太原理工大學
TAIYUAN UNIVERSITY OF TECHNOLOGY

Collaborative Knowledge Distillation for Heterogeneous Information Network Embedding

Can Wang, Sheng Zhou*, Kang Yu, Defang Chen, Bolang Li, Yan Feng, Chun Chen

College of Computer Science, Zhejiang University
ZJU-Bangsun Joint Research Center
School of Software Technology, Zhejiang University
Zhejiang Provincial Key Laboratory of Service Robot

WWW 2022
Code: https://github.com/zhoushengisnoob/CKD

Zhuomin Chen
2022.05.26

Introduction

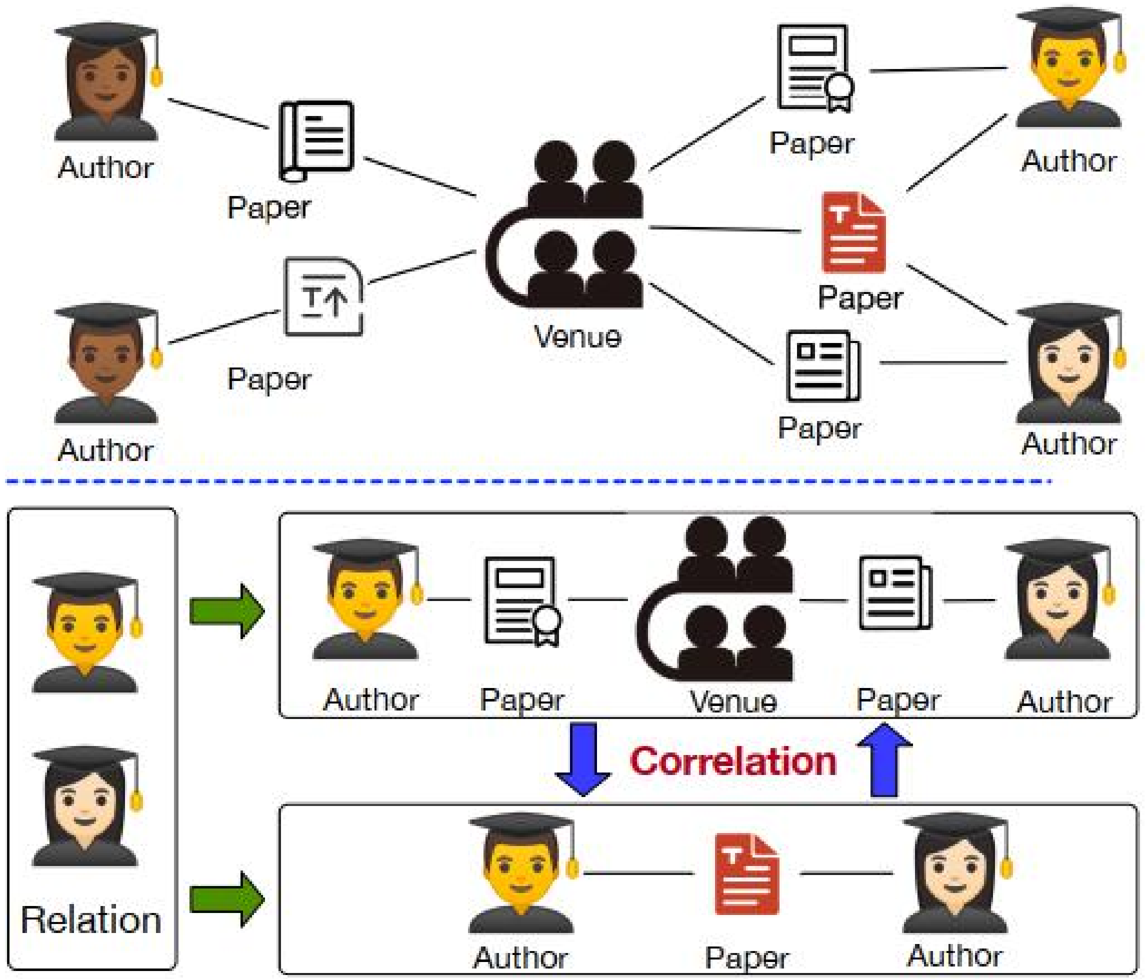
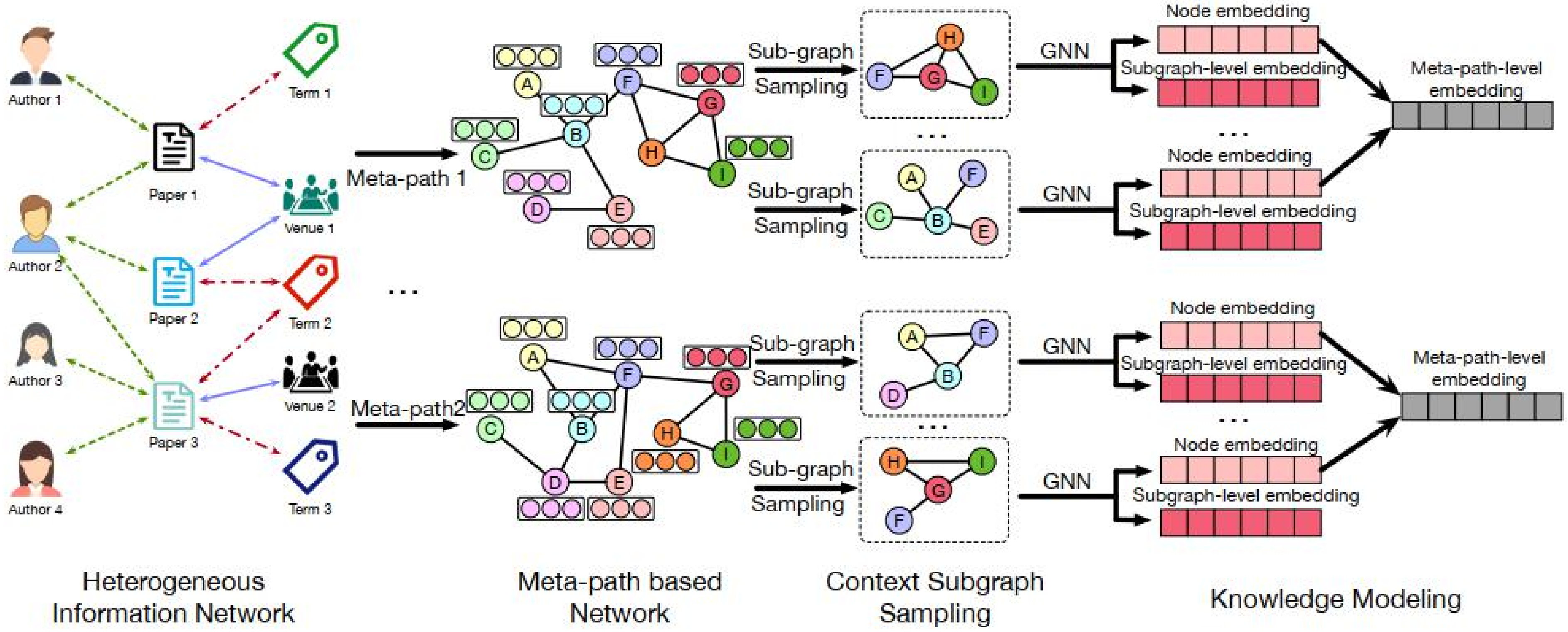
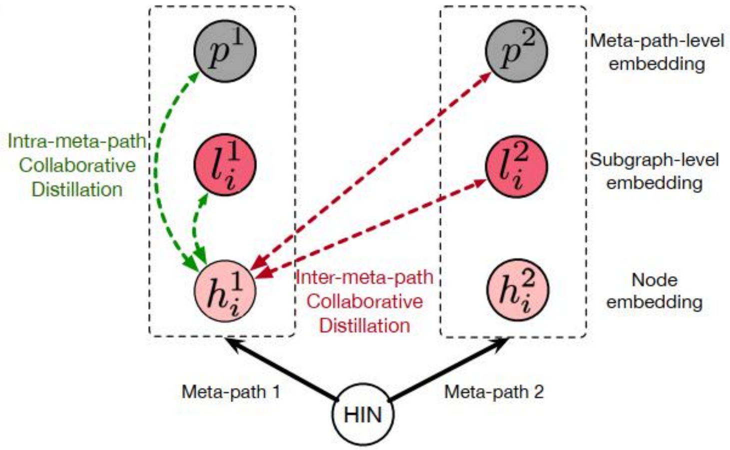


