The 22nd SIAM Conference on Data Mining (SDM 2022)

Structure-Enhanced Heterogeneous Graph Contrastive Learning

Presented by **Yanqiao ZHU**

yanqiao.zhu@cripac.ia.ac.cn

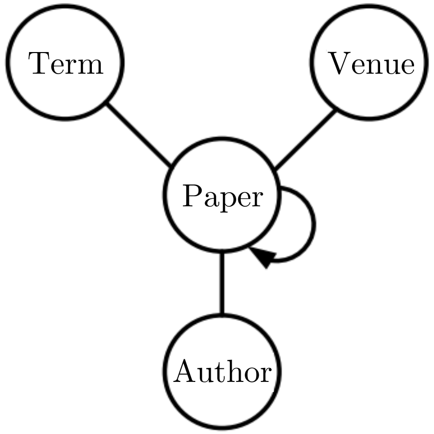
@ https://SXKDZ.github.io

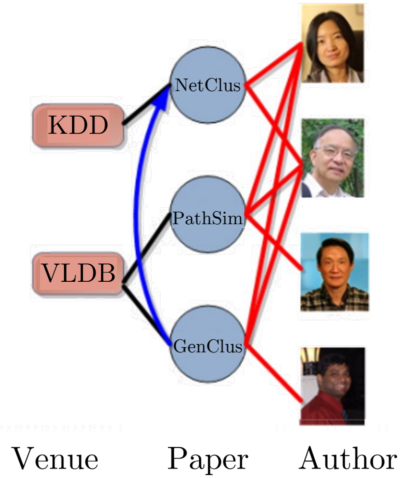
Center for Research on Intelligent Perception and Computing
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

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

Heterogeneous Graphs

- 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.

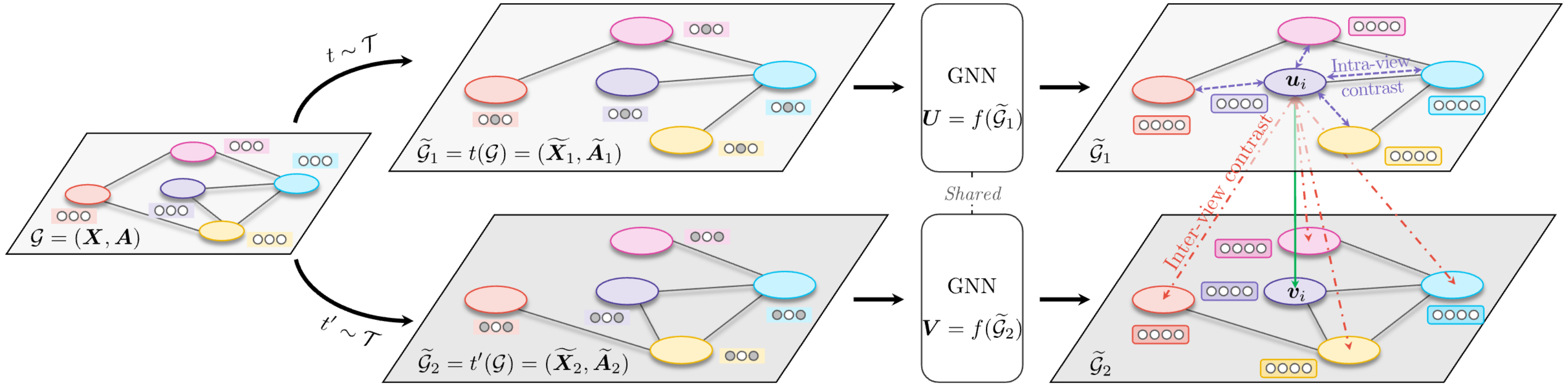




[Sun and Han, 2012] Y. Sun and J. Han, Mining Heterogeneous Information Networks: A Structural Analysis Approach, *SIGKDD Explor. Newsl.*, vol. 14, no. 2, pp. 20–28, 2012.

