Learning from Counterfactual Links for Link Prediction

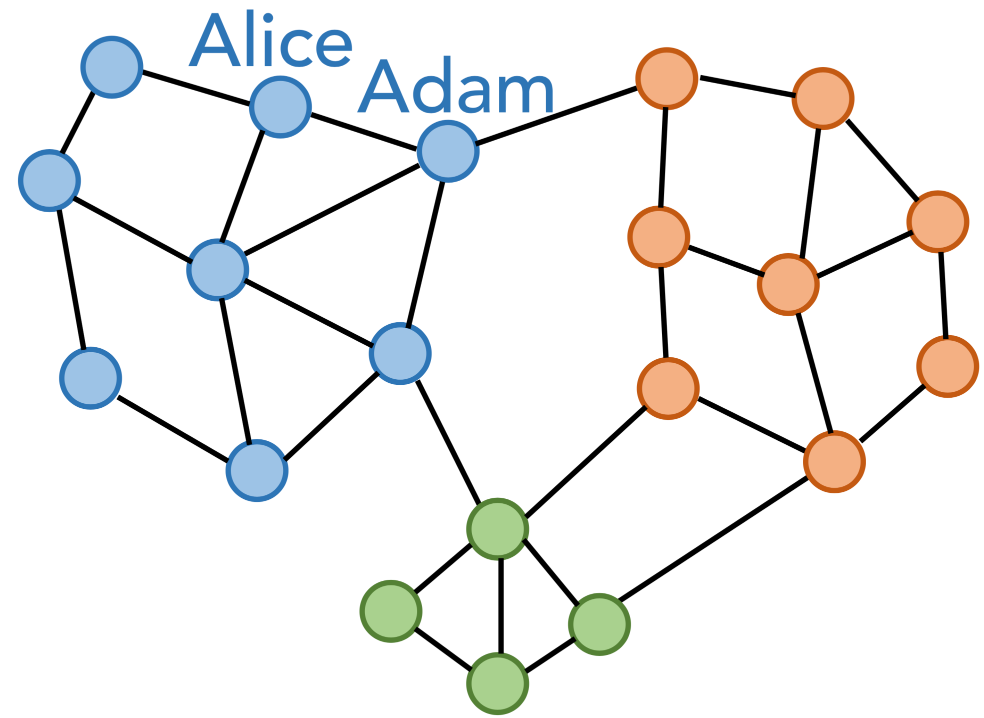
Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, Meng Jiang
University of Notre Dame

UNIVERSITY OF NOTRE DAME | COLLEGE OF ENGINEERING

Link Prediction

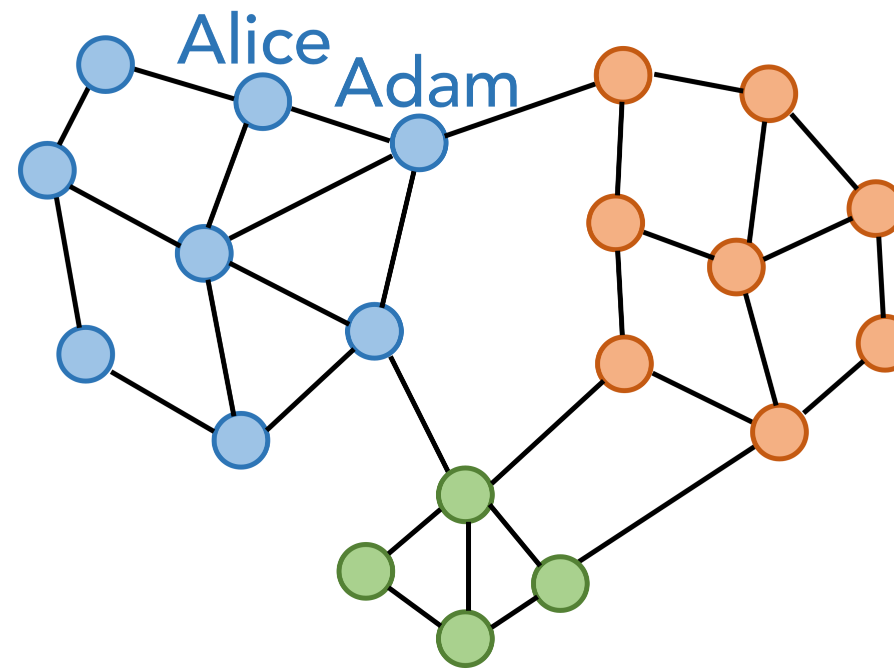
Given: a graph with adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$, raw node features $\mathbf{X} \in \mathbb{R}^{N\times F}$, and binary treatments $\mathbf{T} \in \{0,1\}^{N\times N}$ for each node pair.

Learn: low-dimensional node representations $\mathbf{Z} \in \mathbb{R}^{N\times H}$, which can be used for the prediction of link existences.



