北京郵電大學
Beijing University of Posts and Telecommunications

UIC
UNIVERSITY OF ILLINOIS AT CHICAGO

Recent Developments of Deep Heterogeneous Information Network Analysis --Part III Heterogeneous Information Network Embedding

Chuan Shi

shichuan@bupt.edu.cn

Beijing University of Posts and Telecommunications

Philip S. Yu

psyu@uic.edu

University of Illinois at Chicago

Outline

- Metapath based data mining
- ✓ **Heterogeneous information network embedding**
  - Shallow models
    - MetaPath2Vec (KDD2017), HIN2Vec (CIKM2017)
    - MCRec (KDD2018), HERec (TKDE2018), RHINE (AAAI2018)
    - HHNE (AAAI2018), HeGAN (KDD2019)
  - Deep models
    - NeuACF(IJCAI2018), HetGNN (KDD2019), HAN(WWW2019)
- Applications
- Conclusion and future work

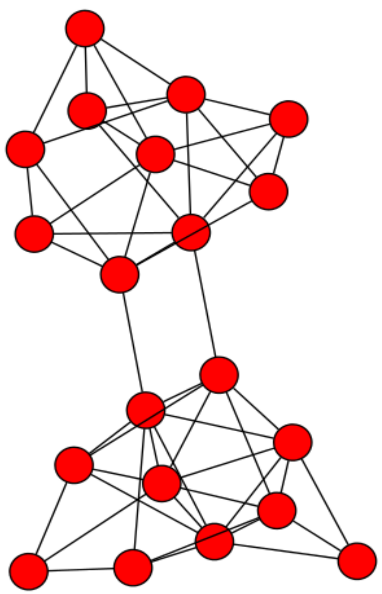
Network Embedding

Network Embedding

Embed each node of a network into a low-dimensional vector space

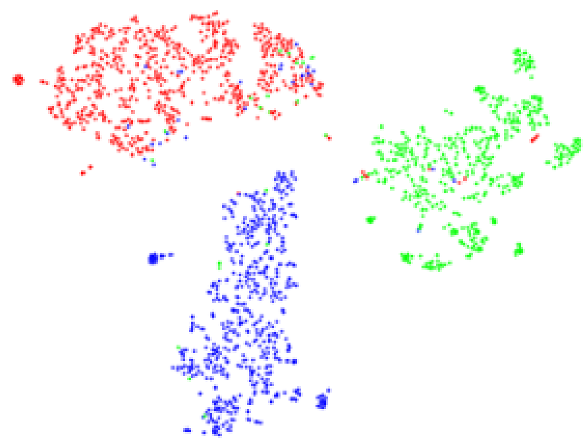
Application

- node classification
- link predication
- community detection
- network evolution
- …



generate

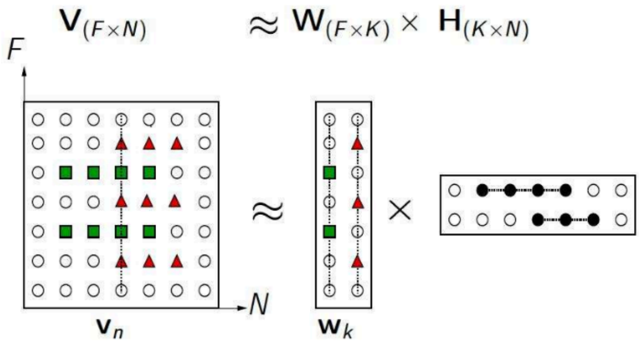
embed

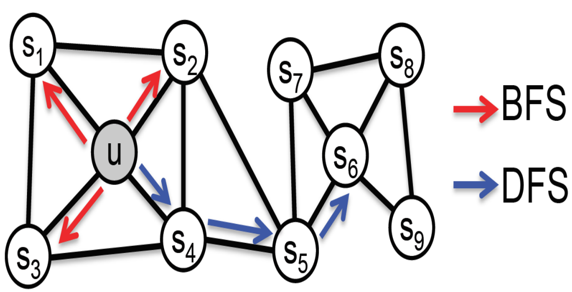


- ☐ Easy to parallel
- ☐ Can apply classical ML methods

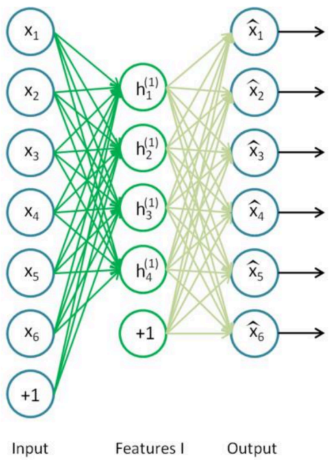
Shallow model

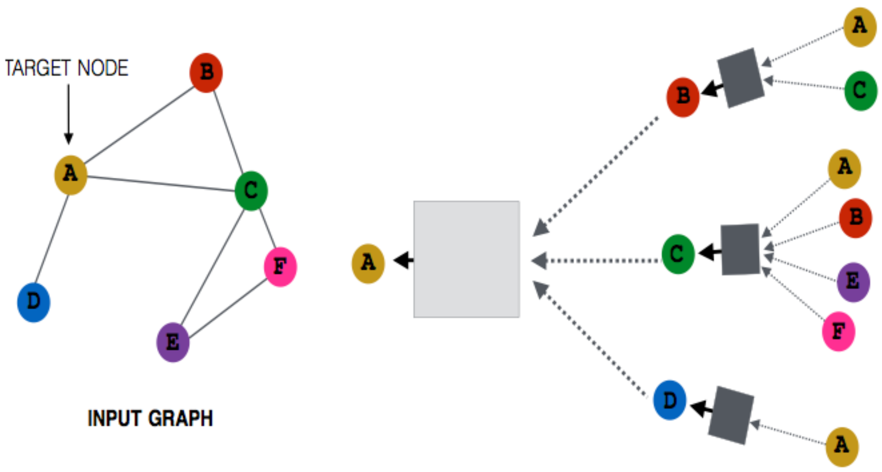
- Factorization-based approaches
  - e.g., Laplacian eigenmaps
- Random walk approaches
  - e.g., DeepWalk, node2vec





- **Deep model**
  - Apply deep neural network for graph
  - Autoencoder approaches: e.g., DNGR and SDNE
  - GNN based approaches
    - Average neighbor information and apply a neural network
    - e.g., GCN, GraphSage, GAT





5

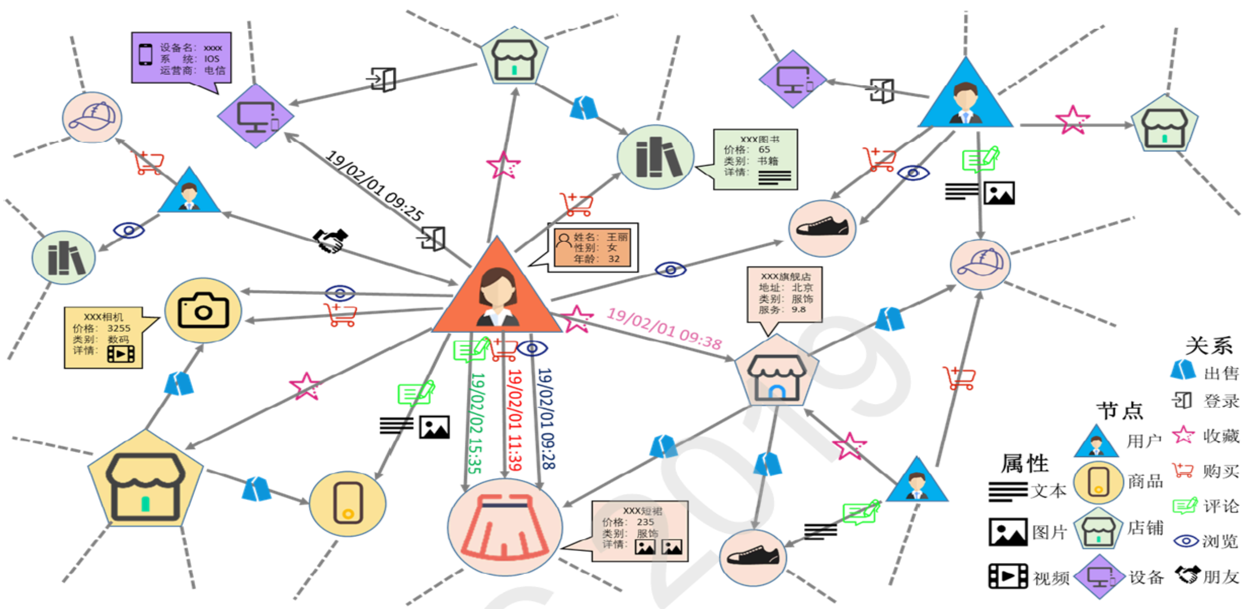
HIN Embedding

Why HIN embedding

- Heterogeneity is ubiquitous
- Information loss
- Rich semantics

Challenges

- How to handle heterogeneity
- How to fuse information
- How to capture rich semantics



Static, Topology

Homogeneous Graph

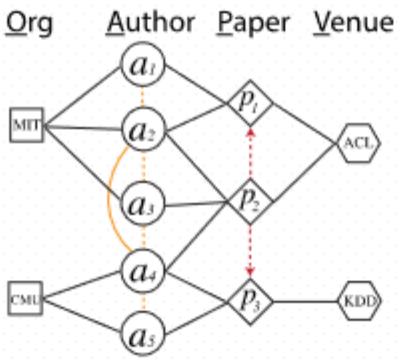
Dynamic, Attributed

Heterogeneous Graph

Outline

- Metapath based data mining
- ✓ **Heterogeneous information network embedding**
  - ✓ **Shallow models**
    - MetaPath2Vec (KDD2017), HIN2Vec (CIKM2017)
    - HERec (TKDE2018), MCRec (KDD2018), RHINE (AAAI2018)
    - HHNE (AAAI2018), HeGAN (KDD2019)
  - Deep models
    - NeuACF(IJCAI2018), HetGNN (KDD2019), HAN(WWW2019)
- Applications
- Conclusion and future work

Basic idea of Metapath2vec



metapath2vec

metapath2vec++

X

latent representation vector

Network Applications

- node attribute inference
- community detection
- similarity search
- link prediction
- social recommendation
- …

metapath2vec

metapath2vec++

meta-path-based random walks

+

skip-gram

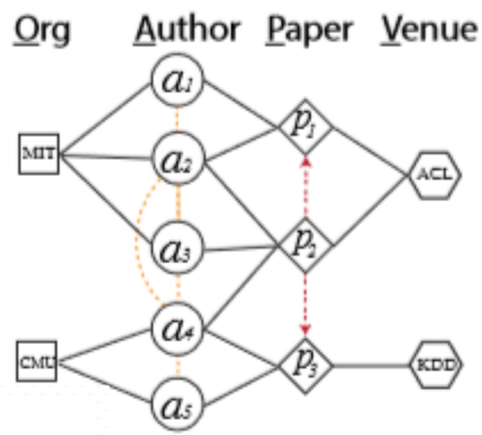
heterogeneous skip-gram

YuXiao Dong, Nitesh V. Chawla, Ananthram Swami. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. KDD 2017.

Meta-Path-Based Random Walks

- Meta-Path-Based Random Walks
  - Given a meta-path

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



  - The transition probability at step i is defined as

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

  - Recursive guidance for a symmetric path, i.e.,

$$p(v^{i+1}|v_t^i) = p(v^{i+1}|v_1^i), \text{ if } t = l$$

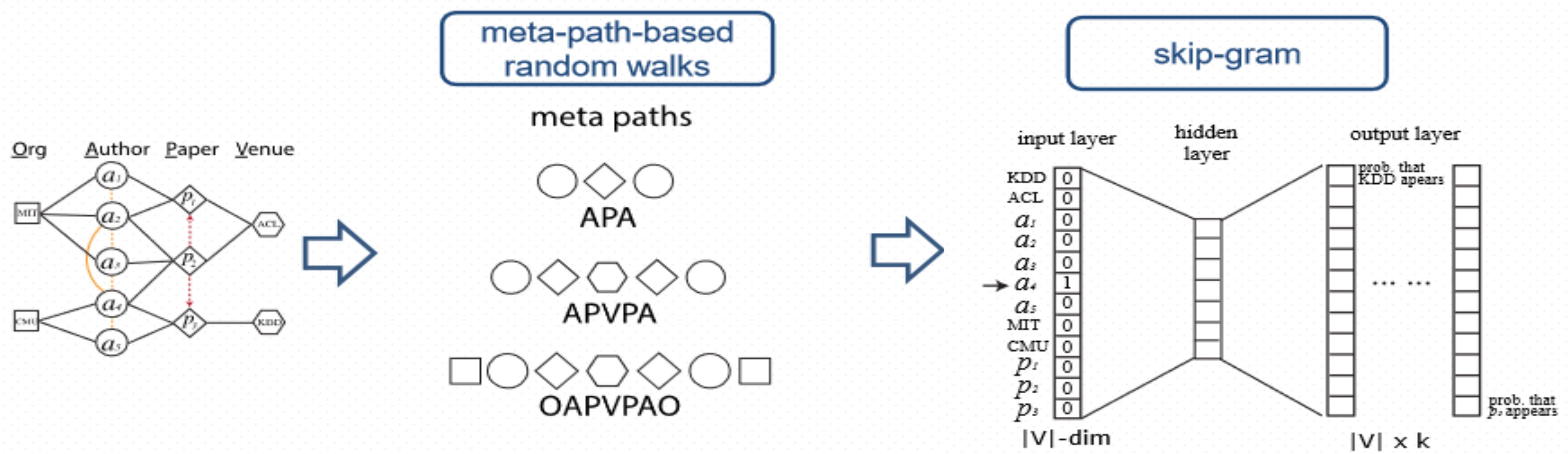
Skip-gram of metapath2vec

- Skip-gram's potential issue

$$\arg\max_{\theta} \prod_{v \in V} \prod_{c \in N(v)} p(c|v;\theta) \Rightarrow \arg\max_{\theta} \sum_{v \in V} \sum_{t \in T_V} \sum_{c_t \in N_t(v)} \log p(c_t|v;\theta)$$

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

$$\log \sigma(X_{c_t} \cdot X_v) + \sum_{m=1}^{M} \mathbb{E}_{u^m \sim P(u)} [\log \sigma(-X_{u^m} \cdot X_v)],$$



