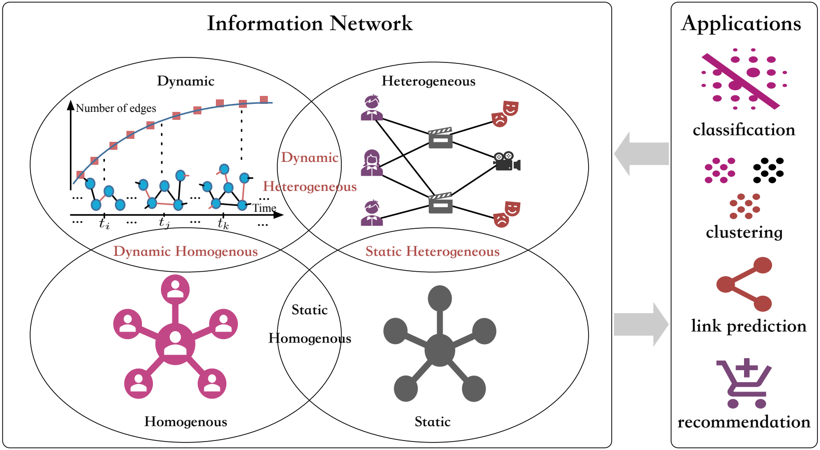
Representation Learning on Dynamic Heterogeneous Information Network

Yuanfu Lu

BUPT

June 19, 2020

Overview


Information Network
Dynamic
Number of edges
Time
t_i
t_j
t_k
Heterogeneous
Dynamic
Heterogeneous
Dynamic Homogenous
Static Heterogeneous
Static
Homogenous
Homogenous
Static
Applications
classification
clustering
link prediction
recommendation

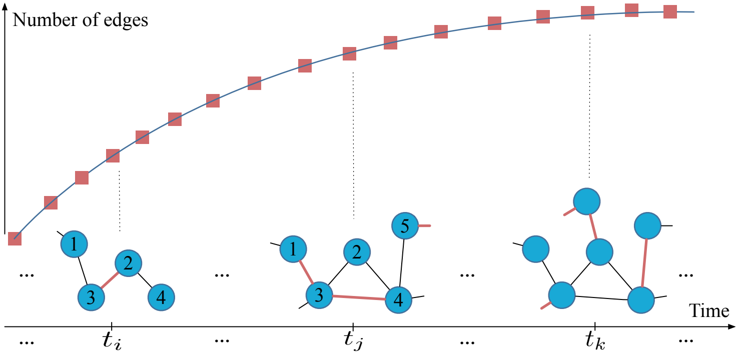
Outline

- Dynamic Homogeneous Information Network (CIKM)
- Static Heterogeneous Information Network (AAAI, TKDE)
- Dynamic Heterogeneous Information Network (TKDE)
- Application (KDD, ECML-PKDD)
- Conclusion

Outline

- Dynamic Homogeneous Information Network (CIKM)
- Static Heterogeneous Information Network (AAAI, TKDE)
- Dynamic Heterogeneous Information Network (TKDE)
- Application (KDD, ECML-PKDD)
- Conclusion

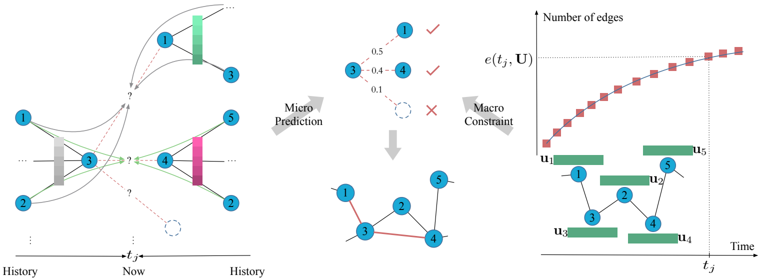
Dynamic Homogeneous Information Network



- Micro-dynamics
  - the formation process of network structures
- Macro-dynamics
  - the evolution pattern of network scale

M^{2}DNE

Temporal Network Embedding with Micro- and Macro-dynamics

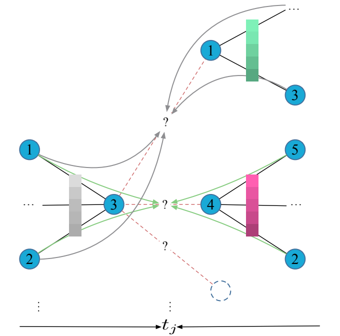


(a) Micro-dynamics preserved embedding (c) Mutual evolution of micro- and macro-dynamics (b) Macro-dynamics preserved embedding