Graph Contrastive Learning

- Graph Contrastive Learning (GCL) is a promising approach to alleviate the label scarcity problem.
- Construct multiple views and train the model using a contrastive loss to maximize the agreement between node embeddings 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 Networks

- Metapath-specific node representations (semantic views)

$$\boldsymbol{h}_i^p = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{v_j \in \mathcal{N}_p(v_i)} \alpha_{ij}^p \boldsymbol{W}^p \boldsymbol{x}_j\right) \qquad \alpha_{ij}^p = \frac{\exp(\sigma(\boldsymbol{a}_p^\top [\boldsymbol{h}_i^p \parallel \boldsymbol{h}_j^p]))}{\sum_{v_k \in \mathcal{N}_p(v_i)} \exp(\sigma(\boldsymbol{a}_p^\top [\boldsymbol{h}_i^p \parallel \boldsymbol{h}_k^p]))}$$

- Aggregated node representations

$$\beta^p = \frac{\exp(w^p)}{\sum_{p' \in \mathcal{P}} \exp(w^{p'})}$$

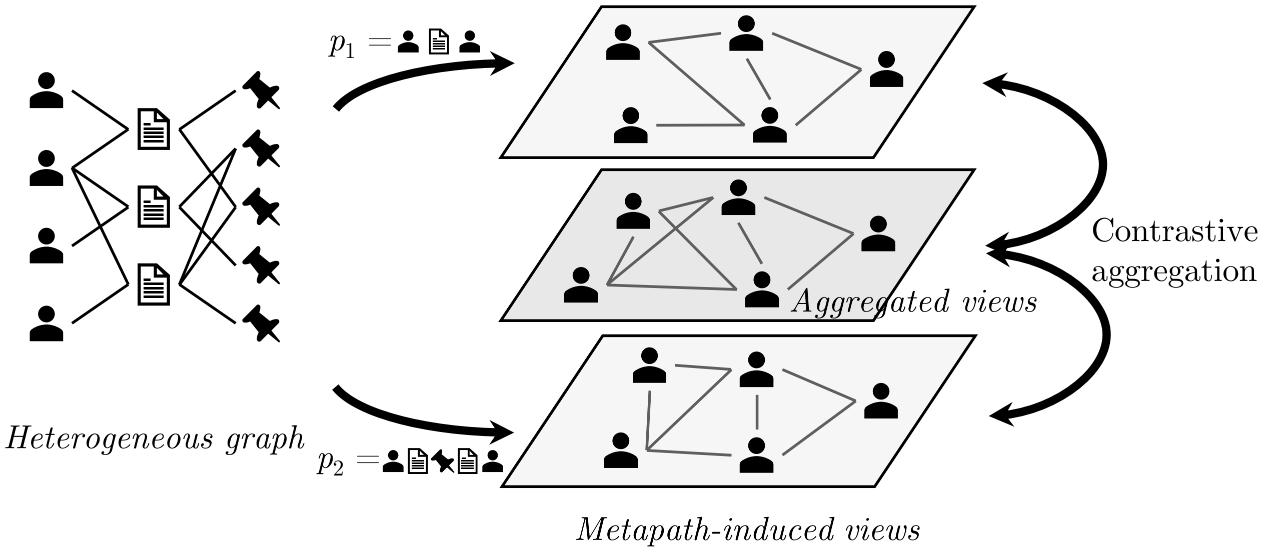
$$\boldsymbol{h}_i = \sum_{p=1}^{|\mathcal{P}|} \beta^p \boldsymbol{h}_i^p$$

$$w^p = \frac{1}{|\mathcal{V}|} \sum_{v_i \in \mathcal{V}} \boldsymbol{q}^\top \cdot \tanh(\boldsymbol{W}\boldsymbol{h}_i^p + \boldsymbol{b})$$

[Wang et al., 2019] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, Heterogeneous Graph Attention Network, WWW, 2019, pp. 2022–2032.

Multiview Contrastive Aggregation

- We first construct semantic views according to metapaths.
- Maximize the agreement between the node representation under a specific metapath view and an aggregated representation for all metapaths.



Multiview Contrastive Aggregation (cont.)