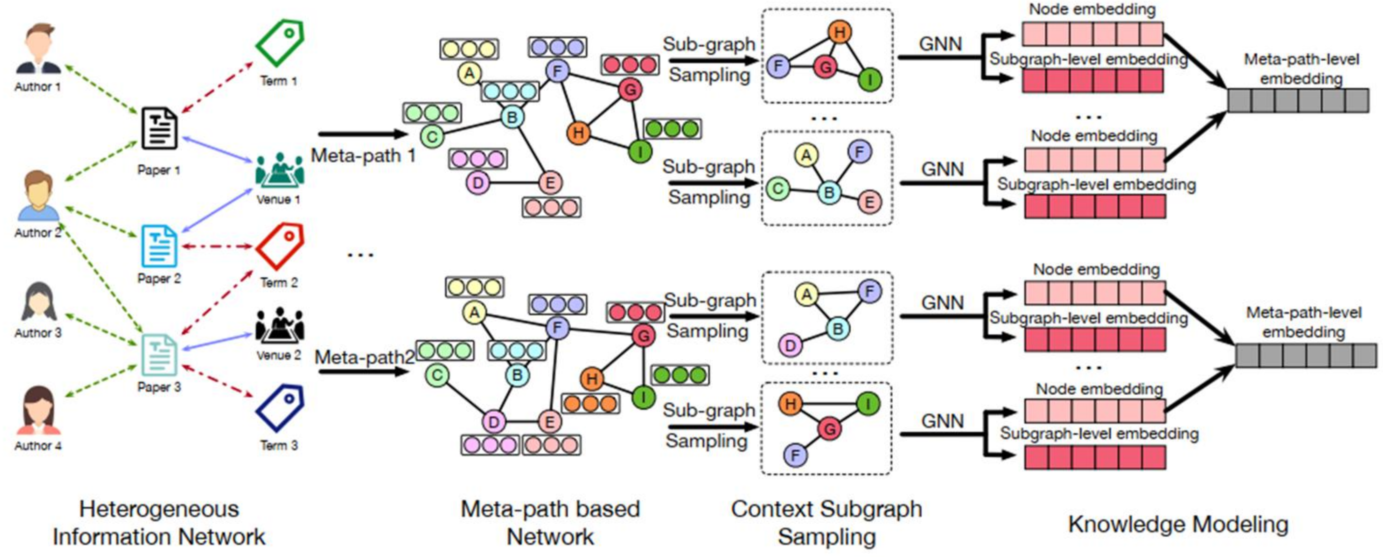
Figure 1: Heterogeneous information network and meta-path.