- Micro-dynamics Preserved Embedding
  - a temporal attention point process
  - $\mathcal{L}_{mi} = -\sum_{t\in\mathcal{T}}\sum_{(i,j,t)\in\mathcal{E}} \log p\left(i,j|\mathcal{H}^i(t),\mathcal{H}^j(t)\right)$
- Macro-dynamics Preserved Embedding
  - a dynamics equation
  - $\mathcal{L}_{ma} = \sum_{t\in\mathcal{T}}\left(\Delta e(t) - \Delta e'(t)\right)^2$

M^{2}DNE—Micro-dynamics Preserved Embedding

A temporal attention point process

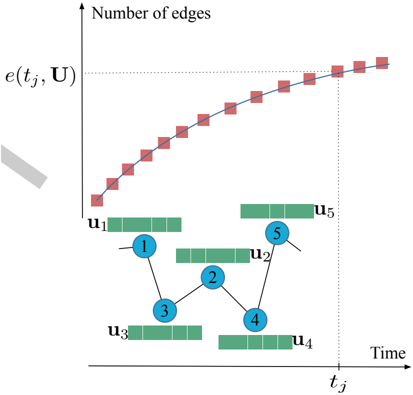


(a) Micro-dynamics preserved embedding

$$\tilde{\lambda}_{i,j}(t) = \underbrace{g(\mathbf{u}_i, \mathbf{u}_j)}_{Base\,Intensity} \underbrace{+ \beta_{ij} \sum_{p \in \mathcal{H}^i(t)} \alpha_{pi}(t) g(\mathbf{u}_p, \mathbf{u}_j) \kappa(t - t_p) + (1 - \beta_{ij}) \sum_{q \in \mathcal{H}^j(t)} \alpha_{qj}(t) g(\mathbf{u}_q, \mathbf{u}_i) \kappa(t - t_q),}_{Neighbor\,Influence}$$

M^2DNE—Macro-dynamics Preserved Embedding

A dynamics equation



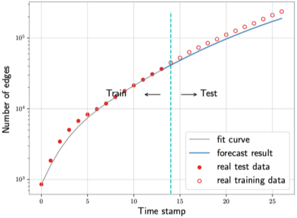
(b) Macro-dynamics preserved embedding

$$\Delta e'(t) = n(t)r(t)(\zeta(n(t)-1)^{\gamma}),$$

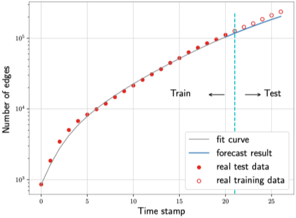
$$r(t) = \frac{\frac{1}{|\mathcal{E}|}\sum_{(i,j,t)\in\mathcal{E}} \sigma(-||\mathbf{u}_i - \mathbf{u}_j||_2^2)}{t^{\theta}},$$

M^2DNE—Experiments

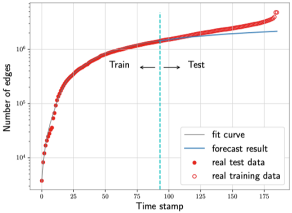
Trend forecast



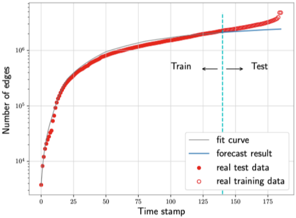
(c) $\frac{1}{2}|\mathcal{T}|$ on DBLP



(d) $\frac{3}{4}|\mathcal{T}|$ on DBLP



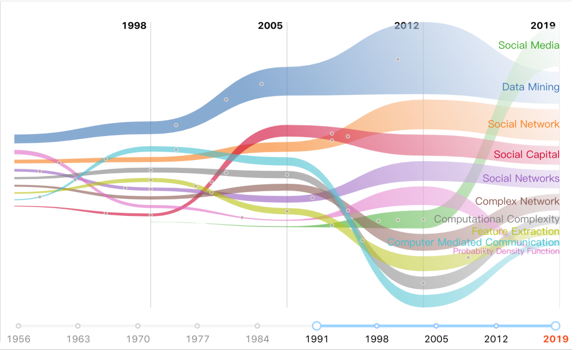
(e) $\frac{1}{2}|\mathcal{T}|$ on Tmall



(f) $\frac{3}{4}|\mathcal{T}|$ on Tmall

M^2DNE—TODO?

Next TODO?



