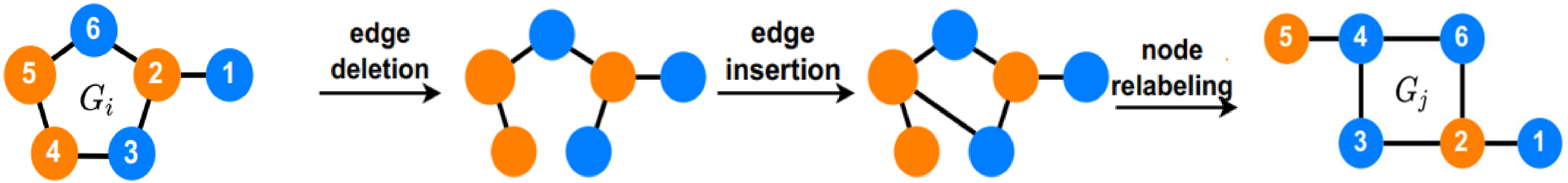
Efficient Graph Similarity Computation with Alignment Regularization

Wei Zhuo, Guang Tan

Shenzhen Campus of Sun Yat-sen University (SYSU), China

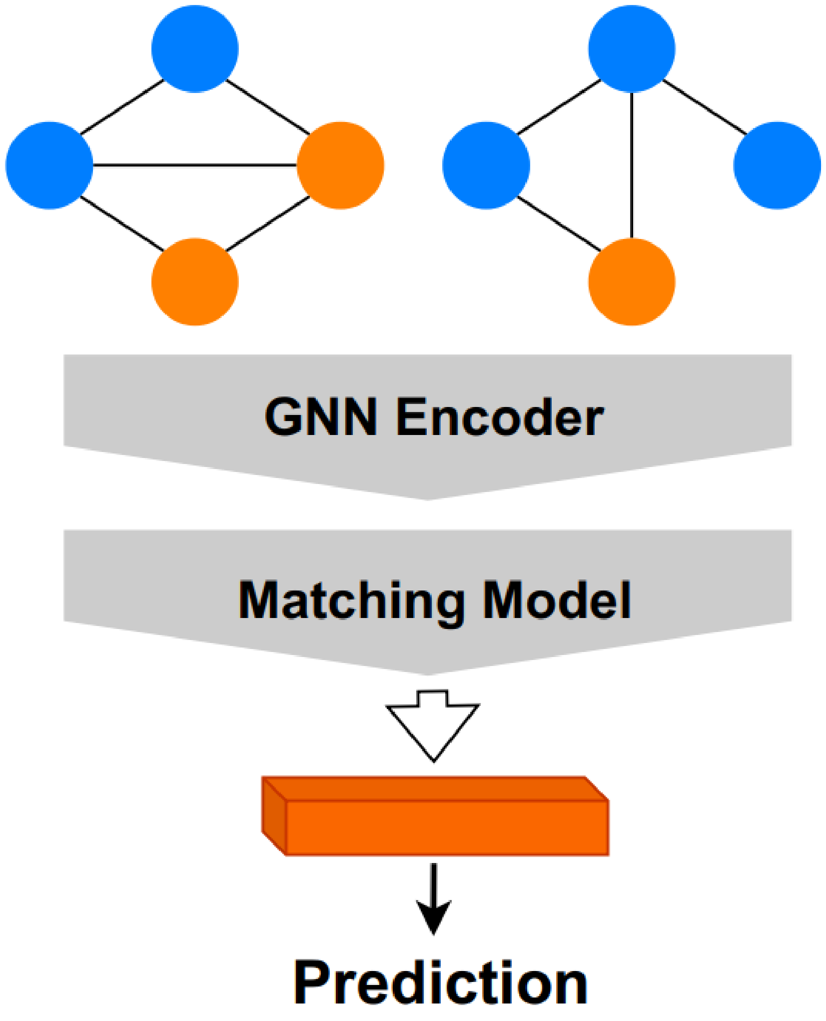
zhuow5@mail2.sysu.edu.cn

Graph Similarity Computation: Graph Edit Distance (GED)



$$c = \min_{\pi} \sum_{k,l} \left| \left(\boldsymbol{A}_i - \pi(\boldsymbol{A}_j) \right) [k, l] \right|$$