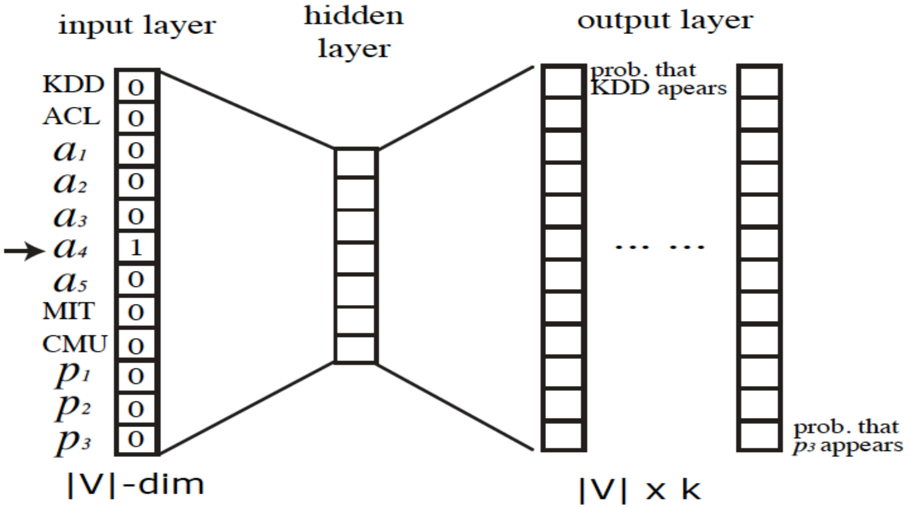
metapath2vec vs metapath2vec++

- Metapath2vec

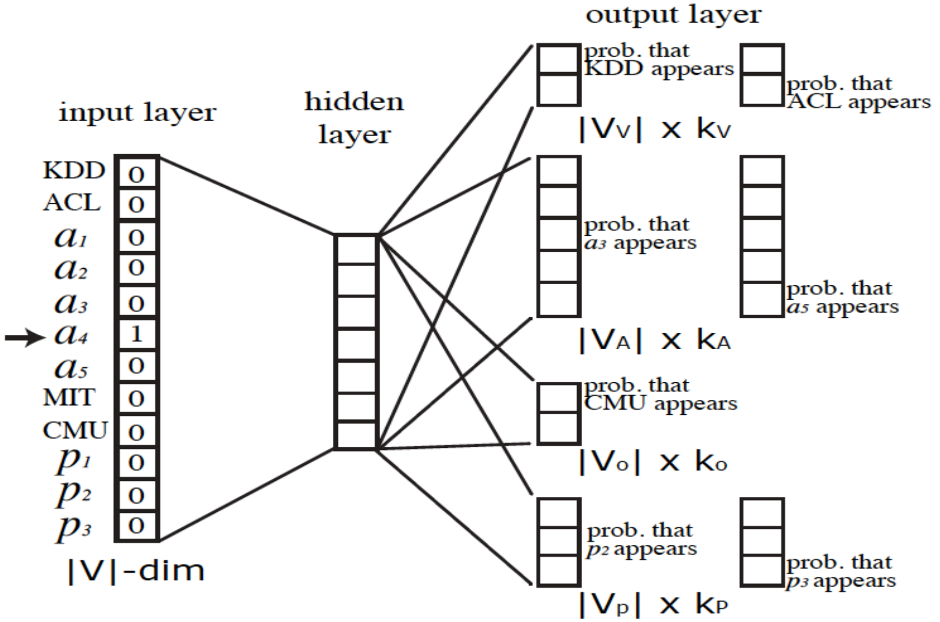
$$p(c_t|v;\theta) = \frac{e^{X_{c_t}} \cdot e^{X_v}}{\sum_{u \in V} e^{X_u} \cdot e^{X_v}}$$

metapath2vec++

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



(b) Skip-gram in *metapath2vec*, node2vec, & DeepWalk



(c) Skip-gram in *metapath2vec++*

Experimental Setup

- Heterogeneous Data

  AMiner Academic Network

  ✓ 9 1.7 million authors

  ✓ 3800+ venues

  ✓ 8 research areas

  ✓ 3 million papers

- Baselines

  ✓ DeepWalk

  ✓ node2vec

  ✓ LINE

  ✓ PTE

Effectiveness Experiments

Application 1: Multi-Class Node Classification

AMiner Academic Network

- ✓ 9 1.7 million authors
- ✓ 3800+ venues
- ✓ 8 research areas
- ✓ 3 million papers

- Baselines
  - ✓ DeepWalk
  - ✓ node2vec
  - ✓ LINE
  - ✓ PTE

Table 2: Multi-class venue node classification results in AMiner data.

Metric	Method	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Macro-F1	DeepWalk/node2vec	0.0723	0.1396	0.1905	0.2795	0.3427	0.3911	0.4424	0.4774	0.4955	0.4457
	LINE (1st+2nd)	0.2245	0.4629	0.7011	0.8473	0.8953	0.9203	0.9308	0.9466	0.9410	0.9466
	PTE	0.1702	0.3388	0.6535	0.8304	0.8936	0.9210	0.9352	0.9505	0.9525	0.9489
	metapath2vec	0.3033	0.5247	0.8033	0.8971	0.9406	0.9532	0.9529	0.9701	0.9683	0.9670
	metapath2vec++	0.3090	0.5444	0.8049	0.8995	0.9468	0.9580	0.9561	0.9675	0.9533	0.9503
Micro-F1	DeepWalk/node2vec	0.1701	0.2142	0.2486	0.3266	0.3788	0.4090	0.4630	0.4975	0.5259	0.5286
	LINE (1st+2nd)	0.3000	0.5167	0.7159	0.8457	0.8950	0.9209	0.9333	0.9500	0.9556	0.9571
	PTE	0.2512	0.4267	0.6879	0.8372	0.8950	0.9239	0.9352	0.9550	0.9667	0.9571
	metapath2vec	0.4173	0.5975	0.8327	0.9011	0.9400	0.9522	0.9537	0.9725	0.9815	0.9857
	metapath2vec++	0.4331	0.6192	0.8336	0.9032	0.9463	0.9582	0.9574	0.9700	0.9741	0.9786

Table 3: Multi-class author node classification results in AMiner data.

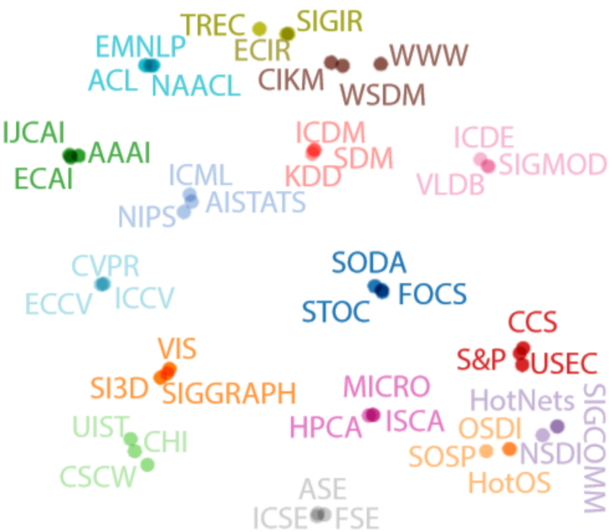
Metric	Method	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Macro-F1	DeepWalk/node2vec	0.7153	0.7222	0.7256	0.7270	0.7273	0.7274	0.7273	0.7271	0.7275	0.7275
	LINE (1st+2nd)	0.8849	0.8886	0.8911	0.8921	0.8926	0.8929	0.8934	0.8936	0.8938	0.8934
	PTE	0.8898	0.8940	0.897	0.8982	0.8987	0.8990	0.8997	0.8999	0.9002	0.9005
	metapath2vec	0.9216	0.9262	0.9292	0.9303	0.9309	0.9314	0.9315	0.9316	0.9319	0.9320
	metapath2vec++	0.9107	0.9156	0.9186	0.9199	0.9204	0.9207	0.9207	0.9208	0.9211	0.9212
Micro-F1	DeepWalk/node2vec	0.7312	0.7372	0.7402	0.7414	0.7418	0.7420	0.7419	0.7420	0.7425	0.7425
	LINE (1st+2nd)	0.8936	0.8969	0.8993	0.9002	0.9007	0.9010	0.9015	0.9016	0.9018	0.9017
	PTE	0.8986	0.9023	0.9051	0.9061	0.9066	0.9068	0.9075	0.9077	0.9079	0.9082
	metapath2vec	0.9279	0.9319	0.9346	0.9356	0.9361	0.9365	0.9365	0.9365	0.9367	0.9369
	metapath2vec++	0.9173	0.9217	0.9243	0.9254	0.9259	0.9261	0.9261	0.9262	0.9264	0.9266

Effectiveness Experiments

Application 2: Node Clustering

Node clustering results (NMI) in AMiner data.

methods	venue	author
DeepWalk/node2vec	0.1952	0.2941
LINE (1st+2nd)	0.8967	0.6423
PTE	0.9060	0.6483
metapath2vec	0.9274	0.7470
metapath2vec++	0.9261	0.7354



Application 3: Similarity Search

Table 1: Case study of similarity search in the heterogeneous DBIS data used in [26].

Method	PathSim [26]		DeepWalk / node2vec [8, 22]		LINE (1st+2nd) [30]		PTE [29]		*metapath2vec*		*metapath2vec++*	
Input	meta-paths		heterogeneous random walk paths		heterogeneous edges		heterogeneous edges		probabilistic meta-paths		probabilistic meta-paths	
Query	PKDD	C. Faloutsos	PKDD	C. Faloutsos	PKDD	C. Faloutsos	PKDD	C. Faloutsos	PKDD	C. Faloutsos	PKDD	C. Faloutsos
1	ICDM	J. Han	R. S.	J. Pan	W. K.	C. Aggarwal	KDD	C. Aggarwal	A. S.	C. Aggarwal	KDD	R. Agrawal
2	SDM	R. Agrawal	M. N.	H. Tong	S. A.	P. Yu	ICDM	P. Yu	M. B.	J. Pei	PAKDD	J. Han
3	PAKDD	J. Pei	R. P.	H. Yang	A. B.	D. Gunopulos	SDM	Y. Tao	P. B.	P. Yu	ICDM	J. Pei
4	KDD	C. Aggarwal	G. G.	R. Filho	M. S.	N. Koudas	DMKD	N. Koudas	M. S.	H. Cheng	DMKD	C. Aggarwal
5	DMKD	H. Jagadish	F. J.	R. Chan	S. A.	M. Vlachos	PAKDD	R. Rastogi	M. K.	V. Ganti	SDM	P. Yu

Framework of HIN2Vec

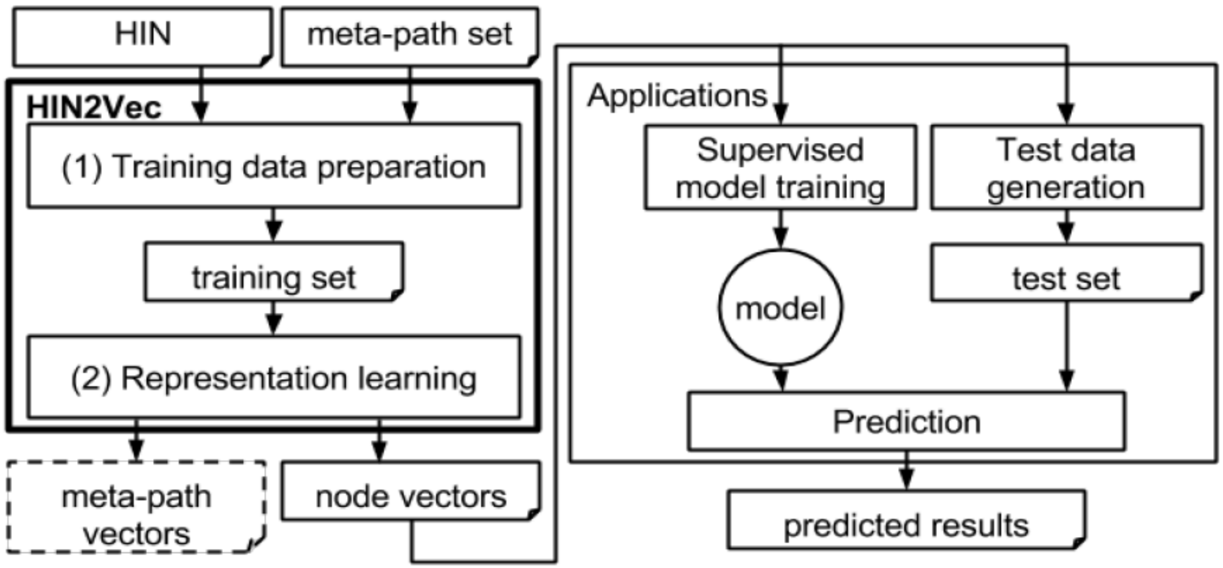


Figure 1: Overview of the HIN2Vec framework

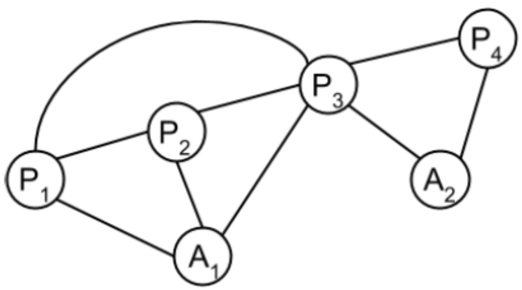


Figure 3: A paper-author HIN

$\langle x, y, r, L(x,y,r)\rangle$ is extracted from the HIN

P_1 as $\langle P_1, P_2, \text{P-P}\rangle$ and $\langle P_1, A_1, \text{P-P-A}\rangle$,

- Classification problem: whether two nodes, x and y, have a specific relationship r.

Tao-yang Fu, Wang-Chien Lee, Zhen Lei. HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning. CIKM 2017.

HIN2Vec

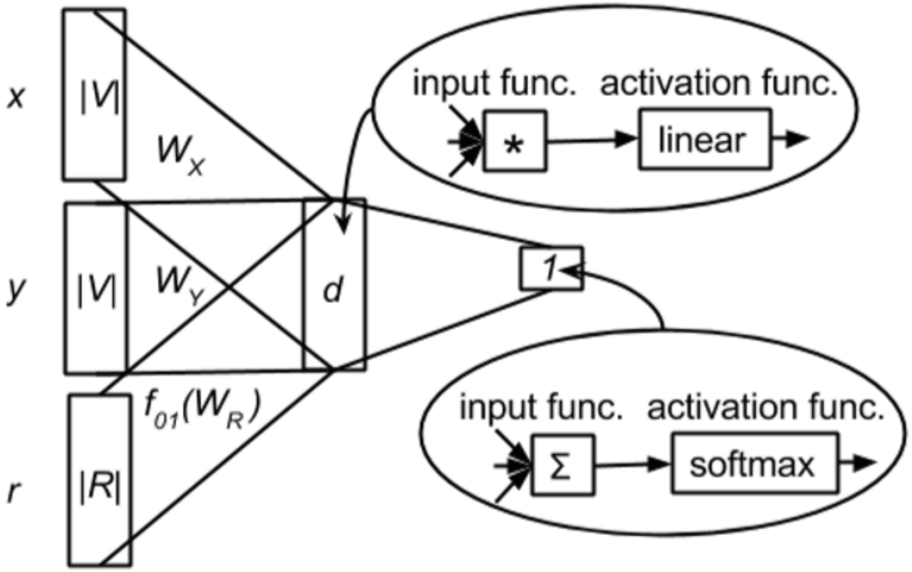


Figure 4: The HIN2Vec NN model

Representation learning

$$O_{x,y,r}(x,y,r) = \begin{cases} P(r|x,y), & \text{if } L(x,y,r) = 1 \\ 1 - P(r|x,y), & \text{if } L(x,y,r) = 0 \end{cases}$$

$$\begin{aligned} \log O_{x,y,r}(x,y,r) =& L(x,y,r) \log P(r|x,y) \\ &+ [1 - L(x,y,r)] \log[1 - P(r|x,y)] \end{aligned}$$

$$P(r|x,y) = sigmoid\left(\sum W'_X \vec{x} \odot W'_Y \vec{y} \odot f_{01}(W'_R \vec{r})\right)$$

Training data

Random walks are used to sample positive data

Generate negative data through randomly replacing some nodes

Experimental Setup

- Dataset

Table 1: Statistics of Datasets

Blogcatalog	User	Group		
	10312	39		
Yelp	User	Business	City	Category
	630639	86810	10	807
DBLP	Paper	Author	Venue	
	53464	54949	20	
U.S. Patents	Patent	Inventor	Assignee	Class
	295145	293848	31805	14

- Tasks
  - Multi-label Classification of Nodes
  - Link Prediction

- Baselines
  - Deepwalk
  - LINE
  - Node2vec
  - PTE
  - HINE
  - ESIM

Effectiveness Experiments

Table 2: Performance Evaluation of Node Classification

	Blogcatalog		Yelp		DBLP		U.S. Patents	
	micro-f1	macro-f1	micro-f1	macro-f1	micro-f1	macro-f1	micro-f1	macro-f1
DeepWalk	0.244	0.140	0.276	0.165	0.481	0.463	0.675	0.676
LINE	0.239	0.128	0.270	0.163	0.449	0.429	0.66	0.663
node2vec	0.246	0.141	0.276	0.166	0.491	0.470	0.676	0.677
PTE	0.179	0.096	0.222	0.130	0.417	0.394	0.547	0.555
HINE	*0.250	*0.144	*0.278	*0.169	0.475	0.461	*0.681	*0.685
ESim	0.207	0.102	0.229	0.132	*0.514	*0.496	0.610	0.562
HIN2Vec	**0.272(9.9%)**	**0.158(11.3%)**	**0.302(7.9%)**	**0.192(12.0%)**	**0.605(23.8%)**	**0.594(20.1%)**	**0.729(6.6%)**	**0.732(6.4%)**

Table 6: Performance Evaluation of Link Prediction

	Blogcatalog		Yelp		DBLP		U.S. Patents	
	MAP	recall@100	MAP	recall@100	MAP	recall@100	MAP	recall@100
DeepWalk	0.124	0.227	*0.021	0.110	0.230	*0.710	0.093	0.500
LINE	*0.134	*0.249	0.017	0.104	0.086	0.580	0.091	0.400
node2vec	0.125	0.229	*0.021	*0.111	*0.231	*0.710	0.095	*0.503
PTE	0.067	0.139	0.004	0.034	0.071	0.324	0.030	0.243
HINE	0.085	0.179	0.016	0.097	0.205	0.697	*0.103	0.495
ESim	0.132	0.185	x	x	0.179	0.633	x	x
MPE	**0.141(5.0%)**	**0.279(10.8%)**	**0.028(31.8%)**	**0.138(24.3%)**	**0.265(12.8%)**	**0.751(5.8%)**	**0.176(70.8%)**	**0.602(19.9%)**

Effectiveness Experiments

Table 3: Clusters of Meta-paths

Cluster	Yelp	U.S. Patents
1	U-U, U-U-U, U-U-U-U-U	P→P←P, P←P→P
2	B-U-B, B-U-U-U-B	P←P, P←P←P
3	B-U-U, B-U-U-U-U	I-P→P, I-P←P
4	B-U	P-A
5	B-C, B-C-B, U-B-C	P-A-P, P-I-P, P-I
6	...	...