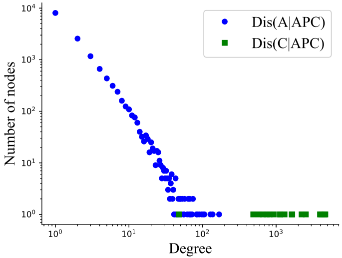
- sudden/bursty evolution
  - 2019-IJCAI-Evolving Graphs with Burst Detection
- heterogeneous point process
  - 2017-NeurIPS-The Neural Hawkes Process

Outline

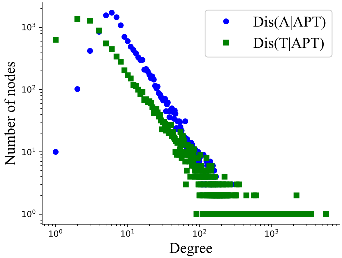
- Dynamic Homogeneous Information Network (CIKM)
- Static Heterogeneous Information Network (AAAI, TKDE)
- Dynamic Heterogeneous Information Network (TKDE)
- Application (KDD, ECML-PKDD)
- Conclusion

Static Heterogeneous Information Network

Data Observation



(a) APC



(b) APT

- two types of nodes connected via *APC* are unbalanced, meaning an inequivalent structure
- two types of nodes connected via *APT* indicate the similar and compatible structural roles

Static Heterogeneous Information Network

Quantitative Analysis

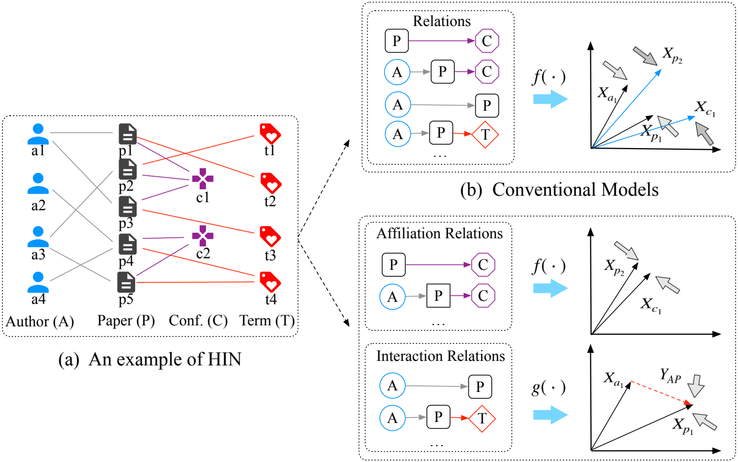
Datasets	Nodes	Number of Nodes	Relations ($t_u \sim t_v$)	Number of Relations	Avg. Degree of t_u	Avg. Degree of t_v	Measures $D(r)$	Measures $S(r)$	Relation Category
DBLP	Term (T) Paper (P) Author (A) Conference (C)	8,811 14,376 14,475 20	PC	14,376	1.0	718.8	718.8	0.05	AR
			APC	24,495	2.9	2089.7	720.6	0.085	AR
			AP	41,794	2.9	2.9	1.0	0.0002	IR
			PT	88,683	6.2	10.7	1.7	0.0007	IR
			APT	260,605	18.0	29.6	1.6	0.002	IR
Yelp	User (U) Service (S) Business (B) Star Level (L) Reservation (R)	1,286 2 2,614 9 2	BR	2,614	1.0	1307.0	1307.0	0.5	AR
			BS	2,614	1.0	1307.0	1307.0	0.5	AR
			BL	2,614	1.0	290.4	290.4	0.1	AR
			UB	30,838	23.9	11.8	2.0	0.009	IR
			BUB	528,332	405.3	405.3	1.0	0.07	IR
AMiner	Paper (P) Author (A) Reference (R) Conference (C)	127,623 164,472 147,251 101	PC	127,623	1.0	1263.6	1263.6	0.01	AR
			APC	232,659	2.2	3515.6	1598.0	0.01	AR
			AP	355,072	2.2	2.8	1.3	0.00002	IR
			PR	392,519	3.1	2.7	1.1	0.00002	IR
			APR	1,084,287	7.1	7.9	1.1	0.00004	IR
Amazon	Tag (T) Item (I) User (U) Brand (B)	22,140 8,493 15,619 22	IB	8,493	1.0	386.1	386.1	0.05	AR
			UIB	16,789	1.5	1067.9	711.9	0.05	AR
			UI	23,493	1.5	2.7	1.8	0.0002	IR
			IT	39,528	4.6	1.8	2.6	0.0002	IR
			UIT	117,618	7.8	5.5	1.4	0.0003	IR

- a degree-based measure $D(r) = \frac{\max[\bar{d}_{t_u}, \bar{d}_{t_v}]}{\min[\bar{d}_{t_u}, \bar{d}_{t_v}]}$
- a sparsity-based measure $S(r) = \frac{N_r}{N_{t_u} \times N_{t_v}}$

Affiliation Relations and Interaction Relations

- Affiliation Relations
  - ARs indicate one-centered-by-another structures, where the average degrees of the types of end nodes are extremely different. They imply an affiliation relationship between nodes.
- Interaction Relations
  - IRs describe peer-to-peer structures, where the average degrees of the types of end nodes are compatible. They suggest an interaction relationship between nodes.