GNN for GED Computation

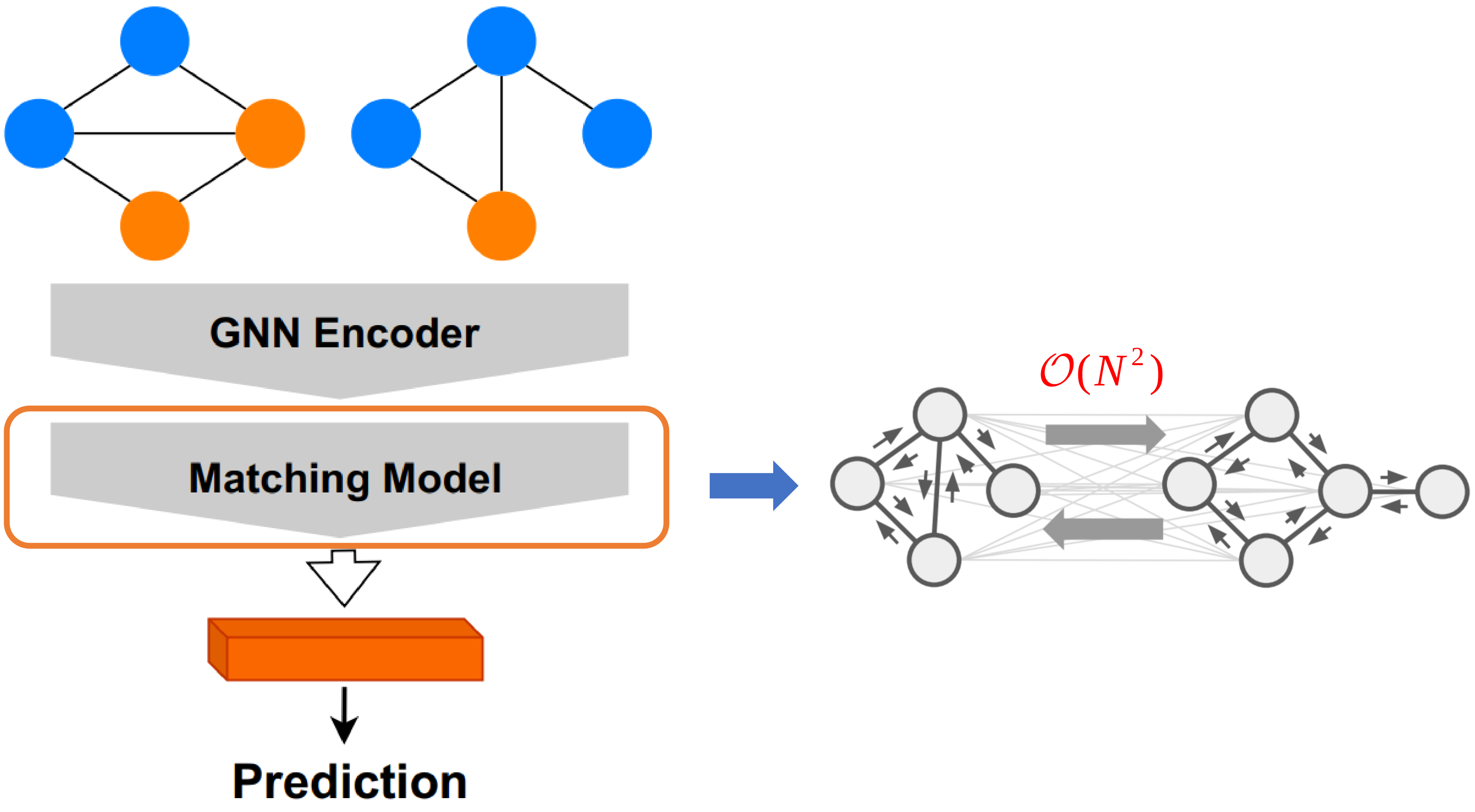


GNN-based Models

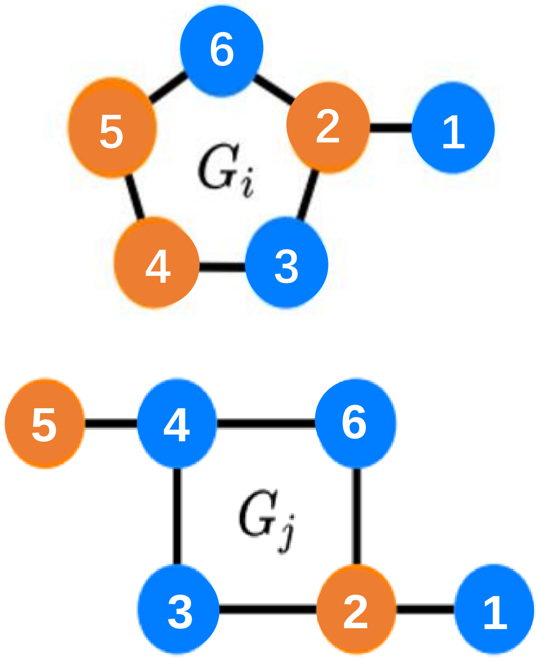
SimGNN, GraphSim, MGMN, GMN, GSimCNN, GCN-Mean

