Heterogeneous Graph Attention Network

Xiao Wang[1], **Houye Ji[1]**, Chuan Shi[1], Bai Wang[1], Peng Cui[2], P. Yu[2], Yanfang Ye[3]

[1]Beijing University of Posts and Telecommunications

[2]Tsinghua University

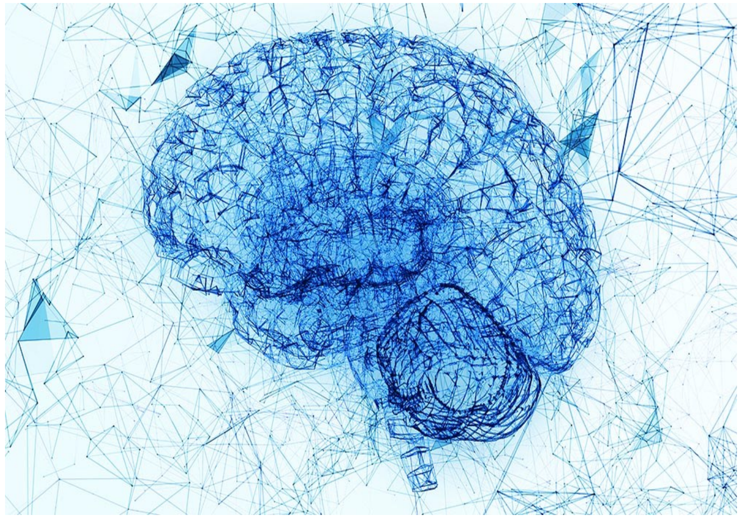
[3]West Virginia University

CONTENTS

CONTENTS

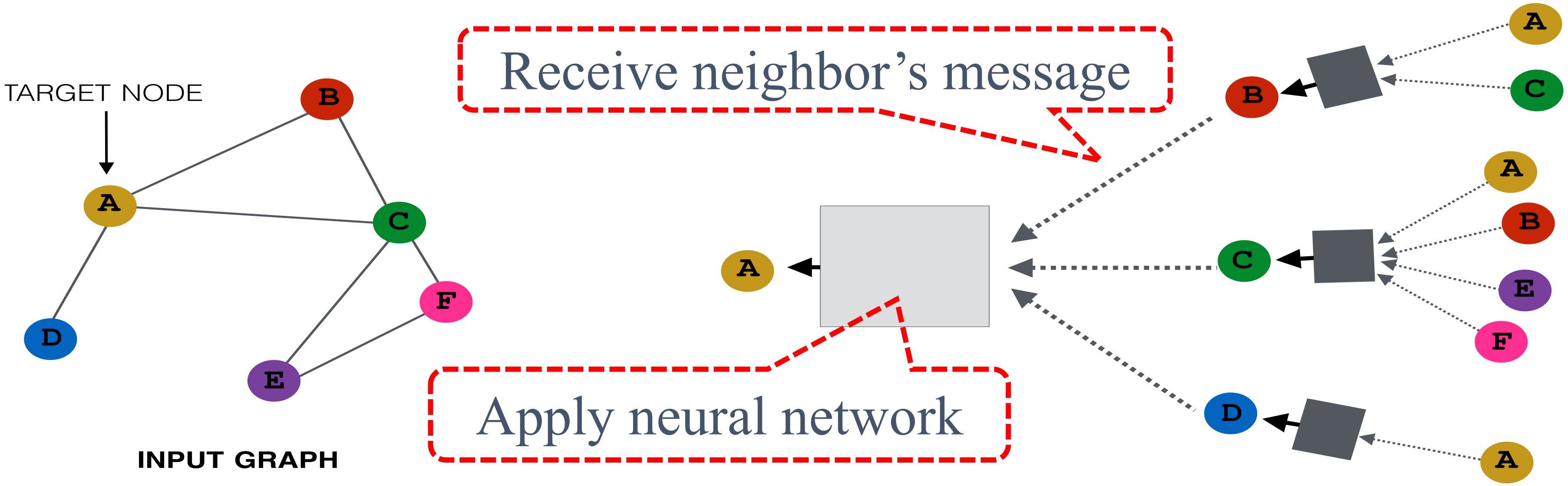
Graph-structured Data

- Graph-structured data are ubiquitous.
- Graph-structured data are flexible to model complex interactions.



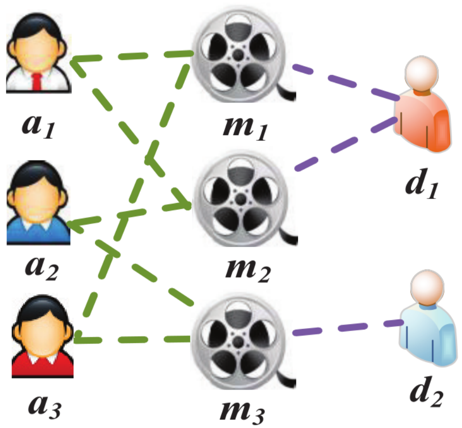
Graph Neural Network

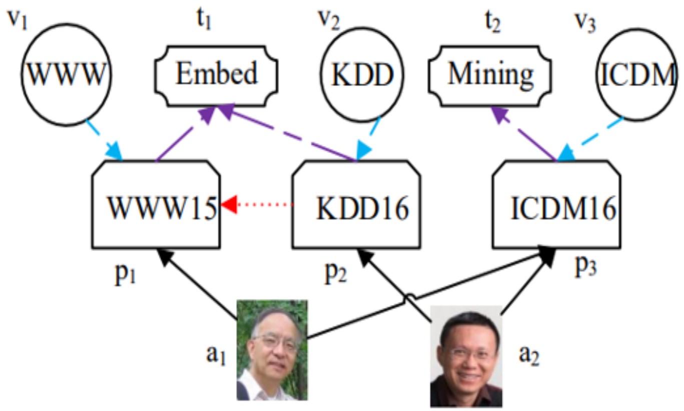
- Neural networks for processing graph-structured inputs.
- Flexible to characterize non-Euclidean data.
- For example, graph convolutional network and graph attention network.

