Imbalanced Graph Classification via Graph-of-Graph Neural Networks

Yu Wang[1] 

Yuying Zhao[1] 

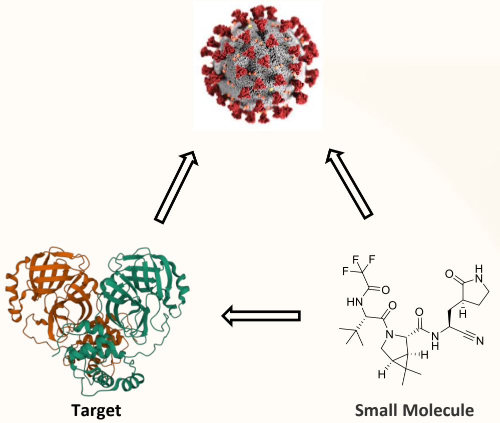
Neil Shah[2] 

Tyler Derr[1] 

1. Network and Data Science Lab, Vanderbilt University
2. Snap Research

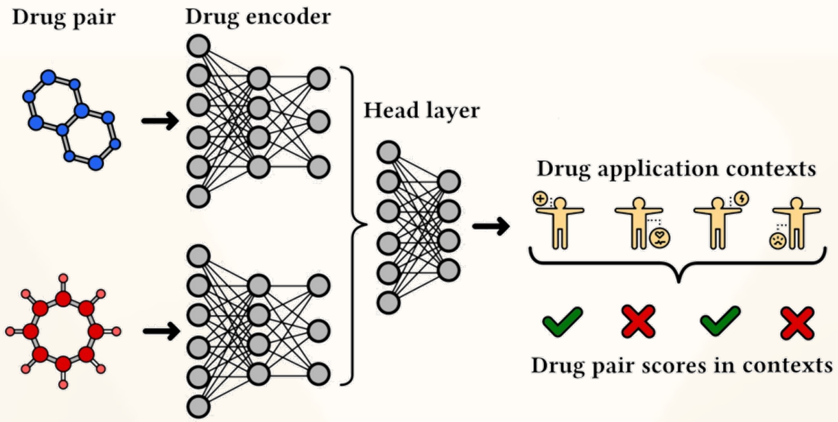
Background – Graph Classification

Drug Discovery



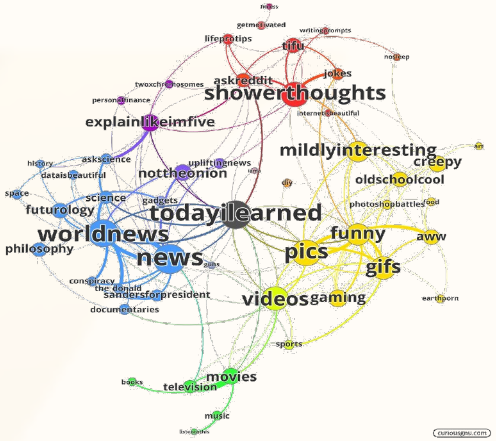
[1] Liana Zucco et al. (2020)

Drug-Drug Interaction



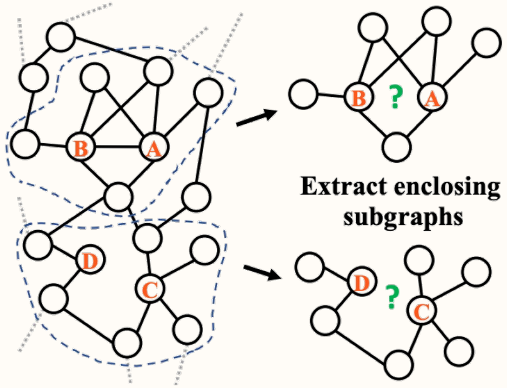
[2] Rozemberczik et al. (2022)

Social Topic Classification



[3] Hamilton et al. (2017)

Link Prediction

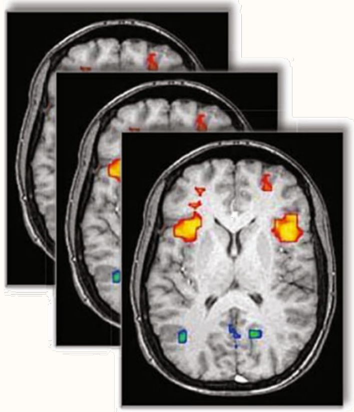


[4] Zhang et al. (2018)

Image Classification

[5] Vasudevan et al. (2022)

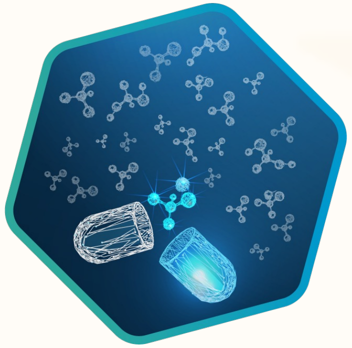
Brain Classification



[6] Bi et al. (2020)

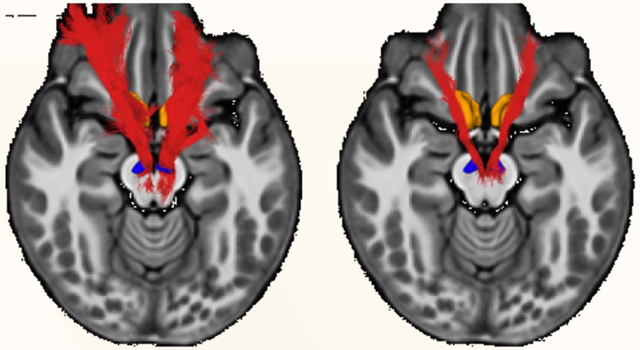
Problem – Imbalanced Graph Classification

