Learning from Counterfactual Links for Link Prediction

Presenter: Zhanke Zhou

2023 / 02 / 09

About the paper

Learning from Counterfactual Links for Link Prediction

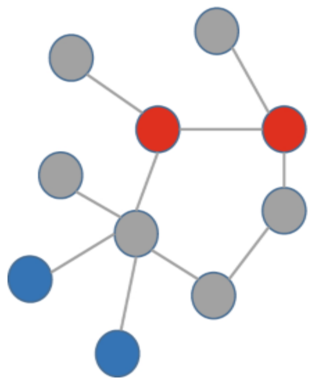
- affiliation: Department of Computer Science and Engineering, University of Notre Dame
- authors: Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, Meng Jiang
- conference: ICML 2022
- paper: https://arxiv.org/pdf/2106.02172.pdf
- official codes: https://github.com/DM2-ND/CFLP
- official slides: https://icml.cc/media/icml-2022/Slides/16774.pdf

Outline

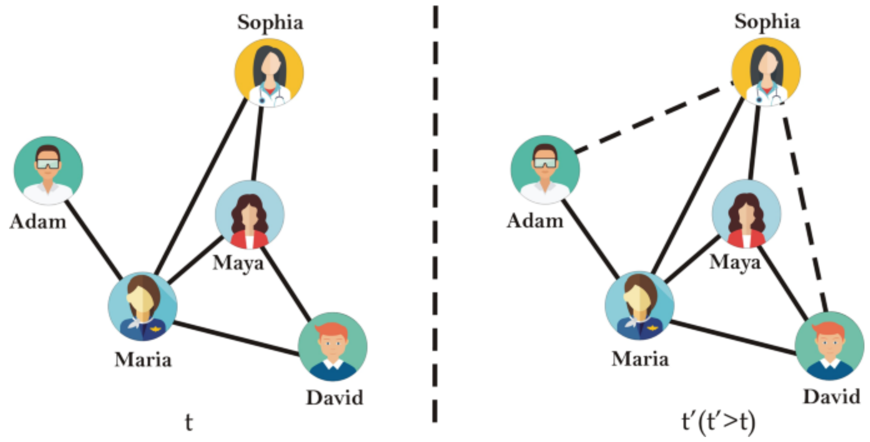
- Background
- Methods
- Experiment
- Summary