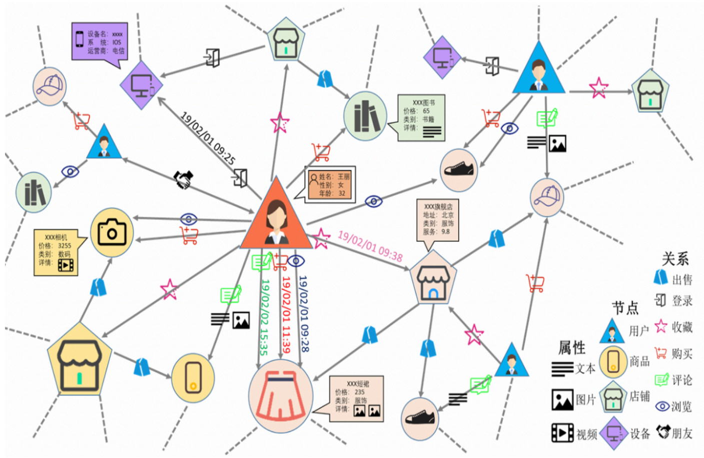


Heterogeneous Graph

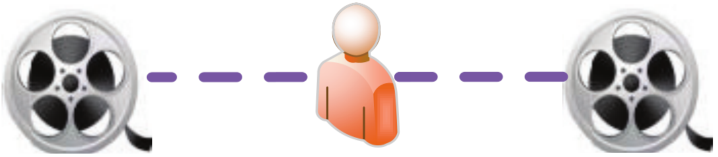
- Multiple types of nodes or links







- Rich semantic information
  - Meta-path: a relation sequence connecting two objects (e.g., Movie-Actor-Movie).



Movie-Director-Moive

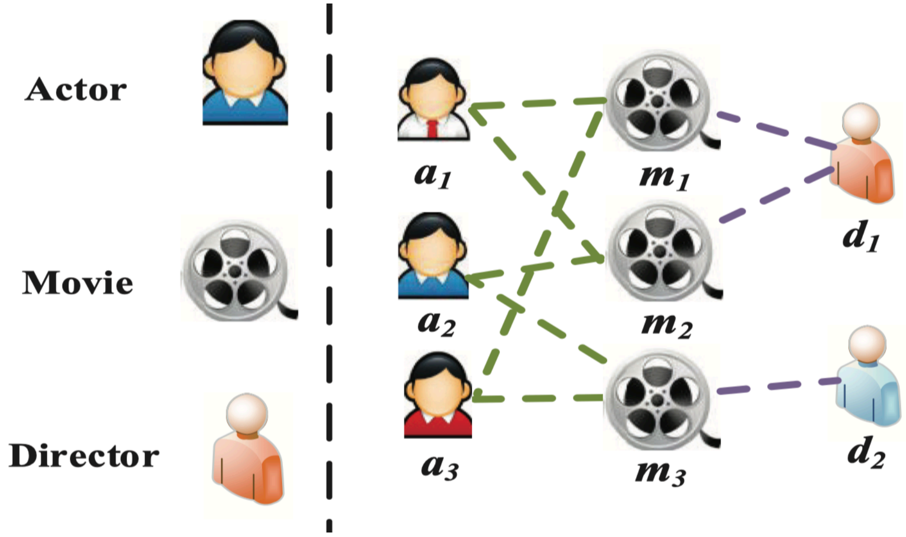
Movie-Actor-Moive

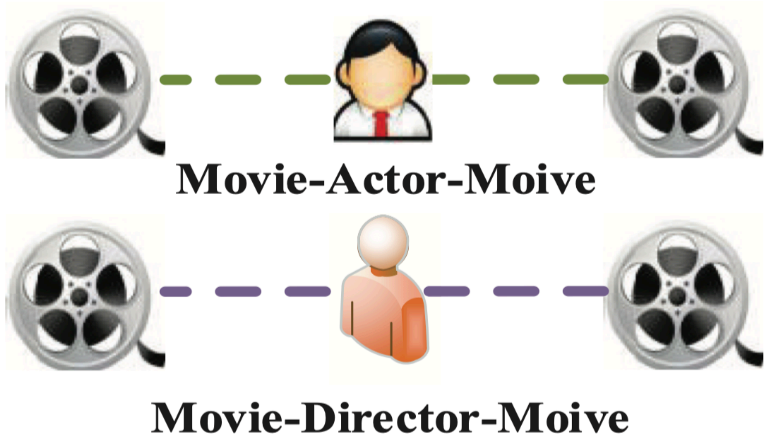
Two movies directed by the same director.