Drug Discovery



HTS Hit Ratio
0.05% to 0.5%
[7] Bajorath et al. 2002

ASD Brain Classification

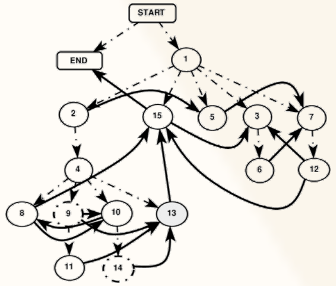


Typical : Autism
46 : 1
Autism Statistics. 2021

Fake News Detection

0.15%
[8] Dou et al. 2021

Malware Detection



0.01% Google, 2% Android,
[9] Oak et al. 2019

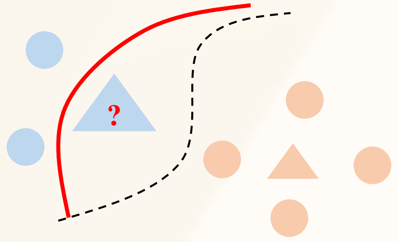
Biased Learning

Minority

$$\mathcal{L} = \mathcal{L}_{G_1} + \boxed{\mathcal{L}_{G_2}} + ..$$

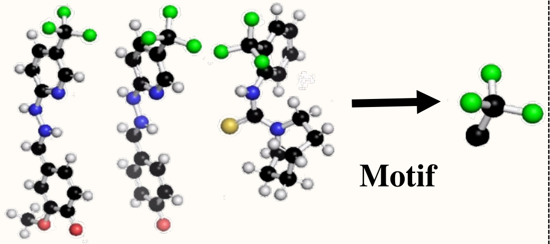
Imbalanced Graph Issue

Population Risk



Decision boundary

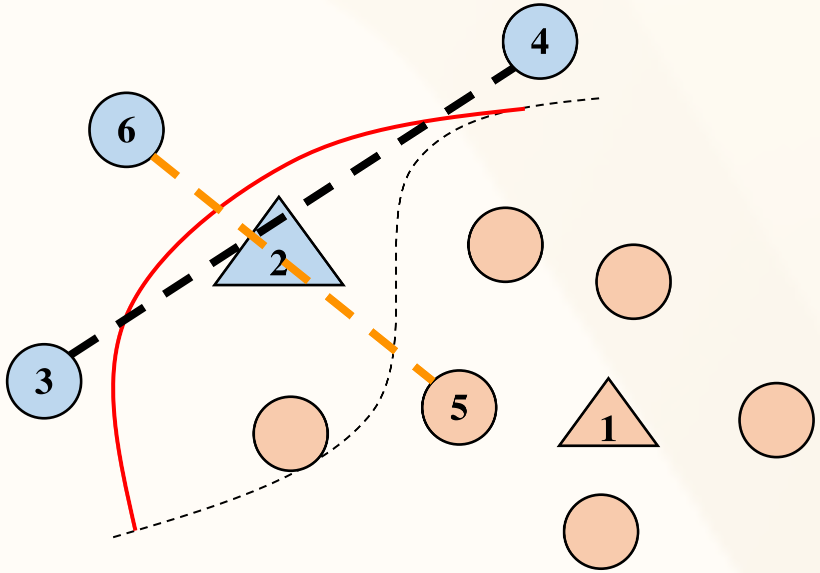
Imbalanced Topology



[10] Liu et al. 2022

Method – Mitigating Population risk

Interpolation

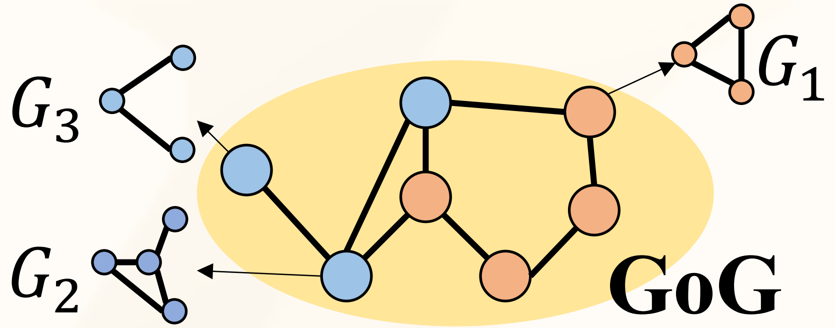


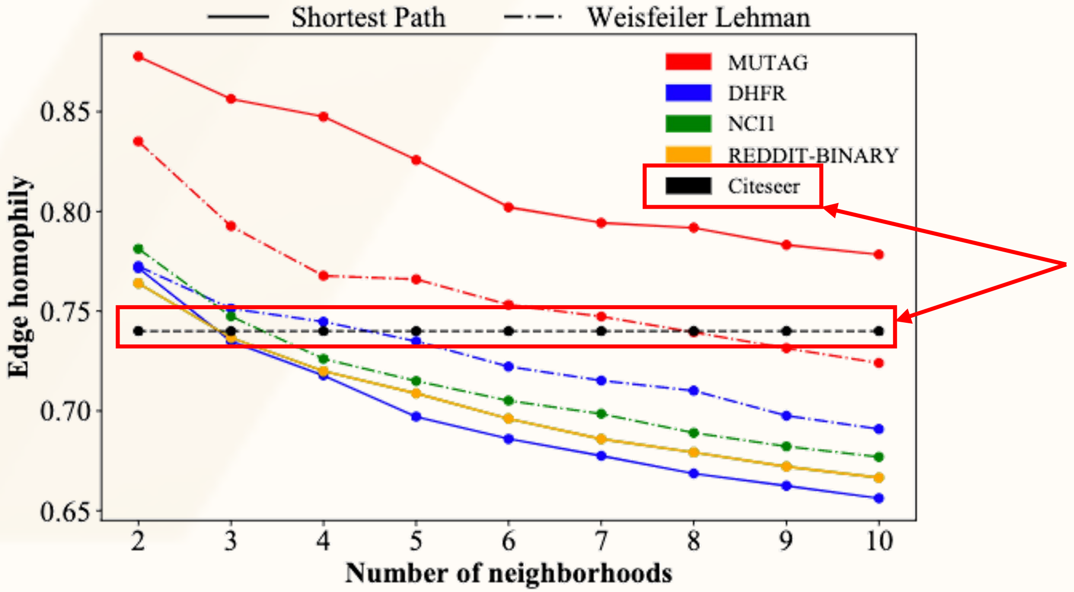
Train – Major
Train – Minor
Test – Major
Test – Minor

SMOTE
[11] Chawla et al. 2002

Mixup
[12] Zhang et al. 2017

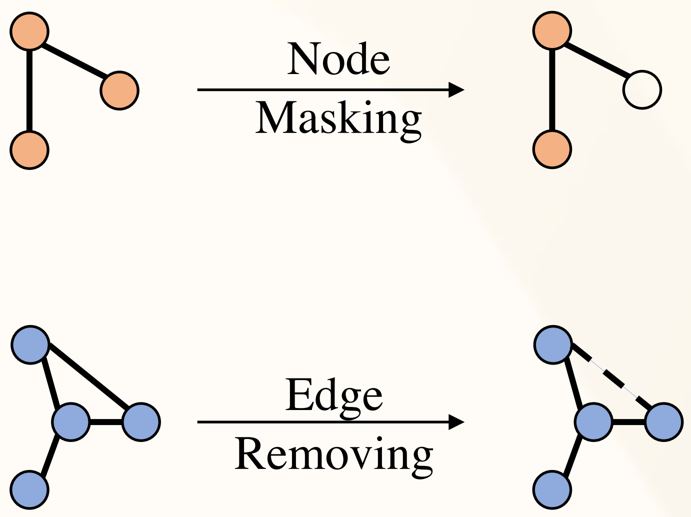
Graph-of-Graphs (GoG)





