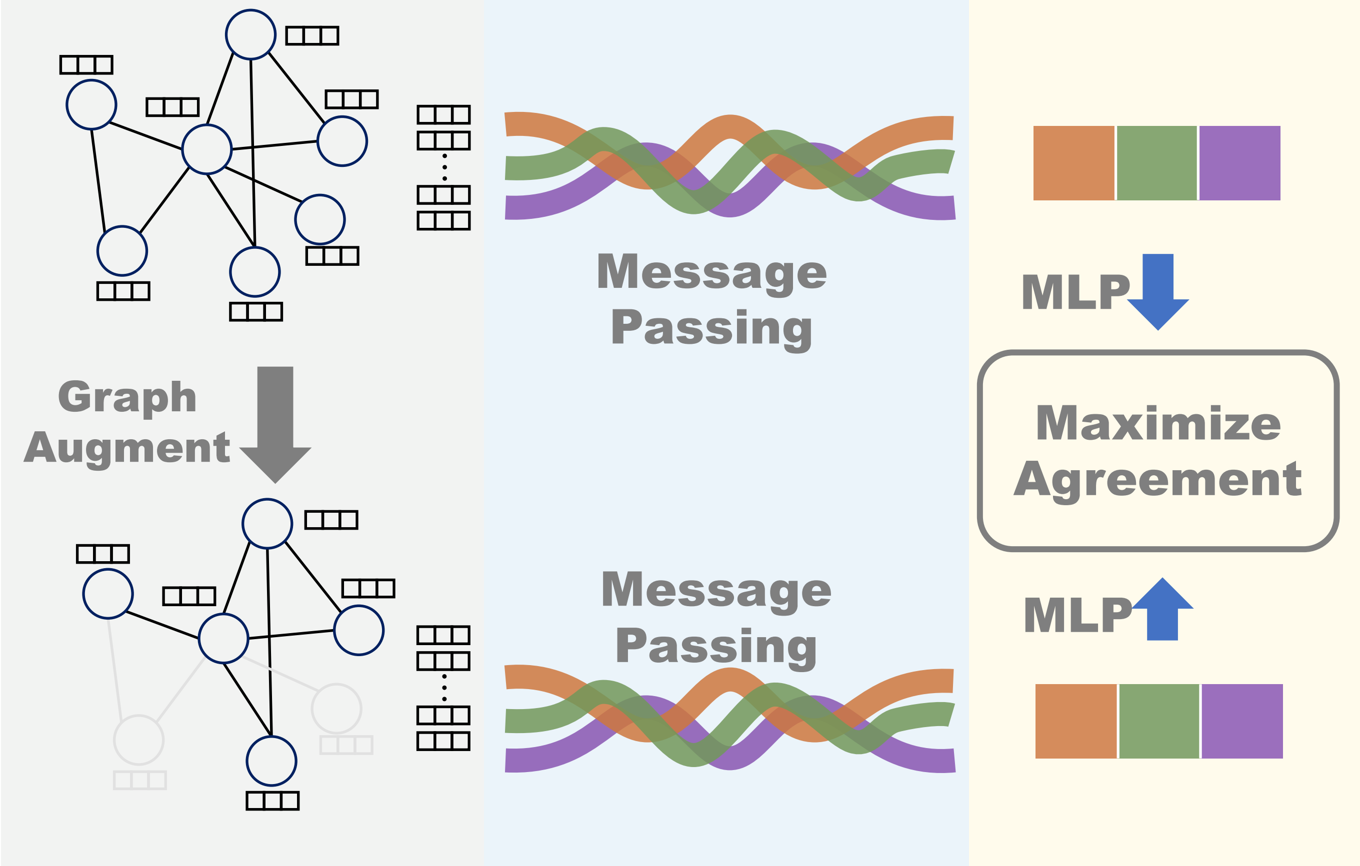
GALOPA: Graph Transport Learning with Optimal Plan Alignment

Yejiang Wang[1], Yuhai Zhao[1], Zhengkui Wang[2], Ling Li[1]

[1] Northeastern University, China
[2] Singapore Institute of Technology, Singapore

Graph Contrastive Learning[1](GCL)

- Maximize the agreement of representations under augmentation (positive pair);
- Minimize the agreement of representation of different graphs (negative pair);



[1] You Y, et al. “Graph contrastive learning with augmentations”. *NeurIPS*. 2020.

