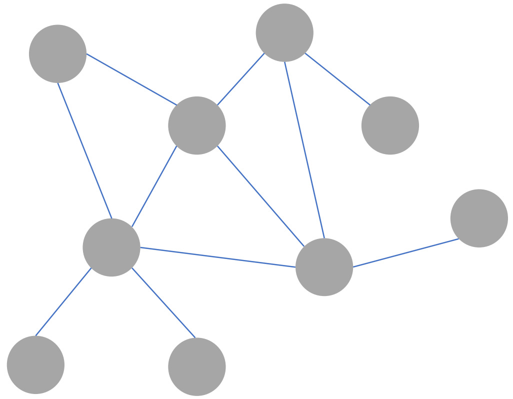
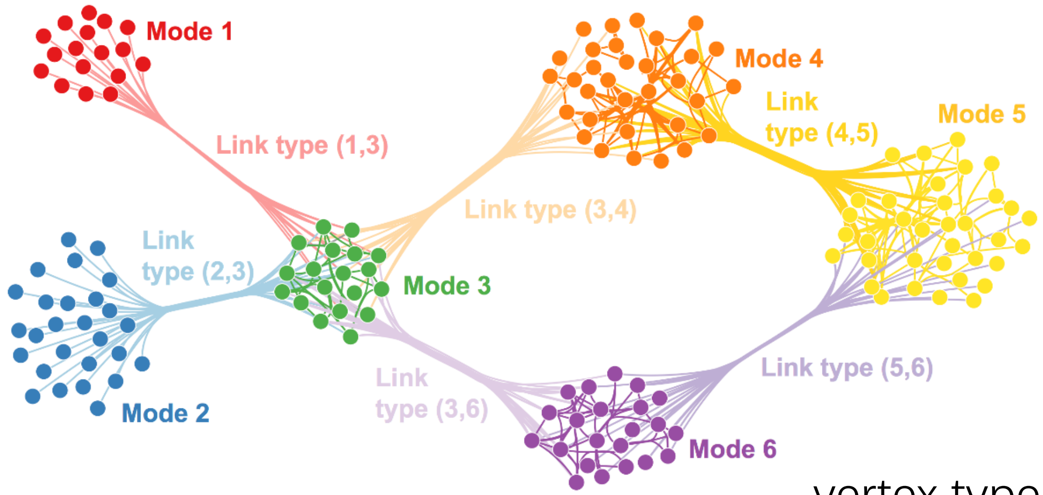
Heterogeneous Networks

