Deep Learning on Graphs: Method and Applications (DLG@KDD)

Structure-Aware Hard Negative Mining for Heterogeneous Graph Contrastive Learning

Presented by **Yanqiao ZHU**

yanqiao.zhu@cripac.ia.ac.cn

@ https://SXKDZ.github.io

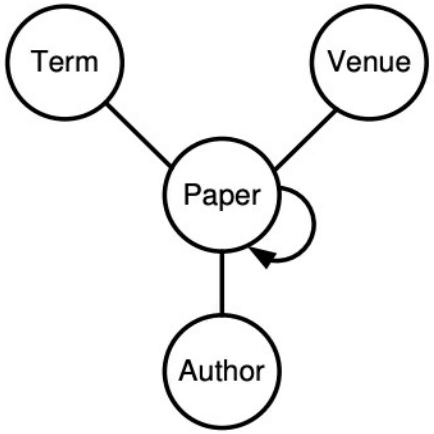
Center for Research on Intelligent Perception and Computing
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

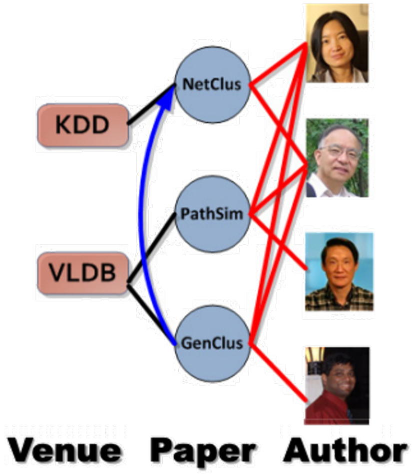
CRIPAC 智能感知与计算研究中心 Center for Research on Intelligent Perception and Computing

Joint work with Yichen XU, Hejie CUI, Carl YANG, Qiang LIU, and Shu WU

Background

- Many real-world complex interactive objects can be represented in a form of Heterogeneous Graphs (HGs).
- Existing graph neural networks for HGs require a relatively large amount of labeled data for proper training.

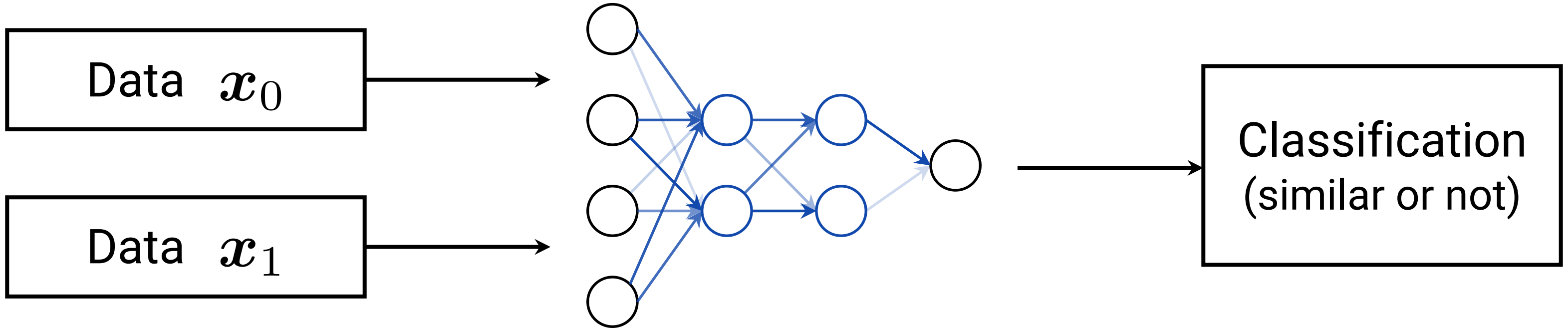




[Sun and Han, 2012] Y. Sun and J. Han, Mining Heterogeneous Information Networks: A Structural Analysis Approach, *SIGKDD Explor. Newsl.*, vol. 14, no. 2, pp. 20–28, 2012.

Graph Contrastive Learning

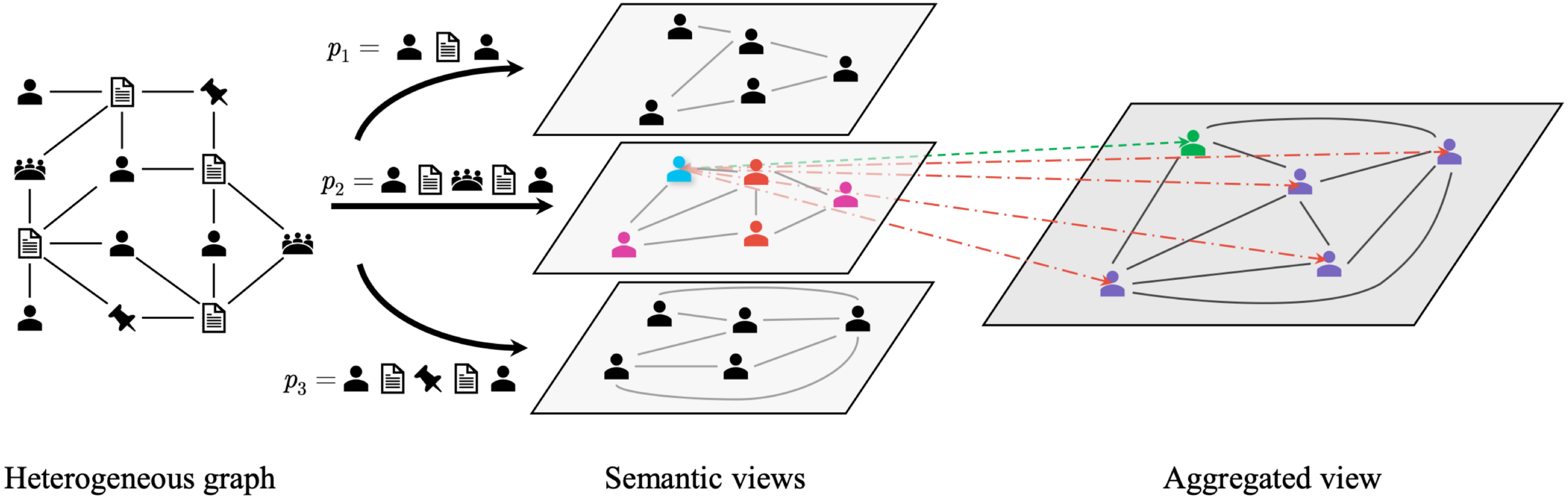
- To alleviate the label scarcity problem, Contrastive Learning (CL) learns representations by distinguishing semantically similar samples (positives) over dissimilar samples (negatives) in the latent space.