Framework of HERec

Heterogeneous information network
Embedding for **Rec**ommendation (HERec)

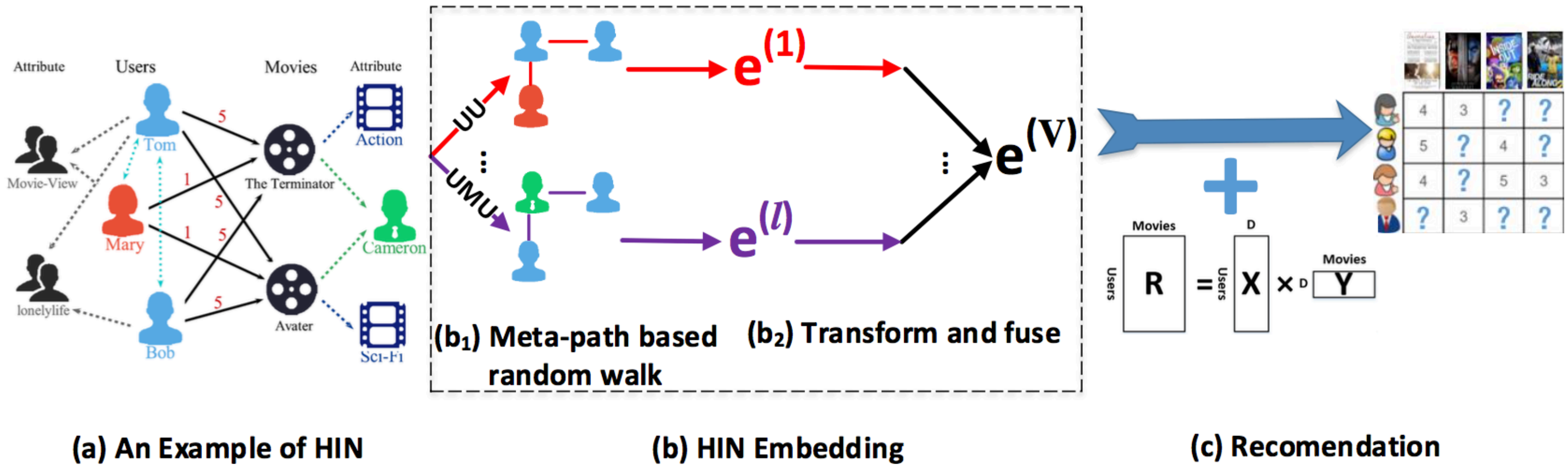


Figure 1: The schematic illustration of the proposed HERec approach.

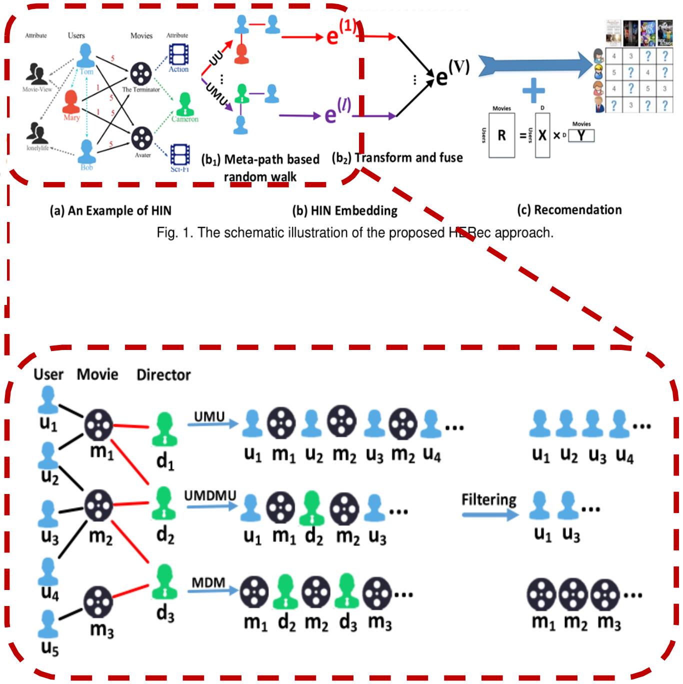
HERec
- **Heterogeneous information network embedding**
- **Fuse embeddings into MF for recommendation**

Chuan Shi, Binbin Hu, Wayne Xin Zhao, Philip S. Yu. Heterogeneous Information Network Embedding for Recommendation. TKDE 2018

HIN Embedding – Single Meta-path

Mete-path based Random Walk

$$P(n_{t+1}=x|n_t=v,\rho) = \begin{cases} \frac{1}{|\mathcal{N}^{A_{t+1}}(v)|}, & (v,x)\in\mathcal{E} \text{ and } \phi(x)=A_{t+1} \\ 0, & \text{otherwise,} \end{cases}$$



Type Constraints and Filtering

- Uncover heterogeneous information with homogeneous node embedding objective
- Utilize more neighbor information in a fixed length window

Embedding with Skip-Gram for Single Meta-path

$$\max_f \sum_{u\in\mathcal{V}} \log Pr(\mathcal{N}_u|f(u))$$

HIN Embedding – Fusion

Embedding Fusion

- Integrate information of various meta-paths
- A good fusion function should be learned according to the specific task

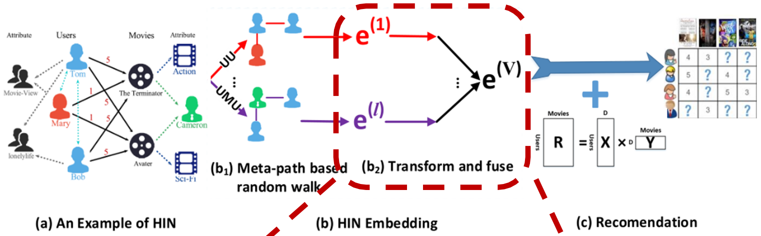


Fig. 1. The schematic illustration of the proposed HERec approach.

- Simple linear fusion.

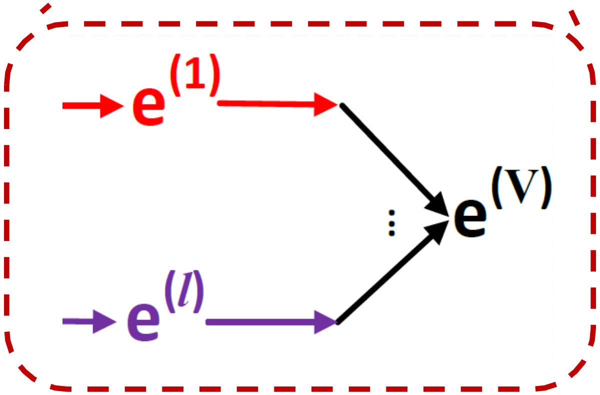
$$g(\{\boldsymbol{e}_u^{(l)}\}) = \frac{1}{|\mathcal{P}|}\sum_{l=1}^{|\mathcal{P}|}(\mathbf{M}^{(l)}\boldsymbol{e}_u^{(l)} + \boldsymbol{b}^{(l)})$$

- Personalized linear fusion.

$$g(\{\boldsymbol{e}_u^{(l)}\}) = \sum_{l=1}^{|\mathcal{P}|} w_u^{(l)}(\mathbf{M}^{(l)}\boldsymbol{e}_u^{(l)} + \boldsymbol{b}^{(l)})$$

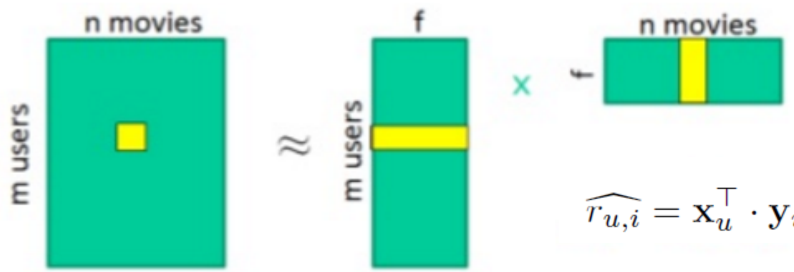
- Personalized non-linear fusion.

$$g(\{\boldsymbol{e}_u^{(l)}\}) = \sigma\left(\sum_{l=1}^{|\mathcal{P}|} w_u^{(l)}\sigma(\mathbf{M}^{(l)}\boldsymbol{e}_u^{(l)} + \boldsymbol{b}^{(l)})\right)$$



Extended MF

Basic Rating Prediction



$$\widehat{r_{u,i}} = \mathbf{x}_u^\top \cdot \mathbf{y}_i,$$

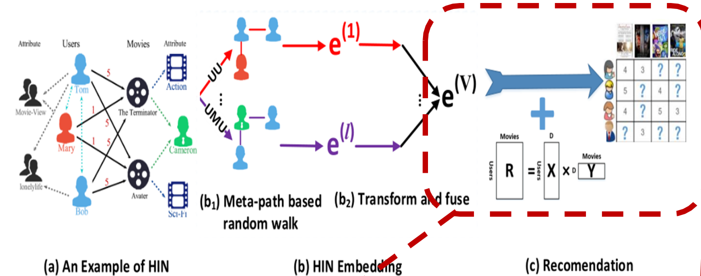
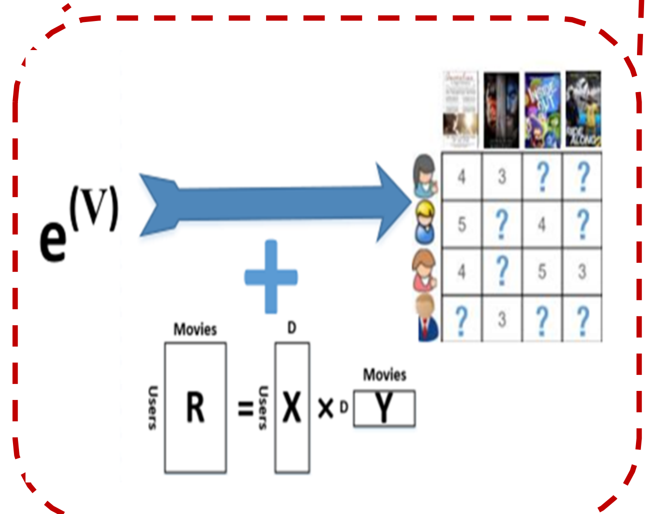


Fig. 1. The schematic illustration of the proposed HERec approach.



Extended Rating Prediction

- Integrated with HIN Embeddings

$$\widehat{r_{u,i}} = \mathbf{x}_u^\top \cdot \mathbf{y}_i + \alpha \cdot \boldsymbol{e}_u^{(U)\top} \cdot \boldsymbol{\gamma}_i^{(I)} + \beta \cdot \boldsymbol{\gamma}_u^{(U)\top} \cdot \boldsymbol{e}_i^{(I)},$$

Model Optimization Objective

$$\pounds = \sum_{\langle u,i,r_{u,i}\rangle \in \mathcal{R}} (r_{u,i} - \widehat{r_{u,i}})^2 + \lambda \sum_u (\|\mathbf{x}_u\|_2 + \|\mathbf{y}_i\|_2 + \|\boldsymbol{\gamma}_u^{(U)}\|_2 + \|\boldsymbol{\gamma}_i^{(I)}\|_2 + \|\boldsymbol{\Theta}^{(U)}\|_2 + \|\boldsymbol{\Theta}^{(I)}\|_2), \qquad (9)$$

Experimental Setup

Dataset

Dataset (Density)	Relations (A-B)	Number of A	Number of B	Number of (A-B)	Ave. degrees of A	Ave. degrees of B	Meta-paths
Douban Movie (0.63%)	User-Movie	13,367	12,677	1,068,278	79.9	84.3	UMU, MUM UMDMU, MDM UMAMU, MAM UMTMU, MTM
	User-User	2,440	2,294	4,085	1.7	1.8	
	User-Group	13,337	2,753	570,047	42.7	207.1	
	Movie-Director	10,179	2,449	11,276	1.1	4.6	
	Movie-Actor	11,718	6,311	33,587	2.9	5.3	
	Movie-Type	12,678	38	27,668	2.2	728.1	
Douban Book (0.27%)	User-Book	13,024	22,347	792,026	60.8	35.4	UBU, BUB UBPBU, BPB UBYBU, BYB UBABU
	User-User	12,748	12,748	169,150	13.3	13.3	
	Book-Author	21,907	10,805	21,905	1.0	2.0	
	Book-Publisher	21,773	1,815	21,773	1.0	11.9	
	Book-Year	21,192	64	21,192	1.0	331.1	
Yelp (0.08%)	User-Business	16,239	14,284	198,397	12.2	13.9	UBU, BUB UBCiBU, BCiB UBCaBU, BCaB
	User-User	10,580	10,580	158,590	15.0	15.0	
	User-Compliment	14,411	11	76,875	5.3	6988.6	
	Business-City	14,267	47	14,267	1.0	303.6	
	Business-Category	14,180	511	40,009	2.8	78.3	

Metrics

- Mean absolute error **(MAE)**
- Root mean square error **(RMSE)**

Baselines

Typical methods

- PMF
- SoMF

HIN-based Methods

- FMHIN
- HeteMF
- SemRec
- DSR

Our Methods

- HERec_{dw}
- HERec_{mp}
- HERec

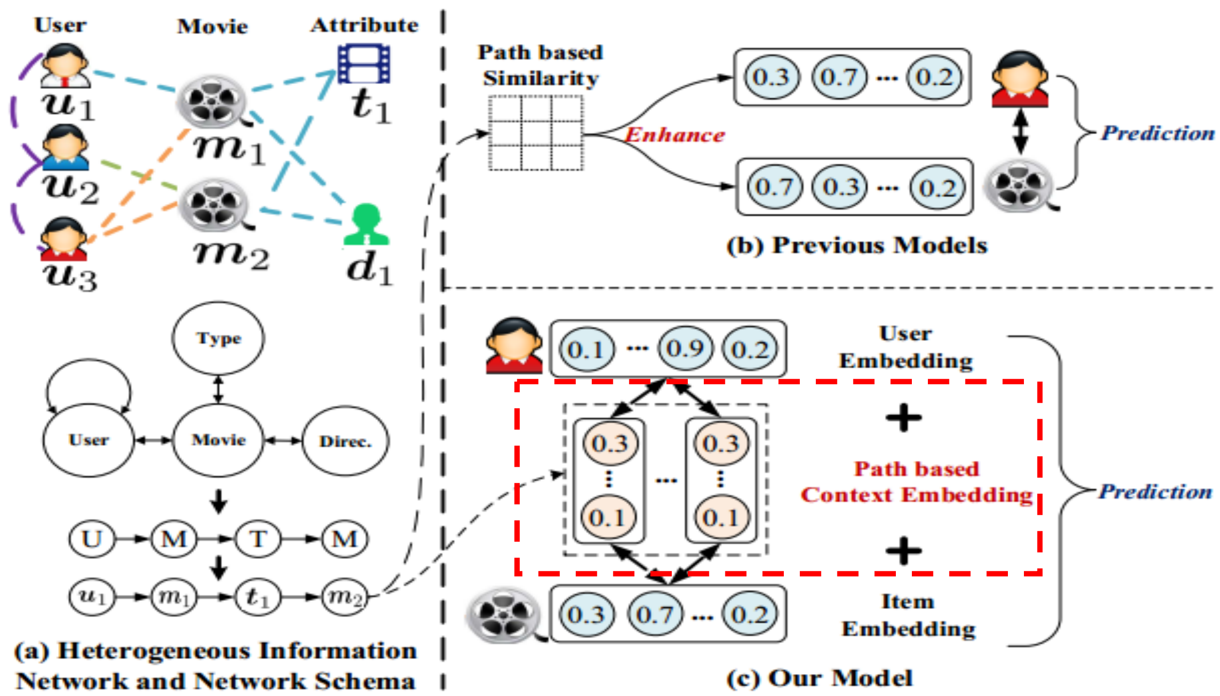
Effectiveness Experiments

Table 3: Results of effectiveness experiments on three datasets. A smaller value indicates a better performance.