Homogeneous networks

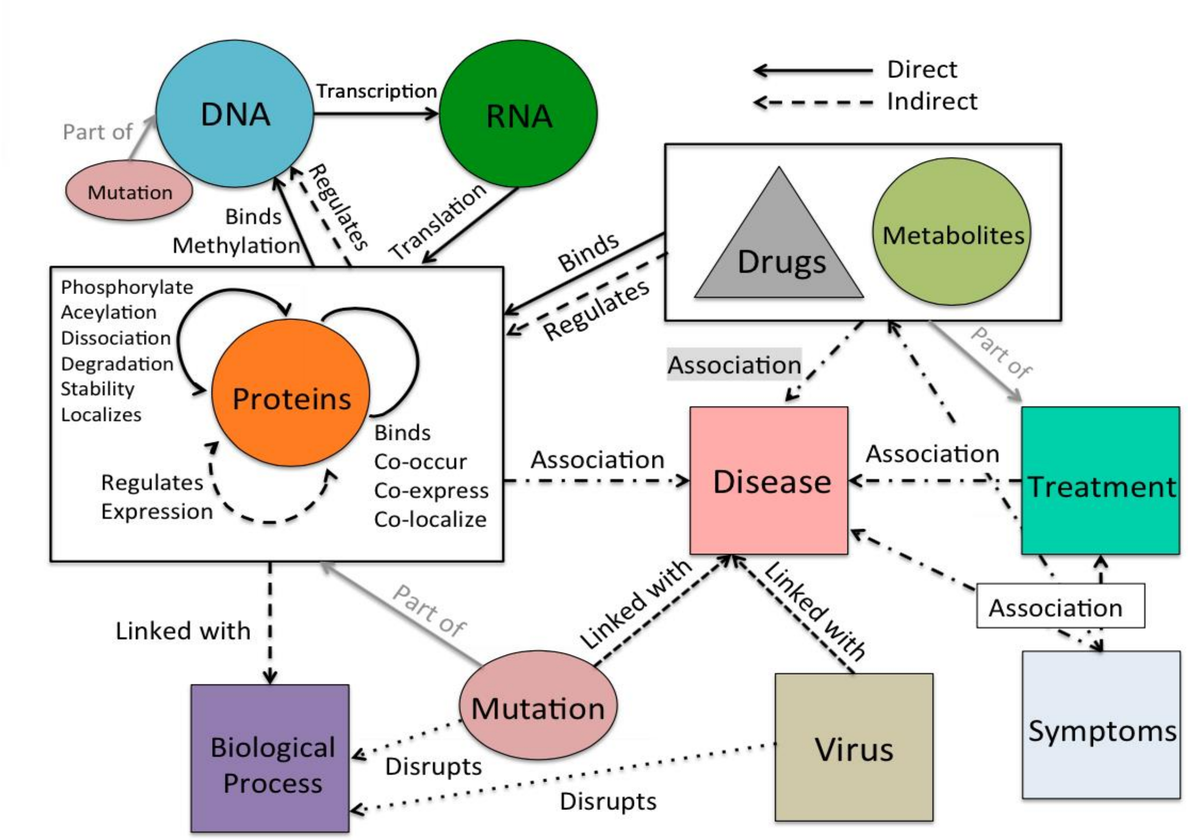


Heterogeneous networks



vertex type + edge type > 2

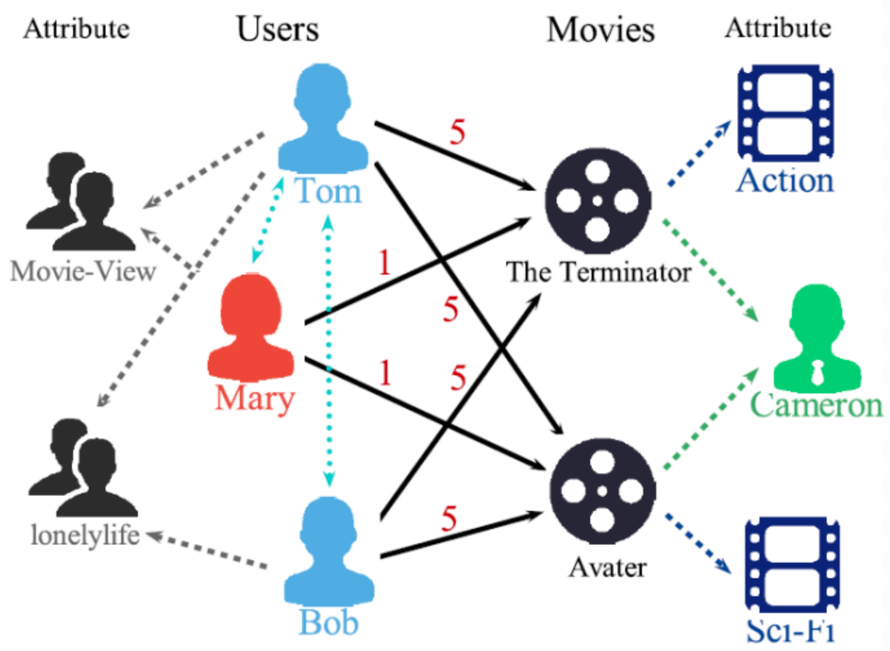
Heterogeneous Networks



Biology Regulatory Network

http://snap.stanford.edu/mambo

arXiv:1711.10730



Recommendation System

Hot Topic

https://cseweb.ucsd.edu/~jmcauley/datasets.html

Bioinformatics, 33, 2017, i190-i198
doi: 10.1093/bioinformatics/btx252
ISMB/ECCB 2017
This is the Author's Original Version (AOV). This article has been accepted for publication in *Bioinformatics* Published by Oxford University Press

OXFORD

Predicting multicellular function through multi-layer tissue networks

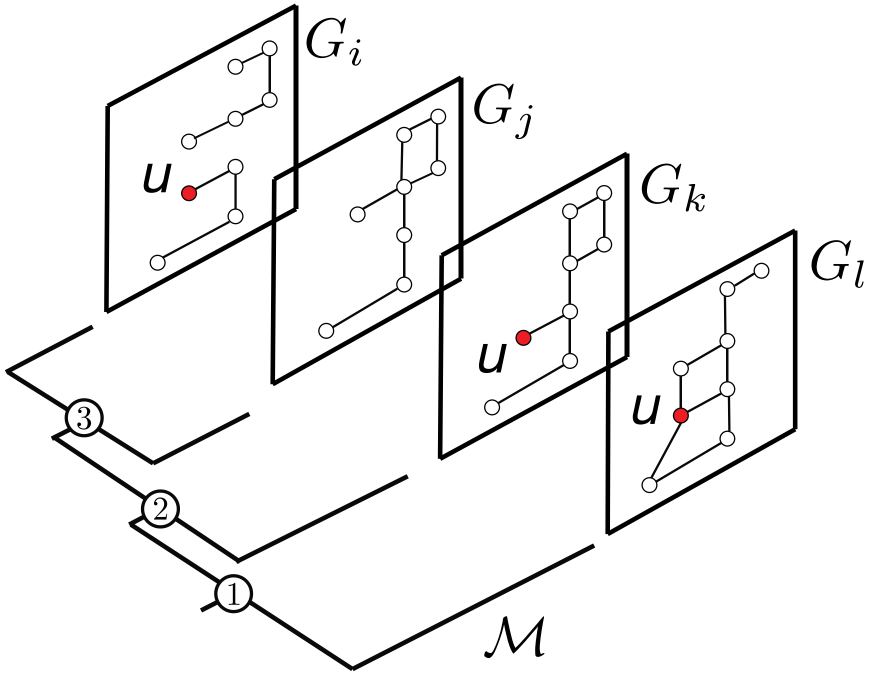
Marinka Zitnik and Jure Leskovec *

Department of Computer Science, Stanford University, Stanford, 94305, USA

*To whom correspondence should be addressed.