Two movies are starred by the same actor.

Existing Graph Neural Networks focus on homogeneous graph

- Cannot handle multiple types of nodes and edges.
- Cannot capture rich semantic information.



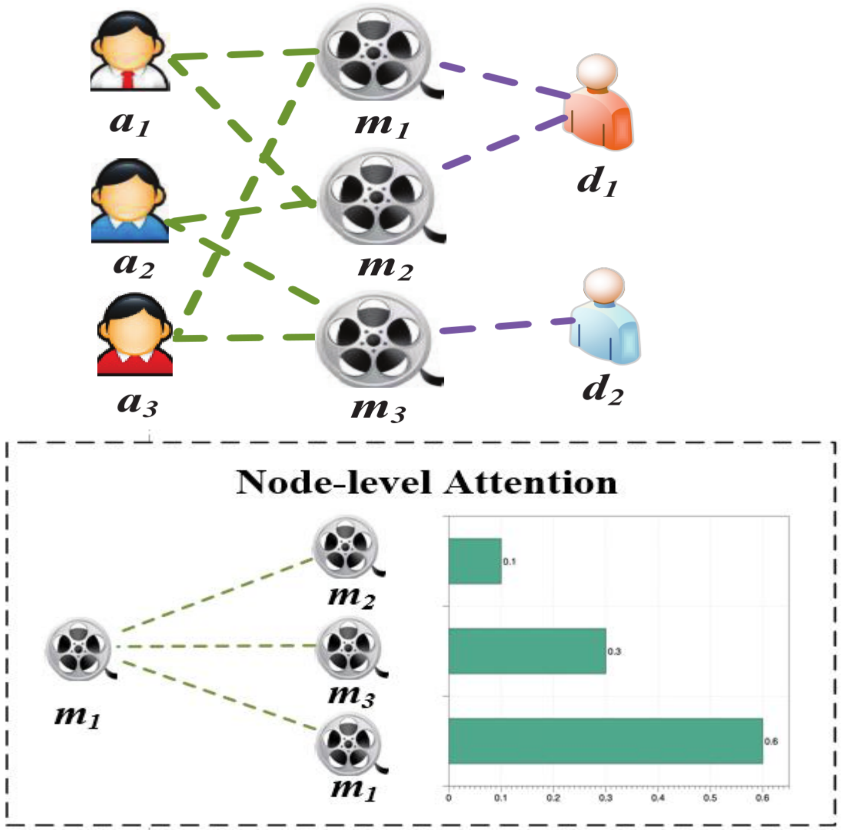


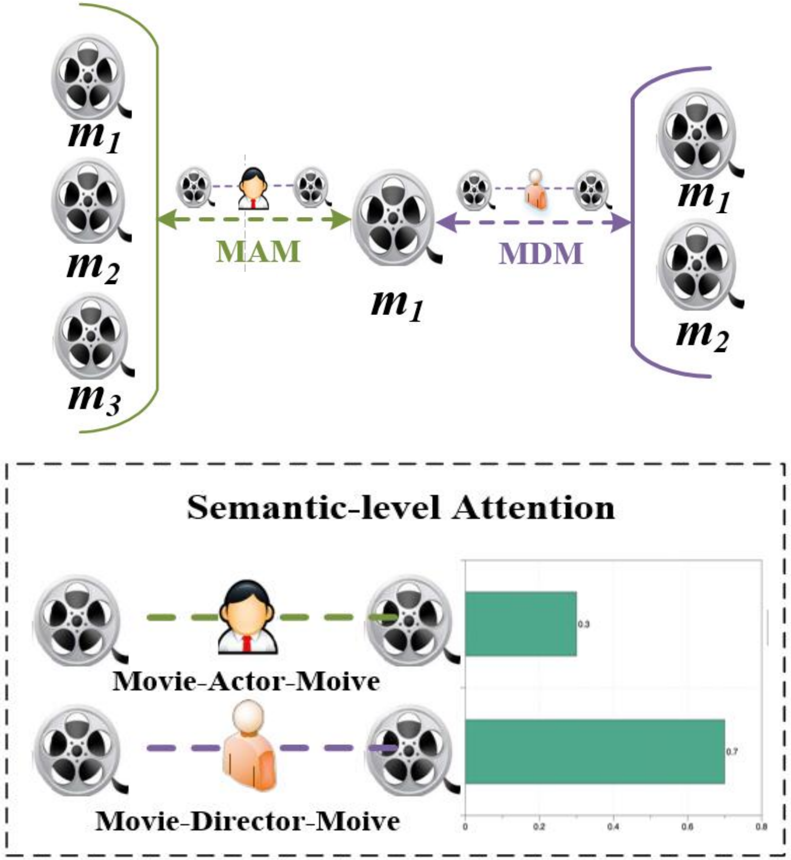
Heterogeneous Structure
Multiple types of nodes or links