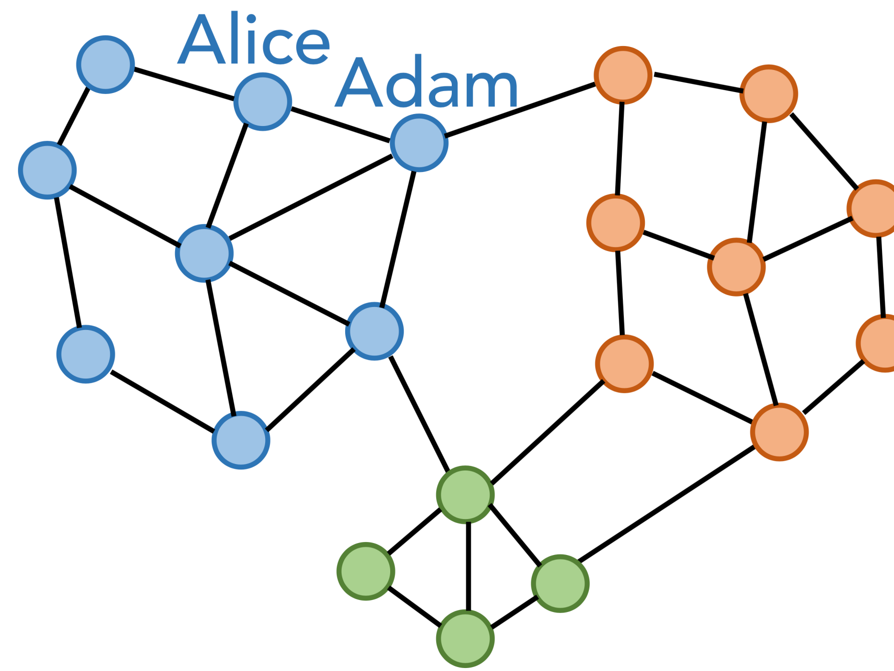
Counterfactual Outcomes to Balance Training Data

- Counterfactual question:
  - Would Alice and Adam still be friends if they were not living in the same neighborhood?



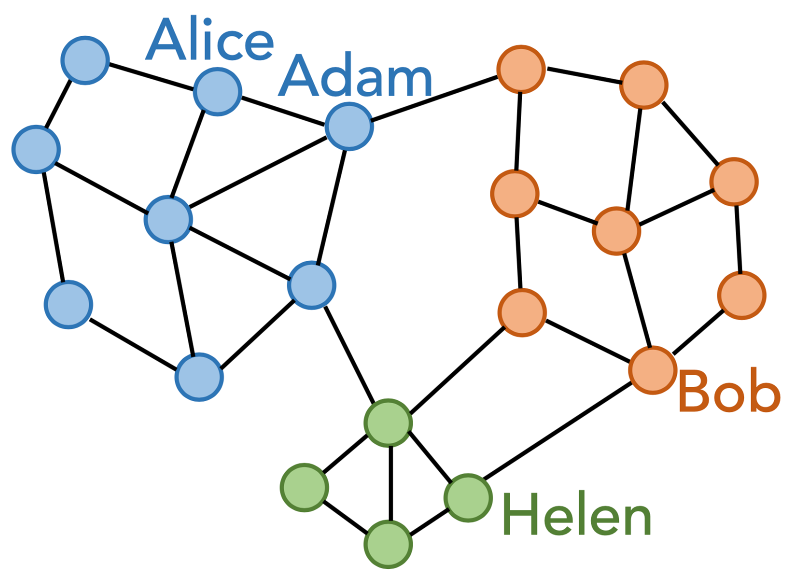
Counterfactual Outcomes to Balance Training Data

- Counterfactual question:
  - Would Alice and Adam still be friends if they were not living in the same neighborhood?



- Idea:
  - Generate *counterfactual links* to help the model learn better node representations for link prediction.