Background | Graph Learning

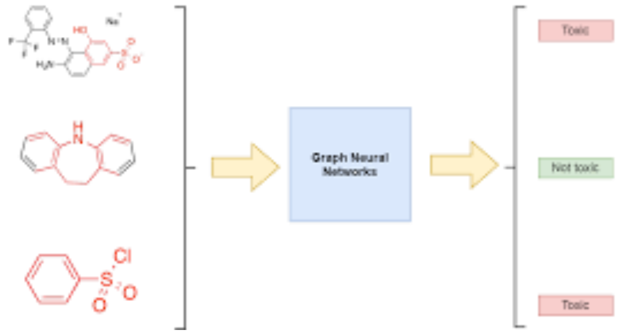
node-level



link-level

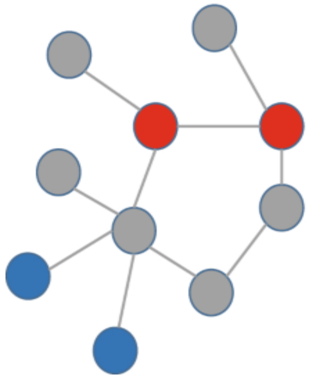


graph-level

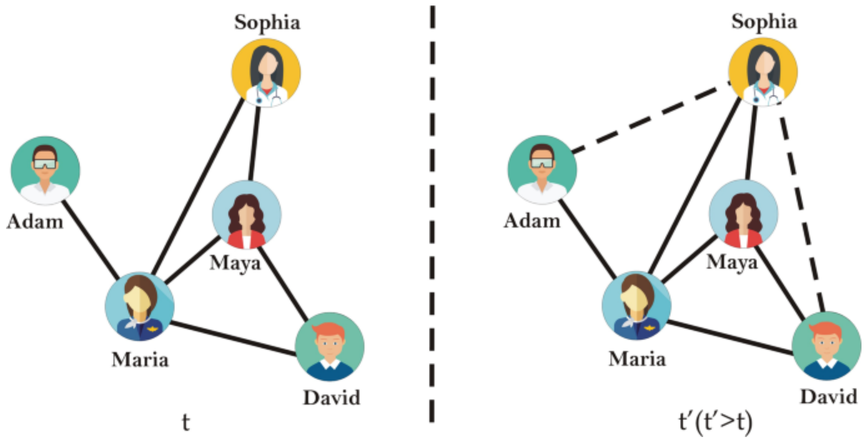


Background | Link Prediction

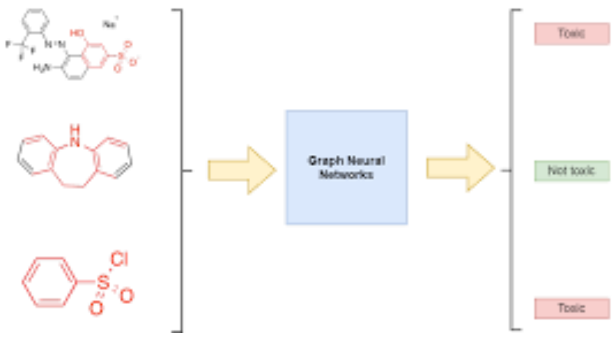
node-level



link-level



graph-level

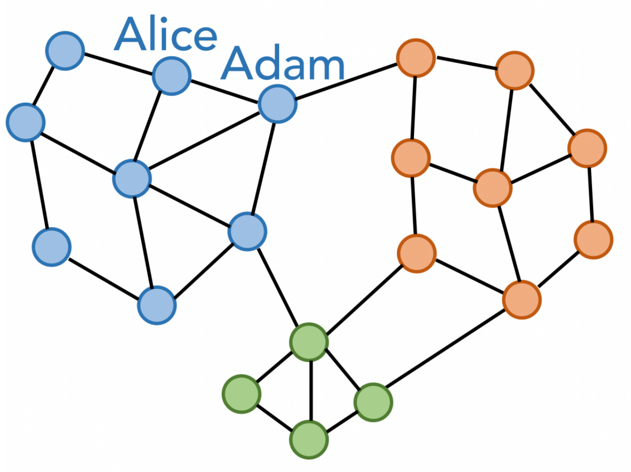


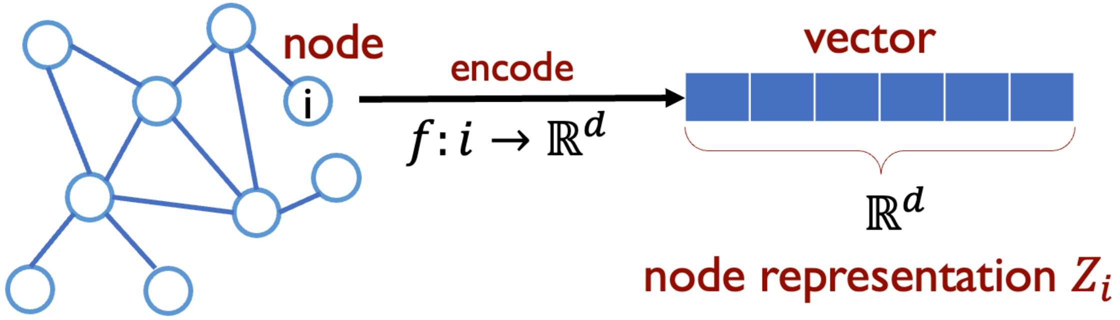
Background | Link Prediction

- Given: a graph with adjacency matrix $\boldsymbol{A} \in \{0,1\}^{N \times N}$ and raw node features $\boldsymbol{X} \in \mathbb{R}^{N \times D}$

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

- i.e., $\underbrace{f_{GNN}(\boldsymbol{A}, \boldsymbol{X}) = \boldsymbol{Z}}_{\text{encode}} \underbrace{\rightarrow \widehat{\boldsymbol{A}}}_{\text{decode}} \underbrace{\leftrightarrow \boldsymbol{A}_{GT}}_{\text{optimize}}$





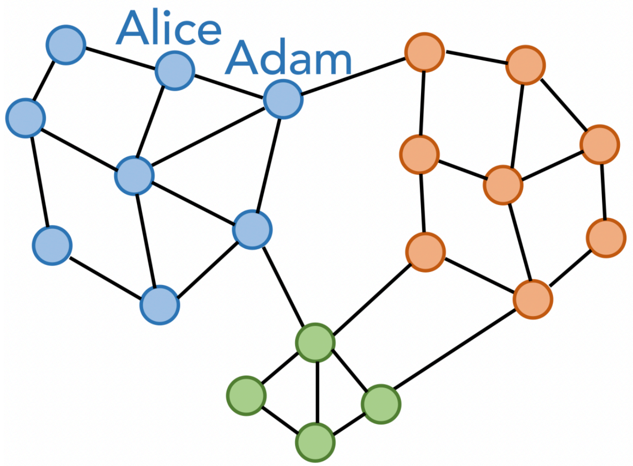
Background | Link Prediction

$$f_{GNN}(\boldsymbol{A}, \boldsymbol{X}) = \boldsymbol{Z} \to \widehat{\boldsymbol{A}}$$

A potential problem

- existing methods that learn from *association*
  - may not capture *essential factors* to predict missing links
- the *causal relationship* between <u>graph structure</u> and <u>link existence</u>
  - was largely *ignored* in previous work

“dog” class



Background | Link Prediction

A potential problem

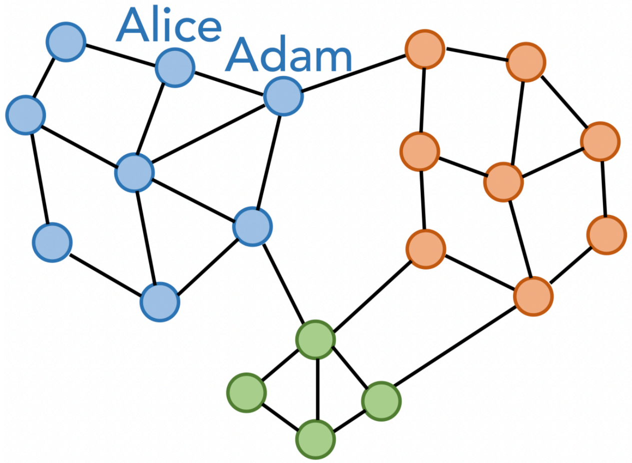
- existing methods that learn from *association*
  - may not capture *essential factors* to predict missing links
- the *causal relationship* between graph structure and link existence
  - was largely *ignored* in previous work

an example

Alice and Adam live in the same neighborhood and they are close friends

The association between ***neighborhood belonging*** and ***friendship*** could be too strong to discover the ***essential factors*** of friendship

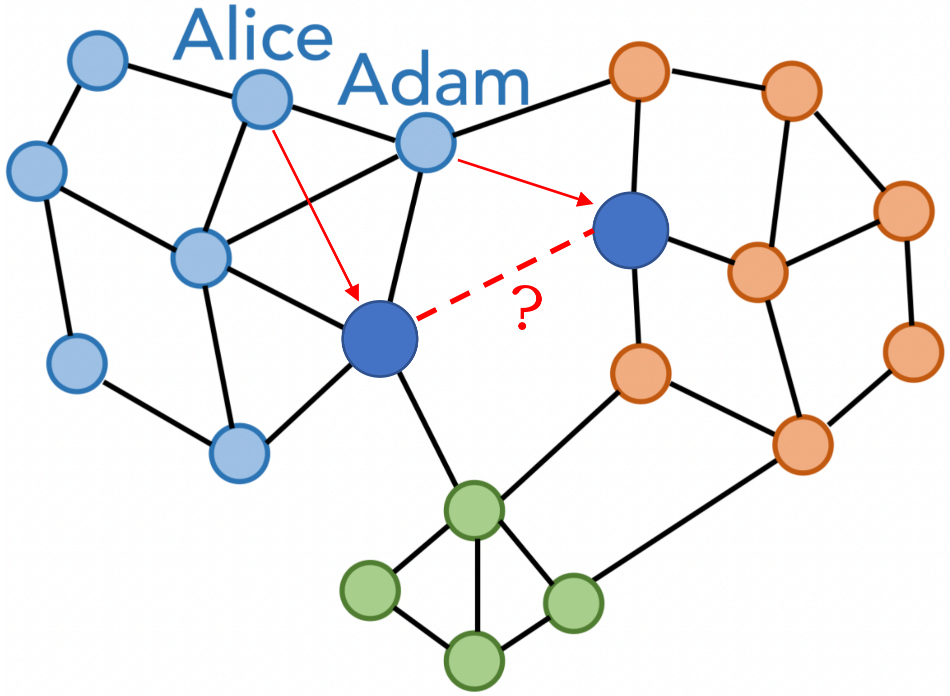
- such as common interests or family relationships (i.e., intrinsic properties)
- such factors may be the cause of them living in the same neighborhood