Dataset	Training	Metrics	PMF	SoMF	FM_{HIN}	HeteMF	SemRec	DSR	$HERec_{dw}$	$HERec_{mp}$	$HERec_{sl}$	$HERec_{pl}$	$HERec_{pnl}$
Douban Movie	80%	MAE	0.5741	0.5817	0.5696	0.5750	0.5695	0.5681	0.5703	0.5515	0.5617	0.5523	**0.5519**
		RMSE	0.7641	0.7680	0.7248	0.7556	0.7399	0.7225	0.7446	0.7121	0.7216	**0.7024**	0.7053
	60%	MAE	0.5867	0.5991	0.5769	0.5894	0.5738	0.5831	0.5838	0.5611	0.5711	0.5606	**0.5587**
		RMSE	0.7891	0.7950	0.7342	0.7785	0.7551	0.7408	0.7670	0.7264	0.7336	**0.7142**	0.7148
	40%	MAE	0.6078	0.6328	0.5871	0.6165	0.5945	0.6170	0.6073	0.5747	0.5832	0.5732	**0.5699**
		RMSE	0.8321	0.8479	0.7563	0.8221	0.7836	0.7850	0.8057	0.7429	0.7514	0.7334	**0.7315**
	20%	MAE	0.7247	0.6979	0.6080	0.6896	0.6392	0.6584	0.6699	0.6063	0.5953	0.5965	**0.5900**
		RMSE	0.9440	0.9852	0.7878	0.9357	0.8599	0.8345	0.9076	0.7877	0.7916	0.7674	**0.7660**
Douban Book	80%	MAE	0.5774	0.5756	0.5716	0.5740	0.5675	0.5740	0.5875	0.5591	0.5578	0.5556	**0.5502**
		RMSE	0.7414	0.7302	0.7199	0.7360	0.7283	0.7206	0.7450	0.7081	0.7079	0.7093	**0.6811**
	60%	MAE	0.6065	0.5603	0.5812	0.5823	0.5833	0.6020	0.6203	0.5666	0.5690	0.5669	**0.5600**
		RMSE	0.7908	0.7518	0.7319	0.7466	0.7505	0.7552	0.7905	0.7318	0.7251	0.7274	**0.7123**
	40%	MAE	0.6800	0.6161	0.6028	0.5982	0.6025	0.6271	0.6976	0.5954	0.5838	**0.5638**	0.5774
		RMSE	0.9203	0.7936	0.7617	0.7779	0.7751	0.7730	0.9022	0.7703	0.7490	0.7549	**0.7400**
	20%	MAE	1.0344	0.6327	0.6396	0.6311	0.6481	0.6300	1.0166	0.6785	**0.6232**	0.6347	0.6450
		RMSE	1.4414	0.8236	0.8188	0.8304	0.8350	0.8200	1.3205	0.8869	**0.8168**	0.8382	0.8581
Yelp	90%	MAE	1.0412	1.0095	0.9013	0.9487	0.9043	0.9054	1.0388	0.8822	0.8643	0.8506	**0.8395**
		RMSE	1.4268	1.3392	1.1417	1.2549	1.1637	1.1186	1.3581	1.1309	1.1204	1.0948	**1.0907**
	80%	MAE	1.0791	1.0373	0.9038	0.9654	0.9176	0.9098	1.0750	0.8953	0.8789	0.8578	**0.8475**
		RMSE	1.4816	1.3782	1.1497	1.2799	1.1771	1.1208	1.4075	1.1516	1.1403	1.1139	**1.1117**
	70%	MAE	1.1170	1.0694	0.9108	0.9975	0.9407	0.9429	1.1196	0.9043	0.8889	0.8650	**0.8580**
		RMSE	1.5387	1.4201	1.1651	1.3229	1.2108	1.1582	1.4632	1.1639	1.1599	**1.1229**	1.1256
	60%	MAE	1.1778	1.1135	0.9435	1.0368	0.9637	1.0043	1.1691	0.9257	0.9042	0.8818	**0.8759**
		RMSE	1.6167	1.4748	1.2039	1.3713	1.2380	1.2257	1.5182	1.1887	1.1817	**1.1451**	1.1488
Average Rank			10.33	8.79	5.25	7.83	6.54	6.13	9.5	4.17	3.2	2.4	**1.79**

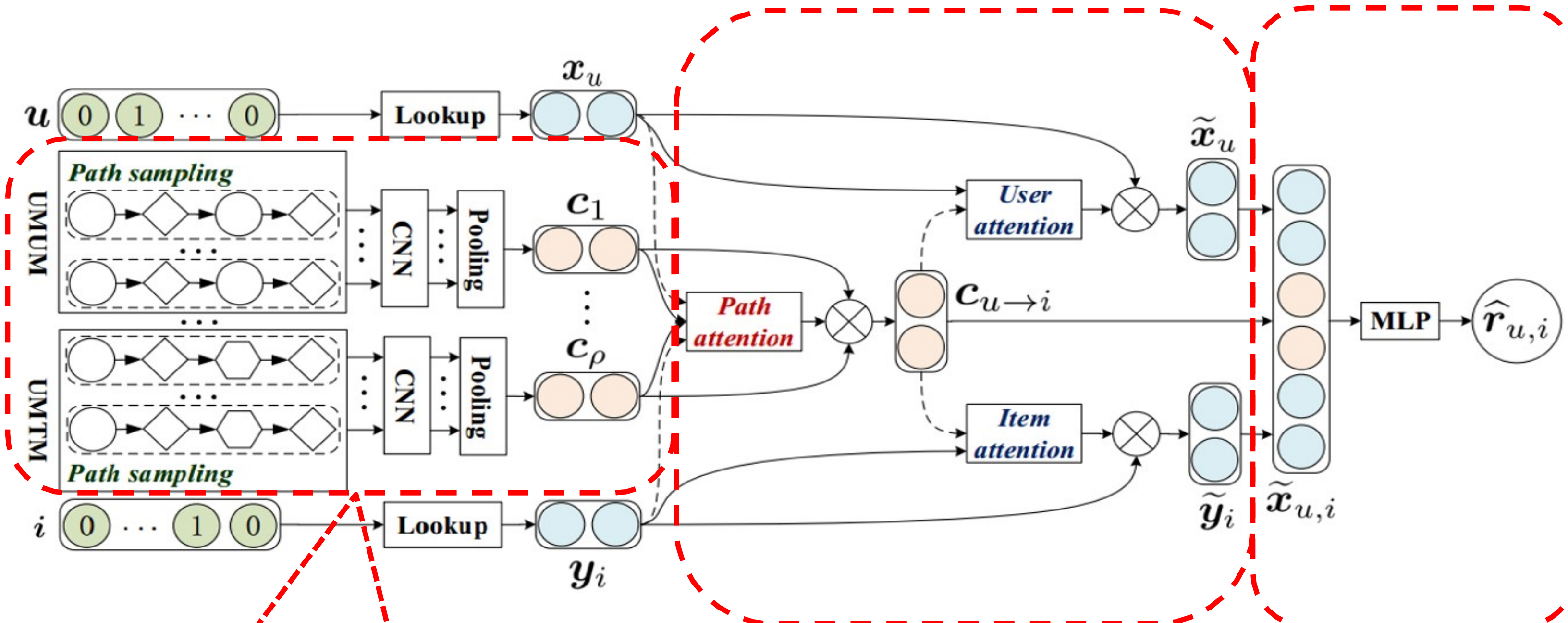
Basic idea

- Learn explicit representations for meta-path based context tailored for the recommendation task
- Characterize a three-way interaction ⟨user, meta-path, item⟩



Binbin Hu, Chuan Shi, Wayne Xin Zhao, Philip S. Yu. Leveraging Meta-path based Context for Top-N Recommendation with A Neural Co-Attention Model. KDD 2018

Meta-path based Context for Recommendation (MCRec)



Interpretability

Meta-path based context embedding

Mutual Effect

Neural co-attention mechanism

Rank

Ranking predication model

Experimental Setup

Dataset

Datasets	Relations (A-B)	#A	#B	#A-B	Meta-paths
Movielens	*User-Movie*	943	1,682	100,000	UMUM
	User-User	943	943	47,150	UMGM
	Movie-Movie	1,682	1,682	82,798	UUUM
	Movie-Genre	1,682	18	2861	UMMM
LastFM	*User-Artist*	1,892	17,632	92,834	UATA
	User-User	1,892	1,892	18,802	UAUA
	Artist-Artist	17,632	17,632	153,399	UUUA
	Artist-Tag	17,632	11,945	184,941	UUA
Yelp	*User-Business*	16,239	14,284	198,397	UBUB
	User-User	16,239	16,239	158,590	UBCaB
	Business-City (Ci)	14,267	47	14,267	UUB
	Business-Category (Ca)	14,180	511	40,009	UBCiB

Metrics

- Perc@10
- Recall@10
- NDCG@10

Baselines

- **CF based Methods**
  - ItemKNN
  - BPR
  - MF
  - NeuMF
- **HIN based Methods**
  - SVDFeature_{hete}
  - SVDFeature_{mp}
  - HeteRS
  - FMG_{rank}

Methods

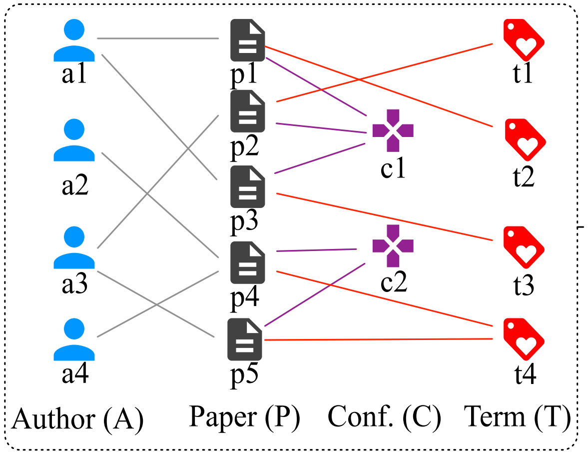
- MCRec_{rand}
- MCRec_{avg}
- MCRec_{mp}
- MCRec

Effectiveness Experiments

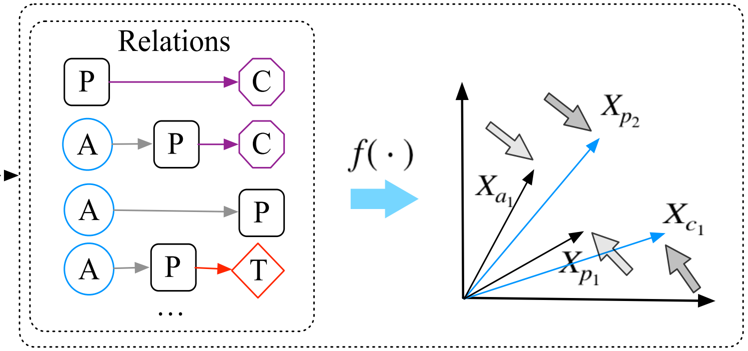
Model	Movielens			LastFM			Yelp		
	Prec@10	Recall@10	NDCG@10	Prec@10	Recall@10	NDCG@10	Prec@10	Recall@10	NDCG@10
ItemKNN	0.2578	0.1536	0.5692	0.4160	0.4513	0.7981	0.1386	0.5421	0.5378
BRP	0.3010	0.1946	0.6459	0.4129	0.4492	0.8099	0.1474	0.5504	0.5549
MF	0.3247	0.2053	0.6511	0.4364	0.4634	0.7921	0.1503	0.5350	0.5322
NeuMF	0.3293*	0.2090	0.6587	0.4540	0.4678	0.8104	0.1504	0.5857	0.5713
SVDFeature_{hete}	0.3171	0.2021	0.6445	0.4576	0.4841	0.8290*	0.1404	0.5613	0.5289
SVDFeature_{mp}	0.3109	0.1929	0.6536	0.4391	0.4651	0.8116	0.1524	0.5932	0.5974*
HeteRS	0.2485	0.1674	0.5967	0.4276	0.4489	0.8026	0.1423	0.5613	0.5600
FMG_{rank}	0.3256	0.2165*	0.6682*	0.4630*	0.4916*	0.8263	0.1538*	0.5951*	0.5861
MCRec_{rand}	0.3223	0.2104	0.6650	0.4540	0.4795	0.8002	0.1510	0.5842	0.5718
MCRec_{avg}	0.3270	0.2111	0.6631	0.4645	0.4914	0.8311	0.1595	0.5933	0.6021
MCRec_{mp}	0.3401	0.2200	0.6828	0.4662	0.4924	0.8428	0.1655	0.6303	0.6228
MCRec	**0.3451**#	**0.2256**#	**0.6900**#	**0.4807**#	**0.5068**#	**0.8526**#	**0.1686**#	**0.6326**#	**0.6301**#

MCRec significantly outperforms CF, NN, and HIN based recommendations

Background of RHINE



(a) An example of HIN



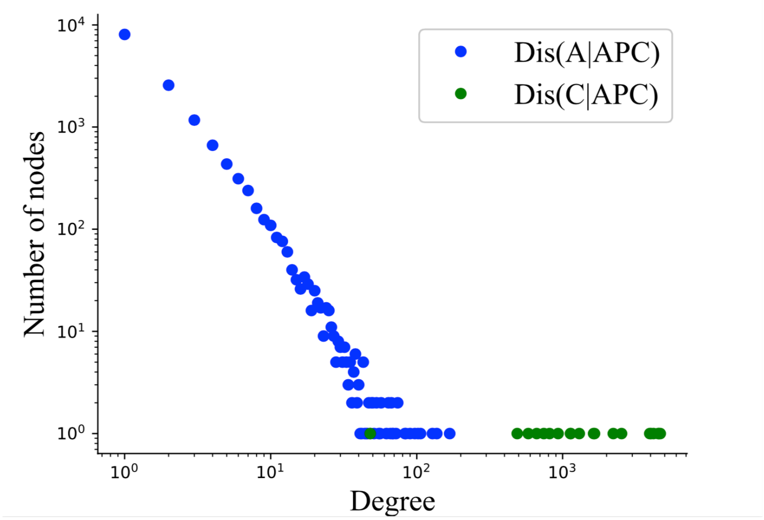
(b) Conventional Models

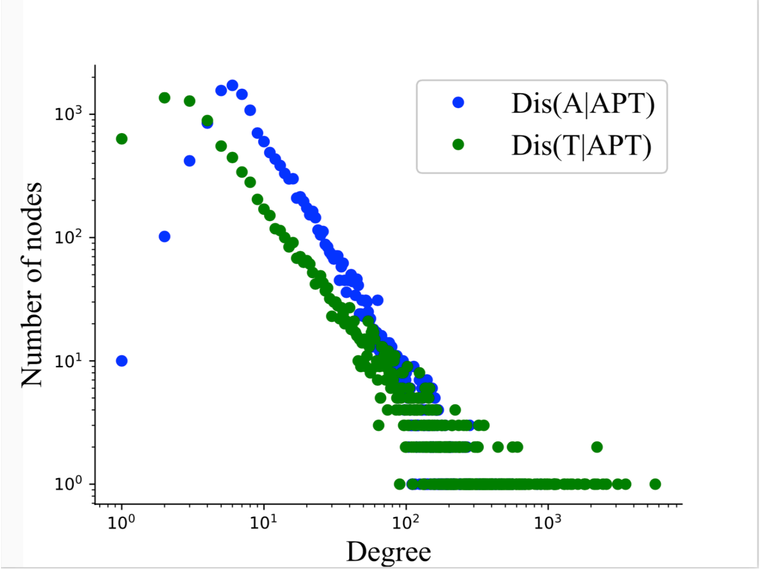
- **Existing methods**
  - Meta-path based random walk (metapath2vec, HIN2vec, MCRec, …)
  - Decompose based models (PTE, EOE, HERec, …)
  - Neural network based methods (HNE, SHINE , BL-MNE, …)

- **Drawbacks**
  - An assumption that one single model can handle all relations and nodes.
  - Without distinguishing the structural characterizes of relations.

