NEURAL INFORMATION PROCESSING SYSTEMS

Advances in Neural Information Processing Systems (NeurIPS), 2024

Beyond Redundancy: Information-aware Unsupervised Multiplex Graph Structure Learning

Zhixiang Shen[1], Shuo Wang[1], Zhao Kang[1]

[1]University of Electronic Science and Technology of China

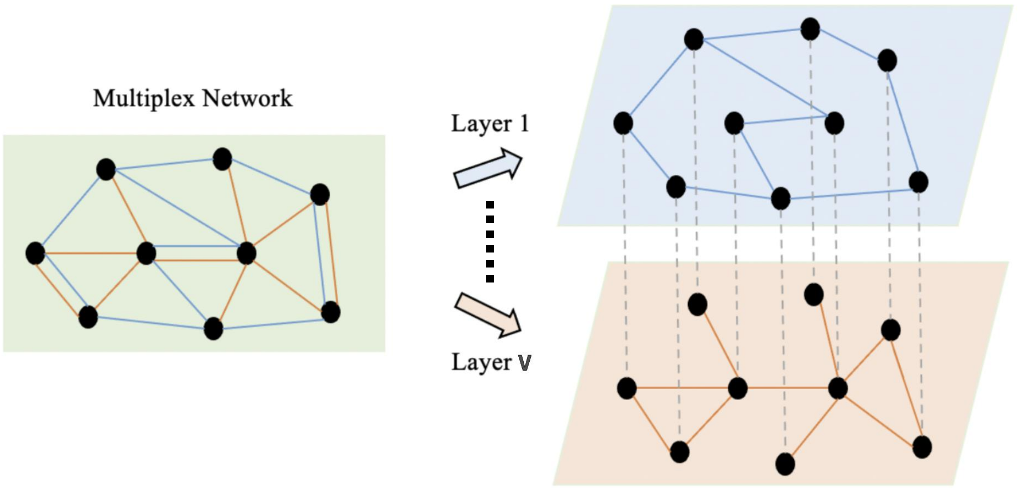
Paper

Github InfoMGF

Backgrounds of Multiplex Graph

❖ What is multiplex graph?

*A special type of **multi-relational heterogeneous graph** with multiple graph layers span across a common set of nodes.*



Unsupervised Multiplex Graph Learning (UMGL):
Learn node representations by leveraging diverse graph structures and features without manual labeling.

Applications：
Biological Network Analysis, Social Network Mining, Recommendation Systems......

[1] Jing B, Park C, Tong H. Hdmi: High-order deep multiplex infomax. Proceedings of the Web Conference (WWW), 2021.

[2] Qian X, Li B, Kang Z. Upper Bounding Barlow Twins: A Novel Filter for Multi-Relational Clustering. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2024.

Backgrounds of Multiplex Graph

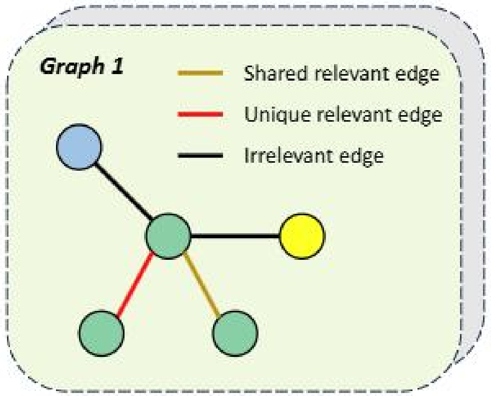
❖ Overlooked Aspects of Unsupervised Multiplex Graph Learning

1) The reliability of graph structure

- **Task-irrelevant information:**
  irrelevant, heterophilic, and missing edges → *beyond graph-fixed methods*

2) Multiplex graph non-redundancy

- **Shared task-relevant information:**
  homophilic edges common to all graphs
- **Unique task-relevant information:**
  homophilic edges appear only in a certain graph