Background | Link Prediction

The counterfactual question:

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



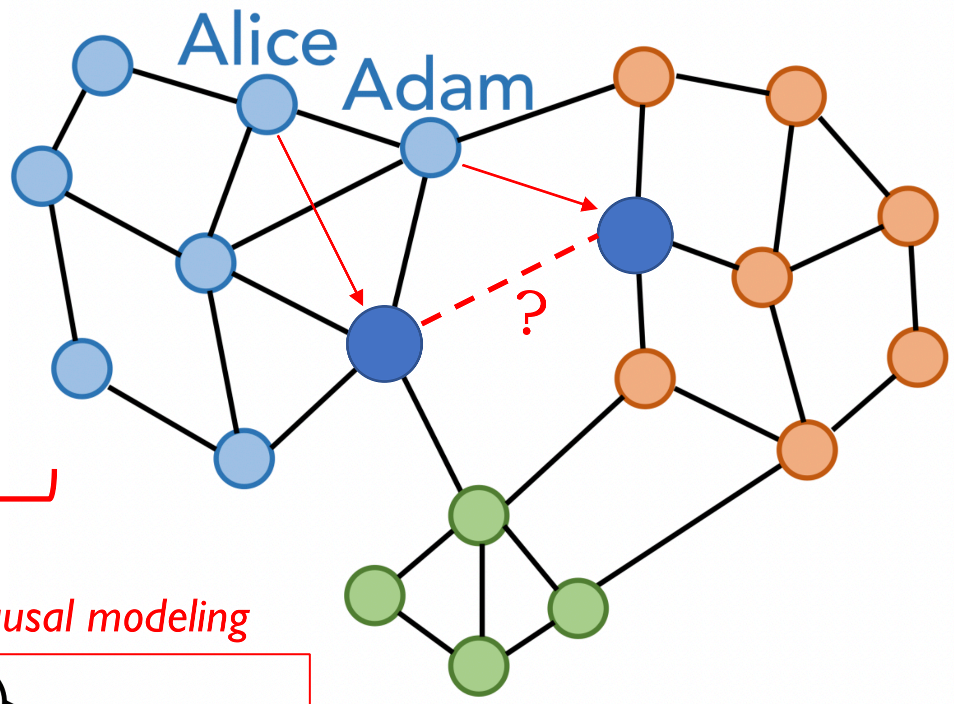
It is a good question, but how to find the answer, i.e., the counterfactual link? 🤔

Background | Link Prediction

A counterfactual question

Would Alice and Adam still be friends if they were not <u>living in the same neighborhood</u>?

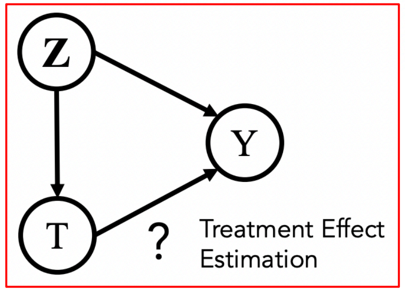
It is a good question, but how to find the answer, i.e., the counterfactual link?



The ***binary case*** is called the ***treatment (T)*** [1]

- (T=1) *<u>living in the same neighborhood</u>*
- (T=0) *<u>not living in the same neighborhood</u>*

Causal modeling



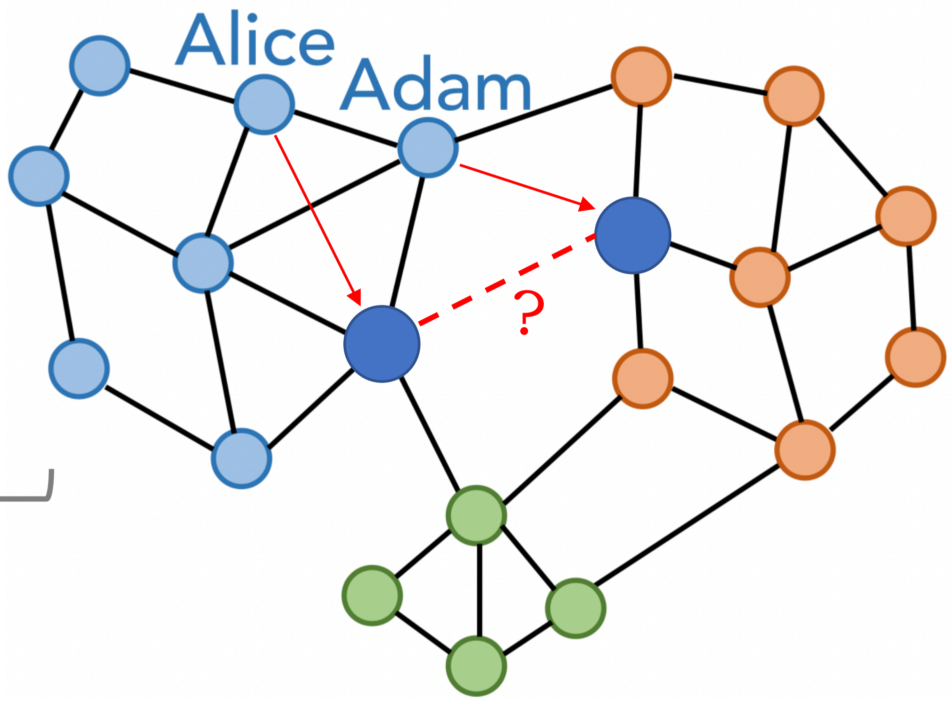
[1]: here, the ***treatment*** can be seen as the ***topological context*** of two nodes

Background | Link Prediction

A counterfactual question

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

It is a good question, but how to find the answer, i.e., the counterfactual link?