Rich Semantic
Various semantic relations

Challenges

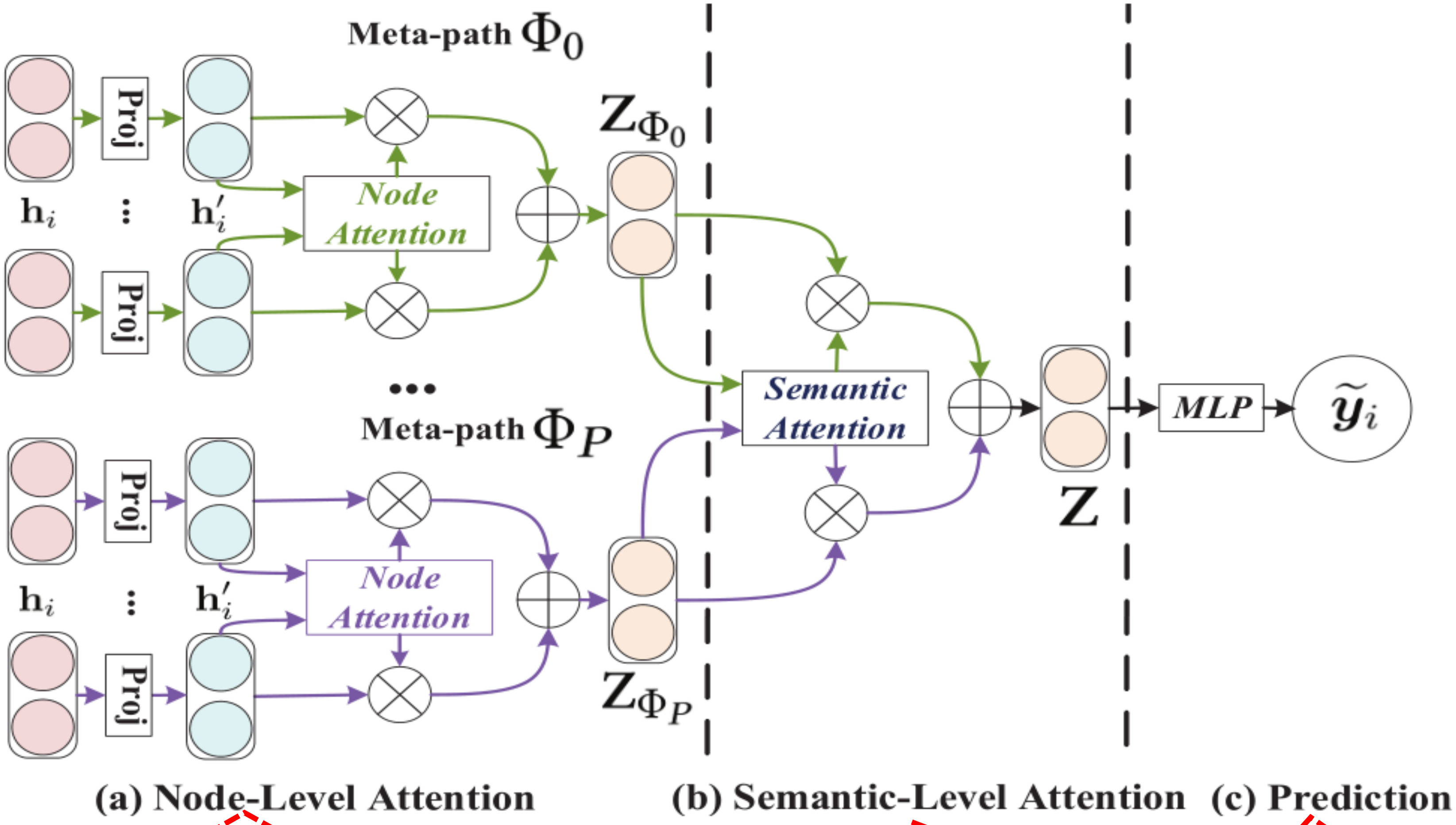
- How to handle the heterogeneity of graph?
- How to discover the differences of meta-path based neighbors?
- How to find some meaningful meta-paths?





