Imbalanced Graph Representation Learning via Graph of Graph Neural Networks

Yu Wang[1] Yuying Zhao[1] Neil Shah[2] Tyler Derr[1]

1. Vanderbilt University, 2. Snap

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

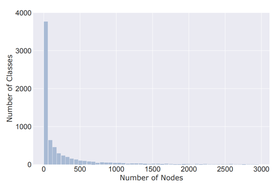
The 15th ACM International Conference on Web Search and Data Mining
wsdm 2022 Phoenix, AZ, USA
February 21–25, 2022

Acknowledgements

This presentation is jointly supported by NSF and SIGIR

Motivation

Imbalanced classification



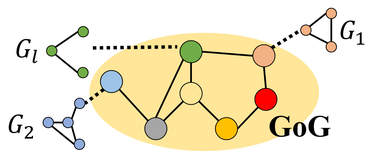
Imbalanced training loss

$\mathcal{L} = \mathcal{L}_1 + \boxed{\mathcal{L}_2}$ Minority

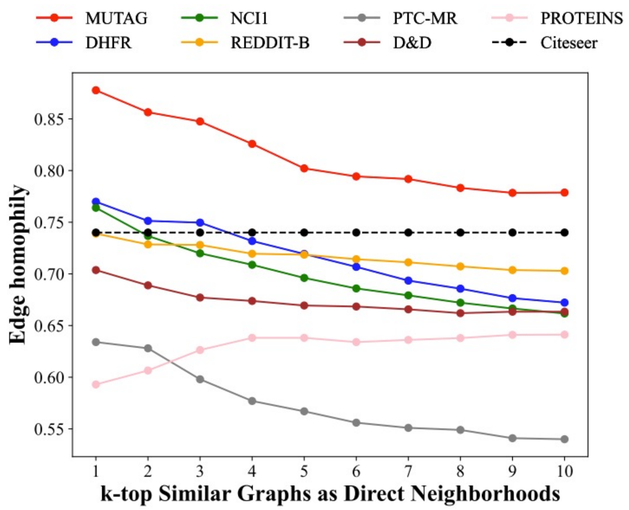
Poor generalibility

Imbalanced issue widely exists in real-world applications

Graph of graphs (GoG)

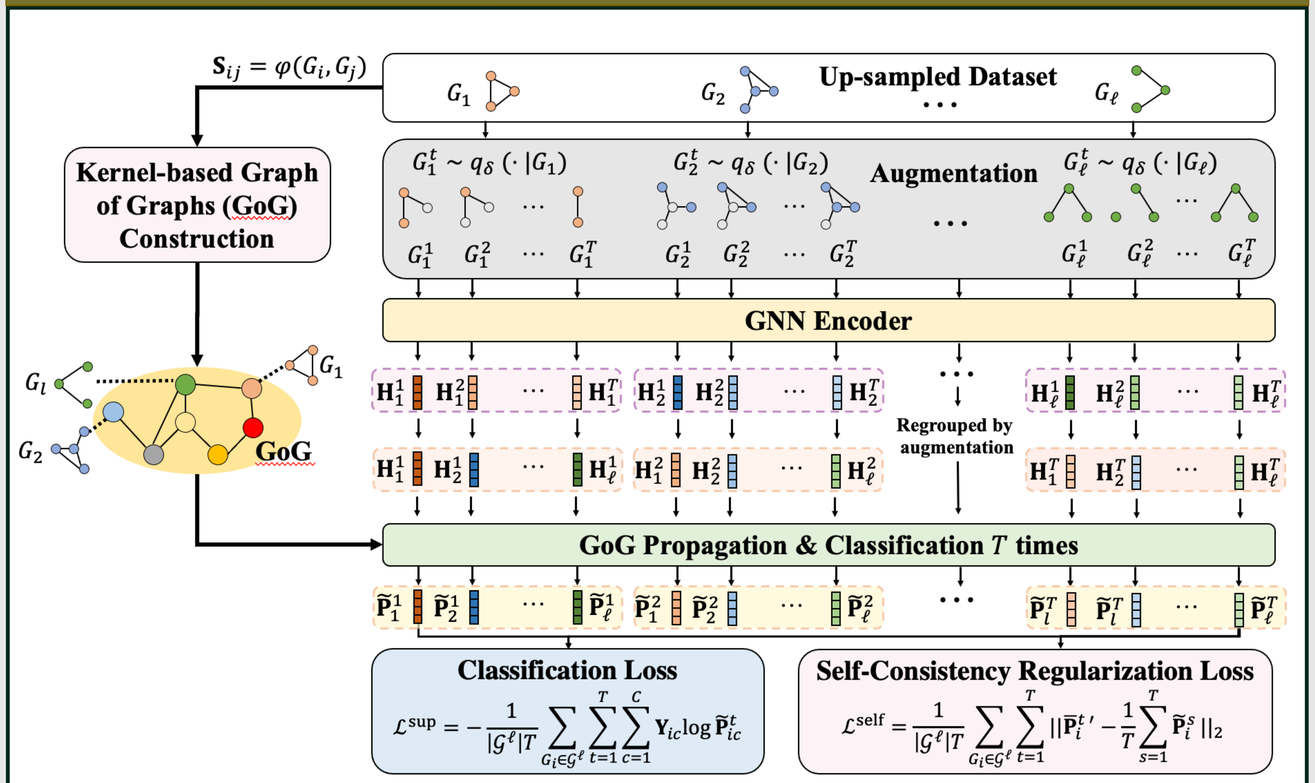


Graph of Graphs: low-level graphs are nodes in the high-level graph and similar graphs are connected by edges