Relation Structure-Aware HIN Embedding



(a) An example of HIN

(b) Conventional Models

(c) Our Model

RHINE—Different Models for ARs and IRs

- Euclidean Distance for Affiliation Relations
  - Nodes connected via ARs share similar properties[1,2]
  - Euclidean distance meets the triangle inequality[3]

$$f(p,q) = w_{pq} \left\| \mathbf{X}_p - \mathbf{X}_q \right\|_2^2 \tag{1}$$

- Translation-based Distance for Interaction Relations
  - Strong interaction relationships between compatible nodes
  - IRs themselves contain structural information of two nodes[4]

$$g(u,v) = w_{uv} \left\| \mathbf{X}_u + \mathbf{Y}_r - \mathbf{X}_v \right\| \tag{2}$$

[1] K. Faust, Centrality in affiliation networks, Social networks, vol.19, no.2, pp.157-191, 1997.
[2] J. Yang and J. Leskovec, Community-affiliation graph model for overlapping network community detection, in ICDM, 2012, pp.1170-1175.
[3] C.-K. Hsieh, L. Yang, Y. Cui, T.-Y. Lin, S. Belongie, and D. Estrin, Collaborative metric learning, in WWW, 2017, pp.193-201.
[4] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, Translating embeddings for modeling multirelational data, in NeurIPS, 2013, pp.2787-2795.

RHINE—A Unied Model for HIN Embedding

$$\mathcal{L}_{EuAR} = \sum_{s \in R_{AR}} \sum_{\langle p,s,q \rangle \in P_{AR}} \sum_{\langle p',s,q' \rangle \in P'_{AR}} \max[0, \gamma + f(p,q) - f(p',q')],$$

$$\mathcal{L}_{TrIR} = \sum_{r \in R_{IR}} \sum_{\langle u,r,v \rangle \in P_{IR}} \sum_{\langle u',r,v' \rangle \in P'_{IR}} \max[0, \gamma + g(u,v) - g(u',v')]$$

$$\mathcal{L} = \mathcal{L}_{EuAR} + \mathcal{L}_{TrIR}$$

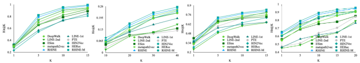
RHINE—Experiments

TABLE 3
Performance Evaluation of Link Prediction.

Methods	DBLP (A-A) AUC	DBLP (A-A) F1	DBLP (A-C) AUC	DBLP (A-C) F1	Yelp (U-B) AUC	Yelp (U-B) F1	AMiner (A-A) AUC	AMiner (A-A) F1	AMiner (A-C) AUC	AMiner (A-C) F1	Amazon (U-I) AUC	Amazon (U-I) F1
DeepWalk	0.9131	0.8246	0.7634	0.7047	0.8476	0.6397	0.9122	0.8471	0.7701	0.7112	0.9045	0.8978
LINE-1st	0.8264	0.7233	0.5335	0.6436	0.5084	0.4379	0.6665	0.6274	0.7574	0.6983	0.8810	0.8619
LINE-2nd	0.7448	0.6741	0.8340	0.7396	0.7509	0.6809	0.5808	0.4682	0.7899	0.7177	0.8879	0.8793
PTE	0.8853	0.8331	0.8843	0.7720	0.8061	0.7043	0.8119	0.7319	0.8442	0.7587	0.9014	0.8991
ESim	0.9077	0.8129	0.7736	0.6795	0.6160	0.4051	0.8970	0.8245	0.8089	0.7392	0.8992	0.8875
HIN2Vec	0.9160	0.8475	0.8966	0.7892	0.8653	0.7709	0.9141	0.8566	0.8099	0.7282	0.9012	0.8981
metapath2vec	0.9153	0.8431	0.8987	0.8012	0.7818	0.5391	0.9111	0.8530	0.8902	0.8125	0.9388	0.9145
HERec	0.9178	0.8523	0.8874	0.8132	0.7923	0.5742	0.9058	0.8514	0.8834	0.8067	0.9295	0.9017
JUST	0.9018	0.8103	0.8174	0.7532	0.7711	0.5932	0.8823	0.8141	0.8045	0.7603	0.8955	0.8912
RHINE	0.9315	0.8664	0.9148	0.8478	0.8762	0.7912	0.9316	0.8664	0.9173	0.8262	0.9561	0.9207
RHINE-M	**0.9421**	**0.8726**	**0.9218**	**0.8508**	**0.8796**	**0.7994**	**0.9460**	**0.8714**	**0.9207**	**0.8322**	**0.9634**	**0.9371**