CONTENTS

Heterogeneous Graph Attention Network（HAN）

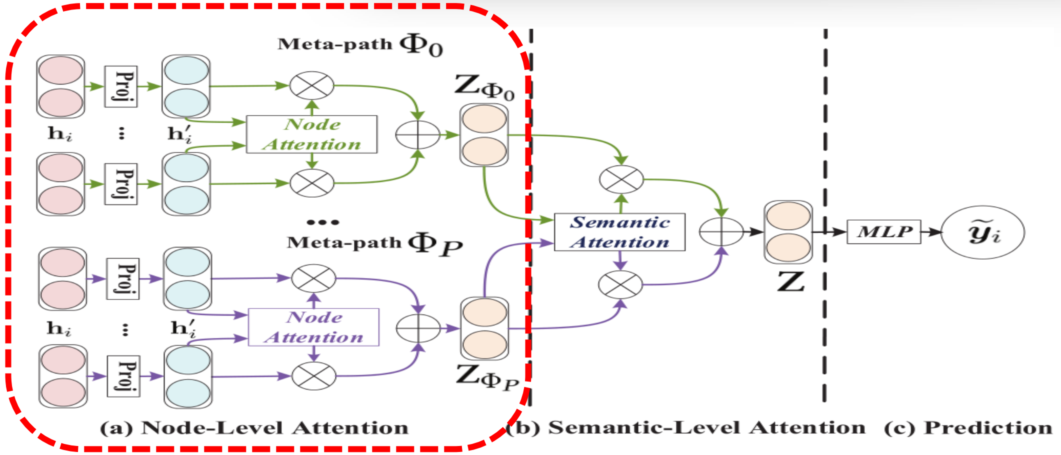


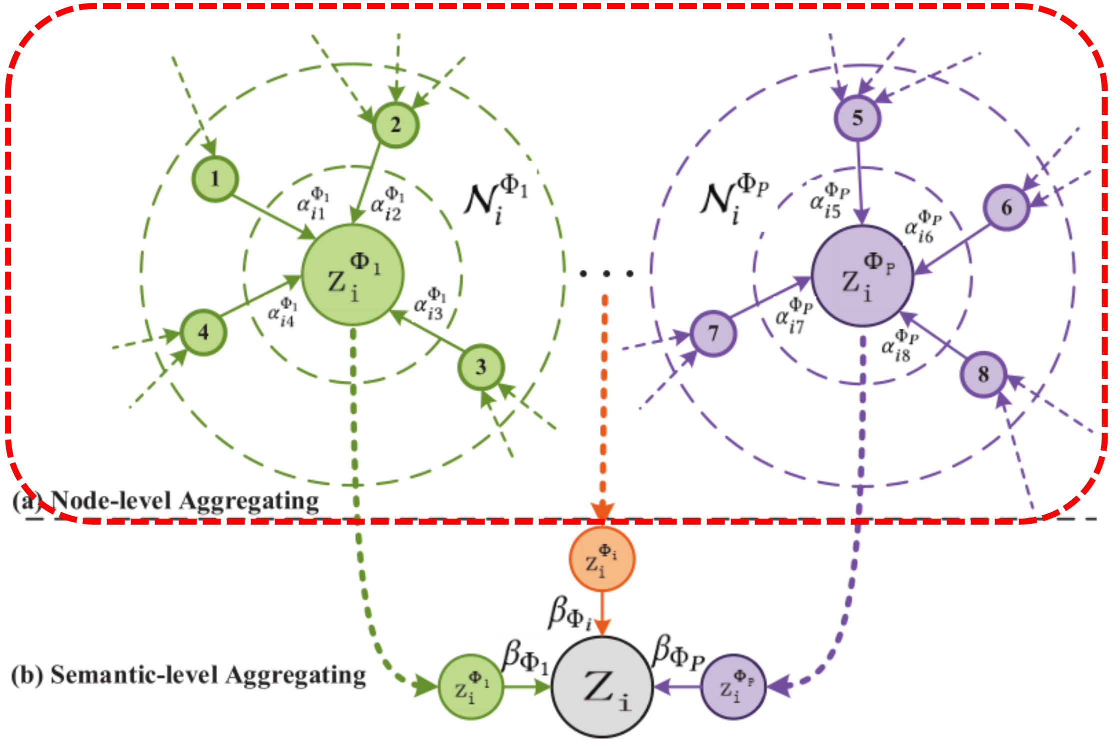
Model heterogeneous structure.

Capture rich semantics.

Task-specific loss.

Node-Level Attention and Aggregating





- **Type-Specific Transformation**

$$\mathbf{h}'_i = \mathbf{M}_{\phi_i} \cdot \mathbf{h}_i,$$

Type-specific transformation matrix

- **Importance of Neighbors**

$$e_{ij}^{\Phi} = att_{node}\left(\mathbf{h}'_i, \mathbf{h}'_j; \Phi\right)$$

$$e_{ij}^{\Phi} = \sigma\left(\mathbf{a}_{\Phi}^{\mathrm{T}} \cdot \left[\mathbf{h}'_i \| \mathbf{h}'_j\right]\right)$$