High edge homophily means neighborhoods share the same class and aggregate their information would be helpful for prediction

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

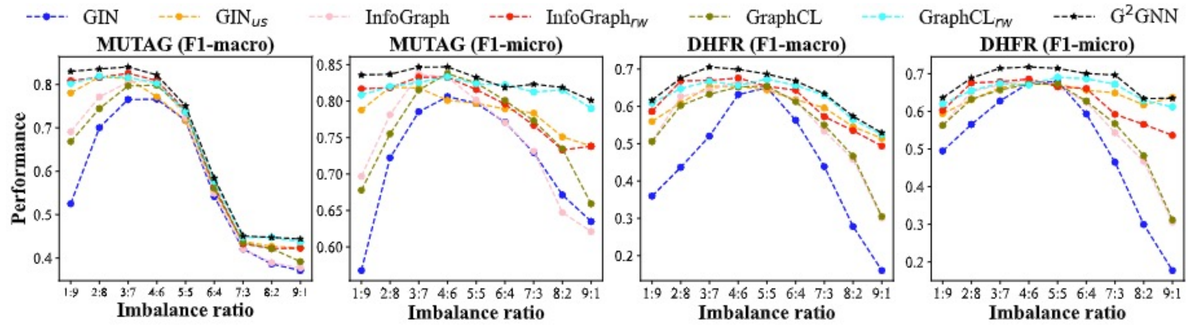


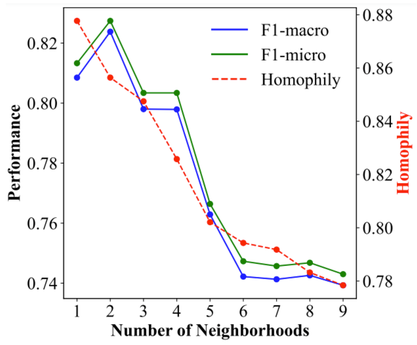
Experimental Results

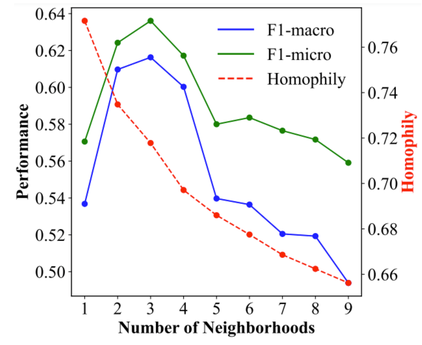
Model	MUTAG (5:45)		PROTEINS (30:270)		D&D (30:270)		NCI1 (100:900)	
	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro
GIN	52.50 ± 18.70	56.77 ± 14.14	25.33 ± 7.53	28.50 ± 5.82	9.99 ± 7.44	11.88 ± 9.49	18.24 ± 7.58	18.94 ± 7.12
GIN_{us}	78.03 ± 7.62	78.77 ± 7.67	65.64 ± 2.67	71.55 ± 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

Model	PTC-MR (9:81)		DHFR (12:108)		REDDIT-B (50:450)		Ave. Rank	
	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro
GIN	17.74 ± 6.49	20.30 ± 6.06	35.96 ± 8.87	49.46 ± 4.90	33.19 ± 14.26	36.02 ± 17.38	12.00	12.00
GIN_{us}	44.78 ± 8.01	55.43 ± 14.25	55.96 ± 10.06	59.39 ± 6.52	66.71 ± 3.92	83.00 ± 5.18	5.00	4.43
GIN_{rw}	36.96 ± 14.08	43.09 ± 20.01	55.16 ± 9.47	57.78 ± 6.69	45.17 ± 8.46	51.92 ± 12.29	8.86	8.86
GIN_{st}	36.30 ± 11.45	40.04 ± 15.32	56.06 ± 9.60	58.48 ± 6.42	60.05 ± 4.14	73.59 ± 6.05	8.29	8.43
InfoGraph	25.85 ± 6.14	26.71 ± 6.50	50.62 ± 8.33	56.28 ± 4.58	57.67 ± 3.80	67.10 ± 4.91	10.00	10.14
InfoGraph_{us}	44.29 ± 4.69	48.91 ± 7.49	59.49 ± 5.20	61.62 ± 4.18	67.01 ± 3.34	78.68 ± 3.71	5.00	5.00
InfoGraph_{rw}	44.09 ± 5.62	49.17 ± 8.78	58.67 ± 5.82	60.24 ± 4.80	65.79 ± 3.38	77.35 ± 3.96	4.43	4.29
GraphCL	24.22 ± 6.21	25.16 ± 5.25	50.55 ± 10.01	56.31 ± 6.12	53.40 ± 4.06	62.19 ± 5.68	10.71	10.57
GraphCL_{us}	45.12 ± 7.33	53.50 ± 13.31	60.29 ± 9.04	61.71 ± 6.75	62.01 ± 3.97	75.84 ± 3.98	5.29	5.43
GraphCL_{rw}	44.75 ± 7.62	52.22 ± 13.24	60.87 ± 6.33	61.93 ± 5.15	62.79 ± 6.93	76.15 ± 9.15	5.00	5.29
G^2GNN_e	46.40 ± 7.73	56.61 ± 13.72	61.63 ± 10.02	63.61 ± 6.05	68.39 ± 2.97	86.35 ± 2.27	1.71	1.86
G^2GNN_n	46.61 ± 8.27	56.70 ± 14.81	59.72 ± 6.83	61.27 ± 5.40	67.52 ± 2.60	85.43 ± 1.80	1.71	1.71

Further Probe







More about me!

https://yuwvandy.github.io/