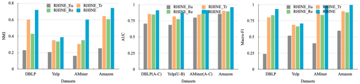
TABLE 4
Performance Evaluation of Multi-class Classification. Tr.Ra is the training ratio. Ma.-F1 and Mi.-F1 mean Macro-F1 and Micro-F1. D.W. L-1st, L-2nd, H2Vec and mp2vec represent DeepWalk, LINE-1st, LINE-2nd, HIN2Vec and metapath2vec, respectively.

Datasets	Metrics	Tr.Ra	D.W.	L-1st	L-2nd	PTE	ESim	H2Vec	mp2vec	HERec	JUST	RHINE	RHINE-M
DBLP	Ma.-F1	40%	0.6754	0.7237	0.7488	0.7723	0.8069	0.9073	0.9084	0.9005	0.8612	0.9261	**0.9301**
		60%	0.7283	0.7311	0.7338	0.7768	0.8161	0.8600	0.8941	0.8963	0.8733	0.9257	**0.9387**
		80%	0.7475	0.8091	0.7559	0.8852	0.8867	0.8631	0.8976	0.9101	0.8745	0.9344	**0.9445**
	Mi.-F1	40%	0.6937	0.7250	0.7583	0.7629	0.8000	0.9083	0.9098	0.9000	0.8567	0.9125	**0.9237**
		60%	0.7125	0.7500	0.7375	0.7875	0.8125	0.8625	0.8950	0.8915	0.8654	0.9051	**0.9254**
		80%	0.7500	0.8250	0.7500	0.8750	0.8750	0.8500	0.9000	0.9021	0.8705	0.9250	**0.9304**
Yelp	Ma.-F1	40%	0.6708	0.4607	0.5393	0.5397	0.6799	0.6104	0.5613	0.5509	0.5583	0.6909	**0.6998**
		60%	0.6717	0.4681	0.5299	0.5407	0.6830	0.6032	0.5478	0.5557	0.5632	0.7021	**0.7110**
		80%	0.6723	0.4872	0.5304	0.5389	0.6836	0.6075	0.5337	0.5517	0.5701	0.7132	**0.7205**
	Mi.-F1	40%	0.6732	0.6680	0.6637	0.7297	0.7247	0.7342	0.7074	0.7265	0.6002	0.7476	**0.7523**
		60%	0.6893	0.6537	0.6857	0.7323	0.7358	0.7189	0.7171	0.7304	0.6201	0.7562	**0.7598**
		80%	0.7012	0.6639	0.7377	0.7342	0.7399	0.7361	0.7208	0.7323	0.6295	0.7572	**0.7634**
AMiner	Ma.-F1	40%	0.9421	0.9473	0.9392	0.9649	0.9898	**0.9955**	0.9895	0.9901	0.9535	0.9798	0.9832
		60%	0.9423	0.9456	0.9467	0.9736	0.9904	**0.9961**	0.9912	0.9942	0.9684	09845	0.9893
		80%	0.9386	0.9494	0.9468	0.9791	0.9910	**0.9962**	0.9934	0.9956	0.9745	0.9884	0.9923
	Mi.-F1	40%	0.9525	0.9471	0.9397	0.9754	0.9936	**0.9958**	0.9901	0.9874	0.9684	0.9804	0.9848
		60%	0.9533	0.9527	0.9484	0.9813	0.9943	**0.9962**	0.9916	0.9953	0.9734	0.9845	0.9897
		80%	0.9512	0.9569	0.9491	0.9874	0.9948	**0.9965**	0.9936	0.9948	0.9844	0.9807	0.9912
Amazon	Ma.-F1	40%	0.9618	0.9596	0.9636	0.9657	0.9713	0.9875	0.9814	0.9822	0.9732	0.9958	**0.9969**
		60%	0.9645	0.9601	0.9623	0.9703	0.9759	0.9896	0.9871	0.9814	0.9766	0.9963	**0.9972**
		80%	0.9703	0.9652	0.9679	0.9721	0.9832	**0.9982**	0.9898	0.9853	0.9801	0.9967	0.9978
	Mi.-F1	40%	0.9634	0.9651	0.9658	0.9634	0.9689	0.9759	0.9866	0.9845	0.9699	0.9864	**0.9870**
		60%	0.9691	0.9672	0.9649	0.9678	0.9702	0.9819	0.9856	0.9878	0.9732	0.9961	**0.9892**
		80%	0.9734	0.9721	0.9714	0.9701	0.9734	**0.9974**	0.9872	0.9882	0.9764	0.9931	0.9939



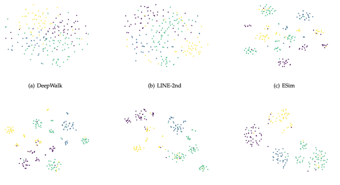
(a) DBLP (b) Yelp (c) Aminer (d) Amazon

Fig. 4. Performance Evaluation of Node Recommendation.