Yuanfu Lu, Chuan Shi, Linmei Hu, Zhiyuan Liu. Relation Structure-Aware Heterogeneous Information Network Embedding. AAAI 2019.

However…





- **Our idea**

➢ Explore the structural characteristics of relations.

➢ Different models specifically tailored to handle various relations.

Relation structure-aware
Heterogeneous Information Network Embedding (RHINE)

RHINE Framework

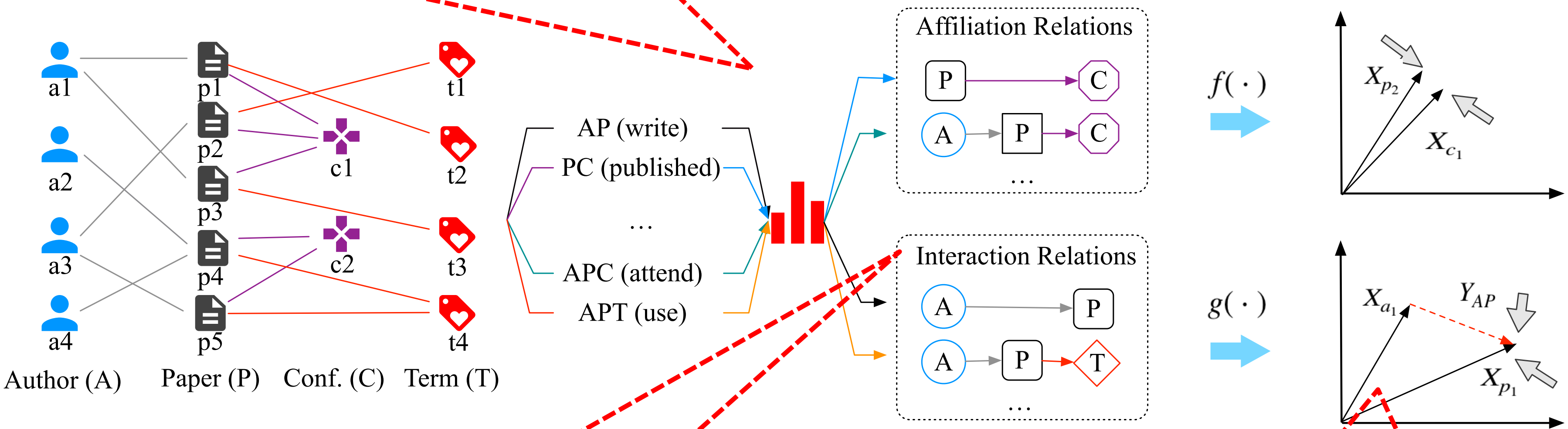
Two structure-related measures

- Degree-based measure

$$D(r) = \frac{\max[\bar{d}_{t_u}, \bar{d}_{t_v}]}{\min[\bar{d}_{t_u}, \bar{d}_{t_v}]}$$

- Sparsity-based measure

$$S(r) = \frac{N_r}{N_{t_u} \times N_{t_v}}$$



(a) An example of HIN (b) Explore the structural characteristic of relations (c) Model various relations

Two categories :
Affiliation Relations (ARs)
Interaction Relations (IRs)

Two models:
Euclidean distance
Translation-based distance

Experimental Setup

- **Datasets**

Datasets	Nodes	Number of Nodes	Relations ($t_u \sim t_v$)	Number of Relations
DBLP	Term (T) Paper (P) Author (A) Conference (C)	8,811 14,376 14,475 20	PC APC AP PT APT	14,376 24,495 41,794 88,683 260,605
Yelp	User (U) Service (S) Business (B) Star Level (L) Reservation (R)	1,286 2 2,614 9 2	BR BS BL UB BUB	2,614 2,614 2,614 30,838 528,332
AMiner	Paper (P) Author (A) Reference (R) Conference (C)	127,623 164,472 147,251 101	PC APC AP PR APR	127,623 232,659 355,072 392,519 1,084,287

- **Baselines**

✓ DeepWalk
✓ LINE
✓ PTE
✓ Esim
✓ HIN2Vec
✓ Metapath2vec

- **Tasks**

✓ Node Clustering
✓ Link Prediction
✓ Multi-Class Classification
✓ Visualization

Effectiveness Experiments

Table 2: Performance Evaluation of Node Clustering (NMI).

Methods	DBLP	Yelp	AMiner
DeepWalk	0.3884	0.3043	0.5427
LINE-1st	0.2775	0.3103	0.3736
LINE-2nd	0.4675	0.3593	0.3862
PTE	0.3101	0.3527	0.4089
ESim	0.3449	0.2214	0.3409
HIN2Vec	0.4256	0.3657	0.3948
metapath2vec	0.6065	0.3507	0.5586
RHINE	**0.7204**	**0.3882**	**0.6024**

- **Compared with the best competitors**
  - **18.79% improvement on DBLP**
  - **6.15% improvement on Yelp**
  - **7.84% improvement on AMiner**

Table 3: Performance Evaluation of Link Prediction.

Methods	DBLP (A-A)		DBLP (A-C)		Yelp (U-B)		AMiner (A-A)		AMiner (A-C)	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1
DeepWalk	0.9131	0.8246	0.7634	0.7047	0.8476	0.6397	0.9122	0.8471	0.7701	0.7112
LINE-1st	0.8264	0.7233	0.5335	0.6436	0.5084	0.4379	0.6665	0.6274	0.7574	0.6983
LINE-2nd	0.7448	0.6741	0.8340	0.7396	0.7509	0.6809	0.5808	0.4682	0.7899	0.7177
PTE	0.8853	0.8331	0.8843	0.7720	0.8061	0.7043	0.8119	0.7319	0.8442	0.7587
ESim	0.9077	0.8129	0.7736	0.6795	0.6160	0.4051	0.8970	0.8245	0.8089	0.7392
HIN2Vec	0.9160	0.8475	0.8966	0.7892	0.8653	0.7709	0.9141	0.8566	0.8099	0.7282
metapath2vec	0.9153	0.8431	0.8987	0.8012	0.7818	0.5391	0.9111	0.8530	0.8902	0.8125
RHINE	**0.9315**	**0.8664**	**0.9148**	**0.8478**	**0.8762**	**0.7912**	**0.9316**	**0.8664**	**0.9173**	**0.8262**

Background

- HIN embedding

HIN Data

In a low dimensional space → Low dimensional Euclidean spaces

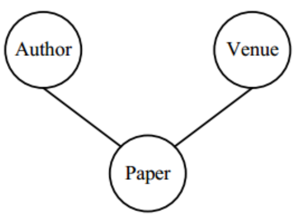
Learn the node representation → Preserving structure and semantics

Euclidean space is optimal for network embedding?

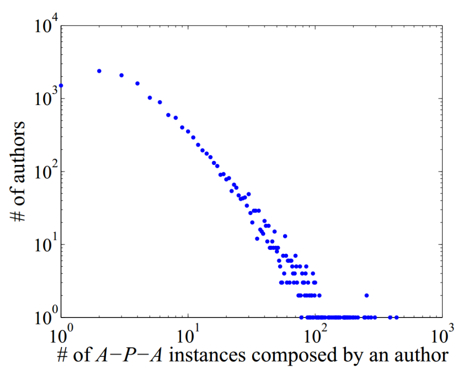
Xiao Wang, Yiding Zhang, Chuan Shi. Hyperbolic Heterogeneous Information Network Embedding. AAAI 2019.

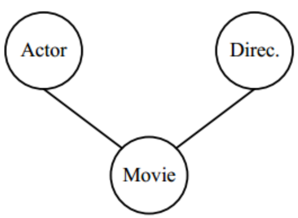
Motivation

- Hyperbolic geometry: a non-Euclidean geometry
  - Model data with power-law distribution naturally
  - Reflect the hierarchy naturally
- Network data
  - Power-law distribution, e.g. citation networks
  - Hierarchical structure

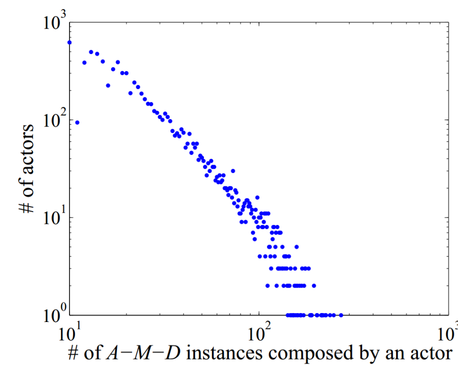


(a) Schema of DBLP





(b) Scheme of MovieLens



Can we embed network in Hyperbolic space?

Overview of HHNE

- Goal
  - learn the representation of nodes in hyperbolic spaces
- Basic process
  - Use meta-path guided random walks to obtain heterogeneous neighborhoods of a node
  - Use distance in hyperbolic spaces to measure the similarity between nodes
  - Design optimization objective based on the skip-gram, and optimize it in hyperbolic spaces

Experiments

- Dataset

DBLP	# A	# P	# V	# P-A	# P-V
	14475	14376	20	41794	14376
MovieLens	# A	# M	# D	# M-A	# M-D
	11718	9160	3510	64051	9160

- Baselines
  - DeepWalk (Perozzi et al. KDD 2014)
  - LINE (Tang et al. WWW 2015)
  - Node2vec (Grover et al. KDD 2016)
  - Metapath2vec (Dong et al. KDD 2017)
  - PoincareEmb (Nickel et al. NIPS 2017)
- Metric
  - AUC

Effectiveness Experiments

- Network Reconstruction

Dataset	Edge	Dimension	Deepwalk	LINE(1st)	LINE(2nd)	node2vec	metapath2vec	PoincaréEmb	HHNE
DBLP	P-A	2	0.6933	0.5286	0.6740	0.7107	0.6686	0.8251	**0.9835**
		5	0.8034	0.5397	0.7379	0.8162	0.8261	0.8769	**0.9838**
		10	0.9324	0.6740	0.7541	0.9418	0.9202	0.8921	**0.9887**
		15	0.9666	0.7220	0.7868	0.9719	0.9500	0.8989	**0.9898**
		20	0.9722	0.7457	0.7600	0.9809	0.9623	0.9024	**0.9913**
		25	0.9794	0.7668	0.7621	0.9881	0.9690	0.9034	**0.9930**
	P-V	2	0.7324	0.5182	0.6242	0.7595	0.7286	0.5718	**0.8449**
		5	0.7906	0.5500	0.6349	0.8019	0.9072	0.5529	**0.9984**
		10	0.8813	0.7070	0.6333	0.8922	0.9691	0.6271	**0.9985**
		15	0.9353	0.7295	0.6343	0.9382	0.9840	0.6446	**0.9985**
		20	0.9505	0.7369	0.6444	0.9524	0.9879	0.6600	**0.9985**
		25	0.9558	0.7436	0.6440	0.9596	0.9899	0.6760	**0.9985**
MoiveLens	M-A	2	0.6320	0.5424	0.6378	0.6402	0.6404	0.5231	**0.8832**
		5	0.6763	0.5675	0.7047	0.6774	0.6578	0.5317	**0.9168**
		10	0.7610	0.6202	0.7739	0.7653	0.7231	0.5404	**0.9211**
		15	0.8244	0.6593	0.7955	0.8304	0.7793	0.5479	**0.9221**
		20	0.8666	0.6925	0.8065	0.8742	0.8189	0.5522	**0.9239**
		25	0.8963	0.7251	0.8123	0.9035	0.8483	0.5545	**0.9233**
	M-D	2	0.6626	0.5386	0.6016	0.6707	0.6589	0.6213	**0.9952**
		5	0.7263	0.5839	0.6521	0.7283	0.7230	0.7266	**0.9968**
		10	0.8246	0.6114	0.6969	0.8308	0.8063	0.7397	**0.9975**
		15	0.8784	0.6421	0.7112	0.8867	0.8455	0.7378	**0.9972**
		20	0.9117	0.6748	0.7503	0.9186	0.8656	0.7423	**0.9982**
		25	0.9345	0.7012	0.7642	0.9402	0.8800	0.7437	**0.9992**

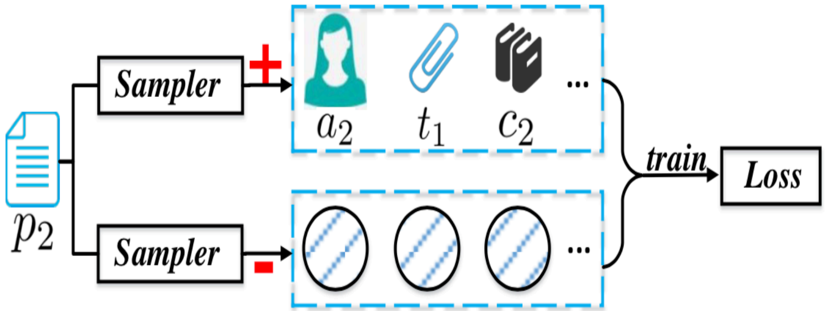
Effectiveness Experiments

- Link Prediction

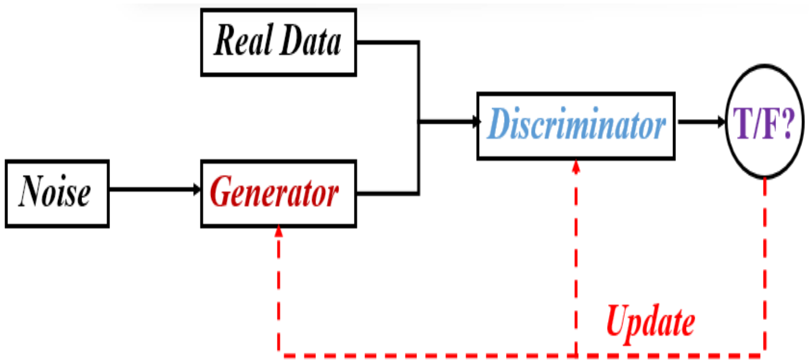
Dataset	Edge	Dimension	Deepwalk	LINE(1st)	LINE(2nd)	node2vec	metapath2vec	PoincaréEmb	HHNE
DBLP	P-A	2	0.5813	0.5090	0.5909	0.6709	0.6536	0.6742	**0.8777**
		5	0.7370	0.5168	0.6351	0.7527	0.7294	0.7381	**0.9041**
		10	0.8250	0.5427	0.6510	0.8469	0.8279	0.7699	**0.9111**
		15	0.8664	0.5631	0.6582	0.8881	0.8606	0.7743	**0.9111**
		20	0.8807	0.5742	0.6644	0.9037	0.8740	0.7806	**0.9106**
		25	0.8878	0.5857	0.6782	0.9102	0.8803	0.7830	**0.9117**
	P-V	2	0.7075	0.5160	0.5121	0.7369	0.7059	0.8257	**0.9331**
		5	0.7197	0.5663	0.5216	0.7286	0.8516	0.8878	**0.9409**
		10	0.7292	0.5873	0.5332	0.7481	0.9248	0.9113	**0.9619**
		15	0.7325	0.5896	0.5425	0.7583	0.9414	0.9142	**0.9625**
		20	0.7522	0.5891	0.5492	0.7674	0.9504	0.9185	**0.9620**
		25	0.7640	0.5846	0.5512	0.7758	0.9536	0.9192	**0.9612**
MoiveLens	M-A	2	0.6278	0.5053	0.5712	0.6349	0.6168	0.5535	**0.7715**
		5	0.6353	0.5636	0.5874	0.6402	0.6212	0.5779	**0.8255**
		10	0.6680	0.5914	0.6361	0.6700	0.6332	0.5984	**0.8312**
		15	0.6791	0.6184	0.6442	0.6814	0.6382	0.5916	**0.8319**
		20	0.6868	0.6202	0.6596	0.6910	0.6453	0.5988	**0.8318**
		25	0.6890	0.6256	0.6700	0.6977	0.6508	0.5995	**0.8309**
	M-D	2	0.6258	0.5139	0.6501	0.6299	0.6191	0.5856	**0.8520**
		5	0.6482	0.5496	0.6607	0.6589	0.6332	0.6290	**0.8967**
		10	0.6976	0.5885	0.7499	0.7034	0.6687	0.6518	**0.8984**
		15	0.7163	0.6647	0.7756	0.7241	0.6702	0.6715	**0.9007**
		20	0.7324	0.6742	0.7982	0.7412	0.6746	0.6821	**0.9000**
		25	0.7446	0.6957	0.8051	0.7523	0.6712	0.6864	**0.9018**