The binary case is called the ***treatment (T)***

- (T=1) *living in the same neighborhood*
- (T=0) *not living in the same neighborhood*

But, in reality, we can only observe the outcome under one particular ***treatment.***
So, how can we find the counterfactual link? 🤔

e.g., effects of a vaccine

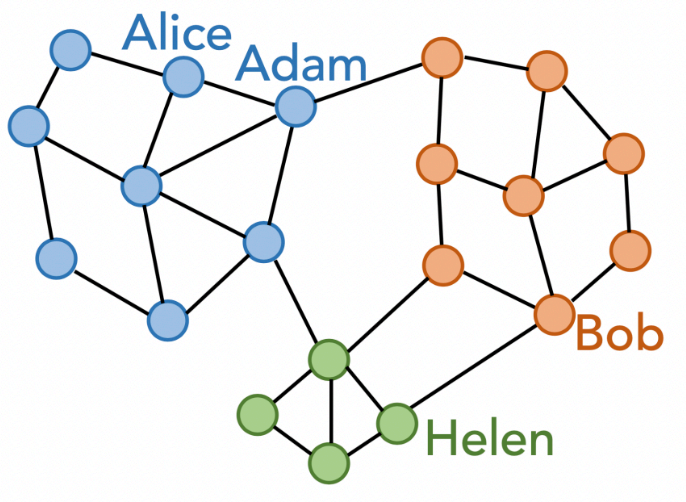
Background | Link Prediction

But, in reality, we can only observe the outcome under one particular ***treatment.***
So, how can we find the counterfactual link? 🤔

➔ **Solution:**
find ***the most similar*** node pair with ***a different treatment*** as the counterfactual link

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

Factual link: 1 Counterfactual link: 1



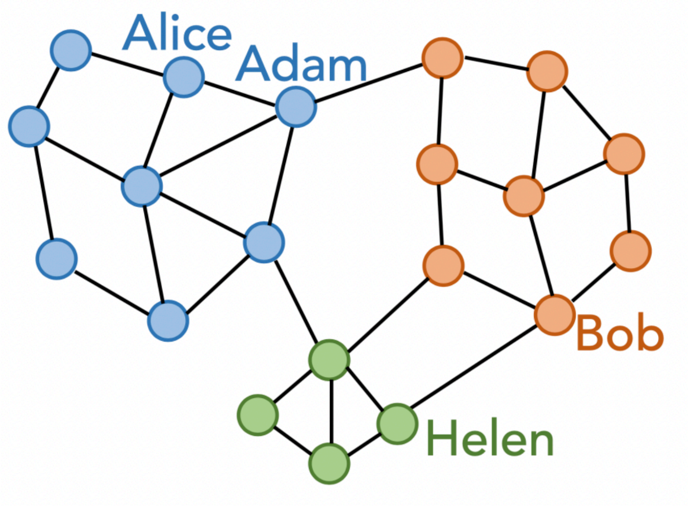
Background | Link Prediction

*But, in reality, we can only observe the outcome under one particular **treatment.***
So, how can we find the counterfactual link? 🤔

➔ **Solution:**
find ***the most similar*** node pair with ***a different treatment*** as the counterfactual link

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

Factual link: 1 Counterfactual link: 1



➔ **core idea:** generate **counterfactual links** to help the model learn **better** node representations for link prediction.

Background | Link Prediction

→ **core idea:** generate **counterfactual links** to help the model learn **better** node representations for link prediction.

Logic behind the idea

- Generally, the ***question*** can be described as "*would the link exist or not if the graph structure became different from observation?*"
- If a model can learn the causal relationship by answering this question, it will improve the prediction with such ***knowledge***.

Background | Experiments

	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	<u>62.98</u>±0.56	<u>57.43</u>±3.71	–	37.86±2.13	–
GCN	49.06±1.72	55.56±1.32	21.84±3.87	<u>53.89</u>±2.14	37.07±5.07
GSAGE	53.54±2.96	53.67±2.94	<u>39.13</u>±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	<u>60.56</u>±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

OGB — 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

Good empirical results

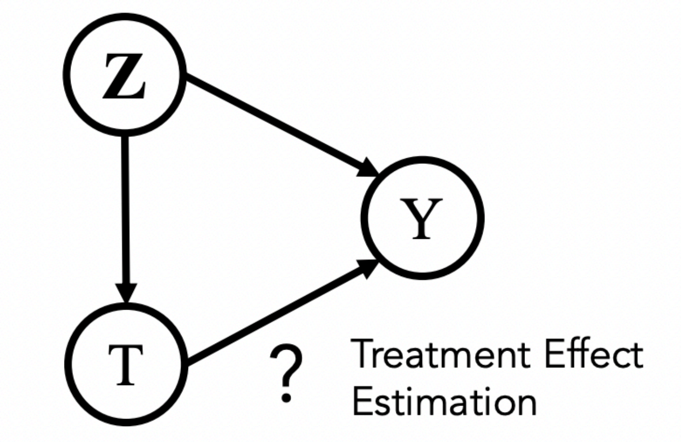
Outline

- Background
- Methods
  - Q1: what are the counterfactual links?
  - Q2: how to generate the counterfactual links?
  - Q3: how to utilize the counterfactual links?
- Experiment
- Summary

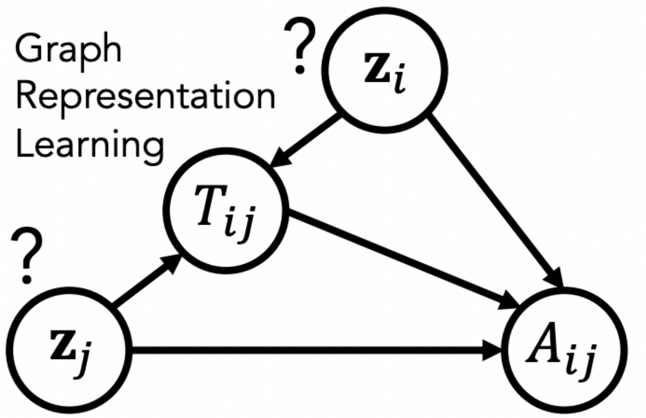
Method | formulation

3 key factors of a counterfactual question

- **node representations (Z)**
  - information of node pairs
- **treatment (T)**
  - global graph structural properties
- **outcome (Y)**
  - link existence



(a) Causal modeling (not the target of our work but related to the idea we propose): Given $\mathbf{Z}$ and observed outcomes, find treatment effect of T on Y.