(a) Node Clustering (b) Link Precdition (c) Multi-Class Classification

Fig. 5. Performance Evaluation of Variant Models.



(a) DeepWalk (b) LINE-2nd (c) ESim

(d) HIN2Vec (e) metapath2vec (f) RHINE

Fig. 6. Visualization of Node Embeddings.

RHINE—TODO?

Next TODO?

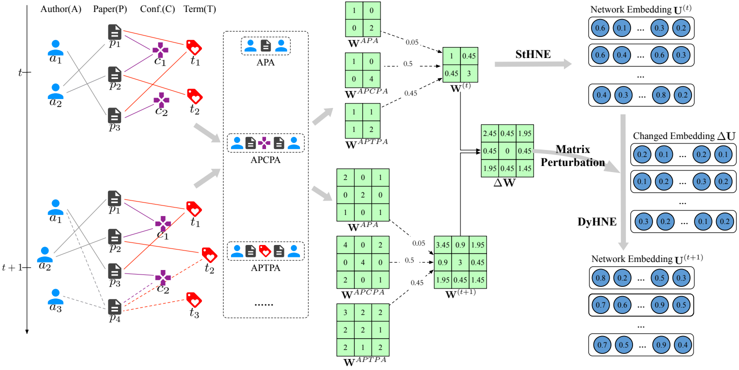
- Integrate measures and model design?
  - 2018-KDD-Easing Embedding Learning by Comprehensive Transcription of Heterogeneous Information Networks

Outline

- Dynamic Homogeneous Information Network (CIKM)
- Static Heterogeneous Information Network (AAAI, TKDE)
- Dynamic Heterogeneous Information Network (TKDE)
- Application (KDD, ECML-PKDD)
- Conclusion

Dynamic Heterogeneous Information Network

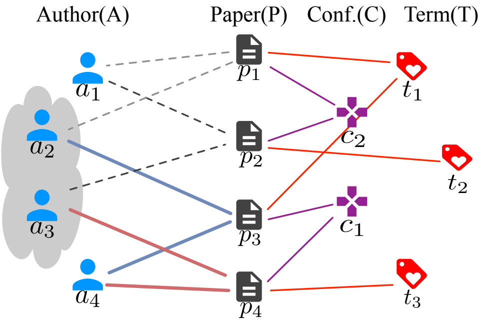
Dynamic HIN Embedding with Meta-path based Proximity



- effectively preserve structure and semantics in a dynamic HIN
- efficiently update node embeddings without retraining

StHNE

Preserve the meta-path based first- and second-order proximities to encode structure and semantics in HINs



- $p_1^m(v_i, v_j) = w_{ij}^m \left\| \mathbf{u}_i - \mathbf{u}_j \right\|_2^2$
- $p_2^m\left(v_p, \mathcal{N}(v_p)^m\right) = \left\| \mathbf{u}_p - \sum_{v_q \in \mathcal{N}(v_p)^m} w_{pq}^m \mathbf{u}_q \right\|_2^2$

Optimization with Spectral Theory

$$\mathcal{L}_1^m = \sum_{v_i, v_j \in \mathcal{V}} w_{ij}^m \|\mathbf{u}_i - \mathbf{u}_j\|_2^2 = 2\,\mathrm{tr}\left(\mathbf{U}^\top \mathbf{L}^m \mathbf{U}\right)$$

$$\mathcal{L}_2^m = \sum_{v_p \in \mathcal{V}} \|\mathbf{u}_p - \sum_{v_q \in \mathcal{N}(v_p)^m} w_{pq}^m \mathbf{u}_q\|_2^2 = 2tr\left(\mathbf{U}^\top \mathbf{H}^m \mathbf{U}\right)$$

$$\mathcal{L} = \mathrm{tr}\left(\mathbf{U}^\top (\mathbf{L} + \gamma \mathbf{H}) \mathbf{U}\right)$$

$$(\mathbf{L} + \gamma \mathbf{H})\mathbf{U} = \mathbf{D}\boldsymbol{\Lambda}\mathbf{U}$$

StHNE reduces to the generalized eigenvalue problem!