Motivation of HeGAN

Negative samples are a basic function in HIN embedding



Adversarial learning can provide better samples



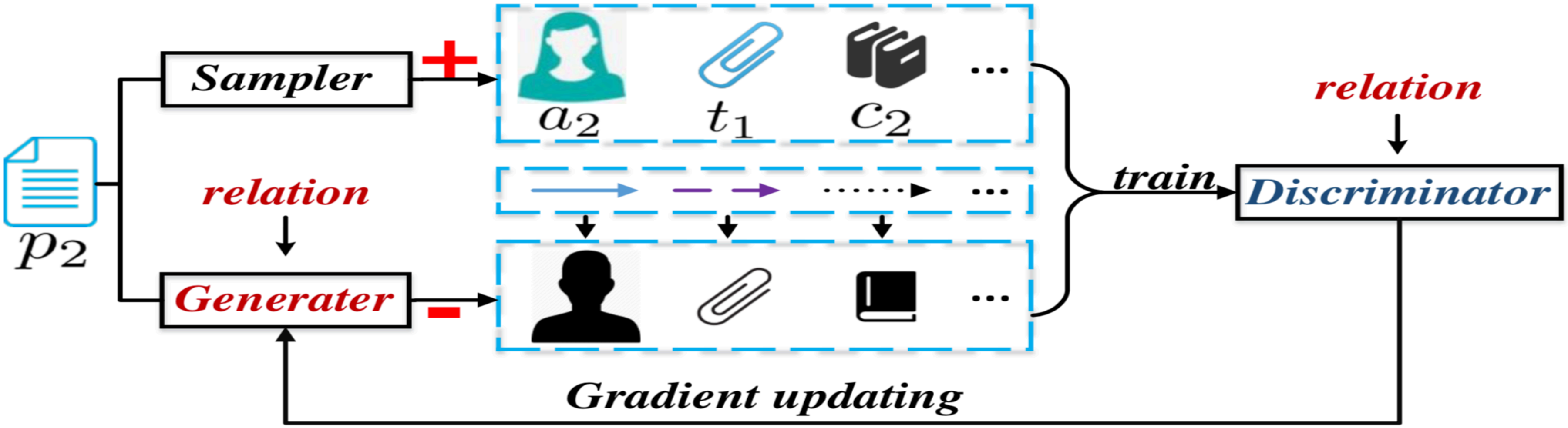
Two Challenges

- How to capture the semantics of multiple types of nodes and relations?
- How to generate fake samples efficiently and effectively?

Binbin Hu, Yuan Fang, Chuan Shi. Adversarial Learning on Heterogeneous Information Networks. KDD 2019.

Basic idea

HIN Embedding with GAN based Adversarial Learning (HeGAN)



Relation-aware Generator and Discriminator → **Challenge 1**

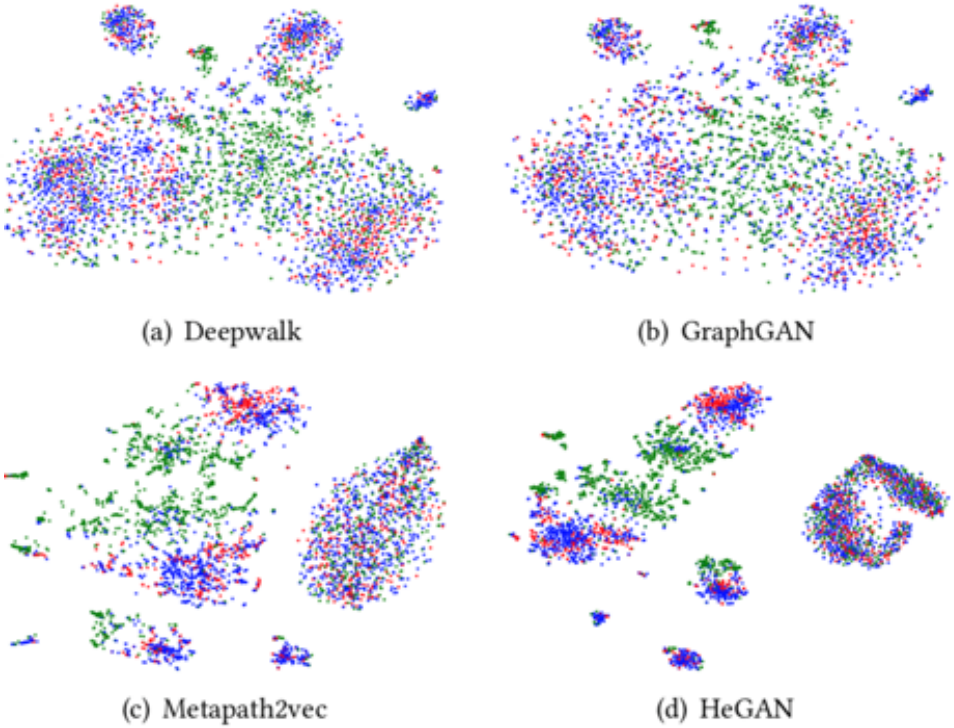
(i) Discriminator tells whether a node pair is real or fake w.r.t relation
(ii) Generator produces fake node pairs that mimic real pairs w.r.t relation

Generalized Generator → **Challenge 2**

(i) Sample latent nodes from a continuous distribution
(ii) no softmax computation and fake samples are not restricted to the existing nodes

Experiments

Methods	DBLP	Yelp	AMiner
Deepwalk	0.7398	0.3306	0.4773
LINE-1st	0.7412	0.3556	0.3518
LINE-2nd	0.7336	0.3560	0.2144
GraphGAN	0.7409	0.3413	-
ANE	0.7138	0.3145	0.4483
HERec-HNE	0.7274	0.3476	0.4635
HIN2vec	0.7204	0.3185	0.2812
Metapath2vec	0.7675	0.3672	0.4726
HeGAN	**0.7920****	**0.4037****	**0.5052****



(a) Deepwalk (b) GraphGAN (c) Metapath2vec (d) HeGAN

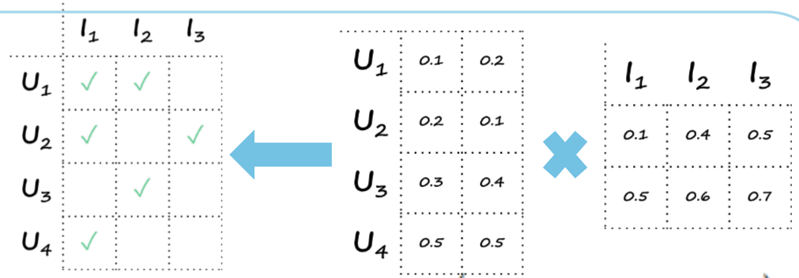
HeGAN learns semantic-preserving representations in a robust manner

HeGAN has a more crisp boundary and denser clusters

Outline

- Metapath based data mining
- ✓ **Heterogeneous information network embedding**
  - ✓ Shadow models
    - MetaPath2Vec (KDD2017), HIN2Vec (CIKM2017)
    - HERec (TKDE2018), MCRec (KDD2018), RHINE (AAAI2019)
    - HHNE (AAAI2019), HeGAN (KDD2019)
  - ✓ Deep models
    - NeuACF(IJCAI2018), HetGNN(KDD2019), HAN(WWW2019)
- Applications
- Conclusion and future work

- **Latent Factor Model**
  - ➢ Factorize user-item interaction matrix
  - ➢ User/item specific latent factor

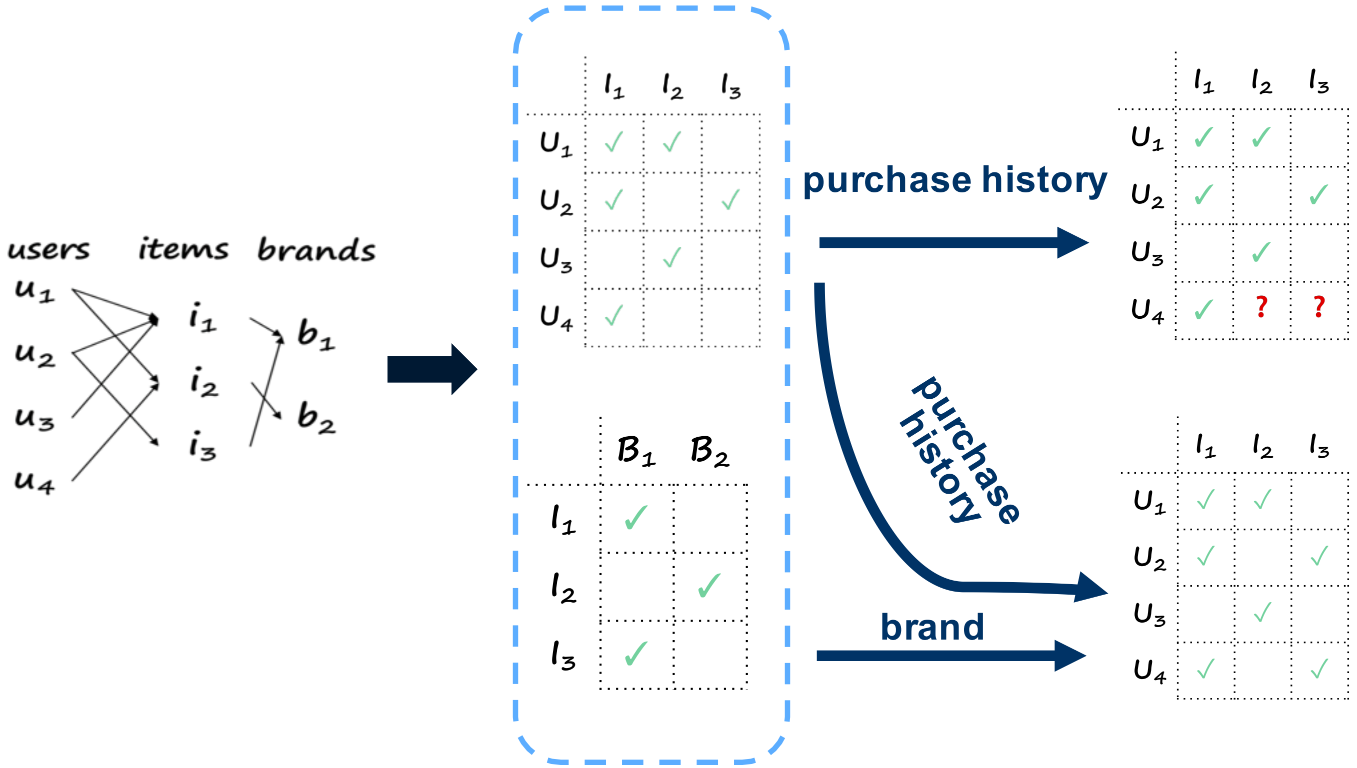


$$\arg\min_{\boldsymbol{u},\boldsymbol{v}} \sum_i \sum_j (\boldsymbol{R}_{i,j} - \boldsymbol{u}_i^T \boldsymbol{v}_j)^2 + \lambda \left(\sum_i ||\boldsymbol{u}_i||_2^2 + \sum_j ||\boldsymbol{v}_j||_2^2 \right)$$

- **Shortcomings**
  - ➢ Only explore the one aspect information
  - ➢ Fail to integrate more information

Xiaotian Han, Chuan Shi, Senzhang Wang, Philip S. Yu, Li Song. Aspect-Level Deep Collaborative Filtering via Heterogeneous Information Networks. IJCAI 2018.
Chuan Shi, Xiaotian Han, Li Song, Xiao Wang, Senzhang Wang, Junping Du, Philip S. Yu.Deep Collaborative Filtering with Multi-Aspect Information in Heterogeneous Networks. TKDE 2019

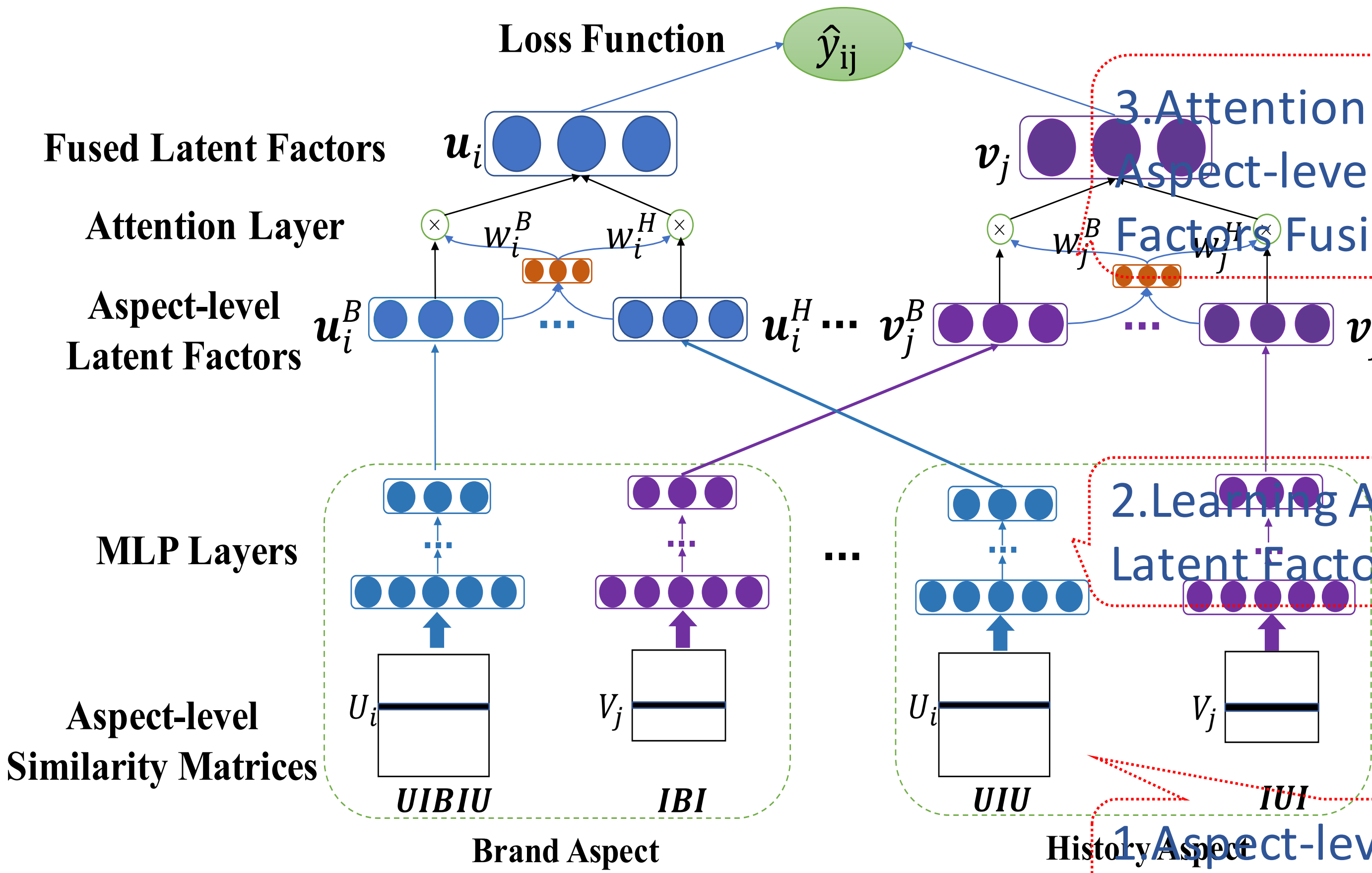
Motivation



- Heterogeneous network contain different-aspect interaction information
- Using different aspects information in HIN to improve the performance