Methodology





Semantic Context Subgraph Sampling



HIN $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{R}\}$ meta-path set $\mathcal{M}$

$\mathcal{V}$ *is the set of typed nodes,* $\mathcal{E}$ *is the set of typed edges.*

$\mathcal{R}$ *is the set of edge types*

for each meta-path $m \in \mathcal{M}$, project the HIN into meta-path based homogeneous network G^m

Personalized PageRank (PPR)

$$\mathbf{S}^m = \alpha \left(\mathbf{I}_n - (1-\alpha)\mathbf{D}_m^{-1/2}\mathbf{A}^m\mathbf{D}_m^{-1/2} \right)^{-1} \quad (1)$$

$\mathbf{S}^m \in \mathbb{R}^{N \times N}$ is the diffusion matrix

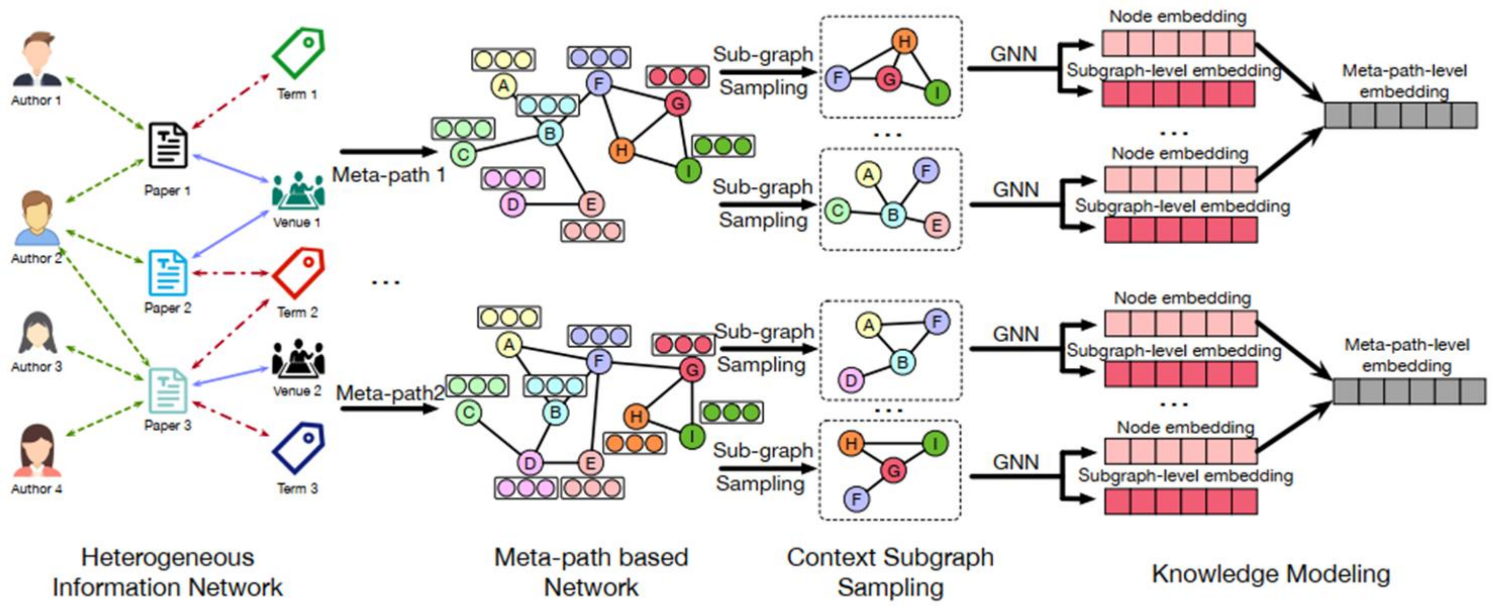
N is the number of target type nodes

$\mathbf{A}^m \in \mathbb{R}^{N \times N}$ is the adjacent matrix

sample the top-K important neighbors

$$G_i^m = \text{top\_rank}(\mathbf{S}^m(i,:), K) \quad (2)$$

Heterogeneous Knowledge Modeling

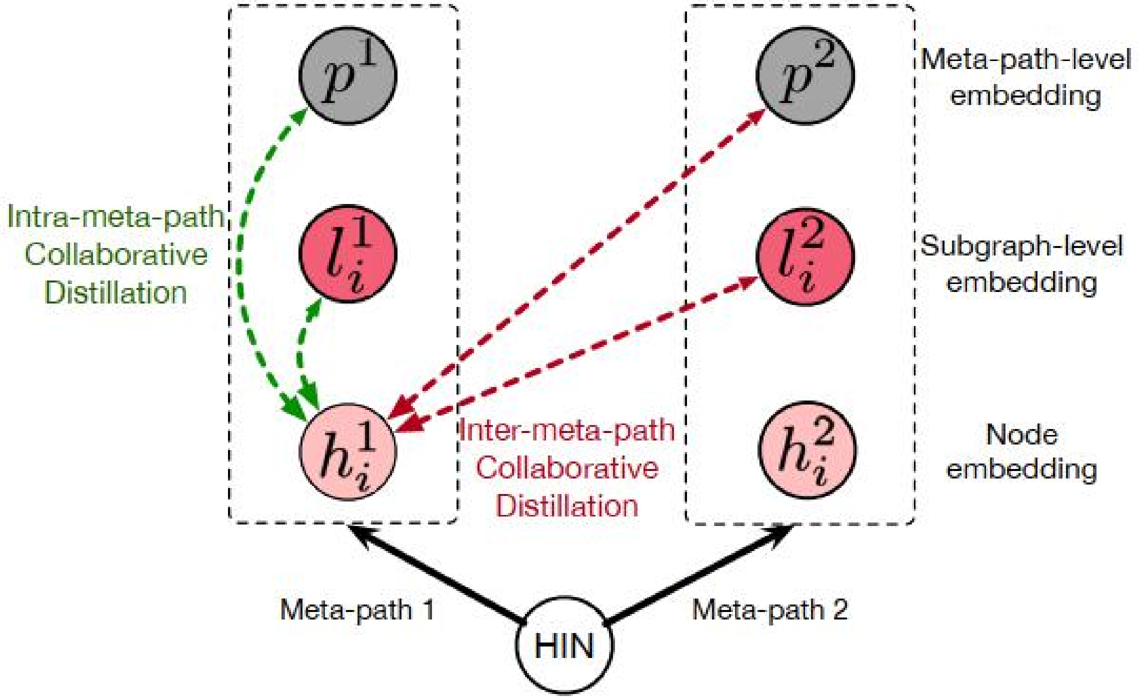


$$\mathbf{H}^m = \left(\widetilde{\mathbf{D}}_m^{-\frac{1}{2}} \tilde{\mathbf{A}}^m \widetilde{\mathbf{D}}_m^{-\frac{1}{2}}\right) \mathbf{X}^m \mathbf{W}^m \quad (3) \qquad \tilde{\mathbf{A}}^m = \mathbf{A}^m + \mathbf{I} \quad \mathbf{H}^m \in \mathcal{R}^{N\times d}$$

$$l_i^m = \mathcal{R}_l\left(G_i^m\right) = \sigma\left(\frac{1}{K}\sum_{j=1}^{K} h_j^m\right) \quad (4) \qquad \text{local readout function } \mathcal{R}_l : \mathbb{R}^{(K)\times d} \to \mathbb{R}^d$$

$$p^m = \mathcal{R}_g\left(H^m\right) = \sigma\left(\frac{1}{N}\sum_{i=1}^{N} h_i^m\right) \quad (5) \qquad \text{global readout function } \mathcal{R}_g : \mathbb{R}^{N\times d} \to \mathbb{R}^d$$