Method – Augmentation with Consistency Regularization

Augmentation

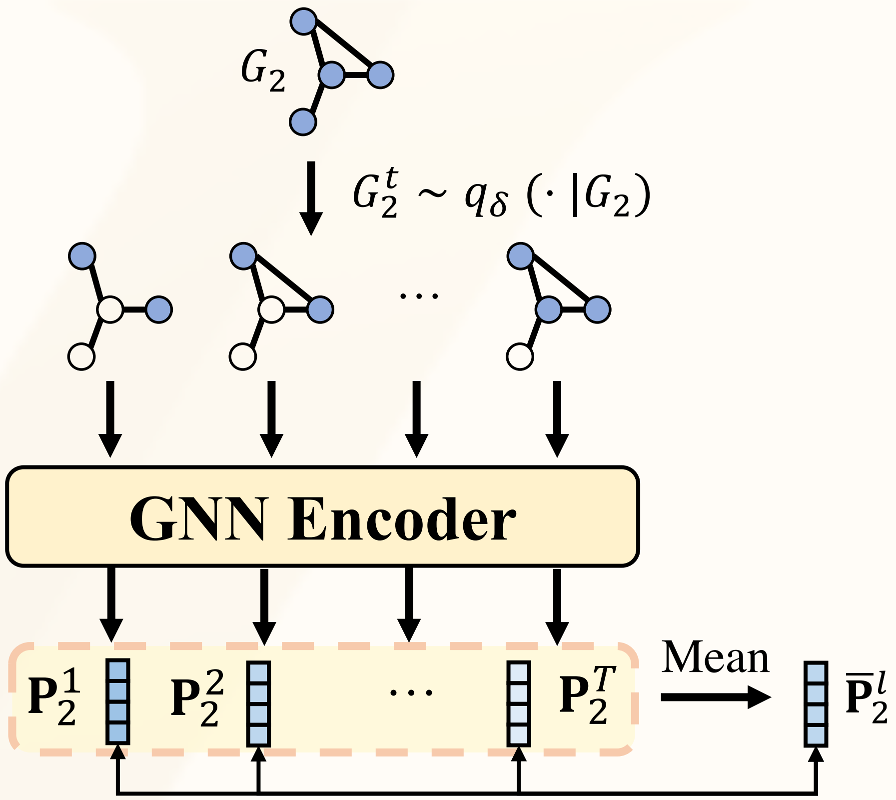


$$\eta_j^{G_i} \sim \text{Bernoulli}(1 - \delta_j^{G_i})$$

$$P(e_{uv} \in \widehat{\mathcal{E}}^{G_i}) = 1 - \delta_{uv}^{G_i}$$

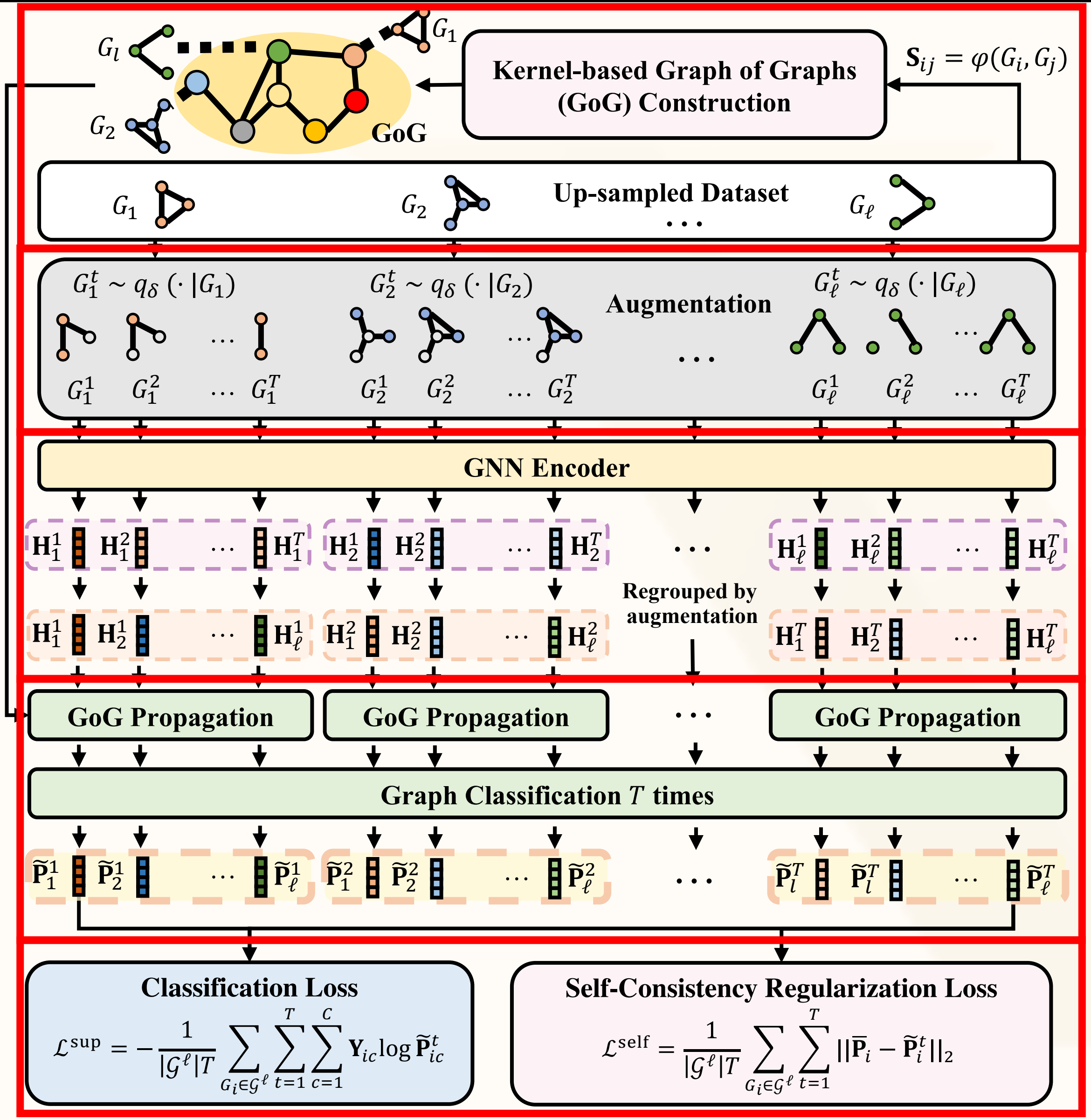
$$\widehat{\mathbf{X}}_j^{G_i} = \eta_j^{G_i} \mathbf{X}_j^{G_i}$$