Method	Time Complexity	
A*	$O(N_1{}^{N_2})$	Combinatorial Search Methods
Beam	subexponential	
Hungarian	$O((N_1 + N_2)^3)$	
VJ	$O((N_1 + N_2)^3)$	
Siamese MPNN	$O(\max(E_1, E_2, N_1 N_2))$	GNN-based Methods
GCNMean	$O(\max(E_1, E_2))$	
GraphSim	$O(\max(N_1, N_2)^2)$	
SimGNN	$O(\max(N_1, N_2)^2)$	
MGMN	$O(\max(N_1, N_2)^2)$	
GMN	$O(\max(N_1, N_2)^2)$	

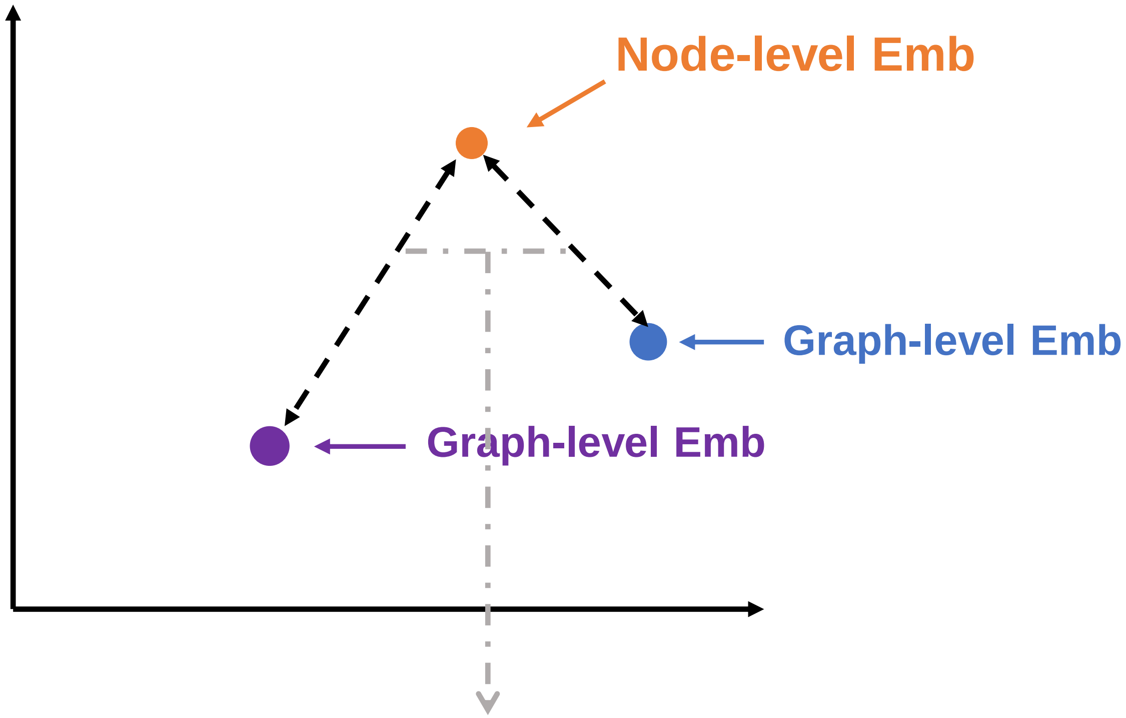
Limitation of GNN-base GSC Model



Analyzing GED in Embedding Space



The best matching between two graphs can be inferred by minimizing the difference between the intra-graph node-graph similarity and cross-graph node-graph similarity.



$$\min \boxed{\mathcal{L}_{\text{AReg}}} = \frac{1}{L}\sum_{\ell}^{L}\left(\gamma_i^{(\ell)} + \gamma_j^{(\ell)} + \left\|\gamma_i^{(\ell)} - \gamma_j^{(\ell)}\right\|_2\right)$$