Collaborative Knowledge Distillation



Intra-meta-path Collaborative Distillation.

$$\mathcal{L}_{intra} = -\sum_{m\in\mathcal{M}}\left(\sum_{i}^{|N|}\left(\mathrm{MI}(h_i^m, l_i^m) + \mathrm{MI}(h_i^m, p^m)\right)\right) \quad (6)$$

Inter-meta-path Collaborative Distillation.

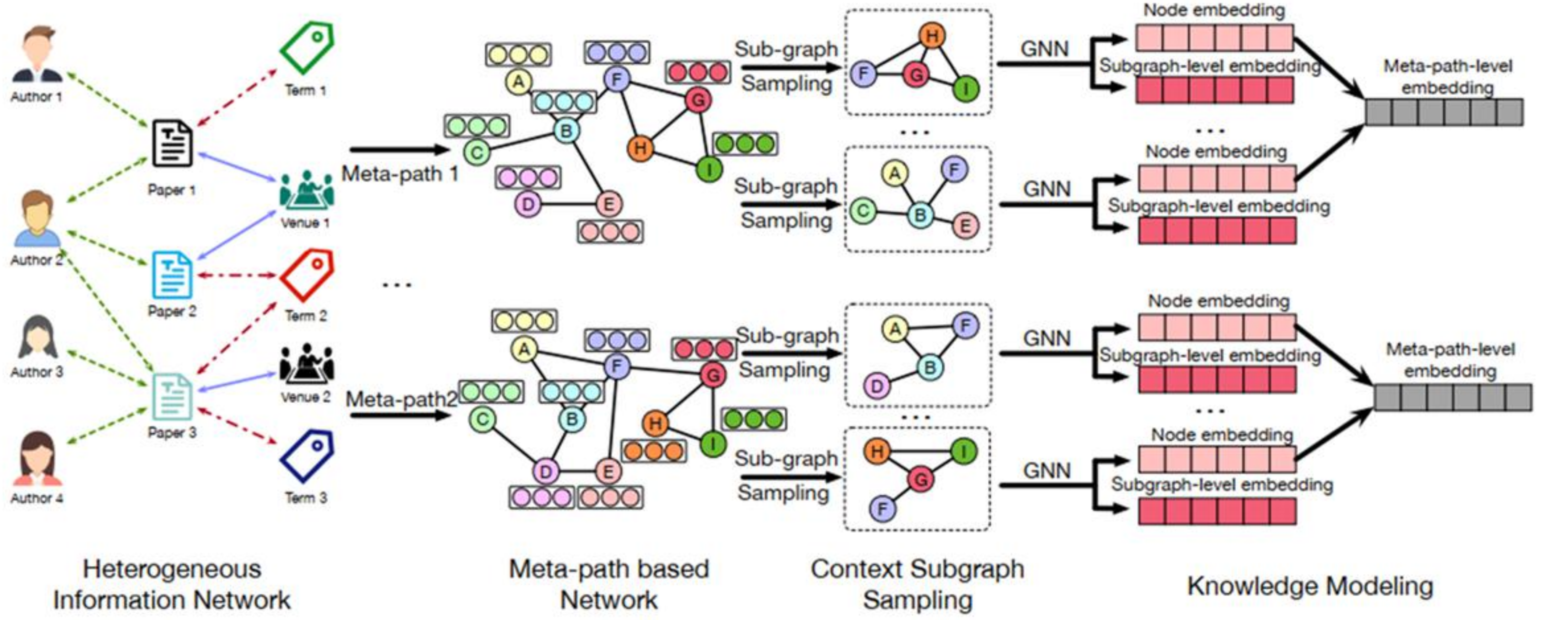
$$\mathcal{L}_{inter} = -\sum_{i}^{|N|}\left(\sum_{m\in\mathcal{M}}\sum_{n\in\mathcal{M},n\neq m}\mathrm{MI}(h_i^m, l_i^n) + \mathrm{MI}(h_i^m, p^n)\right) \quad (7)$$

Mutual Information Estimation.