Framework of NeuACF

Neural network based **A**spect-level **C**ollaborative **F**iltering (NeuACF)



3.Attention based Aspect-level Latent Factors Fusing

2.Learning Aspect-level Latent Factors

1.Aspect-level Similarity Matrix Extraction

Experimental Setup

Dataset

Dataset	#users	#items	#ratings	#density
ML100K	943	1682	100,000	6.304%
ML1M	6040	3706	1,000,209	4.468%
Amazon	3532	3105	57,104	0.521%

Datasets	Aspect	Meta-Paths	
		User	Movie/Item
MovieLens	History	UMU	MUM
	Director	$UMDMU$	MDM
	Actor	$UMAMU$	MAM
Amazon	History	UIU	IUI
	Brand	$UIBIU$	IBI
	Category	$UICIU$	ICI
	Co_view	$UIVIU$	IVI

Metrics

- $HR@k = \frac{\#hits}{\#users}$
- $NDCG@k = \frac{1}{\#users} \sum_{i=1}^{\#users} \frac{1}{log_2(p_i+1)}$

Baselines

Basic methods

- ItemPop
- ItemKNN

LFM methods

- MF
- eALS
- BPR

NN-based methods

- DMF
- NeuMF

HIN-based methods

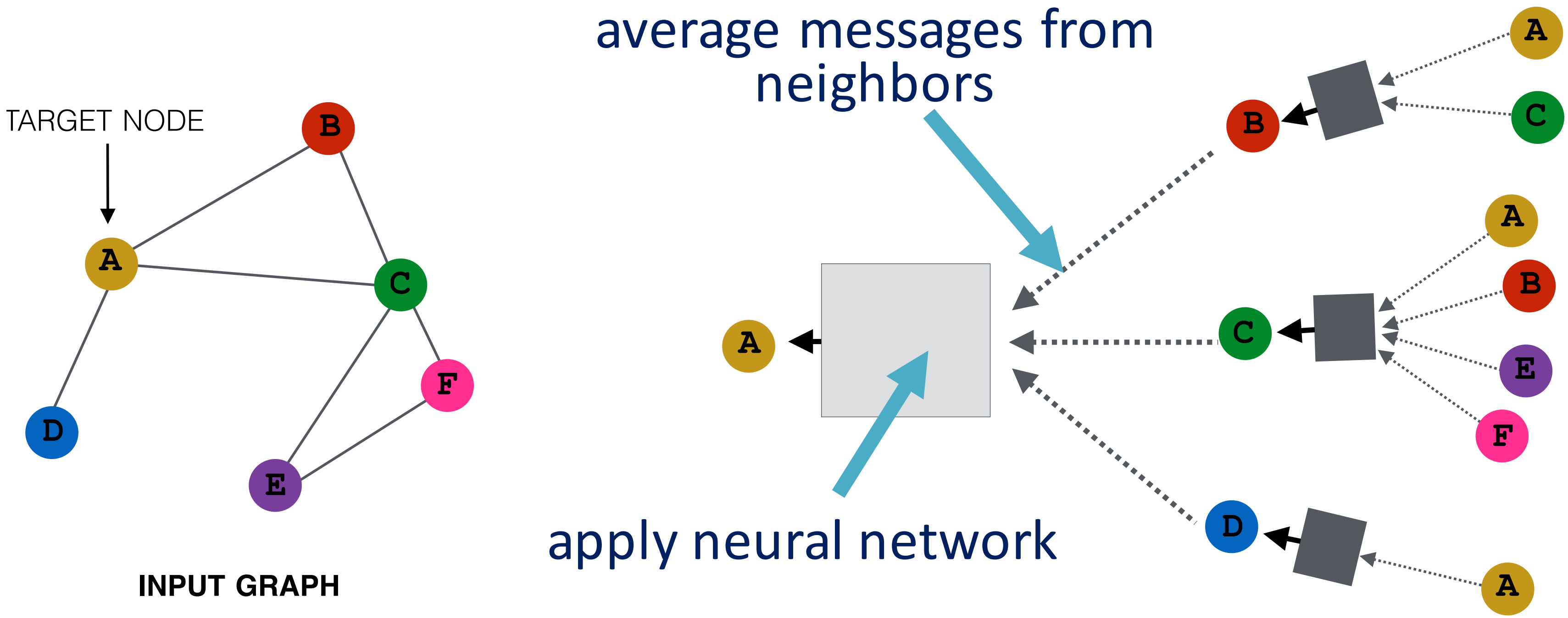
- FMG

Effectiveness Experiments

Datasets	Metrics	ItemPop	ItemKNN	MF	eALS	BPR	DMF	NeuMF	FMG	NeuACF
ML100K	HR@5	0.2831	0.4072	0.4634	0.4698	0.4984	0.3483	0.4942	0.4602	**0.5097**
	NDCG@5	0.1892	0.2667	0.3021	0.3201	0.3315	0.2287	0.3357	0.3014	**0.3505**
	HR@10	0.3998	0.5891	0.6437	0.6638	**0.6914**	0.4994	0.6766	0.6373	0.6846
	NDCG@10	0.2264	0.3283	0.3605	0.3819	0.3933	0.2769	0.3945	0.3588	**0.4068**
	HR@15	0.5366	0.7094	0.7338	0.7529	0.7741	0.5873	0.7635	0.7338	**0.7813**
	NDCG@15	0.2624	0.3576	0.3843	0.4056	0.4149	0.3002	0.4175	0.3844	**0.4318**
	HR@20	0.6225	0.7656	0.8144	0.8155	0.8388	0.6519	0.8324	0.8006	**0.8464**
	NDCG@20	0.2826	0.3708	0.4034	0.4204	0.4302	0.3151	0.4338	0.4002	**0.4469**
ML1M	HR@5	0.3088	0.4437	0.5111	0.5353	0.5414	0.4892	0.5485	0.4732	**0.5630**
	NDCG@5	0.2033	0.3012	0.3463	0.3670	0.3756	0.3314	0.3865	0.3183	**0.3944**
	HR@10	0.4553	0.6171	0.6896	0.7055	0.7161	0.6652	0.7177	0.6528	**0.7202**
	NDCG@10	0.2505	0.3572	0.4040	0.4220	0.4321	0.3877	0.4415	0.3767	**0.4453**
	HR@15	0.5568	0.7118	0.7783	0.7914	0.7988	0.7649	0.7982	0.7536	**0.8018**
	NDCG@15	0.2773	0.3822	0.4275	0.4448	0.4541	0.4143	0.4628	0.4034	**0.4667**
	HR@20	0.6409	0.7773	0.8425	0.8409	0.8545	0.8305	**0.8586**	0.8169	0.8540
	NDCG@20	0.2971	0.3977	0.4427	0.4565	0.4673	0.4296	0.4771	0.4184	**0.4789**
Amazon	HR@5	0.2412	0.1897	0.3027	0.3063	**0.3296**	0.2693	0.3117	0.3216	0.3268
	NDCG@5	0.1642	0.1279	0.2068	0.2049	**0.2254**	0.1848	0.2141	0.2168	0.2232
	HR@10	0.3576	0.3126	0.4278	0.4287	0.4657	0.3715	0.4309	0.4539	**0.4686**
	NDCG@10	0.2016	0.1672	0.2471	0.2441	**0.2693**	0.2179	0.2524	0.2595	0.2683
	HR@15	0.4408	0.3901	0.5054	0.5065	0.5467	0.4328	0.5258	0.5430	**0.5591**
	NDCG@15	0.2236	0.1877	0.2676	0.2647	0.2908	0.2332	0.2774	0.2831	**0.2924**
	HR@20	0.4997	0.4431	0.5680	0.5702	0.6141	0.4850	0.5897	0.6076	**0.6257**
	NDCG@20	0.2375	0.2002	0.2824	0.2797	0.3067	0.2458	0.2925	0.2983	**0.3080**

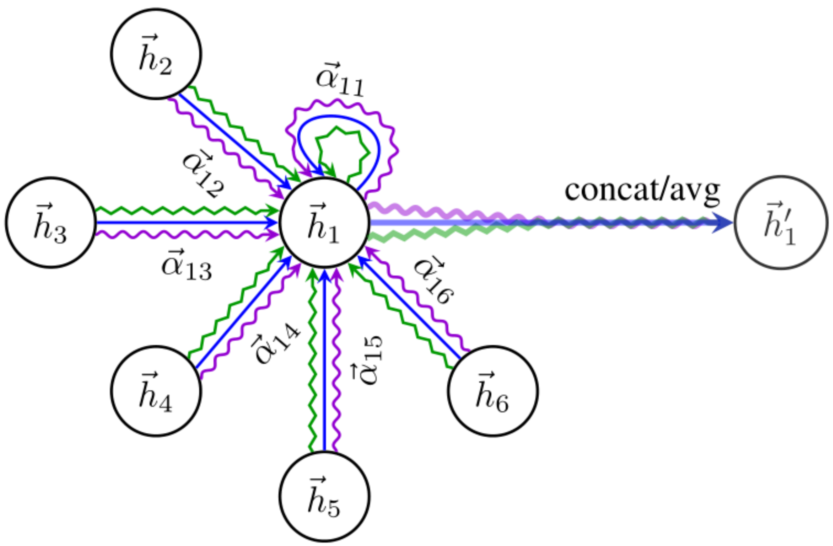
Background

- Graph Neural Network
  - Neural networks for processing graph-structured inputs.
  - Average neighbor information and apply a neural network.



Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Peng Cui, Philip S. Yu, Yanfang Ye. Heterogeneous Graph Attention Network. WWW 2019.

Motivation of HAN



A dog is standing on a hardwood floor.

GAT
for Homogeneous graph
(Petar et al, ICLR2018)

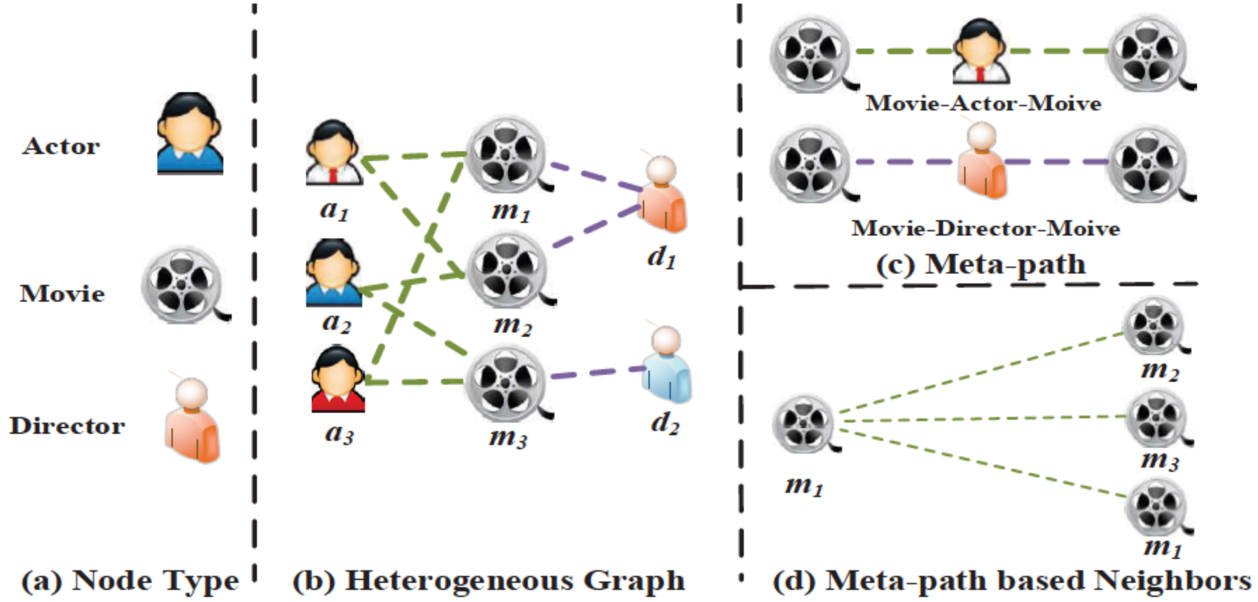
Attention Mechanism

Can we design attention mechanism
for heterogeneous graph?

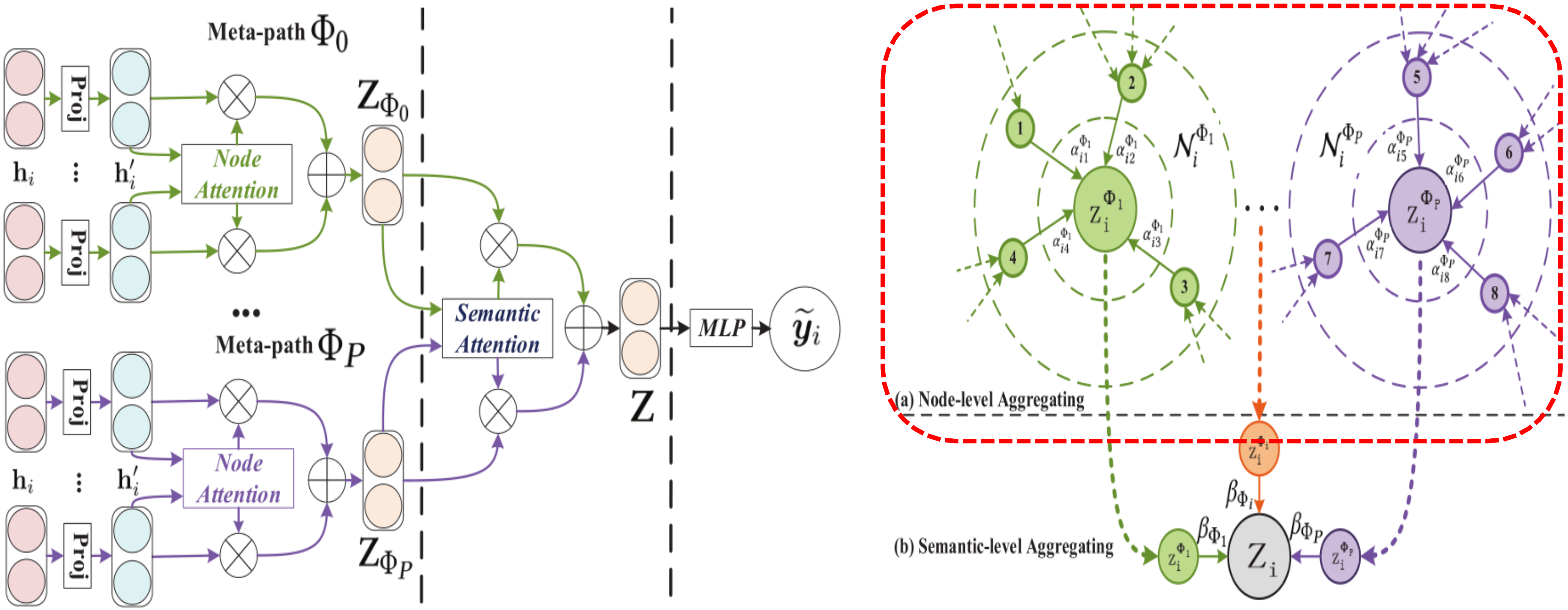
Challenges

Requirements of Heterogeneous Graph Neural Network

- Heterogeneity of graph
- Node-level attention
- Semantic-level attention



Heterogeneous Graph **A**ttention **N**etwork(**HAN**)



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

Experiments Setting

- Baselines
  - Deepwalk
  - Esim
  - Metapath2vec
  - HERec
  - GCN
  - GAT
  - HAN_{nd}
  - HAN_{sem}
- Tasks
  - Classification
  - Clustering
  - Attention Analysis
  - Visualization

Table 2: Statistics of the datasets.