Alignment Regularization

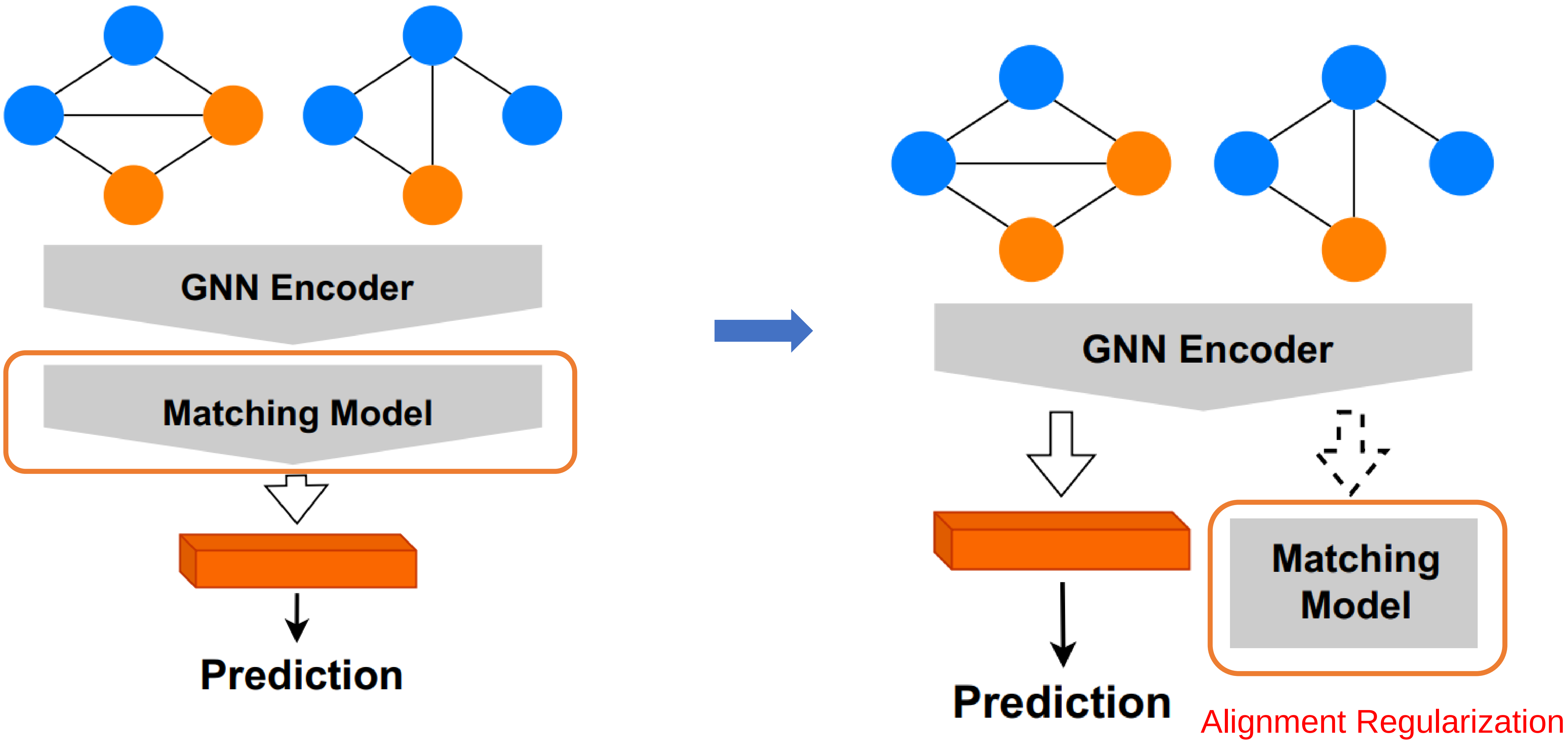
Alignment Regularization

$$\mathcal{L}_{\text{AReg}} = \frac{1}{L}\sum_{\ell}^{L}\left(\boxed{\gamma_i^{(\ell)}} + \gamma_j^{(\ell)} + \left\|\gamma_i^{(\ell)} - \gamma_j^{(\ell)}\right\|_2\right)$$