Counterfactual Links

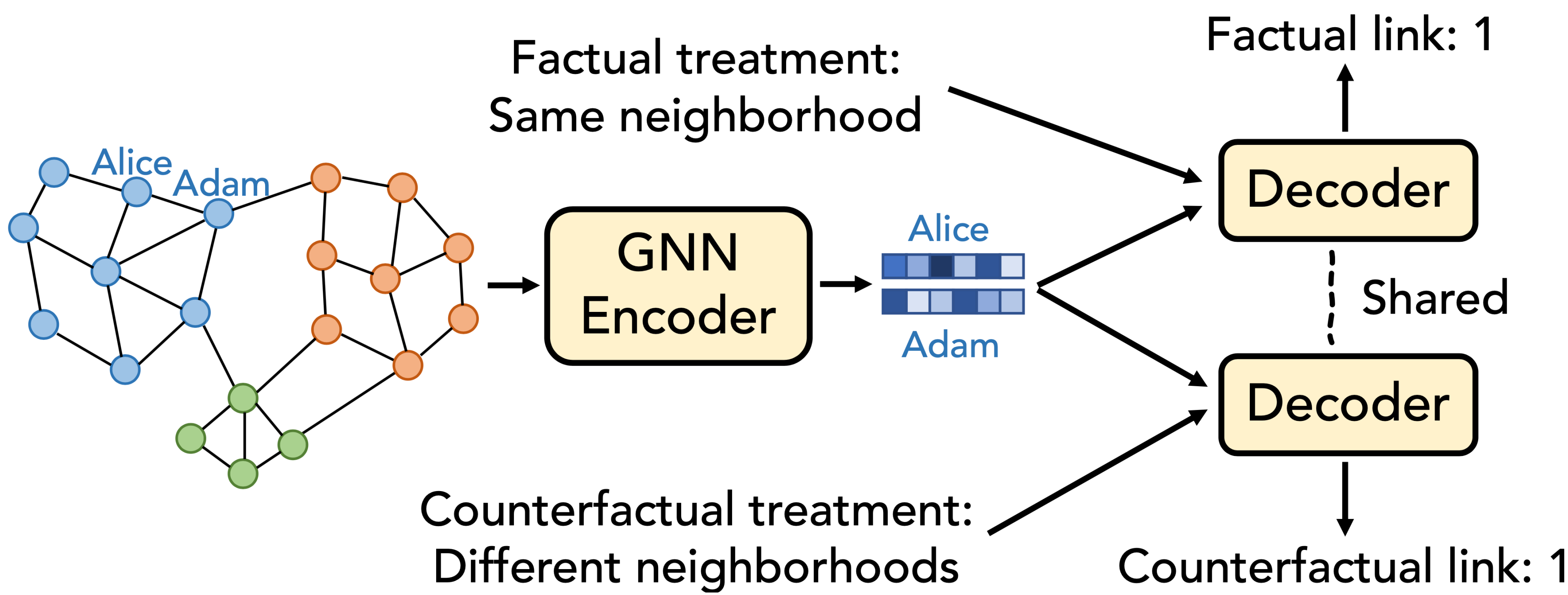


(Alice, Adam) $\xrightarrow[\text{different treatment}]{\text{Most similar with a}}$ (Helen, Bob)

Factual link: 1

Counterfactual link: 1

Learning from Counterfactual Links



Our proposed CFLP learns from both observed and counterfactual link existences.

Results 1

	CORA	CITESEER	PUBMED	FACEBOOK	OGB-DDI
Node2Vec	49.96±2.51	47.78±1.72	39.19±1.02	24.24±3.02	23.26±2.09
MVGRL	19.53±2.64	14.07±0.79	14.19±0.85	14.43±0.33	10.02±1.01
VGAE	45.91±3.38	44.04±4.86	23.73±1.61	37.01±0.63	11.71±1.96
SEAL	51.35±2.26	40.90±3.68	28.45±3.81	40.89±5.70	30.56±3.86
LGLP	62.98±0.56	57.43±3.71	–	37.86±2.13	–
GCN	49.06±1.72	55.56±1.32	21.84±3.87	53.89±2.14	37.07±5.07
GSAGE	53.54±2.96	53.67±2.94	39.13±4.41	45.51±3.22	53.90±4.74
JKNet	48.21±3.86	55.60±2.17	25.64±4.11	52.25±1.48	60.56±8.69
Our proposed CFLP with different graph encoders					
CFLP w/ GCN	60.34±2.33	59.45±2.30	34.12±2.72	53.95±2.29	52.51±1.09
CFLP w/ GSAGE	57.33±1.73	53.05±2.07	43.07±2.36	47.28±3.00	75.49±4.33
CFLP w/ JKNet	**65.57**±1.05	**68.09**±1.49	**44.90**±2.00	**55.22**±1.29	**86.08**±1.98

Consistent improvement against baselines.

Results 2

Get Started | Updates | Large-Scale Challenge | Datasets | Leaderboards | Paper | Team | Github

Leaderboard for ogbl-ddi

The Hits@20 score on the test and validation sets. The higher, the better.

Package: >=1.2.1

Rank	Method	Ext. data	Test Hits@20	Validation Hits@20	Contact	References	#Params	Hardware	Date
1	**PLNLP**	No	0.9088 ± 0.0313	0.8242 ± 0.0253	Zhitao Wang (WeChat@Tencent)	Paper, Code	3,497,473	Tesla-P40(24GB GPU)	Dec 7, 2021
2	**GraphSAGE + Edge Attr**	No	0.8781 ± 0.0474	0.8044 ± 0.0404	Jing Yang	Paper, Code	3,761,665	Tesla V100 (32GB)	Aug 9, 2021
3	**CFLP (w/ JKNet)**	No	0.8608 ± 0.0198	0.8405 ± 0.0284	Tong Zhao	Paper, Code	837,635	GeForce RTX 2080 Ti (11GB GPU)	Nov 17, 2021

Image sources:
https://ogb.stanford.edu/docs/leader_linkprop/#ogbl-ddi

Thank you for listening!

- Feel free to email me at
  - tzhao2@nd.edu (school)
  - tzhao@snap.com (work)

COLLEGE OF ENGINEERING