Consistency Regularization



$$\sum_{t=1}^{T} ||\overline{\mathbf{P}}_i - \mathbf{P}_i^t||_2$$

Framework – Graph-of-Graph Neural Network (G^{2}GNN)



$$\mathbf{S}_{ij} = \phi(G_i, G_j),$$

Framework – Graph-of-Graph Neural Network (G^{2}GNN)

$$\mathbf{X}^{G_i,l+1} = \text{MLP}^l((\mathbf{A}^{G_i} + (1+\epsilon)\mathbf{I})\mathbf{X}^{G_i,l}), \forall l \in \{1, 2, ..., L\}$$

$$\mathbf{P}^{l+1} = (\hat{\mathbf{D}}^{\text{kNN}})^{-1}\hat{\mathbf{A}}^{\text{kNN}}\mathbf{P}^l, l \in \{1, 2, ..., L\}$$

$$\mathcal{L} = \underbrace{-\frac{1}{|\mathcal{G}^\ell|T}\sum_{G_i\in\mathcal{G}^\ell}\sum_{t=1}^{T}\sum_{c=1}^{C}\mathbf{Y}_{ic}\log\widetilde{\mathbf{P}}_{ic}^t}_{\mathcal{L}^{\text{sup}}} + \underbrace{\frac{1}{|\mathcal{G}^\ell|T}\sum_{G_i\in\mathcal{G}^\ell}\sum_{t=1}^{T}||\bar{\mathbf{P}}_i - \widetilde{\mathbf{P}}_i^t||_2}_{\mathcal{L}^{\text{self}}},$$

Experiments - Setting

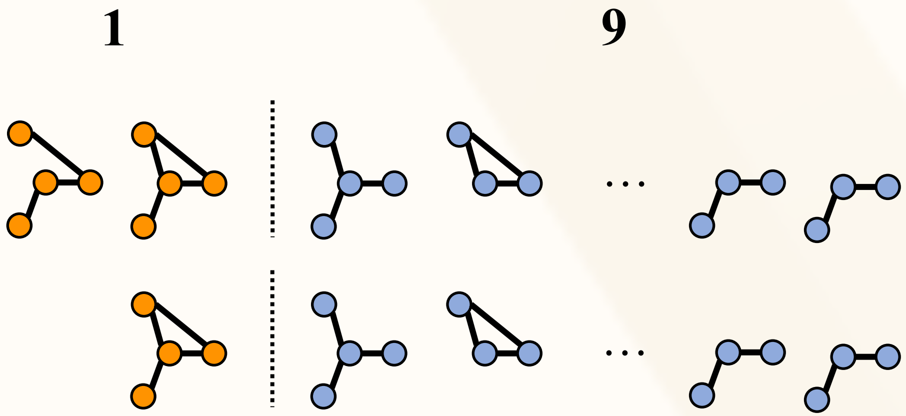
Table 1: Statistics of datasets

Networks	# Graphs	# Avg-Node	# Avg-Edge	# Attr	Time(s)*
PTC-MR [32]	344	14.29	14.69	18	0.257
NCI1 [35]	4110	29.87	32.30	37	11.21
MUTAG [5]	188	17.93	19.79	7	0.212
PROTEINS [43]	1113	39.06	72.82	3	11.36
D&D [30]	1178	284.32	715.66	89	574.71
DHFR [30]	756	42.43	44.54	3	3.70
REDDITB [43]	2000	429.63	497.75	\	3376

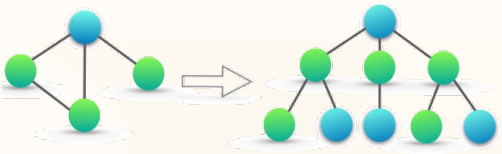
* The column 'time' represents the actual time used for applying Shortest Path kernel to compute S for each dataset.

Train (50%)

Val (25%)

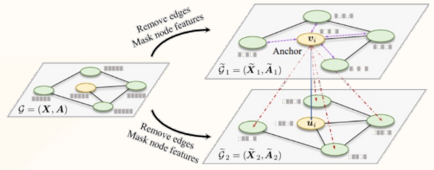


Supervised



GIN [13] Xu et al. 2018

Self-Supervised



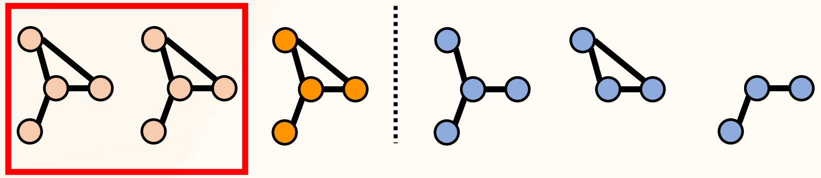
InfoGraph/GraphCL

[14] Sun et al. 2019

[15] You et al. 2021

Imbalanced Strategy:

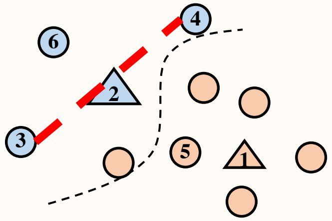
(1) Upsampling – us



(2) Reweighting – rw

$$\mathcal{L} = \mathcal{L}_{G_1} + 3\mathcal{L}_{G_2}$$

(3) SMOTE - st



Experiments - Results

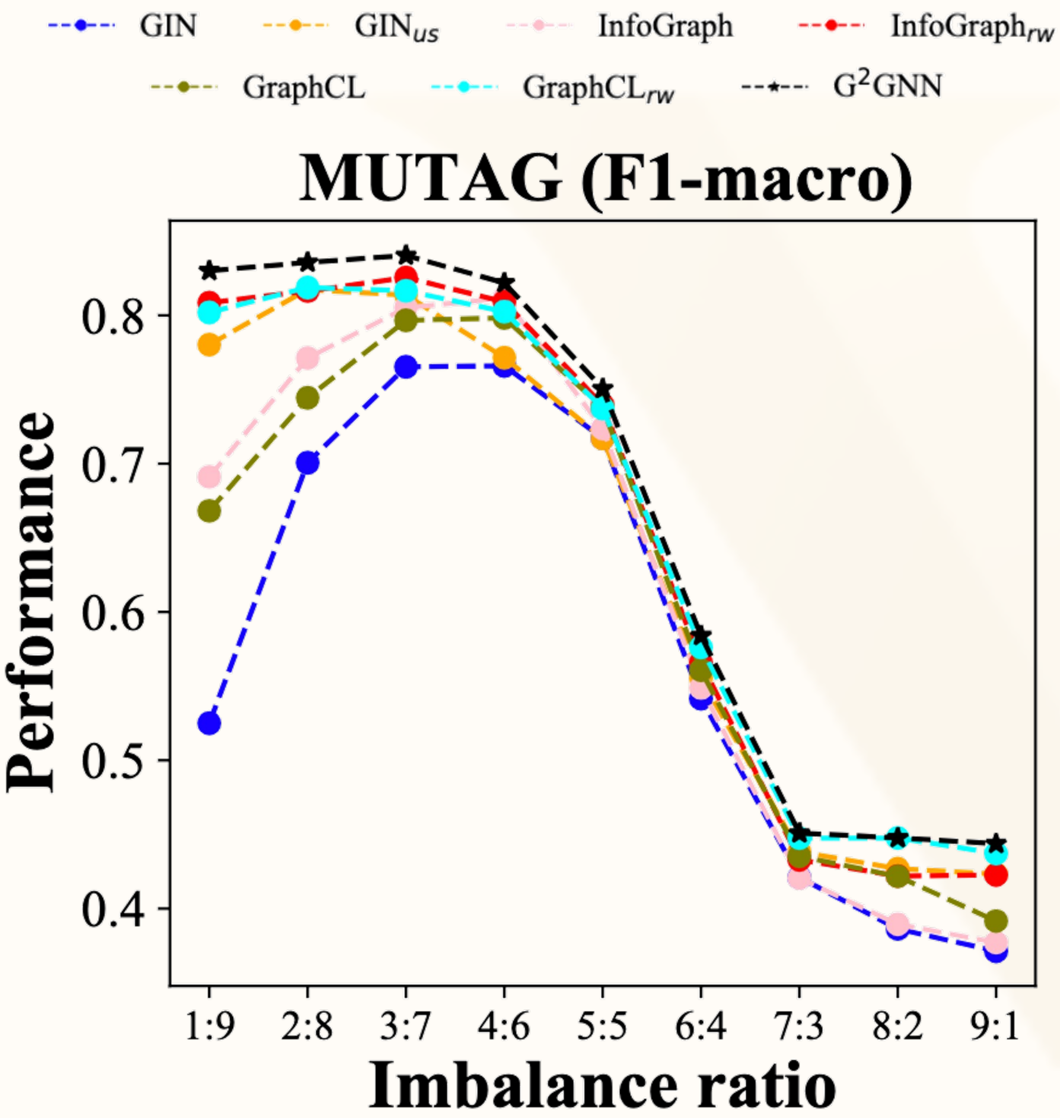
Model	MUTAG (5:45)		PROTEINS (30:270)		D&D (30:270)		NCI1 (100:900)	
	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro
GIN	52.50 ± 18.70	56.77 ± 14.14	25.33 ± 7.53	28.50 ± 5.82	9.99 ± 7.44	11.88 ± 9.49	18.24 ± 7.58	18.94 ± 7.12
GIN_{us}	78.03 ± 7.62	78.77 ± 7.67	65.64 ± 2.67	71.55 ± 3.19	41.15 ± 3.74	70.56 ± 10.28	59.19 ± 4.39	71.80 ± 7.02
GIN_{rw}	77.00 ± 9.59	77.68 ± 9.30	54.54 ± 6.29	55.77 ± 7.11	28.49 ± 5.92	40.79 ± 11.84	36.84 ± 8.46	39.19 ± 10.05
GIN_{st}	74.61 ± 9.66	75.11 ± 9.87	56.07 ± 7.95	57.85 ± 8.70	27.08 ± 8.63	39.01 ± 15.87	40.40 ± 9.63	44.48 ± 12.05
InfoGraph	69.11 ± 9.03	69.68 ± 7.77	35.91 ± 7.58	36.81 ± 6.51	21.41 ± 4.51	27.68 ± 7.52	33.09 ± 3.30	34.03 ± 3.68
InfoGraph_{us}	78.62 ± 6.84	79.09 ± 6.86	62.68 ± 2.70	66.02 ± 3.18	41.55 ± 2.32	71.34 ± 6.76	53.38 ± 1.88	62.20 ± 2.63
InfoGraph_{rw}	80.85 ± 7.75	81.68 ± 7.83	65.73 ± 3.10	69.60 ± 3.68	41.92 ± 2.28	72.43 ± 6.63	53.05 ± 1.12	62.45 ± 1.89
GraphCL	66.82 ± 11.56	67.77 ± 9.78	40.86 ± 6.94	41.24 ± 6.38	21.02 ± 3.05	26.80 ± 4.95	31.02 ± 2.69	31.62 ± 3.05
GraphCL_{us}	80.06 ± 7.79	80.45 ± 7.86	64.21 ± 2.53	65.76 ± 2.61	38.96 ± 3.01	64.23 ± 8.10	49.92 ± 2.15	58.29 ± 3.30
GraphCL_{rw}	80.20 ± 7.27	80.84 ± 7.43	63.46 ± 2.42	64.97 ± 2.41	40.29 ± 3.31	67.96 ± 8.98	50.05 ± 2.09	58.18 ± 3.08
G^2GNN_e	80.37 ± 6.73	81.25 ± 6.87	67.70 ± 2.96	73.10 ± 4.05	43.25 ± 3.91	77.03 ± 9.98	63.60 ± 1.57	72.97 ± 1.81
G^2GNN_n	83.01 ± 7.01	83.59 ± 7.14	67.39 ± 2.99	73.30 ± 4.19	43.93 ± 3.46	79.03 ± 10.78	64.78 ± 2.86	74.91 ± 2.14