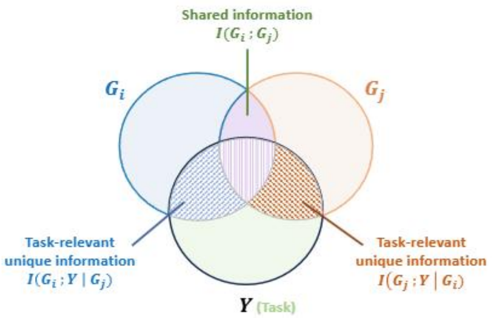
Motivation

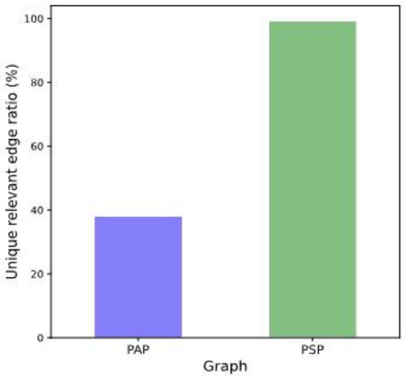
❖ Theoretical Definition

Multiplex Graph Non-redundancy: ***Task-relevant information exists not only in the shared information between graphs but also potentially within the unique information of certain graphs.***

Definition 1. *G_i is considered non-redundant with G_j for Y if and only if there exists $\epsilon > 0$ such that the conditional mutual information $I(G_i; Y \mid G_j) > \epsilon$ or $I(G_j; Y \mid G_i) > \epsilon$.*

❖ Empirical Study





Dataset	Nodes	Relation type	Edges	Unique relevant edge ratio (%)
ACM	3,025	Paper-Author-Paper (PAP) Paper-Subject-Paper (PSP)	26,416 2,197,556	38.08 99.05
DBLP	2,957	Author-Paper-Author (APA) Author-Paper-Conference-Paper-Author (APCPA)	2,398 1,460,724	0 99.82
Yelp	2,614	Business-User-Business (BUB) Business-Service-Business (BSB) Business-Rating Levels-Business (BLB)	525,718 2,475,108 1,484,692	83.12 97.49 93.07
MAG	113,919	Paper-Paper (PP) Paper-Author-Paper (PAP)	1,806,596 10,067,799	64.59 93.48

Motivation

❖ Theoretical Definition

Multiplex Graph Non-redundancy: ***Task-relevant information exists not only in the shared information between graphs but also potentially within the unique information of certain graphs.***

Definition 1. *G_i is considered non-redundant with G_j for Y if and only if there exists $\epsilon > 0$ such that the conditional mutual information $I(G_i; Y \mid G_j) > \epsilon$ or $I(G_j; Y \mid G_i) > \epsilon$.*

