z}\right)\right)$$

$$\tilde{\Delta}^{k}_{z} = \mathbf{A}^{k-2}_{z,*}\mathbf{E}^{(0)} \cdot \sum_{i\in\hat{\mathcal{N}}(z)} \mathrm{A}^{2}_{zi} - \hat{\mathbf{A}}^{\boxed{k-2}}_{z,*}\mathbf{E}^{(0)} \cdot |\hat{\mathcal{N}}(z)|/|\mathcal{N}(z)| \sum_{i\in\hat{\mathcal{N}}(z)} \hat{\mathrm{A}}^{2}_{zi}$$

Embeddings from the original graph

Embeddings from graph after accurate removal method

Main Results

Table 1: Dataset statistics.

Dataset	#User	#Items	#Interactions	Density
Instant	63,884	10,664	174,527	2e-4
Instrument	54,272	33,030	161,105	9.8e-5
Yelp	31,668	38,048	1,561,406	1.3e-3
Gowalla	29,858	40,981	1,027,370	8.4e-4

Table 2: Mean results of recall@20, nDCG@20, and time per epoch (T) in seconds. Bold represent the best and underline represents the second best. Vanila is not considered in the ranking for time.

	Method	Instant			Instrument			Yelp			Gowalla		
		Recall	nDCG	T(s)	Recall	nDCG	T(s)	Recall	nDCG	T(s)	Recall	nDCG	T(s)
2 Layers	Vanilla	0.1698	0.0805	2.96	0.0392	0.0187	2.75	0.0577	0.0467	110.13	0.1623	0.1375	82.65
	DropEdge	0.1656	0.0776	4.43	0.0465	0.0216	4.57	0.0581	**0.0469**	259.93	0.1622	0.1375	130.01
	S&R	**0.2202**	**0.1047**	5.23	**0.0541**	**0.0257**	4.62	0.0576	0.0466	444.94	0.1628	**0.1379**	259.49
	EFLEC	**0.2207**	**0.1029**	**3.11**	**0.0546**	**0.0260**	**3.06**	**0.0583**	**0.0469**	**122.75**	0.1630	**0.1382**	**72.88**
3 Layers	Vanilla	0.1776	0.0874	4.52	0.0471	0.0216	3.55	**0.0604**	**0.0489**	136.23	0.1677	0.1414	67.15
	DropEdge	0.1806	0.0825	**3.65**	0.0521	0.0241	**3.70**	**0.0603**	**0.0487**	219.73	**0.1690**	**0.1422**	105.28
	S&R	**0.2160**	**0.1059**	5.72	**0.0574**	**0.0270**	5.18	0.0600	0.0485	465.03	**0.1687**	**0.1420**	190.42
	EFLEC	**0.2155**	**0.1046**	4.56	**0.0573**	**0.0271**	4.15	0.0602	0.0485	**145.11**	**0.1689**	**0.1422**	**71.06**

[redacted]

Thank you!

Yaochen Hu
Email: yaochen.hu@huawei.com
WeChat: dynamic_relaxing
WeChat QR code: (the bottom right)