Matrix Perturbation in DyHNE

$$(\mathbf{L}+\Delta\mathbf{L}+\gamma\mathbf{H}+\gamma\Delta\mathbf{H})(\mathbf{U}+\Delta\mathbf{U})=(\mathbf{D}+\Delta\mathbf{D})(\mathbf{\Lambda}+\Delta\mathbf{\Lambda})(\mathbf{U}+\Delta\mathbf{U})$$

$$\Delta\lambda_i=\mathbf{u}_i^\top\Delta\mathbf{L}\mathbf{u}_i+\gamma\mathbf{u}_i^\top\Delta\mathbf{H}\mathbf{u}_i-\lambda\mathbf{u}_i^\top\Delta\mathbf{D}\mathbf{u}_i$$

$$\Delta\mathbf{u}_i=\sum_{j=2,j\neq i}^{d+1}\alpha_{ij}\mathbf{u}_j$$

$$\mathbf{\Lambda}^{(t+1)}=\mathbf{\Lambda}^{(t)}+\Delta\mathbf{\Lambda},\quad \mathbf{U}^{(t+1)}=\mathbf{U}^{(t)}+\Delta\mathbf{U}$$

DyHNE—Update

$$\Delta\lambda_i = \mathbf{u}_i^\top \Delta\mathbf{L}\mathbf{u}_i - \lambda_i \mathbf{u}_i^\top \Delta\mathbf{D}\mathbf{u}_i + \gamma\left\{\left[(\mathbf{W}-\mathbf{I})\mathbf{u}_i\right]^\top \Delta\mathbf{W}\mathbf{u}_i + \left(\Delta\mathbf{W}\mathbf{u}_i\right)^\top (\mathbf{W}-\mathbf{I})\mathbf{u}_i\right\}$$

$$\alpha_{ij} = \frac{\mathbf{u}_j^\top \Delta\mathbf{L}\mathbf{u}_i - \lambda_i \mathbf{u}_j^\top \Delta\mathbf{D}\mathbf{u}_i}{\lambda_i - \lambda_j} + \frac{\gamma\left\{\left[(\mathbf{W}-\mathbf{I})\mathbf{u}_j\right]^\top \Delta\mathbf{W}\mathbf{u}_i + \left(\Delta\mathbf{W}\mathbf{u}_j\right)^\top (\mathbf{W}-\mathbf{I})\mathbf{u}_i\right\}}{\lambda_i - \lambda_j}$$

Time complexity of $(\mathbf{W}-\mathbf{I})\mathbf{u}_i$: O(ed)

DyHNE—Acceleration

$$\Delta\lambda_i = \mathbf{C}(i,i) + \gamma\left[\mathbf{A}(:,i)^\top\mathbf{B}(:,i) + \mathbf{B}(:,i)^\top\mathbf{A}(:,i)\right]$$

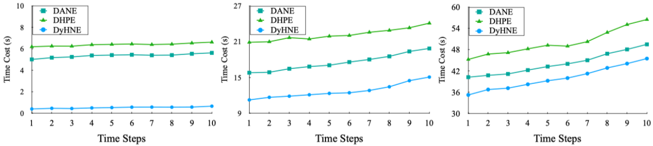
$$\alpha_{ij} = \frac{\mathbf{C}(j,i) + \gamma\left[\mathbf{A}(:,j)^\top\mathbf{B}(:,i) + \mathbf{B}(:,j)^\top\mathbf{A}(:,i)\right]}{\lambda_i - \lambda_j}$$

$$\mathbf{A}^{(t+1)}(:,i) = \sum_{j=2}^{d+1}\beta_{ij}\left(\mathbf{A}^t(:,j) + \mathbf{B}^t(:,j)\right) \tag{3}$$

Time complexity of $\mathbf{A}(:,i) = (\mathbf{W}-\mathbf{I})\mathbf{u}_i$: $\mathrm{O}(d^2)$

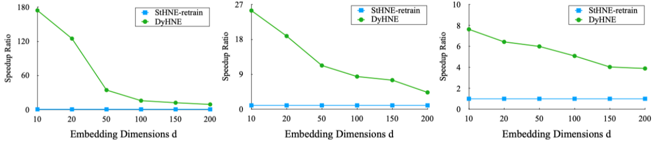
Overall, the time complexity of DyHNE is $\mathrm{O}(\mathrm{T}(\mathrm{f}+\mathrm{g}+N_M)d^2)$.