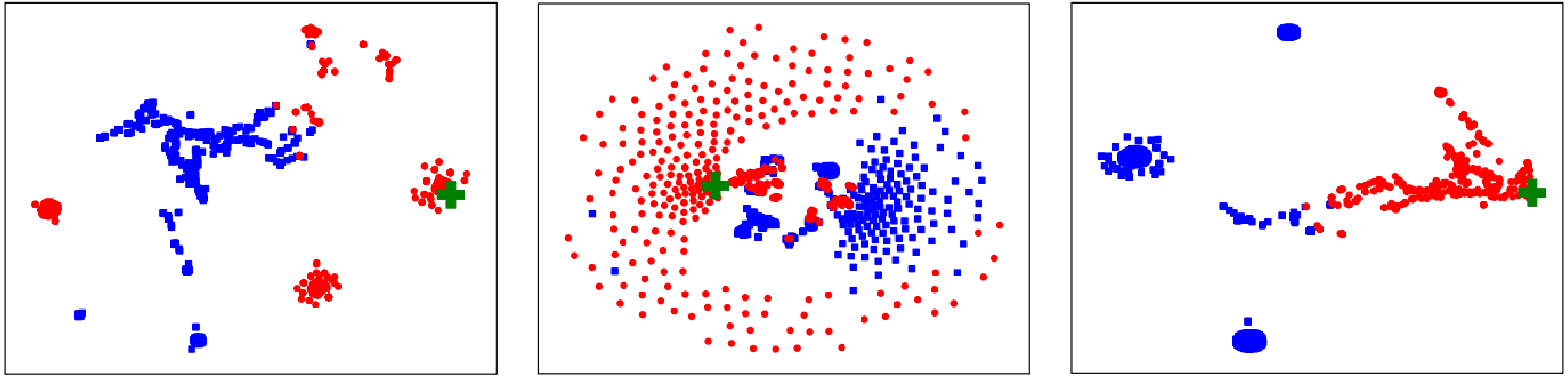
$$\gamma_i = \sum_{k}^{N} \|\text{DIST}\,(\underbrace{f_\theta(\boldsymbol{A}_i[k])}_{\text{Node Level}}, \underbrace{g_\phi(\boldsymbol{A}_i)}_{\text{Graph Level}}) - \text{DIST}\,(f_\theta(\boldsymbol{A}_i[k]), g_\phi(\pi(\boldsymbol{A}_j)))\|_2$$

$$f_\theta^{(\ell)}(\mathbf{A}_i[k]) = \text{MLP}_\theta^{(\ell)}\left(\left(1+\xi^{(\ell)}\right)\mathbf{H}_i^{(\ell-1)}[k] + \mathbf{A}_i[k]\mathbf{H}_i^{(\ell-1)}\right) \qquad g_\phi^{(\ell)}(\mathbf{A}_i) = \text{MLP}_\phi^{(\ell)}\left(\sum_k^N f_\theta^{(\ell)}(\mathbf{A}_i[k])\right)$$

GNN for GED Computation



Multi-Scale GED Discriminator



(a) NTN (b) ℓ_2 distance (c) NTN + ℓ_2 distance

$$\mathcal{L} = \frac{1}{n}\sum_{(G_i,G_j)\in\mathcal{D}\times\mathcal{D}} \mathrm{MSE}\left(\boxed{\alpha s_{\mathrm{NTN}}\left(G_i, G_j\right) + \beta s_p\left(G_i, G_j\right)}, \mathbf{S}_{ij}\right) + \boxed{\lambda\mathcal{L}_{\mathrm{AReg}}}$$

Multi-Scale GED Discriminator

Alignment Regularization
(ignore during inference)

Efficient g**R**aph s**I**milarity **C**omputation (ERIC)

Accuracy and Efficiency Comparison

Table 1: Evaluation on benchmarks. **Bold** : best.