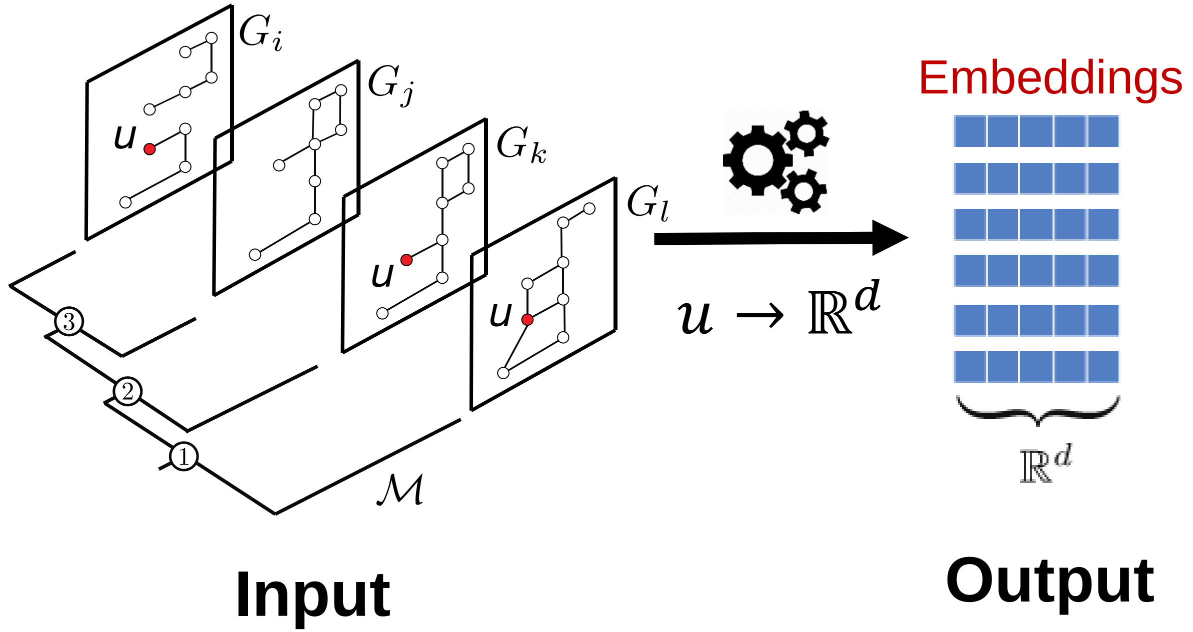
OhmNet: Multi-Layer Graphs

Extending node2vec to **multi-layer graphs**



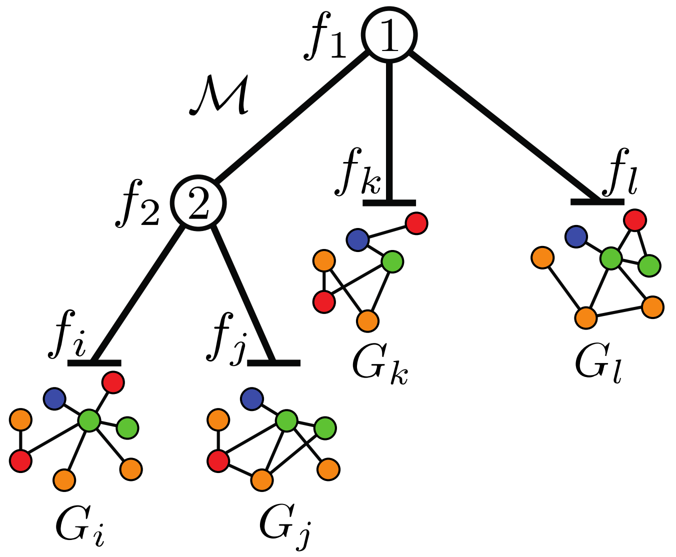
- Zitnik et al., 2017. [Predicting multicellular function through multi-layer tissue networks](). *ISMB & Bioinformatics.*
- [(56条消息) PPI数据集示例项目学习图神经网络_KPer_Yang的博客-CSDN博客]()

OhmNet: Multi-Layer Graphs



OhmNet: Multi-Layer Graphs

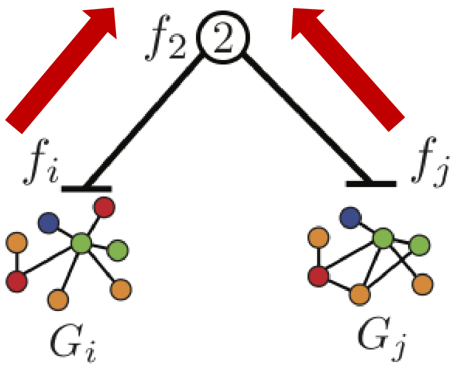
- **Input:** Given graphs G_i and hierarchy M
- **Output:** Embeddings for:
  - Nodes in each graph
  - Nodes in each sub-hierarchy



- Capture hierarchical structure of M

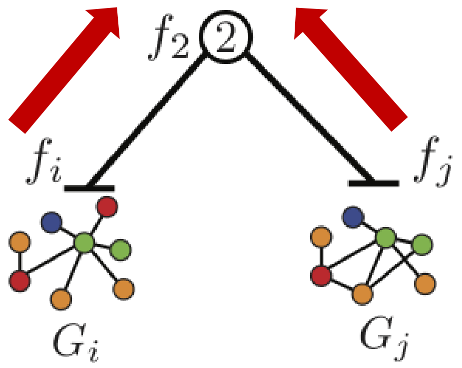
OhmNet: Multi-Layer Graphs

- For graphs G_i:
  - Use node2vec's biased walks
- For hierarchy M:
  - Encode dependencies between graphs
  - **Recursive regularization:** embeddings at level i are encouraged to be similar to embeddings in i's parent in the hierarchy



Random Walk Optimization

- Given simulated random walks for each graph:
  - Optimize node embeddings as described in previous.
  - **Extra:** Include terms for recursive regularization in the loss function