$$\mathrm{MI}(X,Y) = \mathbb{E}_{\mathbb{P}}[-sp(-f(x,y))] - \mathbb{E}_{\mathbb{P}\times\tilde{\mathbb{P}}}[sp(f(x,\tilde{y}))] \quad (8)$$

$$sp(x) = log(1+e^x)$$

Model Training



$$\mathcal{L} = \mathcal{L}_{intra} + \mathcal{L}_{inter} \quad (9)$$

$$h_i = \sum_{m \in \mathcal{M}} h_i^m \quad (10)$$

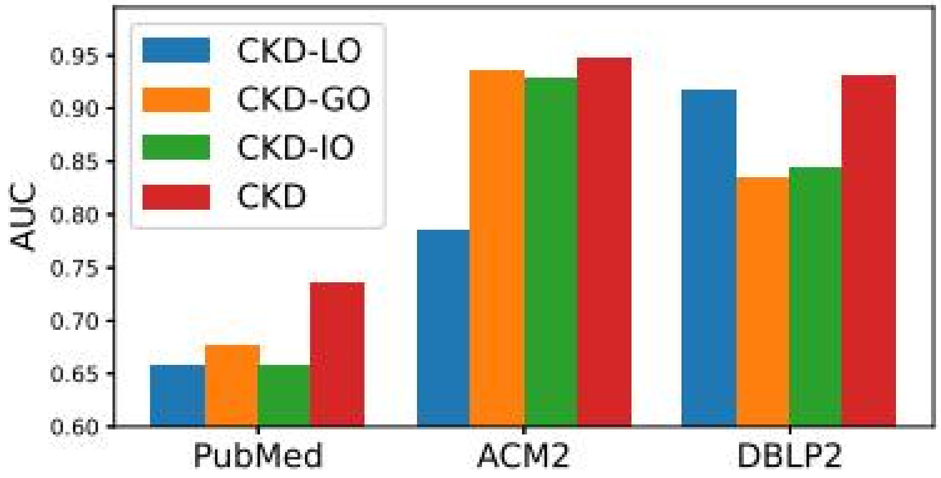
Experiments

Dataset	Nodes	Edges	Features	Labels
ACM	10,942	547,872	100	3
ACM2	29,930	61,770	100	7
DBLP	26,128	239,566	200	4
DBLP2	173,988	20,743,972	300	4
Pubmed	63,109	125,167	200	8
Freebase	79,843	498,508	300	7