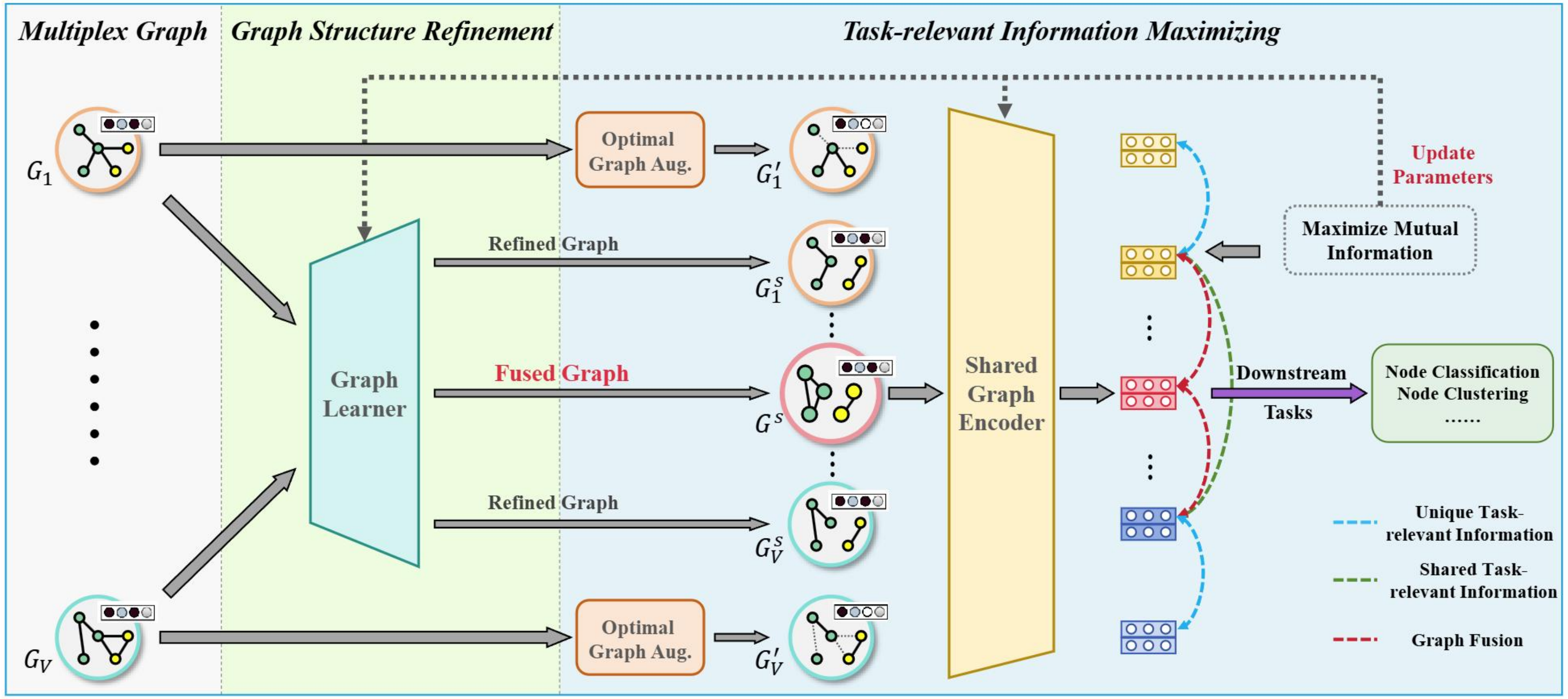
❖ Problem Definition

- **Graph Structure Learning (GSL) Perspective:**

*How can we learn a **fused graph** from the original multiplex graph in an unsupervised manner, mitigating task-irrelevant noise while retaining sufficient task-relevant information?*

Methodology

❖ Information-aware Unsupervised Multiplex Graph Fusion



Input: Multiplex Structures, Node Features

Output: Learned Structure, Representations

- Removing irrelevant noise by GSL
- Maximizing both shared and unique task-relevant information for each graph
- Learnable generative graph augmentation
- Optimization using differentiable lower/upper bound of the mutual information

Minimization Objectives:

$$\mathcal{L}_s = -\frac{2}{V(V-1)} \sum_{i=1}^{V} \sum_{j=i+1}^{V} I(G_i^s; G_j^s)$$

Shared Information

$$\mathcal{L}_u = -\frac{1}{V} \sum_{i=1}^{V} I(G_i^s; G_i')$$

Unique Information

$$\mathcal{L}_f = -\frac{1}{V} \sum_{i=1}^{V} I(G^s; G_i^s)$$

Graph Fusion

Theoretical Contributions

❖ Optimal Graph Augmentation

Definition 2. *G'_i is an optimal augmented graph of G_i if and only if $I(G'_i; G_i) = I(Y; G_i)$, implying that the only information shared between G_i and G'_i is task-relevant without task-irrelevant noise.*

Theorem 1. *If G'_i is the optimal augmented graph of G_i, then $I(G^s_i; G'_i) = I(G^s_i; Y)$ holds.*

Theorem 2. *The maximization of $I(G^s_i; G'_i)$ yields a discernible reduction in the task-irrelevant information relative to the maximization of $I(G^s_i; G_i)$.*

❖ Multiplex Graph Fusion

Theorem 3. *The learned fused graph G^s contains more task-relevant information than the refined graph G^s_i from any single view. Formally, we have:*

$$I(G^s; Y) \geq \max_i I(G^s_i; Y) \tag{7}$$

Theorem 3 theoretically proves that the fused graph G^s can incorporate more task-relevant information than considering each view individually, thus ensuring the effectiveness of multiplex graph fusion.

Experiments

Table 1: Quantitative results (%) on node clustering. The top 3 highest results are highlighted with **red boldface**, red color and **boldface**, respectively. The symbol "OOM" means out of memory.