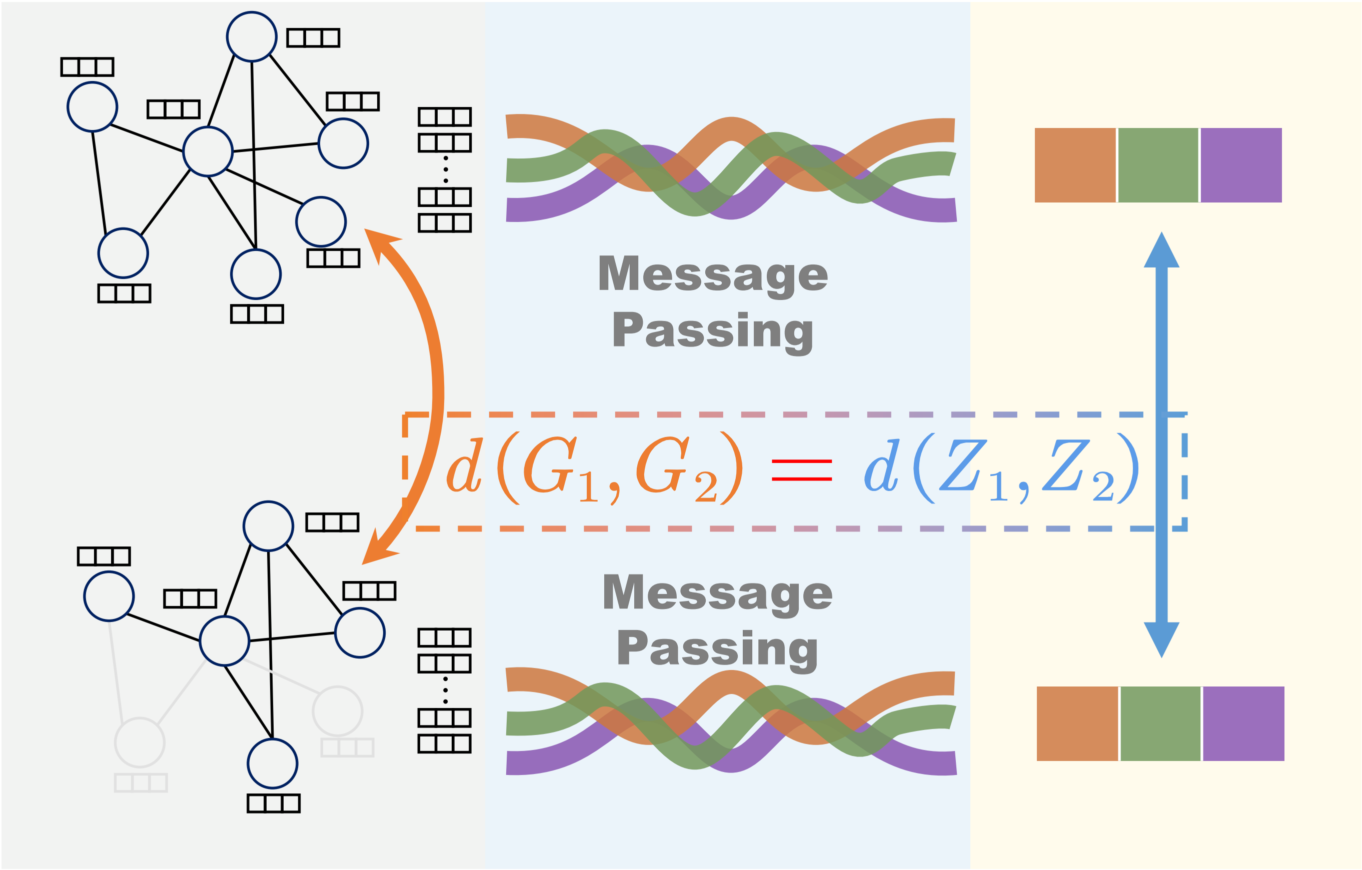
Problem of GCL

- GCL's effectiveness depends on the label-invariant assumption that augmented operations produce consistent labels for original and augmented samples. However, slight perturbations in graph structures can cause significant property variations;