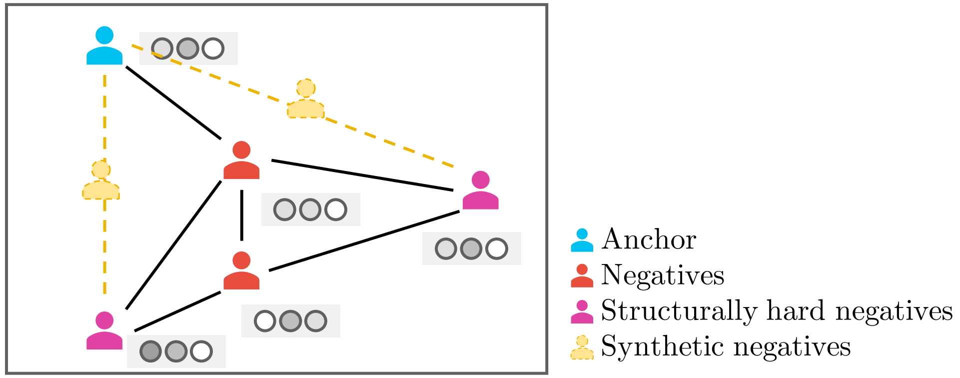
- Node embedding pairs in a semantic view and in the aggregated view constitute positive pairs and all other embedding pairs are naturally negative pairs.

$$\mathcal{L}(\boldsymbol{h}_i^p, \boldsymbol{h}_i) = -\log \frac{e^{\theta(\boldsymbol{h}_i^p, \boldsymbol{h}_i)/\tau}}{e^{\theta(\boldsymbol{h}_i^p, \boldsymbol{h}_i)/\tau} + \sum_{j \neq i} \left(e^{\theta(\boldsymbol{h}_i^p, \boldsymbol{h}_j)/\tau} + e^{\theta(\boldsymbol{h}_i^p, \boldsymbol{h}_j^p)/\tau} \right)}$$

- Ensure global consistency among semantic views and adaptively encode information from each view.

Structure-Enhanced 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.



Metapath-induced view

Structure Embeddings

- We characterize structural properties through the lens of **structure embeddings** to effectively measure the likelihood of each negative sample with respect to the anchor.

- Widely-adopted metrics:
  - Personal PageRank (PPR)
  - Laplacian Positional Embedding (PE)
  - Distance Encoding (DE)
  - …

Negative Mining via Mixup

- Define a structural metric $s(i, j, p)$ representing the distance measure of a negative node v_i to the anchor v_j given structural embeddings in metapath-induced view p.
- Sort negatives according to the hardness metric $s(i, j, p)$ and pick the top-T negatives to form a candidate list for metapath-induced view p.
- Synthesize $M \ll |\mathcal{V}|$ samples by creating a convex linear combination of them:

$$\widetilde{\boldsymbol{h}}_m^p = \alpha_m \boldsymbol{h}_i^p + (1 - \alpha_m)\boldsymbol{h}_j^p$$

  - These interpolated samples will be added into negative bank.

Experimental Configuration

- Datasets: DBLP, ACM, and IMDb

Dataset	Node	Relations	Metapaths
DBLP	Paper (14,328) Author (4,057) Conference (20) Term (8,789)	P–A (19,645) P–C (14,328) P–T (88,420)	APA APCPA APTPA
ACM	Paper (3,025) Author (5,835) Subject (56)	P–A (9,744) P–S (3,025)	PAP PSP
IMDb	Movie (4,780) Actor (5,841) Director (2,269)	M–A (14,340) M–D (4,780)	MAM MDM

- Tasks:
  - Node classification: Macro-F1 and Micro-F1
  - Node clustering: NMI and ARI

Overall Performance

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
HeCo	$\boldsymbol{A}, \boldsymbol{X}$	88.15	88.25	51.69	50.75	91.56	91.02	59.53	57.59	10.11	11.74	70.99	76.67
STENCIL-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**
STENCIL-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

Key Takeaways

- Previous studies focus on homogeneous graphs, failed to consider complex relations in real-world graphs.
- We propose a novel heterogeneous graph contrastive learning framework STENCIL.
  - Model underlying meta-semantics via contrastive aggregation.
  - Improve negative sample selection using structure information.
- Extensive experiments on three real-world heterogeneous datasets demonstrate its effectiveness over both unsupervised and supervised baselines.


THANKS