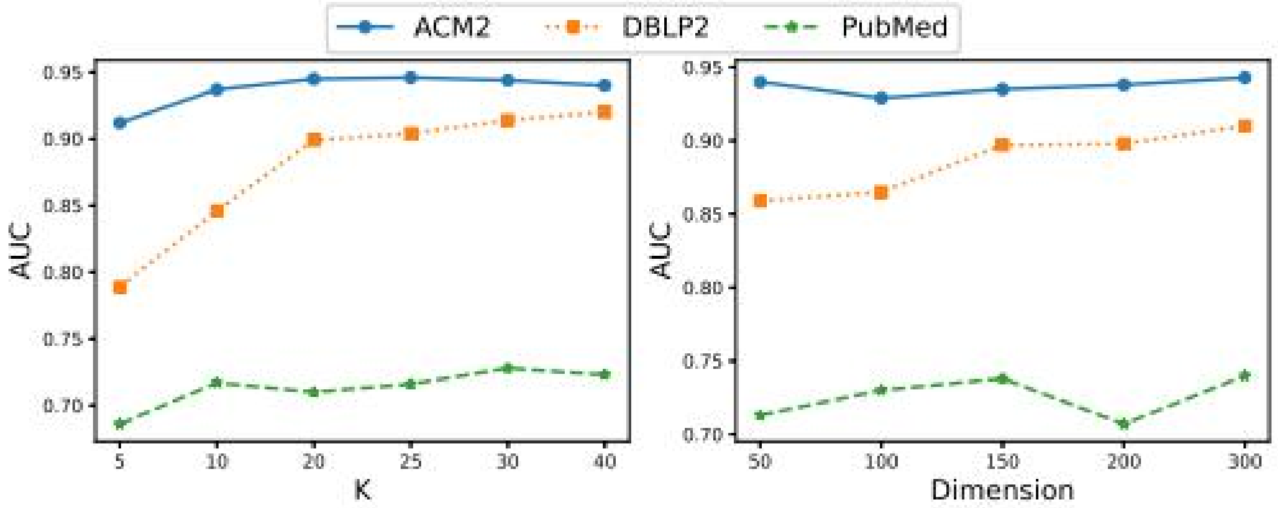
Dataset	ACM		DBLP		ACM2		PubMed		Freebase		DBLP2	
Metric	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
DeepWalk	89.6±0.0	89.6±0.0	91.3±0.0	91.7±0.0	64.8±0.0	75.9±0.0	15.1±0.0	16.8±0.0	**48.9**±0.0	**60.6**±0.0	88.4±0.0	88.3±0.0
	88.8±0.0	88.8±0.0	90.6±0.0	91.0±0.0	64.8±0.0	75.9±0.0	14.7±0.0	16.5±0.0	48.1±0.0	60.1±0.0	88.4±0.0	88.2±0.0
	89.8±0.0	89.8±0.0	90.8±0.0	91.2±0.0	64.6±0.0	76.0±0.0	12.9±0.0	15.7±0.0	**49.3**±0.0	**60.8**±0.0	88.3±0.0	88.1±0.0
Metapath2Vec	91.3±0.3	91.4±0.3	86.3±1.0	87.0±0.9	38.3±1.2	59.0±1.4	13.7±1.2	15.5±1.0	42.2±0.4	54.7±0.2	87.8±0.3	87.6±0.3
	91.7±0.6	91.8±0.6	87.7±1.0	88.3±0.9	38.9±1.1	59.1±1.3	12.4±1.4	14.5±1.2	41.5±1.0	54.6±0.3	88.0±0.2	87.8±0.3
	92.0±0.5	92.1±0.5	89.2±0.5	89.4±0.8	38.8±1.1	59.3±1.5	13.2±1.1	15.2±1.1	41.6±0.3	54.9±0.3	87.9±0.2	87.8±0.1
HIN2Vec	88.5±1.2	88.4±1.3	92.2±0.2	92.5±0.3	23.4±0.6	53.9±0.3	14.8±0.7	18.4±0.5	26.4±0.7	49.3±0.9	86.9±0.4	86.7±0.5
	89.6±1.8	89.8±1.7	91.9±0.2	92.4±0.2	23.7±0.5	54.9±0.6	14.2±0.5	17.8±0.3	25.9±0.4	49.5±0.7	86.6±0.4	86.8±0.3
	89.8±1.6	89.7±1.8	92.5±0.3	93.0±0.2	26.8±0.7	57.4±0.5	14.5±0.8	17.6±0.6	26.0±0.5	49.5±0.8	87.5±0.2	87.3±0.3
HAN	90.4±1.2	90.5±1.2	88.0±0.5	88.5±0.5	59.2±0.9	74.5±0.6	35.1±0.5	37.5±0.3	46.5±0.5	60.1±0.6	88.1±0.6	88.1±0.6
	90.7±1.4	90.8±1.3	87.6±0.7	88.1±0.4	58.7±1.1	74.0±0.8	34.3±0.7	37.1±0.5	46.6±1.1	60.9±0.6	87.5±1.3	87.4±1.4
	90.5±1.0	90.5±1.0	88.4±0.8	88.9±0.8	59.1±0.8	74.5±0.6	35.0±0.8	38.5±0.6	46.7±0.8	60.9±0.4	88.2±0.7	88.2±0.7
HDGI	68.8±2.4	68.9±2.1	74.4±1.0	75.9±1.0	31.5±1.2	57.1±1.2	14.9±0.8	20.3±0.6	-	-	86.8±0.8	87.0±0.8
	68.8±2.2	68.4±2.1	74.5±1.2	75.9±1.1	31.7±1.3	57.2±1.2	15.2±0.7	20.5±0.5	-	-	87.0±0.9	87.2±0.8