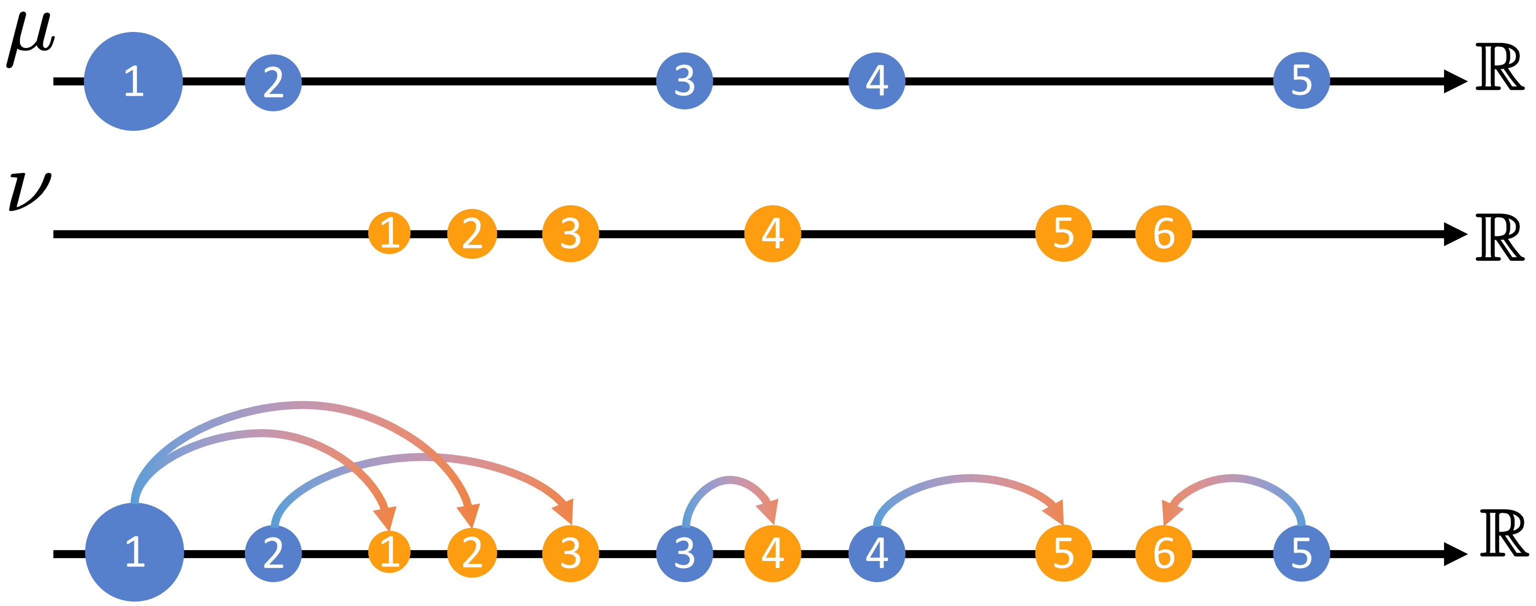
- Maximizing or minimizing the similarity between positive or negative views in contrastive learning lacks clear guidance;

Straightforward Solution

- **Issue**: Graph and Vector are two distinct concepts, making it difficult to agree on their distance metrics;
- **Exmp**: Graph edit distance between graphs and the Euclidean distance between vectors;