(b) Graph learning with causal model (the proposed idea): leverage the estimated $\text{ITE}(A_{i,j}|T_{i,j})$ to improve the learning of $\mathbf{z}_i$ and $\mathbf{z}_j$.

Method | formulation

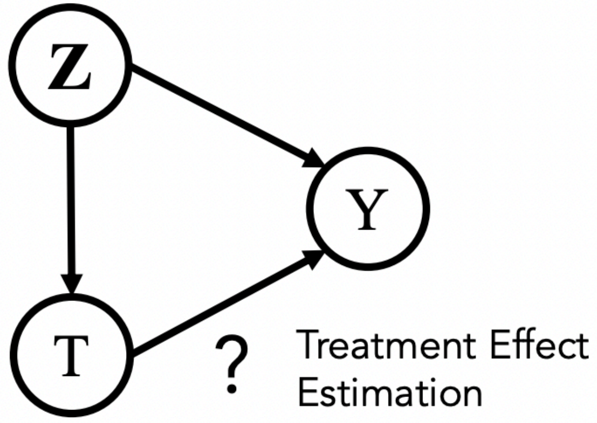
3 key factors of a counterfactual question

- node representations (Z)
  - information of node pairs
- treatment (T)
  - global graph structural properties
- outcome (Y)
  - link existence

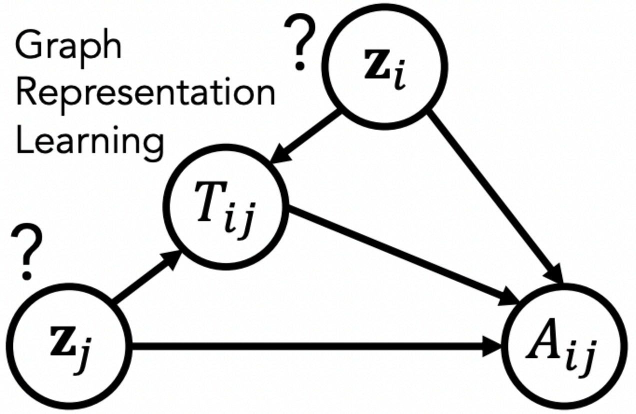
About the treatment (T)

- global graph structural properties
  - id of community/cluster/neighborhood
  - or K-core / Louvain / spectral clustering
  - $T_{ij} = 1$ if $c(v_i) = c(v_j)$ else $T_{ij} = 0$

- $T_{ij} = 1$ means node i and node j
- are ***structurally consistent*** in one aspect



(a) Causal modeling (not the target of our work but related to the idea we propose): Given $\mathbf{Z}$ and observed outcomes, find treatment effect of T on Y.



(b) Graph learning with causal model (the proposed idea): leverage the estimated ITE$(A_{i,j}|T_{i,j})$ to improve the learning of $\mathbf{z}_i$ and $\mathbf{z}_j$.

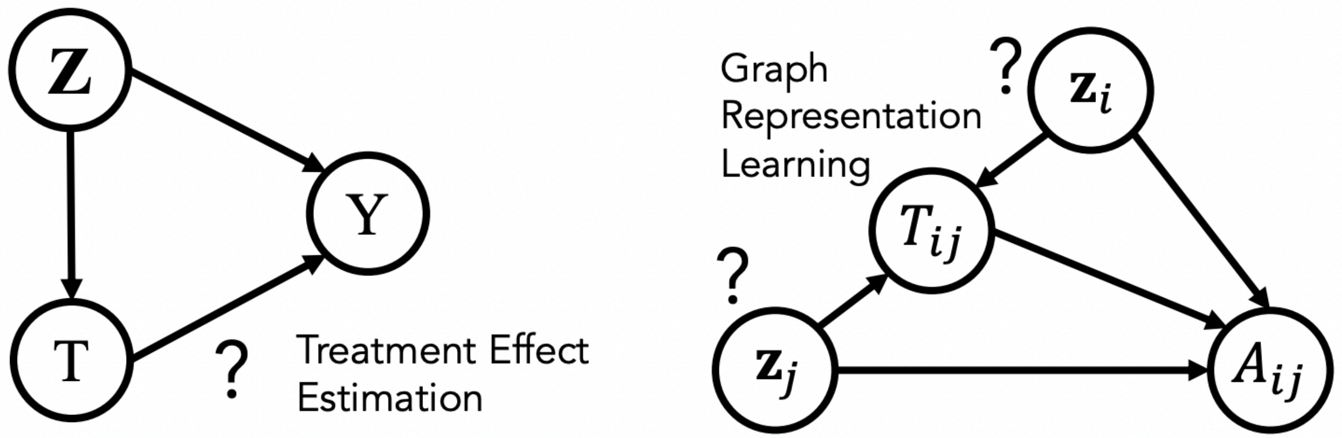
Method | formulation

Q1: what are the counterfactual links?
that is, the link A_{ij}^{CF} that satisfies $T_{ij}^{CF} = 1 - T_{ij}$

when $T_{ij}^{CF} = 1 - T_{ij}$, is the link A_{ij}^{CF} exist?

3 key factors of a counterfactual question

- node representations (Z)
    - information of node pairs
- treatment (T)
    - global graph structural properties
- outcome (Y)
    - link existence



(a) Causal modeling (not the target of our work but related to the idea we propose): Given $\mathbf{Z}$ and observed outcomes, find treatment effect of T on Y.

(b) Graph learning with causal model (the proposed idea): leverage the estimated $\text{ITE}(A_{i,j}|T_{i,j})$ to improve the learning of $\mathbf{z}_i$ and $\mathbf{z}_j$.