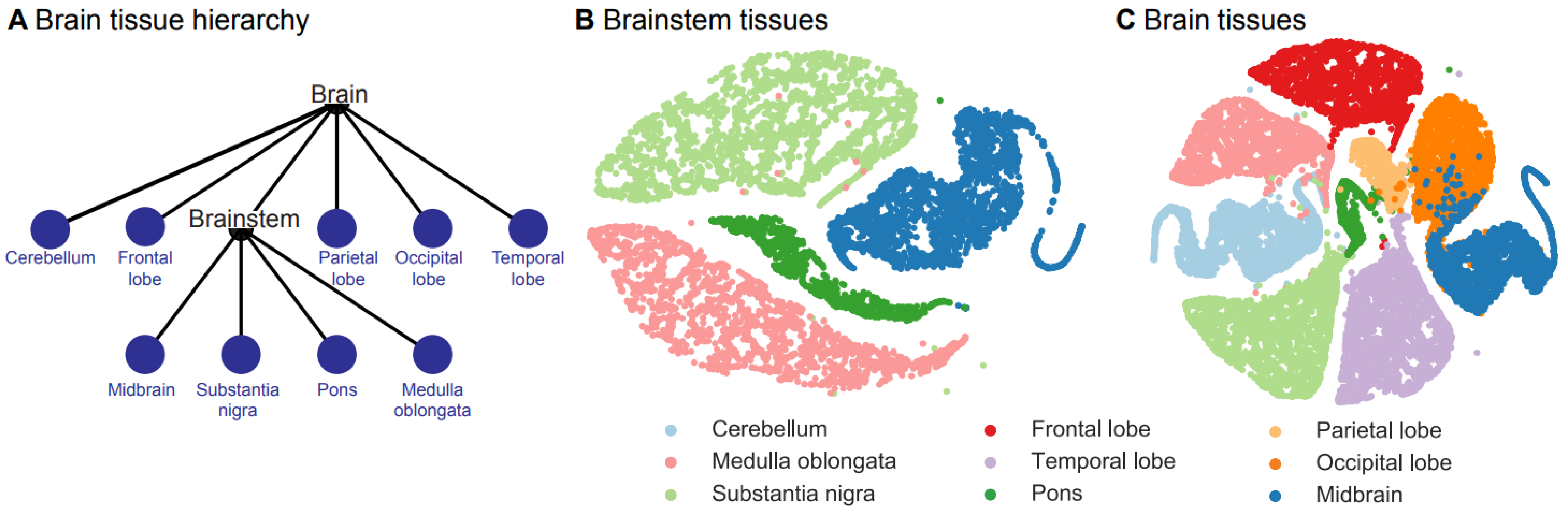
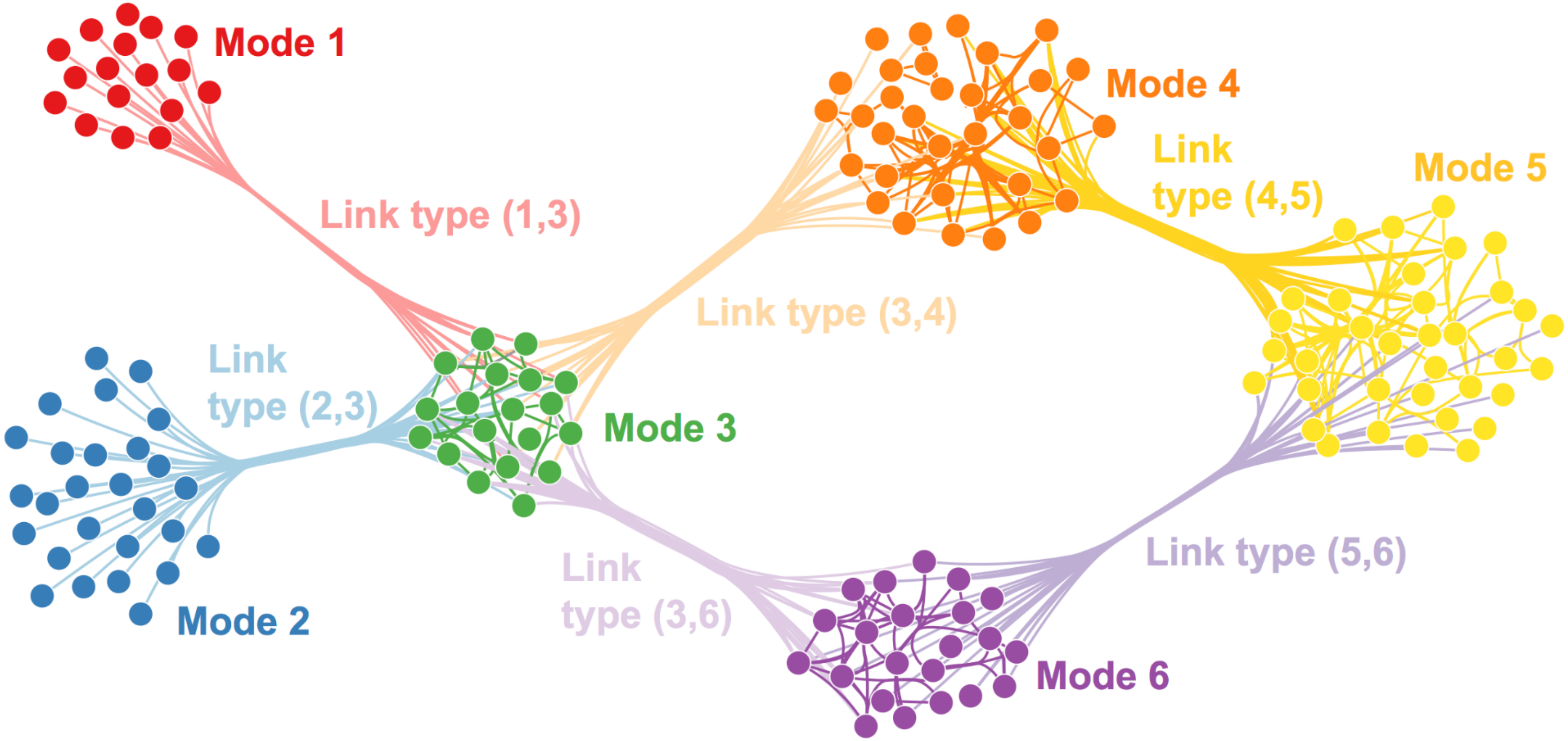