$$\alpha_{ij}^{\Phi} = softmax_j\left(e_{ij}^{\Phi}\right)$$

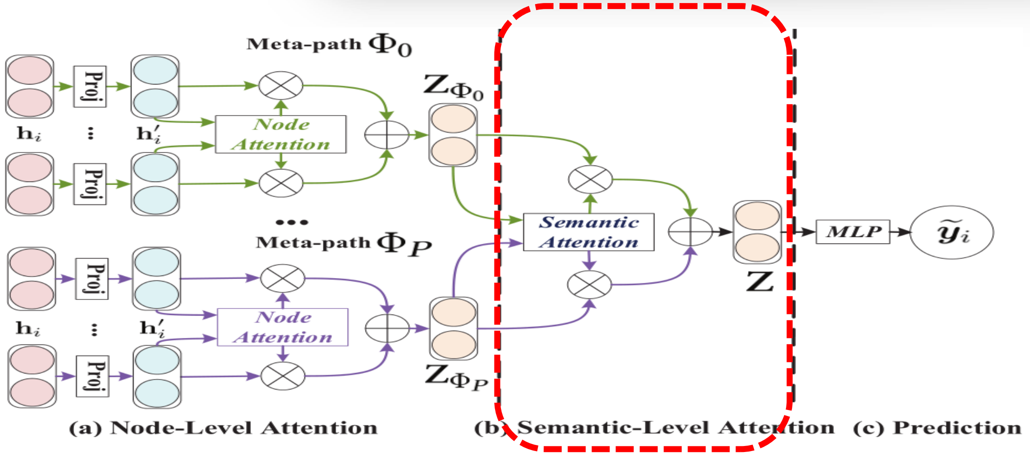
Node-level attention vector

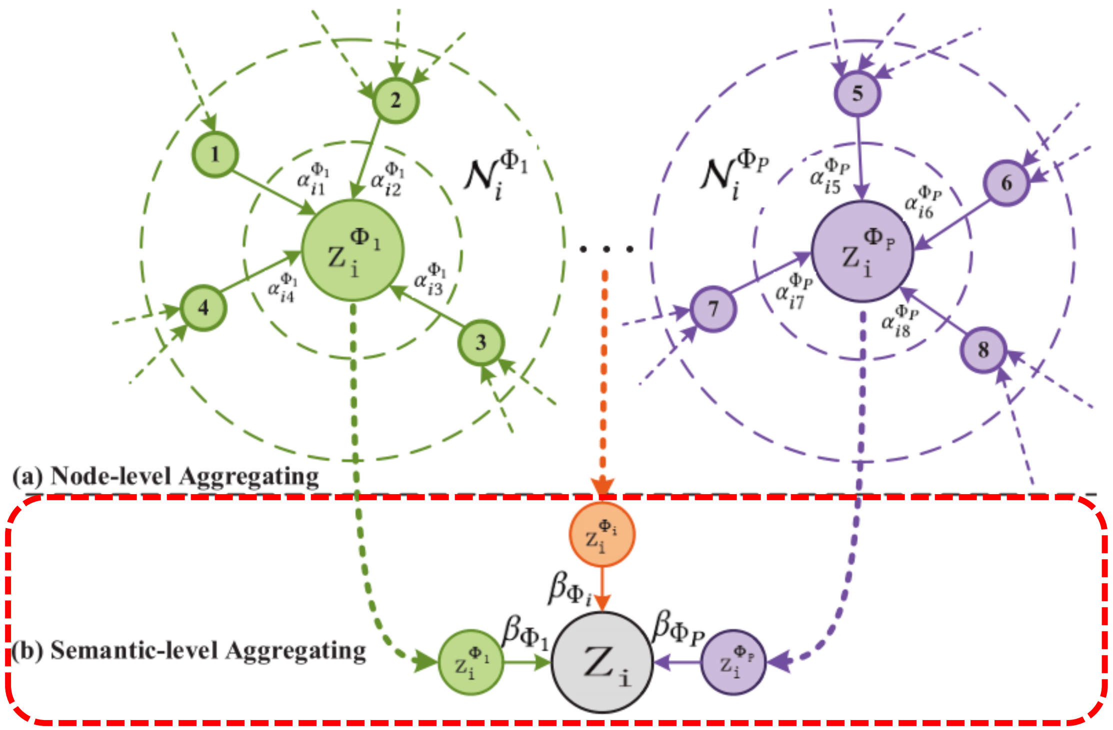
- **Node-Level Aggregating**

Node weight

$$\mathbf{z}_i^{\Phi} = \sigma\left(\sum_{j \in \mathcal{N}_i^{\Phi}} \alpha_{ij}^{\Phi} \cdot \mathbf{h}'_j\right)$$

Semantic-Level Attention and Aggregating





- Semantic-Level Attention

$$(\beta_{\Phi_0}, \beta_{\Phi_1}, \ldots, \beta_{\Phi_P}) = att_{sem}(\mathbf{Z}_{\Phi_0}, \mathbf{Z}_{\Phi_1}, \ldots, \mathbf{Z}_{\Phi_P})$$

- Importance of Meta-path

Semantic-level attention vector

$$w_{\Phi_i} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \mathbf{q}^{\mathrm{T}} \cdot \tanh(\mathbf{W} \cdot \mathbf{z}_i^{\Phi} + \mathbf{b})$$