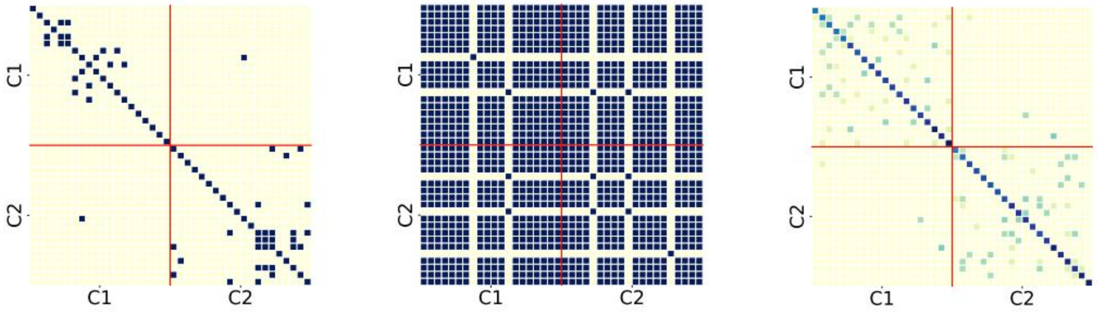
Method	ACM				DBLP				Yelp				MAG			
	NMI	ARI	ACC	F1	NMI	ARI	ACC	F1	NMI	ARI	ACC	F1	NMI	ARI	ACC	F1
VGAE	45.83	41.36	67.93	68.62	61.79	65.56	84.48	83.67	39.19	42.57	65.07	56.74	OOM			
DGI	52.94	47.55	65.36	57.34	65.59	70.35	86.88	86.02	39.42	42.62	65.29	56.79	**53.56**	**42.6**	**59.89**	**57.17**
O2MAC	42.36	46.04	77.92	78.01	58.64	60.01	83.29	82.88	39.02	42.53	65.07	56.74	OOM			
MvAGC	64.49	66.81	87.17	87.21	50.39	51.21	78.39	77.84	24.39	29.25	63.14	56.7	OOM			
MCGC	60.21	50.72	65.62	54.78	65.56	71.51	87.96	87.47	38.35	35.17	65.61	57.49	OOM			
HDMI	65.44	68.87	88.11	88.14	64.85	70.85	87.39	86.75	60.81	59.35	79.56	77.6	48.15	34.92	51.78	49.8
MGDCR	58.8	55.15	73.82	70.34	62.47	62.22	81.91	80.16	44.23	46.47	72.71	54.43	54.43	43.98	61.37	60.53
DMG	64.14	67.21	87.11	87.23	**69.03**	**73.07**	**88.45**	**87.88**	65.66	66.33	88.26	89.27	48.72	39.77	61.61	60.16
BTGF	**68.92**	**73.14**	**90.09**	**90.11**	66.28	72.47	88.05	87.28	**69.97**	**73.53**	**91.39**	**92.32**	OOM			
InfoMGF-RA	74.89	81.09	92.82	92.89	70.19	73.49	88.72	88.31	72.67	74.66	91.85	92.86	**56.65**	**45.25**	**64.13**	**63.09**
InfoMGF-LA	**76.53**	**81.49**	**93.45**	**93.42**	**73.22**	**78.49**	**91.08**	**90.69**	**75.18**	**78.91**	**93.26**	**94.01**	OOM			

Table 2: Quantitative results with standard deviation ($\% \pm \sigma$) on node classification. Available data for GSL during training is shown in the first column, supervised methods depend on Y for GSL. The symbol "-" indicates that the method is structure-fixed, which does not learn a new structure.