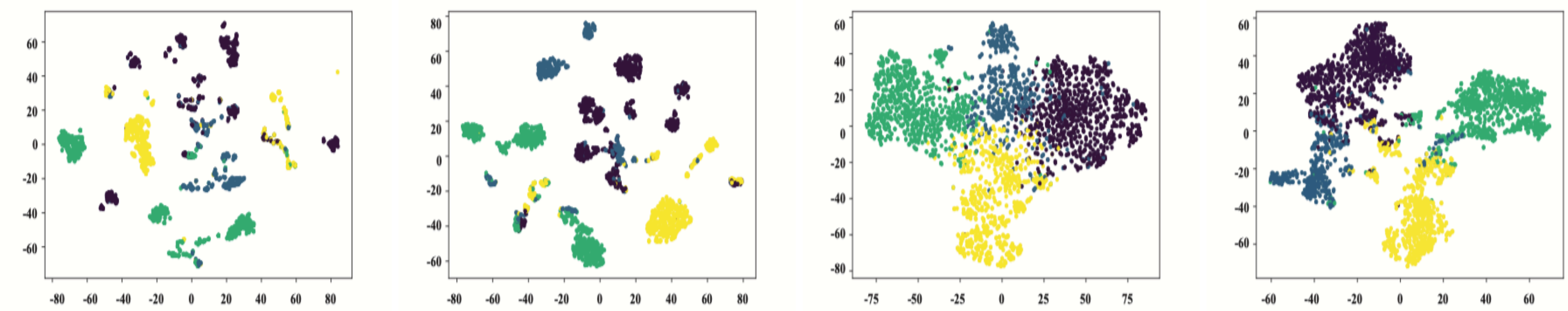
Dataset	Relations(A-B)	Number of A	Number of B	Number of A-B	Feature	Training	Validation	Test	Meta-paths
DBLP	Paper-Author	14328	4057	19645	334	800	400	2857	*APA*
	Paper-Conf	14328	20	14328					*APCPA*
	Paper-Term	14327	8789	88420					*APTPA*
IMDB	Movie-Actor	4780	5841	14340	1232	300	300	2687	*MAM*
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

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**

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**

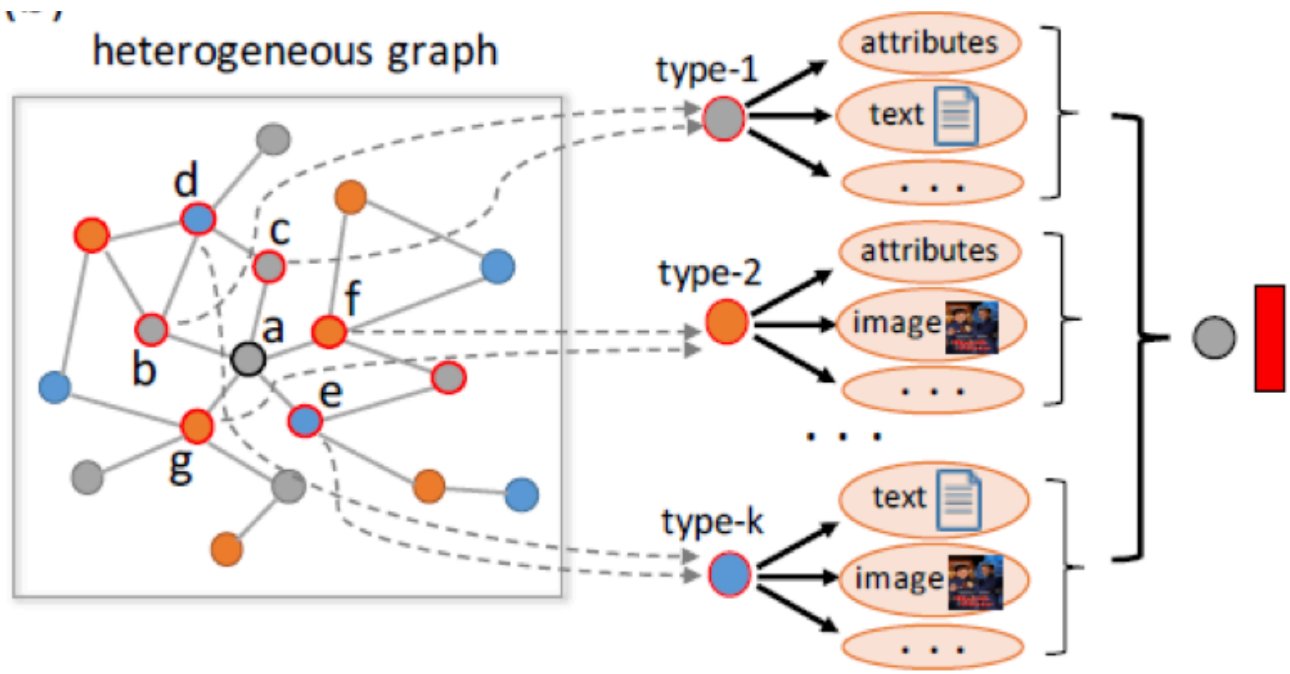


(a) GCN (b) GAT (c) metapath2vec (d) HAN

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

Motivation of HetGNN

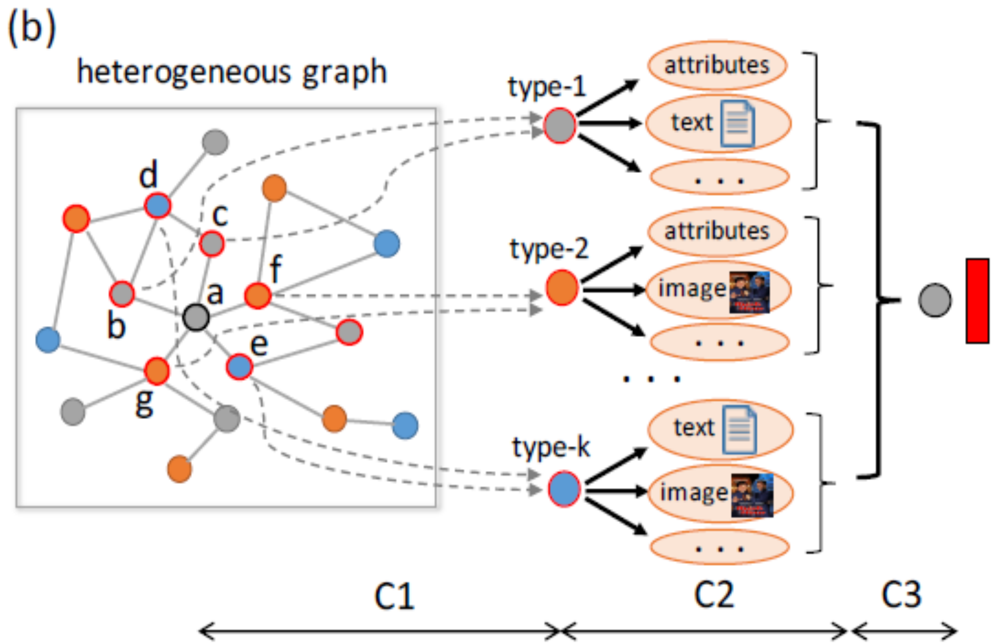
- Can we jointly consider heterogeneous structural (graph) information as well as heterogeneous contents information for node embedding effectively in HIN?



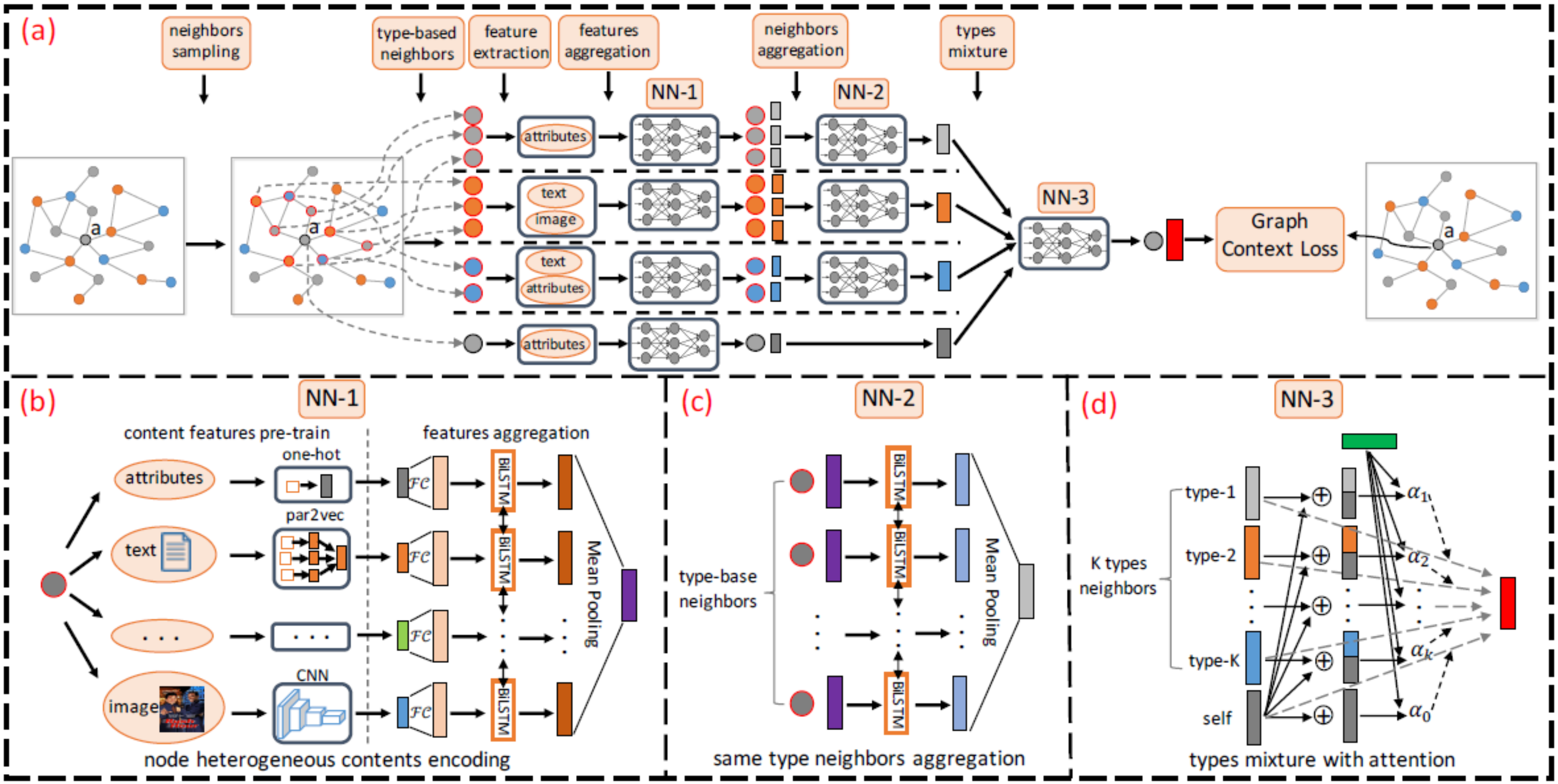
Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, Nitesh V. Chawla. Heterogeneous Graph Neural Network. KDD 2019.

Requirements of Heterogeneous Graph Neural Network

- C1: Sample heterogeneous neighbors are correlated to embedding generation
- C2: Design node content encoder for content heterogeneity
- C3: Aggregate heterogeneous neighbors by considering node types



Heterogeneous Graph Neural Network(HetGNN)



(a) The overall architecture of HetGNN (b) NN-1: node heterogeneous contents encoder; (c) NN-2: type-based neighbors aggregator; (d) NN-3: heterogeneous types combination.

Experimental Setup

Dataset

Data	Node	Edge
Academic I (A-I)	# author: 160,713 # paper: 111,409 # venue: 150	# author-paper: 295,103 # paper-paper: 138,464 # paper-venue: 111,409
Academic II (A-II)	# author: 28,646 # paper: 21,044 # venue: 18	# author-paper: 69,311 # paper-paper: 46,931 # paper-venue: 21,044
Movies Review (R-I)	# user: 18,340 # item: 56,361	# user-item: 629,125
CDs Review (R-II)	# user: 16,844 # item: 106,892	# user-item: 555,050

Baselines

Heterogeneous graph embedding model

- metapath2vec(MP2V)

Attributed graph models

- ASNE
- SHNE

GNN models

- GraphSAGE
- GAT

Effectiveness Experiments

Link prediction task

$Data_{split}$	Metric	MP2V [4]	ASNE [15]	SHNE [34]	GSAGE [7]	GAT [31]	HetGNN
A-I_{2003} (type-1)	AUC F1	0.636 0.435	0.683 0.584	0.696 0.597	0.694 0.586	0.701 0.606	**0.714** **0.620**
A-I_{2003} (type-2)	AUC F1	0.790 0.743	0.794 0.774	0.781 0.755	0.790 0.746	0.821 0.792	**0.837** **0.815**
A-I_{2002} (type-1)	AUC F1	0.626 0.412	0.667 0.554	0.688 0.590	0.681 0.567	0.691 0.589	**0.710** **0.615**
A-I_{2002} (type-2)	AUC F1	0.808 0.770	0.782 0.753	0.795 0.761	0.806 0.772	0.837 0.816	**0.851** **0.828**
A-II_{2013} (type-1)	AUC F1	0.596 0.348	0.689 0.643	0.683 0.639	0.695 0.615	0.678 0.613	**0.717** **0.669**
A-II_{2013} (type-2)	AUC F1	0.712 0.647	0.721 0.713	0.695 0.674	0.714 0.664	0.732 0.705	**0.767** **0.754**
A-II_{2012} (type-1)	AUC F1	0.586 0.318	0.671 0.615	0.672 0.612	0.676 0.573	0.655 0.560	**0.701** **0.642**
A-II_{2012} (type-2)	AUC F1	0.724 0.664	0.726 0.737	0.706 0.692	0.739 0.706	0.750 0.715	**0.775** **0.757**
R-$I_{5:5}$	AUC F1	0.634 0.445	0.623 0.551	0.651 0.586	0.661 0.542	0.683 0.665	**0.749** **0.735**
R-$I_{7:3}$	AUC F1	0.701 0.595	0.656 0.613	0.695 0.660	0.716 0.688	0.706 0.702	**0.787** **0.776**
R-$II_{5:5}$	AUC F1	0.678 0.541	0.655 0.582	0.685 0.593	0.677 0.565	0.712 0.659	**0.736** **0.701**
R-$II_{7:3}$	AUC F1	0.737 0.660	0.695 0.648	0.728 0.685	0.721 0.653	0.742 0.713	**0.772** **0.749**

Effectiveness Experiments

Recommendation

$Data_{split}$	Metric	MP2V [4]	ASNE [15]	SHNE [34]	GSAGE [7]	GAT [31]	HetGNN
A-I_{2003}	Rec	0.158	0.201	0.298	0.263	0.275	**0.319**
	Pre	0.044	0.060	0.081	0.077	0.079	**0.094**
	F1	0.069	0.092	0.127	0.120	0.123	**0.145**
A-I_{2002}	Rec	0.144	0.152	0.279	0.231	0.274	**0.293**
	Pre	0.046	0.050	0.086	0.073	0.087	**0.093**
	F1	0.070	0.075	0.134	0.112	0.132	**0.141**
A-II_{2013}	Rec	0.516	0.419	0.608	0.540	0.568	**0.625**
	Pre	0.207	0.174	0.241	0.219	0.230	**0.252**
	F1	0.295	0.333	0.345	0.312	0.327	**0.359**
A-II_{2012}	Rec	0.468	0.382	0.552	0.512	0.518	**0.606**
	Pre	0.204	0.171	0.233	0.224	0.227	**0.264**
	F1	0.284	0.236	0.327	0.312	0.316	**0.368**

Multi-label Classification

Task	Metric	MP2V [4]	ASNE [15]	SHNE [34]	GSAGE [7]	GAT [31]	HetGNN
MC (10%)	Macro-F1	0.972	0.965	0.939	**0.978**	0.962	**0.978**
	Micro-F1	0.973	0.967	0.940	0.978	0.963	**0.979**
MC (30%)	Macro-F1	0.975	0.969	0.939	0.979	0.965	**0.981**
	Micro-F1	0.975	0.970	0.941	0.980	0.965	**0.982**
NC	NMI	0.894	0.854	0.776	**0.914**	0.845	0.901
	ARI	0.933	0.898	0.813	**0.945**	0.882	0.932

Inductive Multi-label Classification

Task	Metric	GSAGE [7]	GAT [31]	HetGNN
IMC (10%)	Macro-F1	0.938	0.954	**0.962**
	Micro-F1	0.945	0.958	**0.965**
IMC (30%)	Macro-F1	0.949	0.956	**0.964**
	Micro-F1	0.955	0.960	**0.968**
INC	NMI	0.714	0.765	**0.840**
	ARI	0.764	0.803	**0.894**