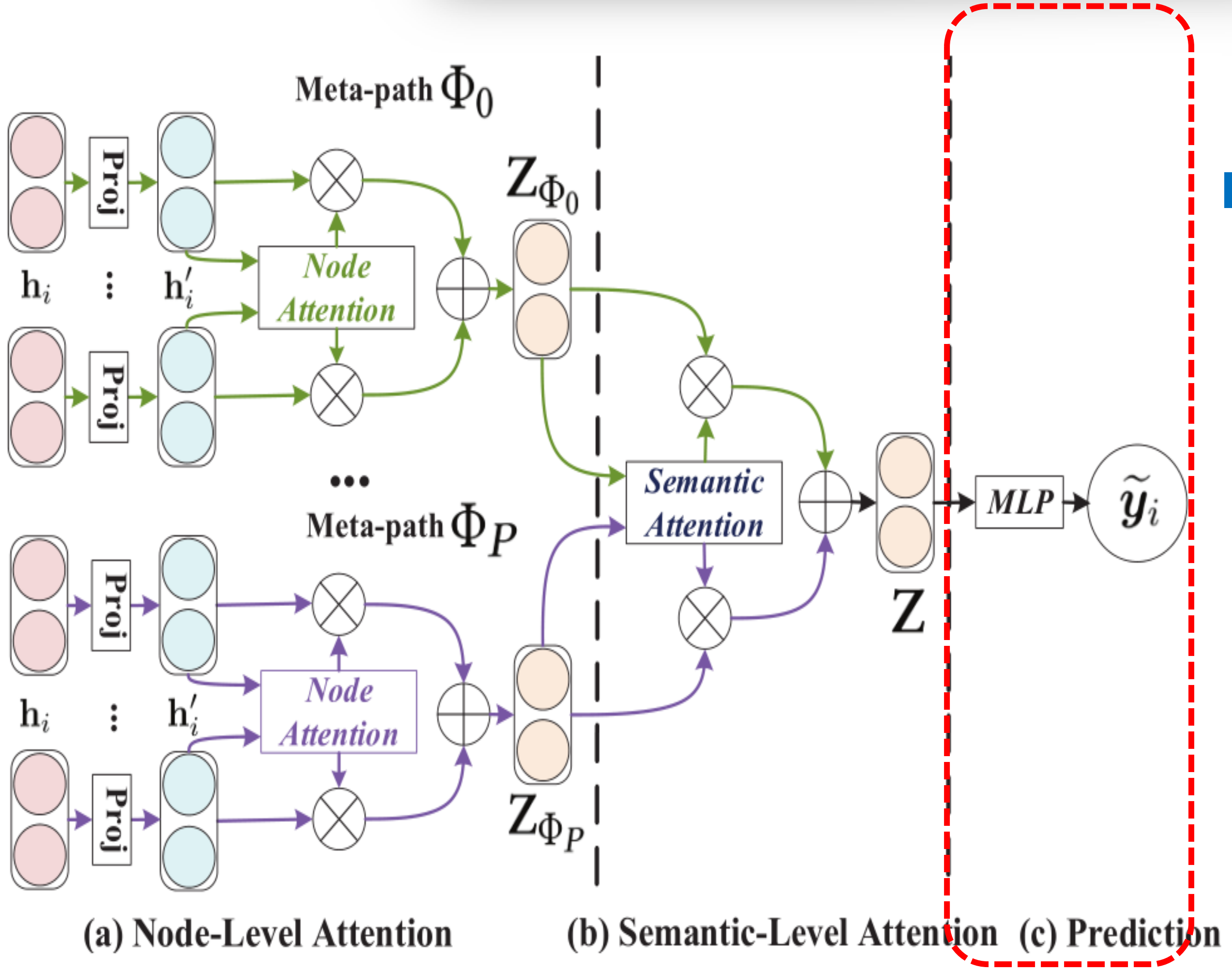
$$\beta_{\Phi_i} = \frac{\exp(w_{\Phi_i})}{\sum_{i=1}^{P} \exp(w_{\Phi_i})}$$

- Semantic-Level Aggregating

$$\mathbf{Z} = \sum_{i=1}^{P} \beta_{\Phi_i} \cdot \mathbf{Z}_{\Phi_i}$$

Semantic weight

Prediction



Prediction

- Semi-supervised Loss

Parameter of classifier

$$L = -\sum_{l \in \mathcal{Y}_L} \mathrm{Y}^l \ln(\mathrm{C} \cdot \mathrm{Z}^l)$$

Labeled data

Optimize for the specific task (e.g., node classification).

CONTENTS

Baselines

- Deepwalk
- Esim
- Metapath2vec
- HERec
- GCN
- GAT
- HAN_{nd}
- HAN_{sem}

Tasks

- Node Classification
- Node Clustering
- Analysis of Attention Mechanism
- Visualization

Table 2: Statistics of the datasets.

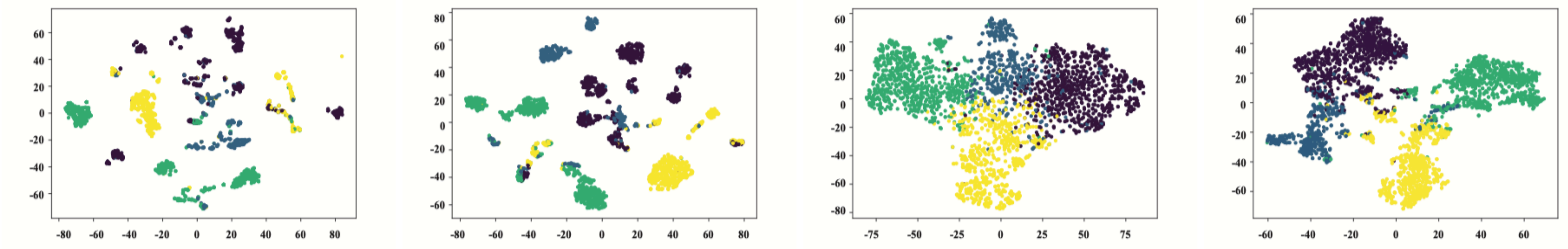
Dataset	Relations(A-B)	Number of A	Number of B	Number of A-B	Feature	Training	Validation	Test	Meta-paths
DBLP	Paper-Author	14328	4057	19645	334	800	400	2857	*APA*
	Paper-Conf	14328	20	14328					*APCPA*
	Paper-Term	14327	8789	88420					*APTPA*
IMDB	Movie-Actor	4780	5841	14340	1232	300	300	2687	*MAM*
	Movie-Director	4780	2269	4780					*MDM*
ACM	Paper-Author	3025	5835	9744	1830	600	300	2125	*PAP*
	Paper-Subject	3025	56	3025					*PSP*

Node Classification

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**

Node Clustering & Visualization

Datasets	Metrics	DeepWalk	ESim	metapath2vec	HERec	GCN	GAT	HAN_{nd}	HAN_{sem}	HAN
ACM	NMI	41.61	39.14	21.22	40.70	51.40	57.29	60.99	61.05	**61.56**
	ARI	35.10	34.32	21.00	37.13	53.01	60.43	61.48	59.45	**64.39**
DBLP	NMI	76.53	66.32	74.30	76.73	75.01	71.50	75.30	77.31	**79.12**
	ARI	81.35	68.31	78.50	80.98	80.49	77.26	81.46	83.46	**84.76**
IMDB	NMI	1.45	0.55	1.20	1.20	5.45	8.45	9.16	10.31	**10.87**
	ARI	2.15	0.10	1.70	1.65	4.40	7.46	7.98	9.51	**10.01**



(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.