Available Data for GSL	Methods	ACM		DBLP		Yelp		MAG	
		Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
-	GCN	90.27±0.59	90.18±0.61	90.01±0.32	90.99±0.28	78.01±1.89	81.03±1.81	75.98±0.07	75.76±0.10
-	GAT	91.52±0.62	91.46±0.62	90.22±0.37	91.13±0.40	82.12±1.47	84.43±1.56	OOM	
-	HAN	91.67±0.39	91.47±0.22	90.53±0.24	91.47±0.22	88.49±1.73	88.78±1.40	OOM	
X,Y,A	LDS	92.35±0.43	92.05±0.26	88.11±0.86	88.74±0.85	75.98±2.35	78.14±1.98	OOM	
X,Y,A	GRCN	**93.04±0.17**	**92.94±0.18**	88.33±0.47	89.43±0.44	76.05±1.05	80.68±0.96	OOM	
X,Y,A	IDGL	91.69±1.24	91.63±1.24	89.65±0.60	90.61±0.56	76.98±5.78	79.15±5.06	OOM	
X,Y,A	ProGNN	90.57±1.03	90.50±1.29	83.13±1.56	84.83±1.36	51.76±1.46	58.39±1.25	OOM	
X,Y,A	GEN	87.91±2.78	87.88±2.61	89.74±0.69	90.65±0.71	80.43±3.78	82.68±2.84	OOM	
X,Y,A	NodeFormer	91.33±0.77	90.60±0.95	79.54±0.78	80.56±0.62	91.69±0.65	90.59±1.21	77.21±0.18	77.08±0.19
X,A	SUBLIME	92.42±0.16	92.13±0.37	**90.98±0.37**	**91.82±0.27**	79.68±0.79	82.99±0.82	75.96±0.05	75.71±0.03
X,A	STABLE	83.54±4.20	83.38±4.51	75.18±1.95	76.42±1.95	71.48±4.71	76.62±2.75	OOM	
X,A	GSR	92.14±1.08	92.11±0.99	76.59±0.45	77.69±0.42	83.85±0.76	85.73±0.54	OOM	
-	HDMI	91.01±0.32	90.86±0.31	89.91±0.49	90.89±0.51	80.73±0.64	84.05±0.91	72.22±0.14	71.84±0.15
-	DMG	90.42±0.36	90.31±0.35	90.42±0.57	91.34±0.49	91.61±0.62	90.24±0.81	**76.34±0.09**	**76.13±0.10**
-	BTGF	91.75±0.11	91.62±0.11	90.71±0.24	91.57±0.21	**92.81±1.12**	**91.37±1.28**	OOM	
X,A	InfoMGF-RA	93.21±0.22	93.14±0.21	90.99±0.36	91.93±0.29	93.09±0.27	92.02±0.34	**77.25±0.06**	**77.11±0.06**
X,A	InfoMGF-LA	**93.42±0.21**	**93.35±0.21**	**91.28±0.31**	**92.12±0.28**	**93.26±0.26**	**92.24±0.34**	OOM	

Experiments

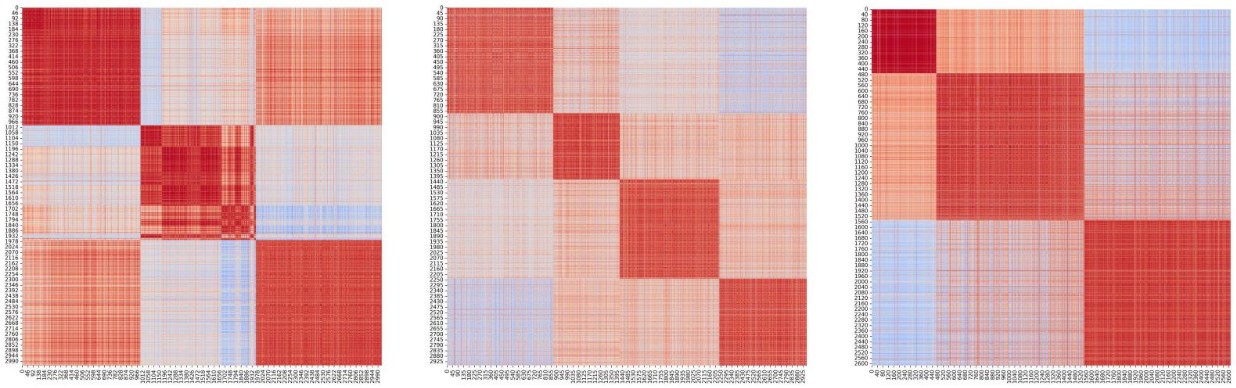
❖ Graph Visualization



(a) PAP (b) PSP (c) G^s

Figure 3: Heatmaps of the subgraph adjacency matrices of the original and learned graphs on ACM.

❖ Node Correlation Visualization



(a) ACM (b) DBLP (c) Yelp

Figure 5: Node correlation maps of representations reordered by node labels.

Summary of InfoMGF

❖ Key takeaways:

- **GSL perspective:**
  Explore graph structure learning in heterogeneous multiplex graph through a data-centric paradigm.

- **Beyond redundancy:**
  Emphasize the importance of unique task-relevant information to better adapt to real-world non-redundant scenarios.

Our github repository contains the source code and datasets of InfoMGF.

Contact me for discussions!

E-mail: zhixiang.zxs@gmail.com