Optimal Transport Plan



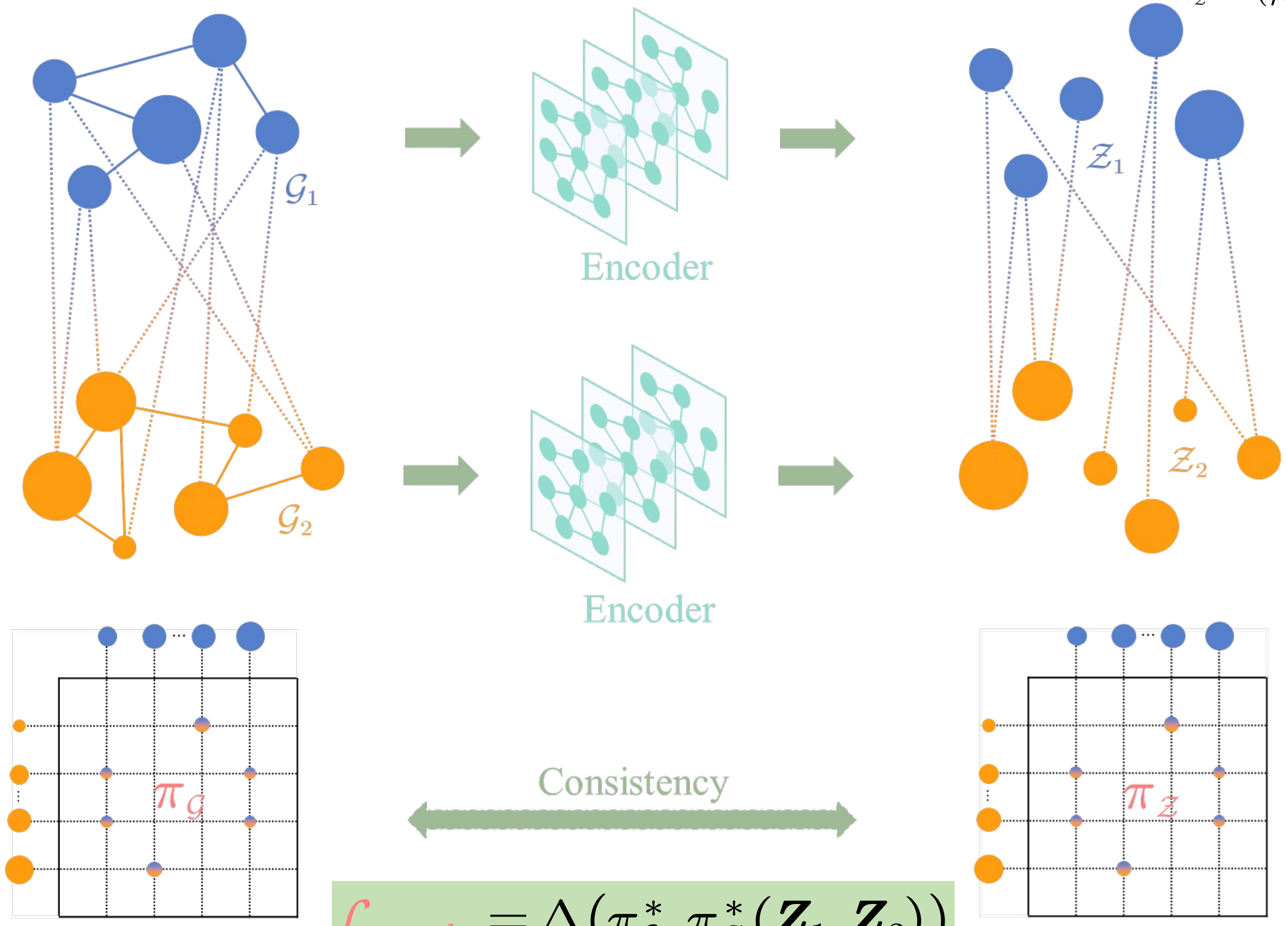
	1	2	3	4	5	6
1	0.4	0.6	0	0	0	0
2	0	0	1	0	0	0
3	0	0	0	1	0	0
4	0	0	0	0	1	0
5	0	0	0	0	0	1

Optimal Plan Matrix π

Plan Alignment Loss

$$\pi_{\mathcal{G}}^* = \underset{\pi_{\mathcal{G}} \in \Pi(\mu,\nu)}{\operatorname{argmin}} \langle \sigma \boldsymbol{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) + (1-\sigma)\boldsymbol{L}(\boldsymbol{A}_1, \boldsymbol{A}_2) \otimes \pi_{\mathcal{G}}, \pi_{\mathcal{G}} \rangle$$

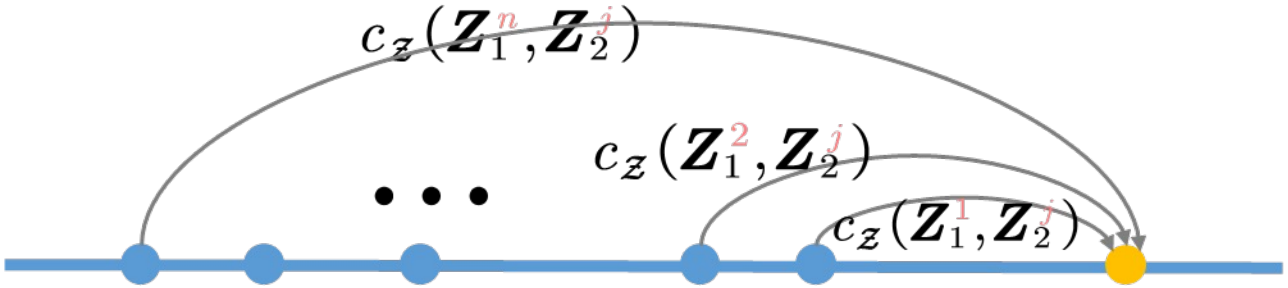
$$\pi_{\mathcal{Z}}^*(\boldsymbol{Z}_1, \boldsymbol{Z}_2) = \underset{\pi_{\mathcal{Z}} \in \Pi(\mu,\nu)}{\operatorname{argmin}} \langle \boldsymbol{J}(\boldsymbol{Z}_1, \boldsymbol{Z}_2), \pi_{\mathcal{Z}} \rangle$$