- Node-Level Attention (e.g., P831[1])

P831 > P699 > … > P2328 > P1973

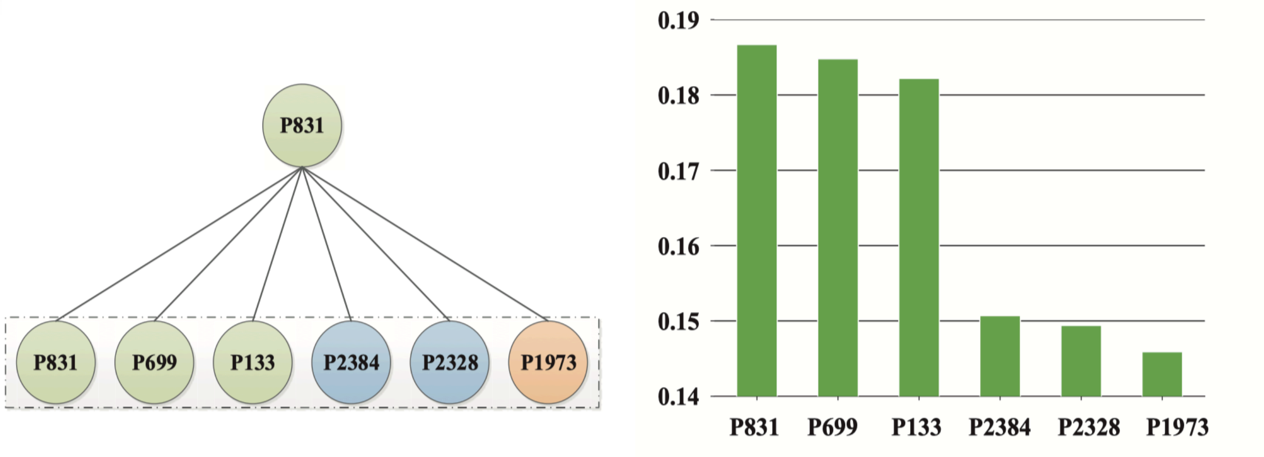
Important neighbors have larger attention values.

- Semantic-Level Attention

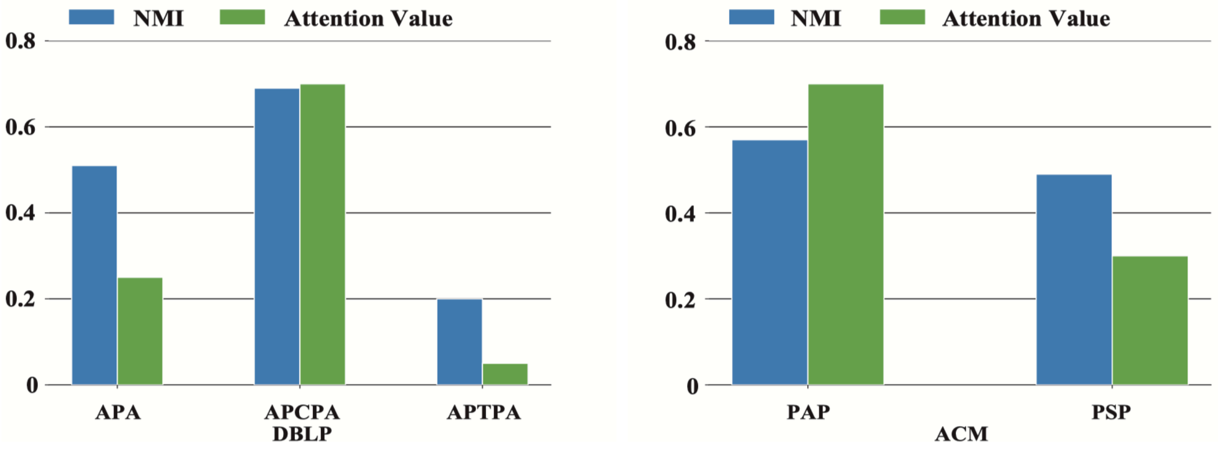
APCPA > APA > APTPA

PAP > PSP

Important meta-paths have larger attention values.



(a) Meta-path based neighbors of P831 (b) Attention values of P831's neighbors



(a) NMI values on DBLP (b) NMI values on ACM

[1]Xintao Wu, et al. Screening and Interpreting Multi-item Associations Based on Log-linear Modeling. KDD 2003

CONTENTS

- The first attempt to study the heterogeneous graph neural network based on attention mechanism.
- A novel heterogeneous graph attention network (HAN) which includes both of the node-level and semantic-level attentions.
- The state-of-art performance and good interpretability.

Thank you !
Q&A

More materials in http://shichuan.org
Code https://github.com/Jhy1993/HAN