[Zhu et al., 2021] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, Graph Contrastive Learning with Adaptive Augmentation, in *WWW*, 2021, pp. 2069–2080.

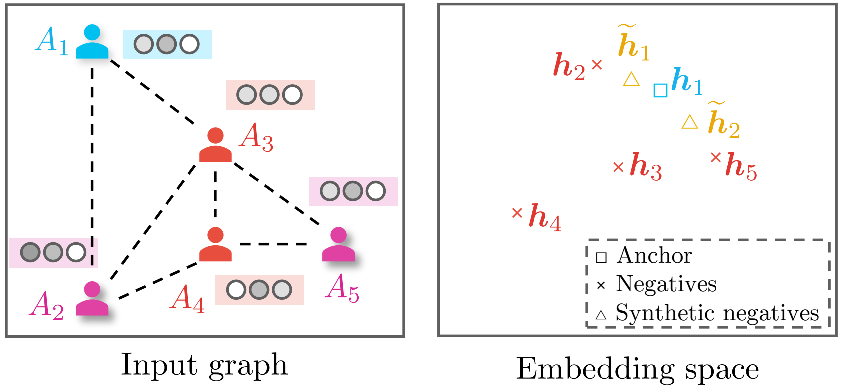
Heterogeneous Graph Contrastive Learning

- We first construct semantic views according to metapaths.
- Multiview contrastive aggregation objective: to ensure global consistency among semantic views and adaptively encode information from each view.



Structure-Aware Hard Negative Mining

- Hard negative sample is of particular concern for effective CL.
  - The more similar a negative sample to its anchor, the more helpful it is for learning effective representatives.
- More global view: negative samples sharing similar structural characteristics should be pushed away.
- Widely-adopted metrics:
  - Personal PageRank (PPR)
  - Laplacian positional embedding (PE)
  - Distance encoding (DE)
  - ...



Experiments

Method	Training Data	Node Classification						Node Clustering					
		ACM		IMDb		DBLP		ACM		IMDb		DBLP	
		Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1	NMI	ARI	NMI	ARI	NMI	ARI
DeepWalk	$\boldsymbol{A}$	76.92	77.25	46.38	40.72	79.37	77.43	41.61	35.10	1.45	2.15	76.53	81.35
ESim	$\boldsymbol{A}$	76.89	77.32	35.28	32.10	92.73	91.64	39.14	34.32	0.55	0.10	66.32	68.31
metapath2vec	$\boldsymbol{A}$	65.00	65.09	45.65	41.16	91.53	90.76	21.22	21.00	1.20	1.70	74.30	78.50
HERec	$\boldsymbol{A}$	66.03	66.17	45.81	41.65	92.69	91.78	40.70	37.13	1.20	1.65	**76.73**	78.50
HAN-U	$\boldsymbol{A}, \boldsymbol{X}$	82.63	81.89	43.98	40.87	90.47	89.65	39.84	32.98	3.92	4.10	74.17	79.98
DGI	$\boldsymbol{A}, \boldsymbol{X}$	89.15	89.09	48.86	45.38	91.30	90.69	58.13	57.18	8.31	11.25	60.62	60.42
GRACE	$\boldsymbol{A}, \boldsymbol{X}$	88.72	88.72	46.64	42.41	90.88	89.76	53.38	54.39	7.52	9.16	62.06	64.13
HORACE-PE	$\boldsymbol{A}, \boldsymbol{X}$	**90.76**	**90.72**	**58.98**	**54.48**	**92.81**	**92.33**	67.93	72.65	**15.09**	**17.23**	76.60	**81.58**
HORACE-PPR	$\boldsymbol{A}, \boldsymbol{X}$	90.75	90.70	58.96	54.47	92.78	92.30	**68.10**	**73.15**	15.03	17.09	76.52	81.49
GCN	$\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}$	86.77	86.81	49.78	45.73	91.71	90.79	51.40	53.01	5.45	4.40	75.01	80.49
GAT	$\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}$	86.01	86.23	55.28	49.44	91.96	90.97	57.29	60.43	8.45	7.46	71.50	77.26
HAN	$\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}$	89.22	89.40	54.17	49.78	92.05	91.17	61.56	64.39	10.31	9.51	79.12	84.76

Concluding Remarks

- A novel heterogeneous graph contrastive learning framework with a multiview contrastive aggregation objective that encodes information adaptively from each semantic view.
- A novel hard negative mining scheme to improve the embedding quality, considering the complex structure of heterogeneous graphs and smoothing nature of heterogeneous GNNs.
- Extensive experiments on three real-world heterogeneous datasets demonstrate its effectiveness over both unsupervised and supervised baselines.


THANKS