About the treatment (T)

- global graph structural properties
    - id of community/cluster/neighborhood
    - or K-core / Louvain / spectral clustering
    - $T_{ij} = 1$ if $c(v_i) = c(v_j)$ else $T_{ij} = 0$

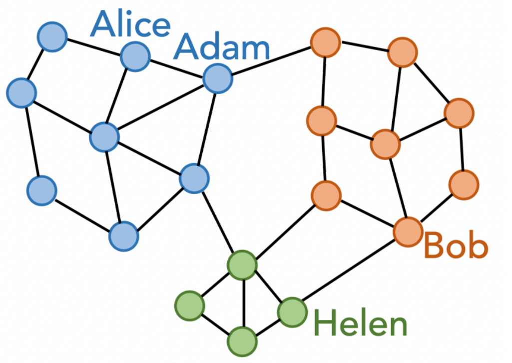
- $T_{ij} = 1$ means node i and node j
- are ***structurally consistent*** in one aspect

Method

- Q2: How to generate the counterfactual links?

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

Factual link: 1 Counterfactual link: 1



To find the counterfactual link (v_a, v_b) of the given link (v_i, v_j)

$$(v_a, v_b) = \underset{v_a, v_b \in \mathcal{V}}{\arg\min} \{h((v_i, v_j), (v_a, v_b)) \mid T_{a,b} = 1 - T_{i,j}\}, \tag{2}$$

where $h(\cdot,\cdot)$ is a metric of measuring the distance between a two edges

Method

- Q2: How to generate the counterfactual links?

$$(v_a, v_b) = \arg\min_{v_a, v_b \in \mathcal{V}} \{h((v_i, v_j), (v_a, v_b)) \mid T_{a,b} = 1 - T_{i,j}\}, \tag{2}$$

$$\Big\downarrow \text{relax}$$

$$\begin{aligned}(v_a, v_b) = \arg\min_{v_a, v_b \in \mathcal{V}} \{ & d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_a) + d(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_b) \mid \\ & T_{a,b} = 1 - T_{i,j}, d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_a) + d(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_b) < 2\gamma\},\end{aligned} \tag{3}$$

$$O(N^4) \longrightarrow O(N^2)$$

Method

- Q2: How to generate the counterfactual links?

$$(v_a, v_b) = \underset{v_a, v_b \in \mathcal{V}}{\arg\min} \{ d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_a) + d(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_b) \mid \qquad (3)$$

$$T_{a,b} = 1 - T_{i,j}, d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_a) + d(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_b) < 2\gamma \},$$

$$\rightarrow \boxed{T^{CF}_{i,j}, A^{CF}_{i,j} = \begin{cases} 1 - T_{i,j}, A_{a,b} & \text{, if } \exists \, (v_a, v_b) \in \mathcal{V} \times \mathcal{V} \\ & \quad \text{satisfies Eq. (3);} \\ T_{i,j}, A_{i,j} & \text{, otherwise.} \end{cases}}$$

Method

- Q2: How to generate the counterfactual links?

$$(v_a, v_b) = \underset{v_a, v_b \in \mathcal{V}}{\arg\min} \{ d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_a) + d(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_b) \mid \qquad (3)$$
$$T_{a,b} = 1 - T_{i,j}, d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_a) + d(\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_b) < 2\gamma \},$$

$$T^{CF}_{i,j}, A^{CF}_{i,j} = \begin{cases} 1 - T_{i,j}, A_{a,b} & \text{, if } \exists\, (v_a, v_b) \in \mathcal{V} \times \mathcal{V} \\ & \quad \text{satisfies Eq. (3);} \\ T_{i,j}, A_{i,j} & \text{, otherwise.} \end{cases}$$

So far, we obtain A^{CF}_{ij}, T^{CF}_{ij} from A_{ij}, T_{ij}.

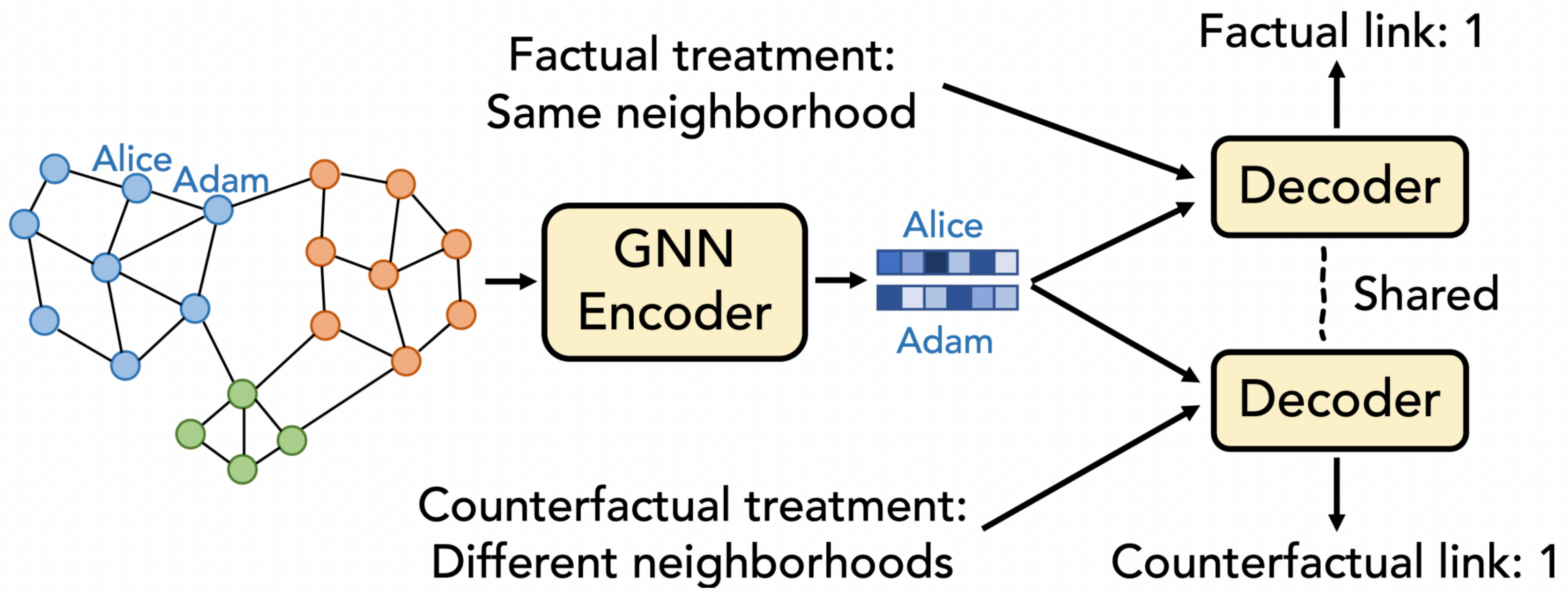
But, how to utilize the A^{CF}_{ij}, T^{CF}_{ij}? 🤔