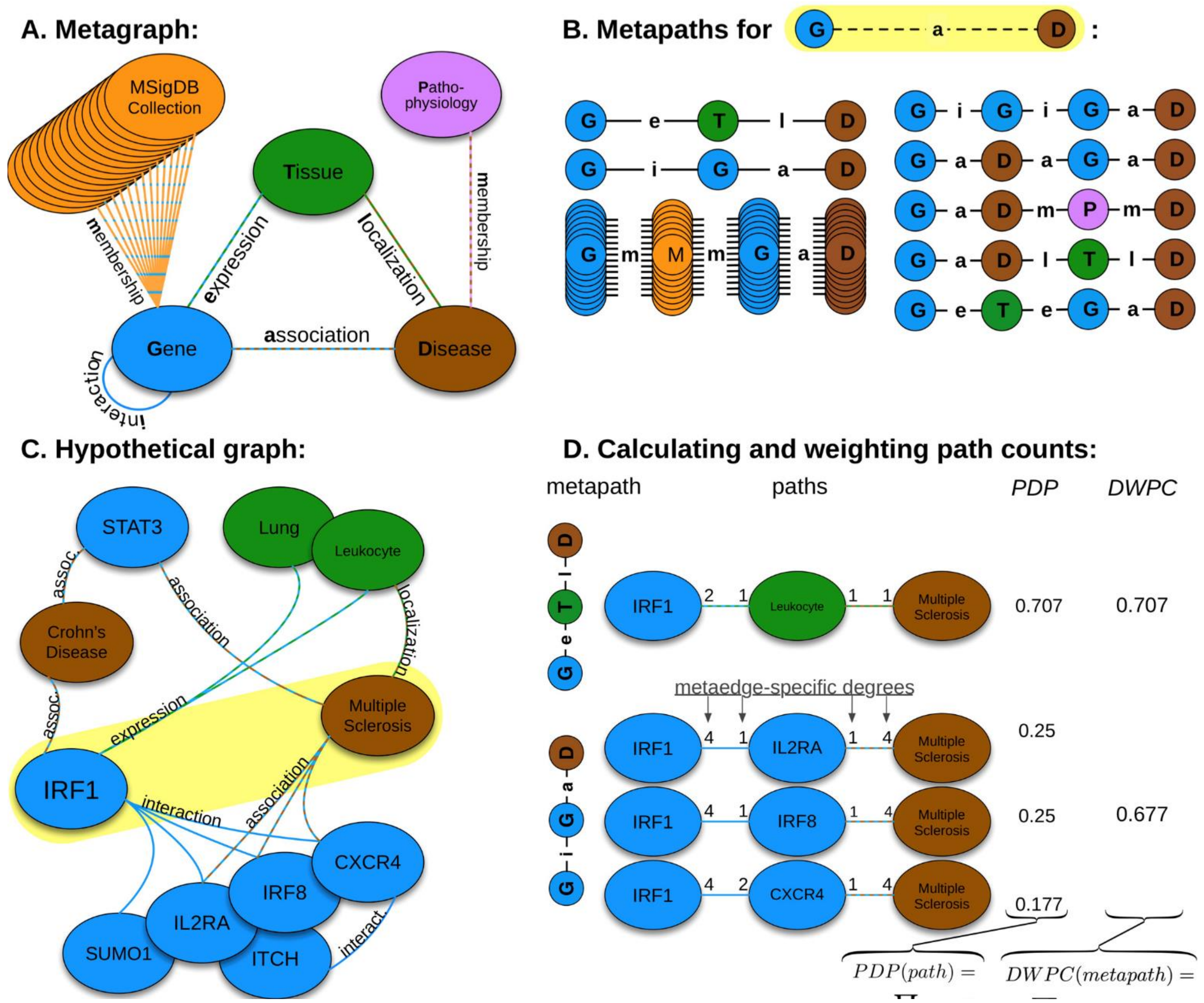
Fig. 5: **Visualization of the brain tissue-specific protein interaction networks. A.** The two-level brain tissue hierarchy as specified by the BRENDA Tissue Ontology (Chang *et al.*, 2014) and used in the case study in Section 5.3. Leaves of the hierarchy (in blue) represent nine brain tissues each of which is associated with a tissue-specific protein interaction network. **B.** Visualization of the brainstem-specific networks. The proteins are mapped to the 2-D space using the t-SNE package with learned features as input. Color of a node indicates the tissue of the protein. **C.** Visualization of the brain-specific networks. The proteins are mapped and colored using the same procedure as in B.