(1) Up-sampling, reweighting and Smote alleviate the imbalance issue

(2) Self-supervised learning could also alleviate the imbalance issue

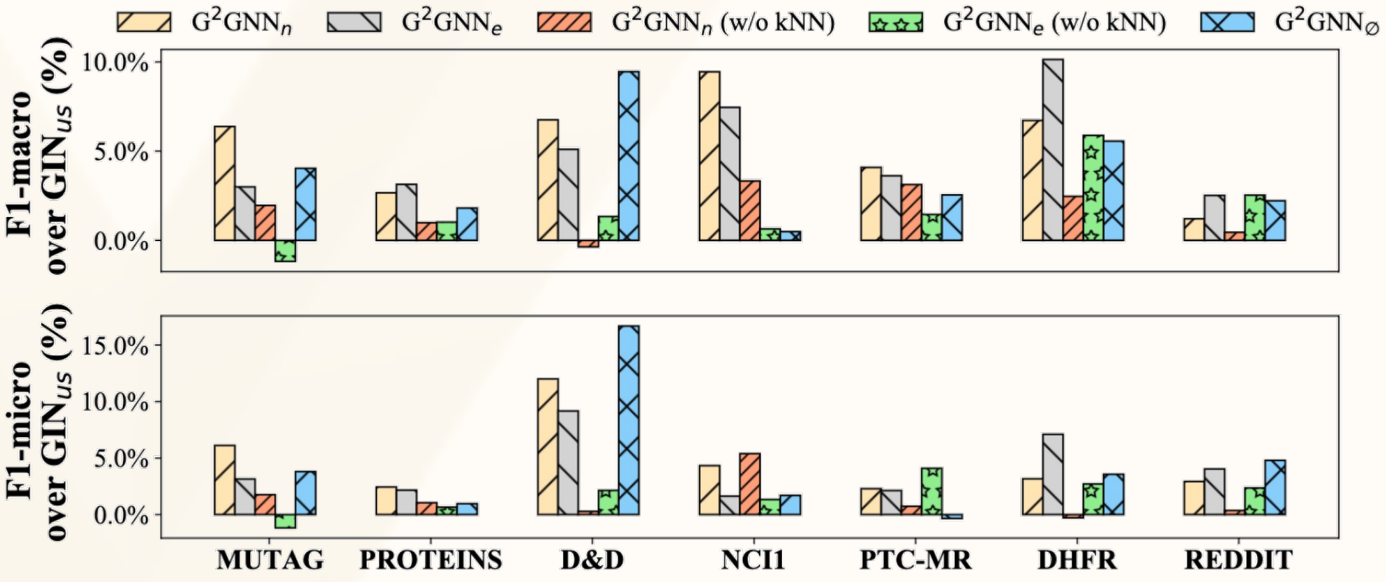
(3) Our $\mathbf{G^2GNN}$ consistently achieves better performance in imbalanced scenario

Experiments - Results



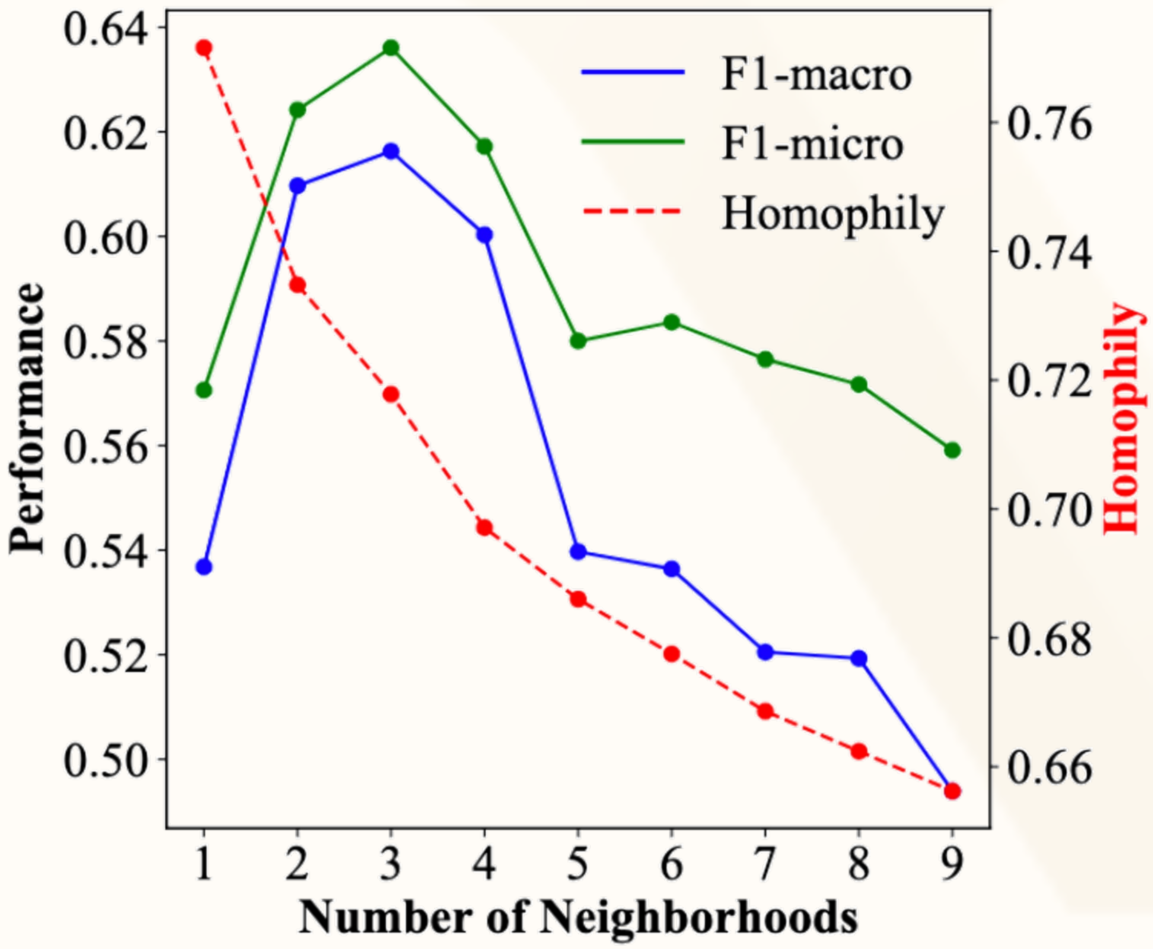
(1) Different minority classes lead to significantly different performance