Outline

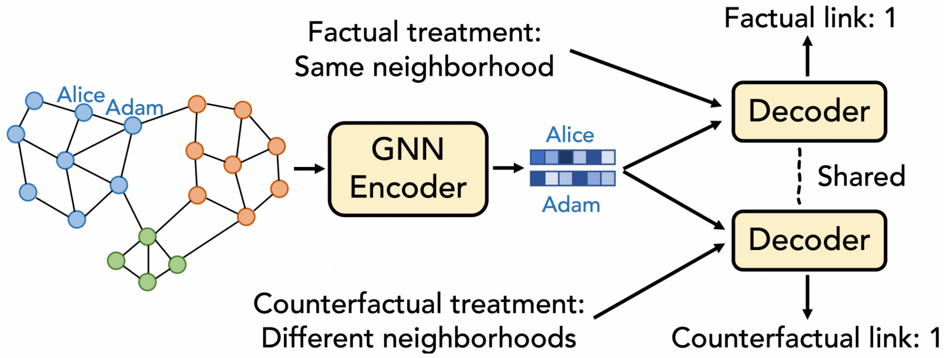
- Background
- Methods
  - Q1: what are the counterfactual links?
  - Q2: how to generate the counterfactual links?
  - Q3: how to utilize the counterfactual links?
- Experiment
- Summary

Method

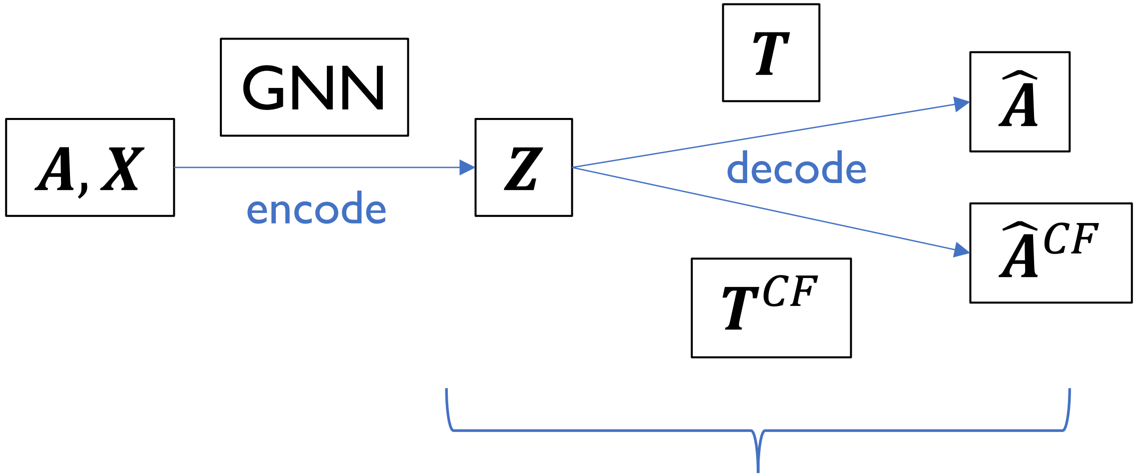
- Learning from Counterfactual Links
  - ***train a GNN to predict*** factual links and counterfactual links
  - given the corresponding ***treatments***



Method



- Learning from Counterfactual Links
  - ***train a GNN to predict*** factual links and counterfactual links
  - given the corresponding ***treatments***



$$\widehat{\mathbf{A}} = g(\mathbf{Z}, \mathbf{T}), \text{ s.t. } \widehat{A}_{i,j} = \mathrm{MLP}([\mathbf{z}_i \odot \mathbf{z}_j, T_{i,j}]), \quad (6)$$

$$\widehat{\mathbf{A}}^{CF} = g(\mathbf{Z}, \mathbf{T}^{CF}), \text{ s.t. } \widehat{A}^{CF}_{i,j} = \mathrm{MLP}([\mathbf{z}_i \odot \mathbf{z}_j, T^{CF}_{i,j}]), \quad (7)$$

Method

- Learning from Counterfactual Links
  - ***optimization***

$$\mathcal{L}_F = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{i,j} \cdot \log \widehat{A}_{i,j} + (1 - A_{i,j}) \cdot \log(1 - \widehat{A}_{i,j}), \tag{8}$$

$$\mathcal{L}_{CF} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} A^{CF}_{i,j} \cdot \log \widehat{A}^{CF}_{i,j} + (1 - A^{CF}_{i,j}) \cdot \log(1 - \widehat{A}^{CF}_{i,j}). \tag{9}$$

$$\mathcal{L}_{disc} = \mathrm{disc}(\hat{P}^F_f, \hat{P}^{CF}_f), \text{ where } \mathrm{disc}(P, Q) = ||P - Q||_F, \tag{10}$$

$$\boxed{\mathcal{L} = \mathcal{L}_F + \alpha \cdot \mathcal{L}_{CF} + \beta \cdot \mathcal{L}_{disc}, \tag{11}}$$

Algorithm 1 CFLP

```
Input: f, g, A, X, n_epochs, n_epoch_ft
Compute T as presented in Section 3.1.
Compute T^CF, A^CF by Eqs. (3) and (4).
// model training
for epoch in range(n_epochs) do
    Z = f(A, X).
    Get Â and Â^CF via g with Eqs. (6) and (7).
    Update Θ_f and Θ_g with L. (Eq. (11))
end for
// decoder fine-tuning
Freeze Θ_f and re-initialize Θ_g.
Z = f(A, X).
for epoch in range(n_epochs_ft) do
    Get Â via g with Eq. (6).
    Update Θ_g with L_F. (Eq. (8))
end for
// inference
Z = f(A, X).
Get Â and Â^CF via g with Eqs. (6) and (7).
Output: Â for link prediction, Â^CF.
```

1. collect data

2. train

3. fine-tune

4. test

Outline

- Background
- Methods
    - Q1: what are the counterfactual links?
    - Q2: how to generate the counterfactual links?
    - Q3: how to utilize the counterfactual links?
- Experiment
    - Q4: how helpful are the counterfactual links?
    - Q5: how to justify the effectiveness of counterfactual links?
- Summary

Experiments

Table 1. Statistics of datasets used in the experiments.