Metapath2vec



Dong et al., 2017. [metapath2vec: Scalable representation learning for heterogeneous networks](). *KDD.*

Metapaths



A. Metagraph:
MSigDB Collection
Patho-physiology
Tissue
membership
expression
localization
membership
Gene
association
Disease
interaction
B. Metapaths for G a D :
C. Hypothetical graph:
STAT3
Lung
Leukocyte
assoc.
association
localization
Crohn's Disease
Multiple Sclerosis
assoc.
expression
IRF1
interaction
association
CXCR4
IRF8
IL2RA
SUMO1
ITCH
interact.
D. Calculating and weighting path counts:
metapath
paths
PDP
DWPC
metaedge-specific degrees
0.707
0.707
0.25
0.25
0.677
0.177
PDP(path) =
DWPC(metapath) =

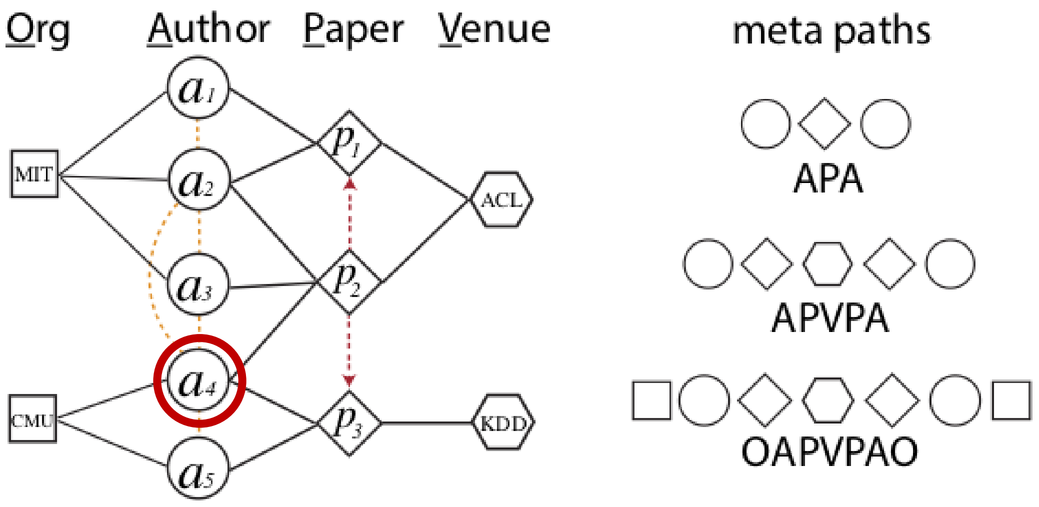
Metapath2vec: Two Main Steps

Extending node2vec to **het nets:**

1. **Metapath-based random walks**
   - Specify a metapath of interest
   - Run random walks that capture structural correlations between different node types
2. **Random walk optimization**
   - Given the random walks, optimize node embeddings

Step 1: Run Random Walks

- Given a metapath:
  - E.g., **OAP**VPAO
  - Generally, it is **symmetrical**

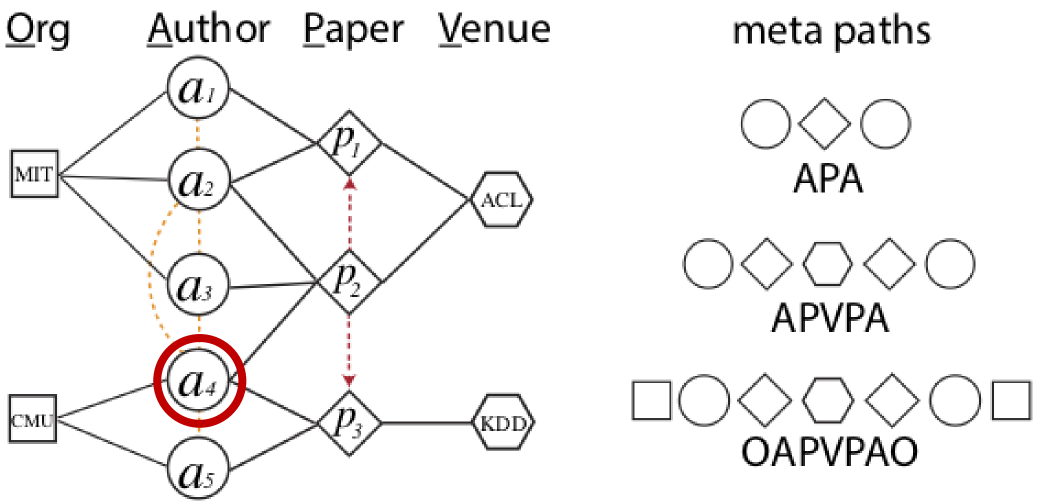


$$V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \cdots V_t \xrightarrow{R_t} V_{t+1} \cdots \xrightarrow{R_{l-1}} V_l$$

$$p(v^{i+1}|v_t^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t+1}(v_t^i)|} & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) = t{+}1 \\ 0 & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) \neq t{+}1 \\ 0 & (v^{i+1}, v_t^i) \notin E \end{cases}$$

Step 1: Run Random Walks

- Given a metapath:
  - E.g., **OAP**VPAO
  - Generally, it is **symmetrical**



- What is the next step of a walker on node a_4 that transitioned from node CMU?
  - **Standard random walk:** The next step can be all types of nodes surrounding it:
    - a_2, a_3, a_5, p_2, p_3, and CMU
  - **Metapath-based random walk:** The next step can only be a paper node (P), given that its current node is an author node a_4 (A) and its previous step was an organization node CMU (O):
    - Follow the semantics of this metapath

Step 2: Optimize

1. Simulate many **metapath-based random walks** starting from each node
2. For each node u, get $N_t(u)$ as a nodes of type t that are visited by random walks starting at u
3. **For each node u,** learn its embedding by **predicting which nodes are in $N_t(u)$:**

$$\mathcal{L} = \sum_{u \in V} \sum_{t \in V_t} \sum_{v \in N_t(u)} -\log(P(v|\mathbf{z}_u))$$

V_t is the vertex set for node type t

Step 2: Optimize

$$\mathcal{L} = \sum_{u \in V} \sum_{t \in V_t} \sum_{v \in N_t(u)} -\log(P(v|\mathbf{z}_u))$$