$$\mathcal{L}_{\text{match}} = \Delta(\pi_{\mathcal{G}}^*, \pi_{\mathcal{Z}}^*(\boldsymbol{Z}_1, \boldsymbol{Z}_2))$$

Cost and Structure

To guide the encoder to learn a representation retaining structural information inside the graph, we also calibrate the cost matrix $J(Z_1, Z_2)$, which implies the implicit structure relationships between nodes, in the representation space.



Implicit Structure Loss

$$\pi_{\mathcal{G}}^* = \underset{\pi_{\mathcal{G}} \in \Pi(\mu,\nu)}{\operatorname{argmin}} \left\langle \sigma \boldsymbol{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) + (1-\sigma) \boldsymbol{L}(\boldsymbol{A}_1, \boldsymbol{A}_2) \otimes \pi_{\mathcal{G}}, \pi_{\mathcal{G}} \right\rangle$$

$$\pi_{\mathcal{Z}}^* = \underset{\pi_{\mathcal{Z}} \in \Pi(\mu,\nu)}{\operatorname{argmin}} \left\langle \boldsymbol{J}(\boldsymbol{Z}_1, \boldsymbol{Z}_2), \pi_{\mathcal{Z}} \right\rangle$$

$$\mathcal{L}_{\text{(im)strc}} = \Delta\left(\sigma \boldsymbol{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) + (1-\sigma) \boldsymbol{L}(\boldsymbol{A}_1, \boldsymbol{A}_2) \otimes \pi_{\mathcal{G}}^*, \;\; \boldsymbol{J}(\boldsymbol{Z}_1, \boldsymbol{Z}_2)\right)$$

Overall Loss

$$\mathcal{L}_{\text{match}} = \Delta\left(\pi^*_{\mathcal{G}}, \pi^*_{\mathcal{Z}}(\boldsymbol{Z}_1, \boldsymbol{Z}_2)\right)$$

$$\mathcal{L}_{\text{(im)strc}} = \Delta\left(\sigma\boldsymbol{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) + (1-\sigma)\boldsymbol{L}(\boldsymbol{A}_1, \boldsymbol{A}_2) \otimes \pi^*_{\mathcal{G}},\ \ \boldsymbol{J}(\boldsymbol{Z}_1, \boldsymbol{Z}_2)\right)$$

$$\boxed{\mathcal{L}_{\text{GALOPA}} = \mathcal{L}_{\text{match}} + \rho\mathcal{L}_{\text{(im)strc}}}$$

Plan vs. Distance

$$\boxed{\begin{matrix}\mathcal{W}(\boldsymbol{Z}_1,\boldsymbol{Z}_2) \\ \| \\ \mathcal{W}_{\mathcal{G}}(\mathcal{G}_1,\mathcal{G}_2)\end{matrix}}\begin{matrix}= \min_{\pi_{\mathcal{Z}}\in\Pi(\mu,\nu)} \langle \boldsymbol{J}(\boldsymbol{Z}_1,\boldsymbol{Z}_2),\pi_{\mathcal{Z}}\rangle \\ \\ = \min_{\pi_{\mathcal{G}}\in\Pi(\mu,\nu)} \langle \sigma\boldsymbol{K}(\boldsymbol{X}_1,\boldsymbol{X}_2)+(1-\sigma)\boldsymbol{L}(\boldsymbol{A}_1,\boldsymbol{A}_2)\otimes\pi_{\mathcal{G}},\ \ \pi_{\mathcal{G}}\rangle\end{matrix}$$

$$\boxed{\mathcal{L}_{\text{dist}} = \left|\mathcal{W}_{\mathcal{G}}(\mathcal{G}_1,\mathcal{G}_2) - \mathcal{W}(\boldsymbol{Z}_1,\boldsymbol{Z}_2)\right|}$$