	AIDS700					LINUX					IMDB					NCI109				
	mse $(\times 10^{-3})\downarrow$	$\rho\uparrow$	$\tau\uparrow$	$p@10\uparrow$	$p@20\uparrow$	mse $(\times 10^{-3})\downarrow$	$\rho\uparrow$	$\tau\uparrow$	$p@10\uparrow$	$p@20\uparrow$	mse $(\times 10^{-3})\downarrow$	$\rho\uparrow$	$\tau\uparrow$	$p@10\uparrow$	$p@20\uparrow$	mse $(\times 10^{-3})\downarrow$	$\rho\uparrow$	$\tau\uparrow$	$p@10\uparrow$	$p@20\uparrow$
Beam	12.090	0.609	0.463	0.481	0.493	9.268	0.827	0.714	0.973	0.924	-	-	-	-	-	-	-	-	-	-
VJ	29.157	0.517	0.383	0.310	0.345	63.86	0.581	0.450	0.287	0.251	-	-	-	-	-	-	-	-	-	-
Hungarian	25.296	0.510	0.378	0.360	0.392	29.81	0.638	0.517	0.913	0.836	-	-	-	-	-	-	-	-	-	-
SimGNN	1.573	0.835	0.678	0.417	0.489	2.479	0.912	0.791	0.635	0.650	1.437	0.871	0.752	0.710	0.769	7.767	0.576	0.435	0.023	0.040
GraphSim	2.014	0.839	0.662	0.401	0.499	0.762	0.953	0.882	0.956	0.951	1.924	0.825	0.821	0.813	0.825	6.752	0.557	0.497	0.086	0.092
GMN	4.610	0.672	0.497	0.200	0.263	2.571	0.906	0.763	0.888	0.856	4.320	0.665	0.601	0.588	0.593	11.710	0.336	0.358	0.017	0.019
EGSC	1.676	0.888	0.723	0.604	0.708	0.214	0.984	0.897	0.987	0.989	0.573	**0.939**	**0.829**	0.872	0.883	9.356	0.545	0.414	0.055	0.078
MGMN	2.297	0.904	0.736	0.456	0.534	2.040	0.965	0.858	0.956	0.920	0.496	0.881	0.803	0.874	0.861	9.631	0.492	0.426	0.015	0.051
ERIC	**1.374**	**0.906**	**0.741**	**0.685**	**0.758**	**0.107**	**0.988**	**0.908**	**0.994**	**0.999**	**0.385**	0.890	0.791	**0.882**	**0.891**	**6.327**	**0.591**	**0.525**	**0.118**	**0.127**

- ERIC consistently achieve state-of-the-arts performance across all evaluation metric.
- Alignment Regularization can be incorporated into existing methods and improve their performance, such as SimGNN and EGSC.
- ERIC is faster than all baseline models in the inference stage.

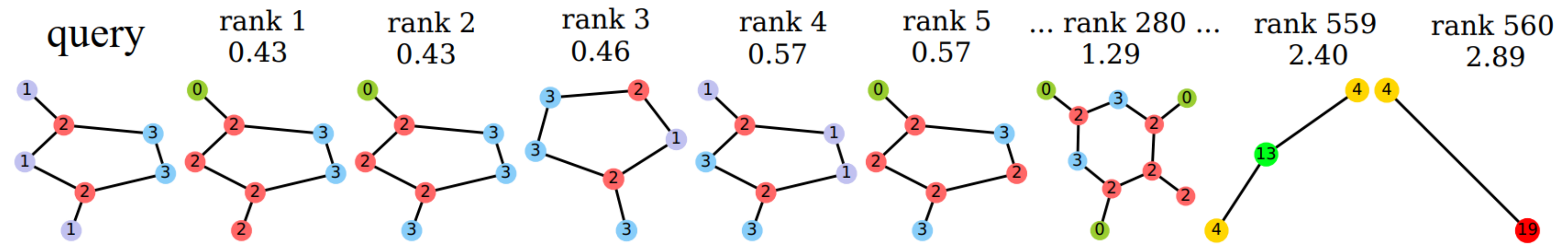
Table 4: Inference time (sec) comparison.

Dataset	SimGNN	GraphSim	GMN	MGMN	EGSC	ERIC
AIDS700	10.773	14.043	23.975	11.337	8.763	**6.662**
LINUX	19.347	31.238	82.489	22.574	21.573	**18.969**
IMDB	225.682	379.480	1253.551	357.933	133.437	**48.750**
NCI109	2913.178	3463.620	$> 10^4$	3726.834	2097.405	**1763.356**

Table 3: Transferability study of AReg on AIDS700 and LINUX.

	AIDS700			LINUX		
	mse	ρ	$p@10$	mse	ρ	$p@10$
SimGNN	1.573	0.835	0.417	2.479	0.912	0.635
SimGNN+AReg	**1.439**	**0.858**	**0.506**	**1.974**	**0.945**	**0.658**
EGSC	1.676	0.888	0.604	0.214	0.984	0.987
EGSC+AReg	**1.478**	**0.904**	**0.643**	**0.142**	**0.989**	**0.992**

Visualization



(a) The exact similarity ranking based on $A^\star$