(2) Topological augmentation cannot guarantee the performance improvement

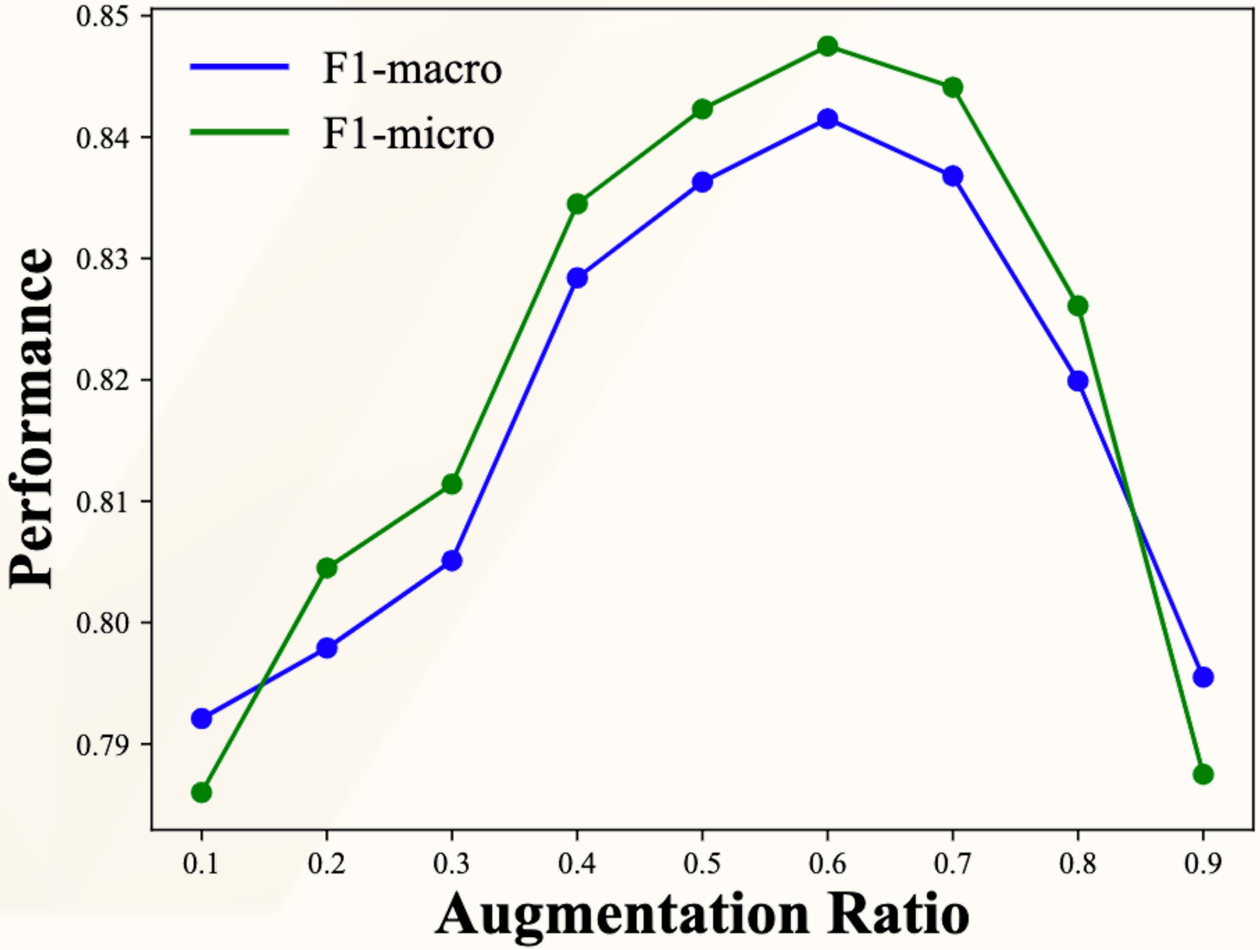


Experiments - Results

Neighbors in GoG



Augmentation ratio

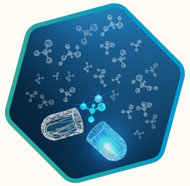


Conclusion

https://github.com/YuWVandy/G2GNN

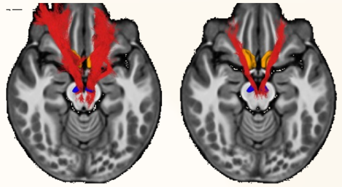
Imbalanced Graph Classification

Drug Discovery



HTS Hit Ratio
0.05% to 0.5%

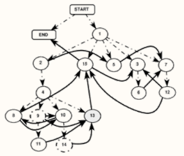
ASD Brain Classification



Typical : Autism
1 : 46

Fake News Detection

Malware Detection

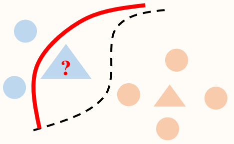


Reasoning

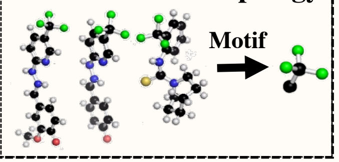
Imbalanced Graph Issue

Biased learning
Minority
$\mathcal{L} = \mathcal{L}_{G_1} + \boxed{\mathcal{L}_{G_l}} + \cdots$