Dataset	CORA	CITESEER	PUBMED	FACEBOOK	OGB-DDI
# nodes	2,708	3,327	19,717	4,039	4,267
# links	5,278	4,552	44,324	88,234	1,334,889
# validation node pairs	1,054	910	8,864	17,646	235,371
# test node pairs	2,110	1,820	17,728	35,292	229,088

Experiments

Link prediction performances measured by ***Hits@20***

	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.

Experiments

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.

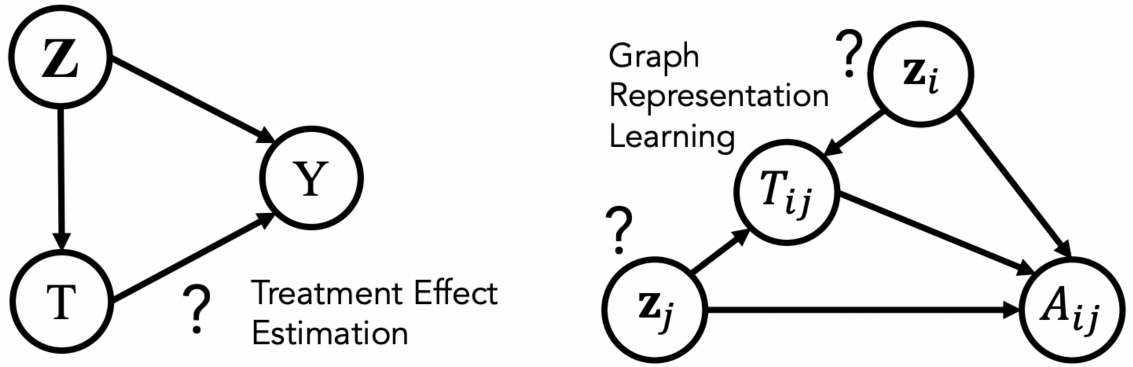
Package: >=1.2.1

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

Experiments

Q5: how to justify the effectiveness of counterfactual links?

when $T_{ij} = 1$ or $T_{ij} = 0$, is the link A_{ij} still exist?



(a) Causal modeling (not the target of our work but related to the idea we propose): Given $\mathbf{Z}$ and observed outcomes, find treatment effect of T on Y.

(b) Graph learning with causal model (the proposed idea): leverage the estimated $\text{ITE}(A_{i,j}|T_{i,j})$ to improve the learning of $\mathbf{z}_i$ and $\mathbf{z}_j$.

The ***individual treatment effect (ITE)*** can be used to quantify the effect of treatment on the outcome.

$$\text{ITE}(\mathbf{z}) = g(\mathbf{z}, 1) - g(\mathbf{z}, 0)$$

$$\text{ITE}_{(v_i, v_j)} = \underbrace{g((\mathbf{z}_i, \mathbf{z}_j), 1)}_{T_{ij}=1} - \underbrace{g((\mathbf{z}_i, \mathbf{z}_j), 0)}_{T_{ij}=0}$$

The ***averaged treatment effect (ATE)*** is the expectation of ITE

$$\text{ATE} = \mathbb{E}_{\mathbf{z}\sim\mathbf{Z}}\, \text{ITE}(\mathbf{z})$$

Experiments

Table 4. Results of CFLP with different treatments on CORA. (sorted by Hits@20)

	Hits@20	$\widehat{\text{ATE}}_{obs}$	$\widehat{\text{ATE}}_{est}$
K-core	65.6±1.1	0.002	0.013±0.003
SBM	64.2±1.1	0.006	0.023±0.015
CommN	62.3±1.6	0.007	0.053±0.021
PropC	61.7±1.4	0.037	0.059±0.065
Ward	61.2±2.3	0.001	0.033±0.012
SpecC	59.3±2.8	0.002	0.033±0.011
Louvain	57.6±1.8	0.025	0.138±0.091
Katz	56.6±3.4	0.740	0.802±0.041

Table 5. Results of CFLP with different treatments on CITESEER. (sorted by Hits@20)

	Hits@20	$\widehat{\text{ATE}}_{obs}$	$\widehat{\text{ATE}}_{est}$
SBM	71.6 ±1.9	0.004	0.005 ±0.001
K-core	68.1±1.5	0.002	0.010±0.002
Ward	67.0±1.7	0.003	0.037±0.009
PropC	64.6±3.6	0.141	0.232±0.113
Louvain	63.3±2.5	0.126	0.151±0.078
SpecC	59.9±1.3	0.009	0.166±0.034
Katz	57.3±0.5	0.245	0.224±0.037
CommN	56.8±4.9	0.678	0.195±0.034

$$\widehat{\text{ATE}}_{obs} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\mathbf{T}\odot(\mathbf{A}-\mathbf{A}^{CF}) + (\mathbf{1}_{N\times N}-\mathbf{T})\odot(\mathbf{A}^{CF}-\mathbf{A})\}_{i,j}$$

$$\widehat{\text{ATE}}_{est} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\mathbf{T}\odot(\hat{\mathbf{A}}-\hat{\mathbf{A}}^{CF}) + (\mathbf{1}_{N\times N}-\mathbf{T})\odot(\hat{\mathbf{A}}^{CF}-\hat{\mathbf{A}})\}_{i,j}$$

Outline

- Background
- Methods
- Experiment
- Summary

Summary

Contributions

1. the ***first*** work that aims at improving link prediction by ***causal inference***

2. introduce CFLP that trains GNNs to ***predict*** both factual and counterfactual links

3. leverage ***causal relationship*** to ***enhance*** link prediction

4. CFLP ***outperforms*** competitive baselines on several benchmark datasets

Q&A

Thanks for your attention!