Plan vs. Distance

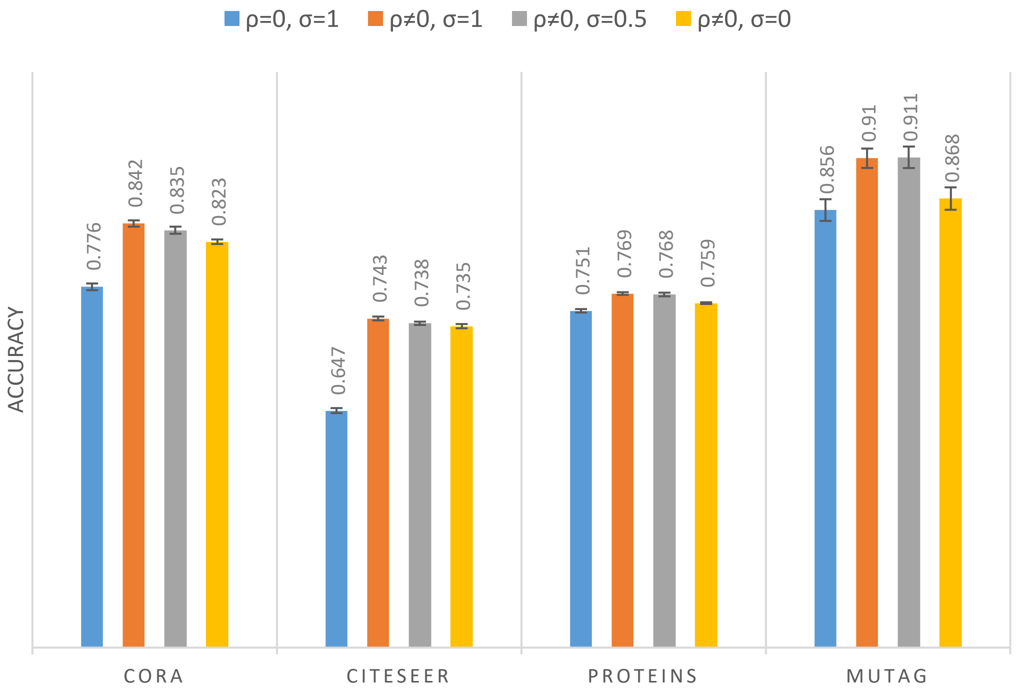
- **Result**: The model using the plan as objective significantly outperforms the counterpart models using the distance;
- **Explanation**: The optimal transport plan for the discrete OT problem is not unique in general and the optimal distance may correspond to several plans.

Node Feature vs. Edge

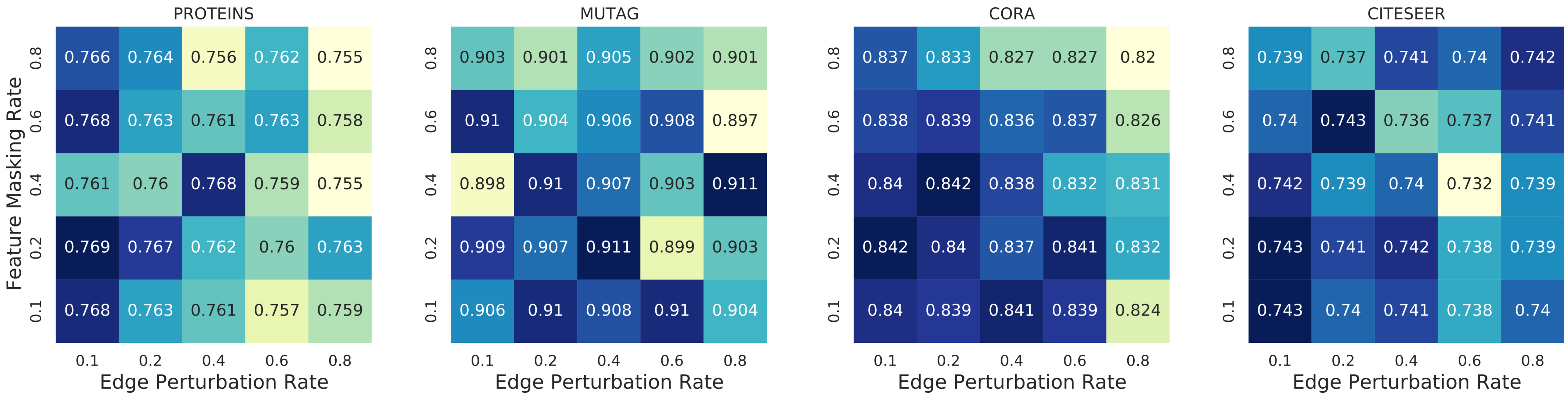
$$\pi_{\mathcal{G}}^{*} = \underset{\pi_{\mathcal{G}} \in \Pi(\mu,\nu)}{\operatorname{argmin}} \langle \sigma \boldsymbol{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) + (1-\sigma) \boldsymbol{L}(\boldsymbol{A}_1, \boldsymbol{A}_2) \otimes \pi_{\mathcal{G}}, \pi_{\mathcal{G}} \rangle$$