V_t is the vertex set for node type t

$$\arg\max_{\theta} \sum_{v \in V} \sum_{t \in T_V} \sum_{c_t \in N_t(v)} logp(c_t|v;\theta)$$

$$p(c_t|v;\theta) = \frac{e^{X_{c_t} \cdot X_v}}{\sum_{u \in V} e^{X_u \cdot X_v}}$$

Heterogeneous Graph Attention Network

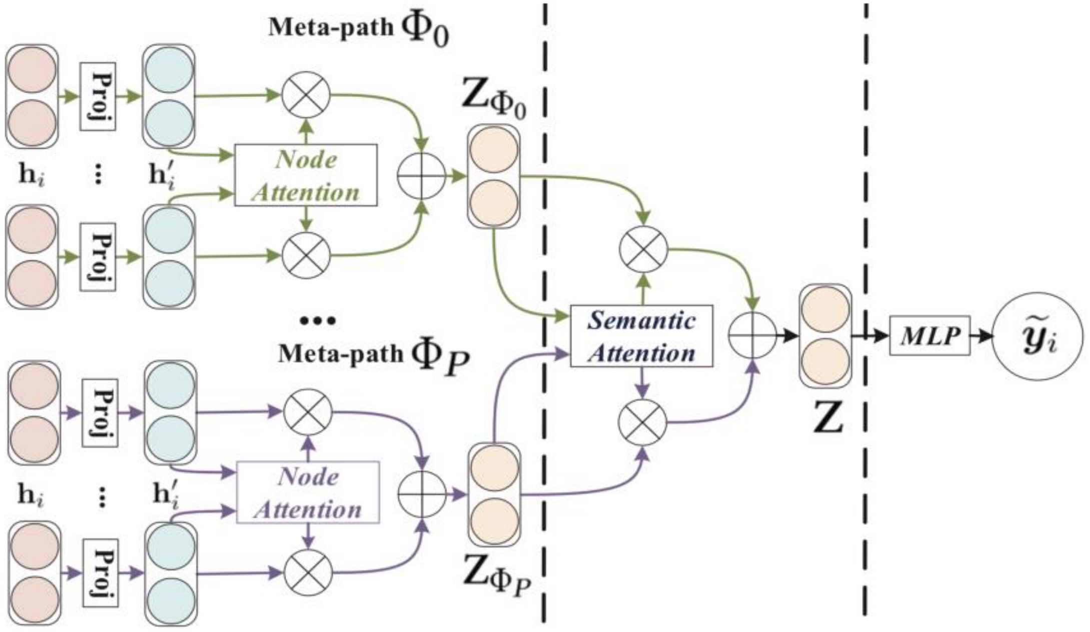
Xiao Wang, Houye Ji
Beijing University of Posts and Telecommunications
Beijing, China
{xiaowang,jhy1993}@bupt.edu.cn

Chuan Shi*, Bai Wang
Beijing University of Posts and Telecommunications
Beijing, China
{shichuan,wangbai}@bupt.edu.cn

Peng Cui, P. Yu
Tsinghua University
Beijing, China
{cuip,psyu}@tsinghua.edu.cn

Yanfang Ye
West Virginia University
WV, USA
yanfang.ye@mail.wvu.edu

Notation	Explanation
Φ	Meta-path
$\mathbf{h}$	Initial node feature
$\mathbf{M}_{\phi}$	Type-specific transformation matrix
$\mathbf{h}'$	Projected node feature
e_{ij}^{Φ}	Importance of meta-path based node pair (i,j)
$\mathbf{a}_{\Phi}$	Node-level attention vector for meta-path Φ
α_{ij}^{Φ}	Weight of meta-path based node pair (i,j)
$\mathcal{N}^{\Phi}$	Meta-path based neighbors
$\mathbf{Z}_{\Phi}$	Semantic-specific node embedding
$\mathbf{q}$	Semantic-level attention vector
w_{Φ}	Importance of meta-path Φ
β_{Φ}	Weight of meta-path Φ
$\mathbf{Z}$	The final embedding



(a) Node-Level Attention (b) Semantic-Level Attention (c) Prediction

Heterogeneous Graph Attention Network

Xiao Wang, Houye Ji
Beijing University of Posts and Telecommunications
Beijing, China
{xiaowang,jhy1993}@bupt.edu.cn

Chuan Shi*, Bai Wang
Beijing University of Posts and Telecommunications
Beijing, China
{shichuan,wangbai}@bupt.edu.cn