	69.8±2.7	69.5±2.8	74.5±1.3	76.0±1.4	31.8±1.4	57.4±1.2	15.4±0.6	20.7±0.4	-	-	87.1±0.9	87.2±0.8
HGT	89.1±0.4	89.3±0.3	50.8±1.0	50.7±1.2	60.9±1.0	75.4±1.2	19.0±0.5	19.9±0.8	-	-	84.1±0.6	84.3±0.6
	89.1±0.5	89.3±0.4	50.9±1.2	51.0±1.1	61.1±1.1	75.7±1.3	20.6±1.9	22.0±1.3	-	-	84.2±0.6	84.4±0.7
	89.2±0.7	89.3±0.7	52.7±0.7	52.8±0.6	61.3±1.2	75.8±1.3	19.4±2.5	20.7±3.7	-	-	84.3±0.9	89.2±0.9
NSHE	90.3±0.3	90.4±0.2	**93.9**±0.1	**94.1**±0.2	62.4±0.6	75.9±0.2	17.1±0.7	22.3±0.9	-	-	-	-
	90.5±0.2	90.6±0.2	**93.8**±0.3	**94.0**±0.3	62.4±0.7	75.9±0.2	17.5±0.8	22.7±0.6	-	-	-	-
	89.7±0.3	89.8±0.3	**93.9**±0.2	**94.1**±0.2	62.5±0.8	76.1±0.2	17.7±0.8	22.9±1.1	-	-	-	-
MAGNN	85.7±0.2	85.7±0.2	87.9±0.3	88.3±0.4	51.0±0.8	70.8±0.4	34.1±1.2	38.3±0.9	47.1±0.6	60.1±0.3	-	-
	87.3±0.4	87.3±0.4	87.5±0.5	88.3±0.2	52.1±0.7	67.8±1.1	36.3±0.6	38.9±0.7	47.6±0.3	60.0±0.5	-	-
	87.9±0.4	88.0±0.4	88.2±0.8	88.9±0.5	53.8±0.6	70.8±0.7	**39.4**±0.7	**42.1**±0.8	47.4±0.7	60.4±0.4	-	-
HeCo	71.0±0.2	71.2±0.1	91.5±0.5	91.8±0.6	57.2±0.8	72.9.0±0.5	16.5±0.5	26.1±1.2	-	-	-	-
	71.2±0.4	71.3±0.3	91.2±0.5	91.4±0.6	56.7±0.9	73.0±0.3	16.8±0.6	25.7±1.1	-	-	-	-
	71.3±0.1	71.3±0.1	91.2±0.4	91.5±0.5	57.5±1.1	72.9±0.7	16.9±0.7	25.9±1.0	-	-	-	-
HetGNN	85.7±0.1	85.6±0.1	92.0±0.6	92.3±0.7	-	-	-	-	-	-	-	-
	86.1±0.1	86.1±0.1	92.3±0.5	92.6±0.5	-	-	-	-	-	-	-	-
	86.6±0.2	86.7±0.2	92.8±0.6	93.1±0.5	-	-	-	-	-	-	-	-
CKD	**91.9**±0.4	**91.9**±0.4	92.5±0.2	92.8±0.2	**69.7**±0.5	**79.7**±0.8	**36.8**±1.1	**39.3**±1.6	48.2±0.7	60.5±0.4	**90.2**±0.3	**90.1**±0.3
	92.9±0.3	**92.9**±0.3	92.5±0.4	92.8±0.4	**65.6**±0.3	**77.9**±0.1	**37.4**±0.9	**40.1**±0.6	**49.6**±0.4	**61.1**±0.7	**90.4**±0.3	**90.3**±0.3
	92.8±0.8	**92.7**±1.0	92.3±0.4	92.6±0.4	**70.4**±0.5	**80.2**±0.6	37.8±1.2	40.4±1.2	48.1±0.8	60.4±0.5	**90.2**±0.2	**90.1**±0.1