$$\mathcal{L}_{\text{GALOPA}} = \mathcal{L}_{\text{match}} + \rho \mathcal{L}_{\text{(im)strc}}$$

- **Result**: With only node attribute for the calculation of the plan, the model achieves outstanding performance. However, if we remove the implicit structure constraint, the model's performance deteriorates dramatically;
- **Explanation**: The implicit structure information provides correction information to the encoder when explicit structural information (edges) is missing;



Robustness



- **Result**: The performance of our model does not change much even when the original graph is perturbed heavily. It validates that the alignment of optimal transport between the source space and target space is indeed free from the label-invariant assumption.

Comparison with the State-of-the-art

Table 1: Mean node classification accuracy (%) for supervised and unsupervised models. The highest performance of unsupervised models is highlighted in **boldface**. OOM indicates Out-Of-Memory.

Model	CORA	CITESEER	PUBMED	WikiCS	Amz-Comp.	Amz-Photo	Coauthor-CS	Average
MLP	47.92 ± 0.41	49.31 ± 0.26	69.14 ± 0.34	71.98 ± 0.42	73.81 ± 0.21	78.53 ± 0.32	90.37 ± 0.19	68.72 ± 0.31
GCN	81.54 ± 0.68	70.73 ± 0.65	79.16 ± 0.25	93.02 ± 0.11	86.51 ± 0.54	92.42 ± 0.22	93.03 ± 0.31	85.20 ± 0.39
DEEPWALK	70.72 ± 0.63	51.39 ± 0.41	73.27 ± 0.86	74.42 ± 0.13	85.68 ± 0.07	89.40 ± 0.11	84.61 ± 0.22	75.64 ± 0.35
NODE2VEC	71.08 ± 0.91	47.34 ± 0.84	66.23 ± 0.95	71.76 ± 0.14	84.41 ± 0.14	89.68 ± 0.19	85.16 ± 0.04	73.67 ± 0.46
GAE	71.49 ± 0.41	65.83 ± 0.40	72.23 ± 0.71	73.97 ± 0.16	85.27 ± 0.19	91.62 ± 0.13	90.01 ± 0.71	78.63 ± 0.39
VGAE	77.31 ± 1.02	67.41 ± 0.24	75.85 ± 0.62	75.56 ± 0.28	86.40 ± 0.22	92.16 ± 0.12	92.13 ± 0.16	80.97 ± 0.38
DGI	82.34 ± 0.71	71.83 ± 0.54	76.78 ± 0.31	75.37 ± 0.13	84.01 ± 0.52	91.62 ± 0.42	92.16 ± 0.62	82.02 ± 0.46
GMI	82.39 ± 0.65	71.72 ± 0.15	79.34 ± 1.04	74.87 ± 0.13	82.18 ± 0.27	90.68 ± 0.18	OOM	–
MVGRL	83.45 ± 0.68	73.28 ± 0.48	80.09 ± 0.62	77.51 ± 0.06	87.53 ± 0.12	91.74 ± 0.08	92.11 ± 0.14	83.67 ± 0.31
GRACE	81.92 ± 0.89	71.21 ± 0.64	80.54 ± 0.36	78.19 ± 0.10	86.35 ± 0.44	92.15 ± 0.25	92.91 ± 0.20	83.32 ± 0.41
GCA	82.38 ± 0.47	71.51 ± 0.32	80.89 ± 0.28	78.29 ± 0.36	87.88 ± 0.26	92.33 ± 0.68	92.64 ± 0.34	83.70 ± 0.39
BGRL	81.30 ± 0.54	72.06 ± 0.63	80.52 ± 0.30	76.13 ± 0.18	$\mathbf{89.09 \pm 0.51}$	92.15 ± 0.32	92.33 ± 0.39	83.37 ± 0.41
GALOPA	$\mathbf{84.21 \pm 0.30}$	$\mathbf{74.34 \pm 0.18}$	$\mathbf{84.57 \pm 0.34}$	$\mathbf{81.23 \pm 0.19}$	88.65 ± 0.11	$\mathbf{92.77 \pm 0.40}$	$\mathbf{93.04 \pm 0.25}$	$\mathbf{85.54 \pm 0.25}$

Comparison with the State-of-the-art

Table 2: Supervised and unsupervised representation learning classification accuracy (%) along with average accuracy of the algorithms on TU datasets. **Bold** indicates the best performance for unsupervised methods on each dataset. '–' means that the results are unavailable.