Peng Cui, P. Yu
Tsinghua University
Beijing, China
{cuip,psyu}@tsinghua.edu.cn

Yanfang Ye
West Virginia University
WV, USA
yanfang.ye@mail.wvu.edu

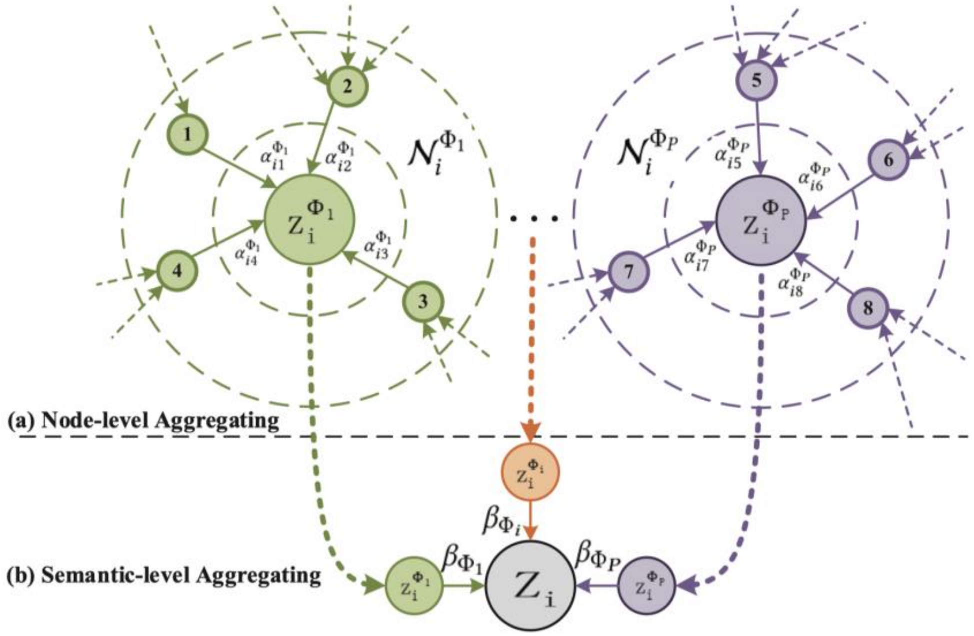
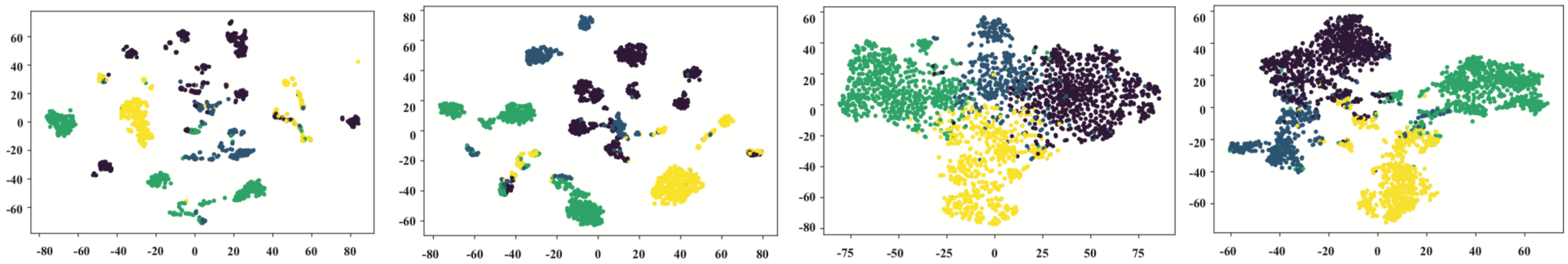


Table 3: Qantitative results (%) on the node classification task.

Datasets	Metrics	Training	DeepWalk	ESim	metapath2vec	HERec	GCN	GAT	HAN_{nd}	HAN_{sem}	HAN
ACM	Macro-F1	20%	77.25	77.32	65.09	66.17	86.81	86.23	88.15	89.04	**89.40**
		40%	80.47	80.12	69.93	70.89	87.68	87.04	88.41	89.41	**89.79**
		60%	82.55	82.44	71.47	72.38	88.10	87.56	87.91	**90.00**	89.51
		80%	84.17	83.00	73.81	73.92	88.29	87.33	88.48	90.17	**90.63**
	Micro-F1	20%	76.92	76.89	65.00	66.03	86.77	86.01	87.99	88.85	**89.22**
		40%	79.99	79.70	69.75	70.73	87.64	86.79	88.31	89.27	**89.64**
		60%	82.11	82.02	71.29	72.24	88.12	87.40	87.68	**89.85**	89.33
		80%	83.88	82.89	73.69	73.84	88.35	87.11	88.26	89.95	**90.54**
DBLP	Macro-F1	20%	77.43	91.64	90.16	91.68	90.79	90.97	91.17	92.03	**92.24**
		40%	81.02	92.04	90.82	92.16	91.48	91.20	91.46	92.08	**92.40**
		60%	83.67	92.44	91.32	92.80	91.89	90.80	91.78	92.38	**92.80**
		80%	84.81	92.53	91.89	92.34	92.38	91.73	91.80	92.53	**93.08**
	Micro-F1	20%	79.37	92.73	91.53	92.69	91.71	91.96	92.05	92.99	**93.11**
		40%	82.73	93.07	92.03	93.18	92.31	92.16	92.38	93.00	**93.30**
		60%	85.27	93.39	92.48	93.70	92.62	91.84	92.69	93.31	**93.70**
		80%	86.26	93.44	92.80	93.27	93.09	92.55	92.69	93.29	**93.99**
IMDB	Macro-F1	20%	40.72	32.10	41.16	41.65	45.73	49.44	49.78	**50.87**	50.00
		40%	45.19	31.94	44.22	43.86	48.01	50.64	52.11	50.85	**52.71**
		60%	48.13	31.68	45.11	46.27	49.15	51.90	51.73	52.09	**54.24**
		80%	50.35	32.06	45.15	47.64	51.81	52.99	52.66	51.60	**54.38**
	Micro-F1	20%	46.38	35.28	45.65	45.81	49.78	55.28	54.17	55.01	**55.73**
		40%	49.99	35.47	48.24	47.59	51.71	55.91	56.39	55.15	**57.97**
		60%	52.21	35.64	49.09	49.88	52.29	56.44	56.09	56.66	**58.32**
		80%	54.33	35.59	48.81	50.99	54.61	56.97	56.38	56.49	**58.51**



(a) GCN (b) GAT (c) metapath2vec (d) HAN

Figure 6: Visualization embedding on DBLP. Each point indicates one author and its color indicates the research area.