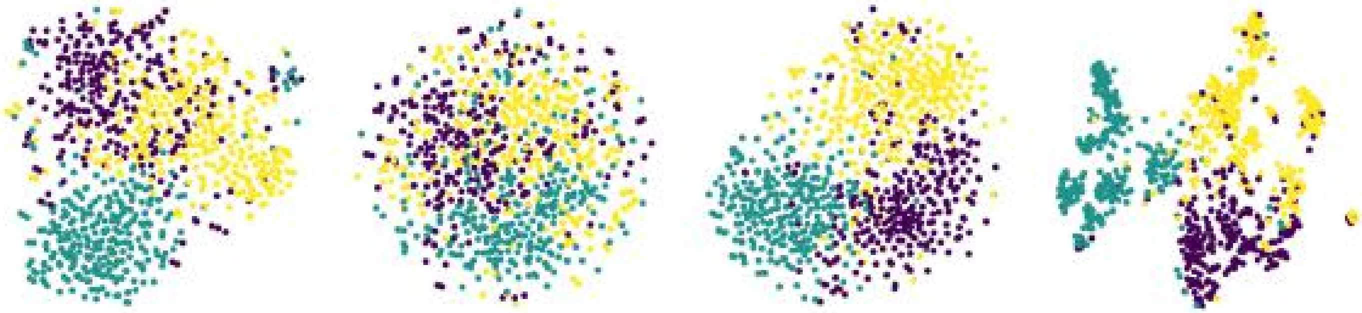
Experiments

Method \ Data	ACM2	DBLP2	PubMed
DeepWalk	0.818	0.789	0.663
Metapath2Vec	0.712	0.915	0.628
HIN2Vec	0.736	0.803	0.649
HAN	0.868	0.711	0.717
HDGI	0.537	0.691	0.594
HGT	0.920	0.868	0.736
NSHE	0.939	-	0.654
MAGNN	0.696	-	0.514
HeCo	0.681	-	0.519
CKD	**0.948**	**0.931**	**0.735**



Experiments





(a) HetGNN (b) MAGNN (c) HIN2Vec (d) CKD