Model	PROTEINS	DD	MUTAG	NCI1	COLLAB	IMDB-B	Average
GCN	74.92 ± 0.33	76.24 ± 0.14	85.63 ± 0.24	80.20 ± 0.14	79.01 ± 0.18	70.45 ± 0.37	77.74 ± 0.23
GIN	76.28 ± 0.28	78.91 ± 0.13	89.47 ± 0.16	82.75 ± 0.19	80.23 ± 0.19	73.70 ± 0.60	80.22 ± 0.25
SP	75.07 ± 0.54	>1d	85.25 ± 0.24	73.53 ± 0.16	–	55.62 ± 0.02	–
GK	71.67 ± 0.55	78.53 ± 0.03	81.71 ± 0.21	66.06 ± 0.12	71.81 ± 0.31	65.93 ± 0.10	72.61 ± 0.22
WL	72.92 ± 0.56	79.78 ± 0.36	80.76 ± 0.30	80.01 ± 0.50	69.30 ± 0.42	72.30 ± 0.44	75.84 ± 0.43
WLPM	–	78.79 ± 0.38	87.13 ± 0.42	$\mathbf{86.32 \pm 0.19}$	–	–	–
FGW	74.50 ± 0.23	–	88.34 ± 0.12	86.24 ± 0.31	–	62.97 ± 0.24	–
DGK	73.21 ± 0.61	74.79 ± 0.32	87.51 ± 0.65	79.98 ± 0.36	64.43 ± 0.48	67.09 ± 0.37	74.50 ± 0.46
MLG	41.23 ± 0.27	>1d	87.94 ± 0.16	>1d	>1d	66.67 ± 0.30	–
NODE2VEC	57.58 ± 0.36	–	72.62 ± 1.02	54.93 ± 0.16	56.12 ± 0.02	50.25 ± 0.09	–
SUB2VEC	53.06 ± 0.56	54.33 ± 0.24	61.17 ± 1.59	52.82 ± 0.15	55.26 ± 0.15	55.34 ± 0.15	55.33 ± 0.47
GRAPH2VEC	73.33 ± 0.21	79.32 ± 0.29	83.28 ± 0.93	73.21 ± 0.18	71.10 ± 0.54	71.16 ± 0.05	75.23 ± 0.36
INFOGRAPH	74.44 ± 0.31	72.85 ± 1.78	89.01 ± 1.13	76.20 ± 1.06	70.05 ± 1.13	$\mathbf{73.03 \pm 0.87}$	75.93 ± 1.04
GRAPHCL	74.39 ± 0.45	78.62 ± 0.40	86.80 ± 1.34	77.87 ± 0.41	71.36 ± 1.15	71.14 ± 0.44	76.69 ± 0.69
AD-GCL	73.28 ± 0.46	75.79 ± 0.87	88.74 ± 1.85	73.91 ± 0.77	72.02 ± 0.56	70.21 ± 0.68	75.65 ± 0.86
JOAOV2	74.13 ± 0.51	77.32 ± 0.29	87.17 ± 1.09	78.40 ± 0.17	69.19 ± 0.16	70.37 ± 0.37	76.09 ± 0.43
RGCL	75.03 ± 0.43	78.86 ± 0.48	87.66 ± 1.01	78.14 ± 1.08	70.92 ± 0.65	71.85 ± 0.84	77.07 ± 0.74
SIMGRACE	75.23 ± 0.19	77.45 ± 1.03	89.27 ± 1.39	79.10 ± 0.25	71.37 ± 0.44	71.45 ± 0.29	77.31 ± 0.59
GALOPA	$\mathbf{76.93 \pm 0.18}$	$\mathbf{83.87 \pm 0.42}$	$\mathbf{91.11 \pm 1.27}$	77.86 ± 0.36	$\mathbf{73.20 \pm 0.37}$	70.72 ± 0.48	$\mathbf{78.94 \pm 0.51}$

Thanks!