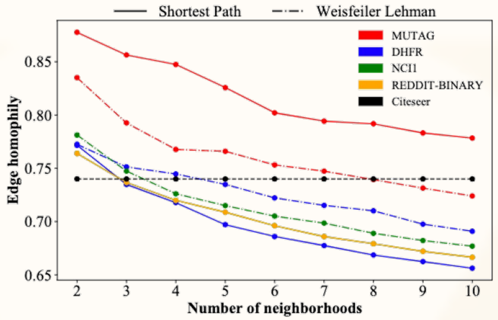
Population risk

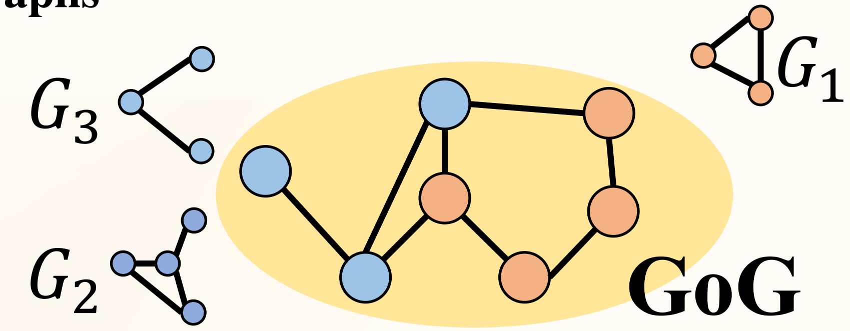


Imbalanced Topology

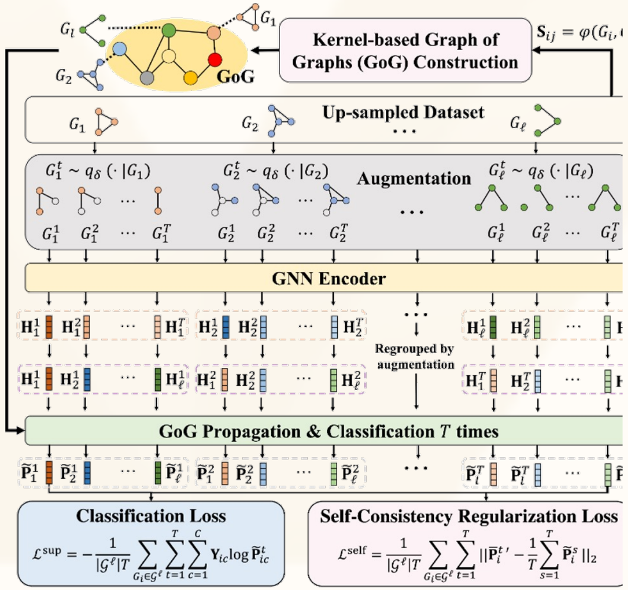


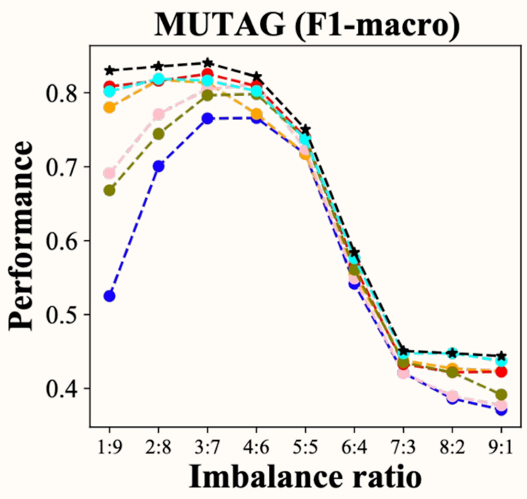
Constructing Graph-of-Graphs





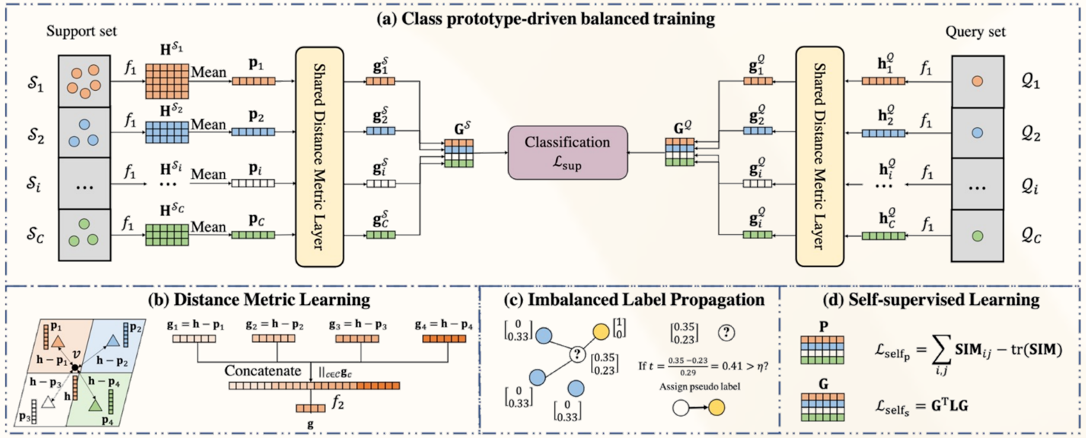
Graph-of-Graph Neural Network (G^{2}GNN)



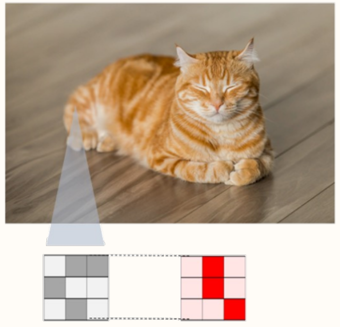


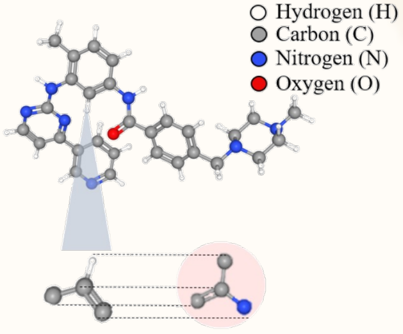
Other related works & Acknowledgement

Distance-wise Prototypical Graph Neural Network for Imbalanced Node Classification [16] Wang et al.



Interpretable Chirality-Aware Graph Neural Network for Quantitative Structure Activity Relationship Modeling in Drug Discovery





[10] Liu et al.

Special Interest Group on Information Retrieval

Yuying Zhao

Neil Shah

Tyler Derr

Network and Data Science Lab

https://nds-vu.github.io/

More about me
https://yuwvandy.github.io/

References

[1] https://www.apsf.org/article/an-update-on-the-perioperative-considerations-for-covid-19-severe-acute-respiratory-syndrome-coronavirus-2-sars-cov-2/

[2] Rozemberczki, B., Hoyt, C.T., Gogleva, A., Grabowski, P., Karis, K., Lamov, A., Nikolov, A., Nilsson, S., Ughetto, M., Wang, Y. and Derr, T., 2022. ChemicalX: A Deep Learning Library for Drug Pair Scoring. arXiv preprint arXiv:2202.05240.

[3] Hamilton, W., Ying, Z. and Leskovec, J., 2017. Inductive representation learning on large graphs. Advances in neural information processing systems, 30.

[4] Zhang, M. and Chen, Y., 2018. Link prediction based on graph neural networks. Advances in neural information processing systems, 31

[5] Vasudevan, V., Bassenne, M., Islam, M.T. and Xing, L., 2022. Image Classification using Graph Neural Network and Multiscale Wavelet Superpixels. arXiv preprint arXiv:2201.12633.

[6] Bi, X., Zhao, X., Huang, H., Chen, D. and Ma, Y., 2020. Functional brain network classification for Alzheimer's disease detection with deep features and extreme learning machine. Cognitive Computation, 12(3), pp.513-527.

[7] Bajorath, J., 2002. Integration of virtual and high-throughput screening. Nature Reviews Drug Discovery, 1(11), pp.882-894.

[8] Dou, Y., Shu, K., Xia, C., Yu, P.S. and Sun, L., 2021, July. User preference-aware fake news detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 2051-2055).

References

[9] Oak, R., Du, M., Yan, D., Takawale, H. and Amit, I., 2019, November. Malware detection on highly imbalanced data through sequence modeling. In Proceedings of the 12th ACM Workshop on artificial intelligence and security (pp. 37-48).

[10] Liu, Y., Wang, Y., Vu, O.T., Moretti, R., Bodenheimer, B., Meiler, J. and Derr, T., 2022. Interpretable Chirality-Aware Graph Neural Network for Quantitative Structure Activity Relationship Modeling in Drug Discovery. bioRxiv.

[11] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, pp.321-357.

[12] Zhang, H., Cisse, M., Dauphin, Y.N. and Lopez-Paz, D., 2017. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

[13] Xu, K., Hu, W., Leskovec, J. and Jegelka, S., 2018. How powerful are graph neural networks?. arXiv preprint arXiv:1810.00826.

[14]Sun, F.Y., Hoffmann, J., Verma, V. and Tang, J., 2019. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. arXiv preprint arXiv:1908.01000.

[15] You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z. and Shen, Y., 2020. Graph contrastive learning with augmentations. Advances in Neural Information Processing Systems, 33, pp.5812-5823.

[16] Wang, Y., Aggarwal, C. and Derr, T., 2021. Distance-wise Prototypical Graph Neural Network in Node Imbalance Classification. arXiv preprint arXiv:2110.12035.