DyHNE—Experiments



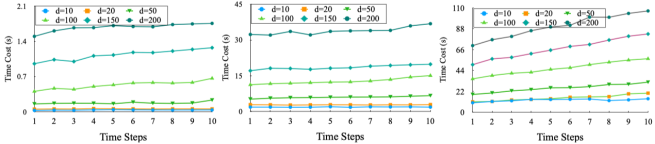
(a) Yelp (b) DBLP (c) AMiner

Fig. 4: Efficiency of the DyHNE compared to baselines.



(a) Yelp (b) DBLP (c) AMiner

Fig. 5: The speedup ratio of DyHNE.



(a) Yelp (b) DBLP (c) AMiner

Fig. 6: The running time w.r.t embedding dimensions.

DyHNE—TODO?

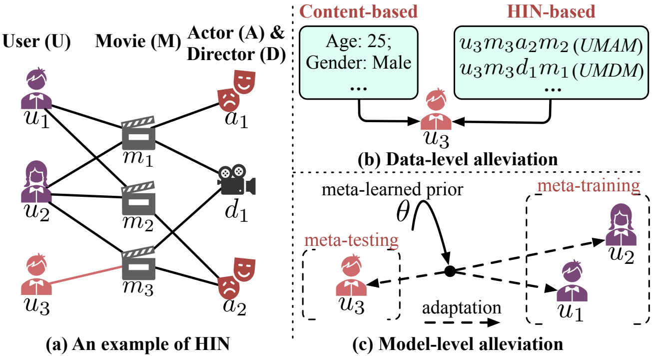
Next TODO?

- the dynamic evolution of heterogeneous information networks
  - different types of nodes have different evolution trends

Outline

- Dynamic Homogeneous Information Network (CIKM)
- Static Heterogeneous Information Network (AAAI, TKDE)
- Dynamic Heterogeneous Information Network (TKDE)
- Application (KDD, ECML-PKDD)
- Conclusion

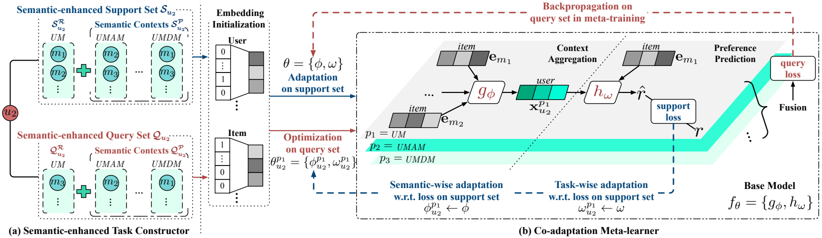
Application—Cold-start Rec.



- Data-level alleviation
- Model-level alleviation

MetaHIN

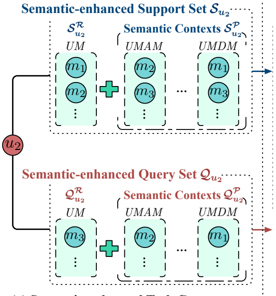
Meta-learning on HINs for Cold-start Recommendation



- capture the semantics on HINs in the meta-learning setting
- learn the prior knowledge that can be generalized to work with multifaceted heterogeneous semantics

MetaHIN—Semantic-enhanced Task Constructor

Each task is to learn the preferences of one user.



(a) Semantic-enhanced Task Constructor

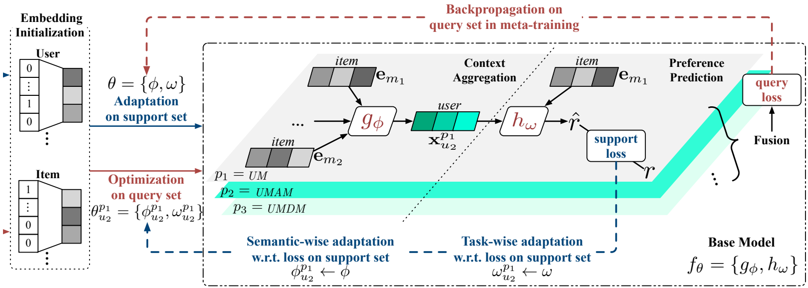
$$\mathcal{S}_u = (\mathcal{S}_u^{\mathcal{R}}, \mathcal{S}_u^{\mathcal{P}}),$$

$$\mathcal{S}_u^p = \bigcup_{i \in \mathcal{S}_u^{\mathcal{R}}} C_{u,i}^p.$$

$$C_{u,i}^p = \{j : j \in \text{items reachable along } p \text{ starting from } u\text{–}i\}.$$

- support set: calculate loss on the support set for the adaptation of the global prior
- query set: backward propagated w.r.t. the loss on the query set and evaluate performance

MetaHIN—Co-adaptation Meta-learner



(b) Co-adaptation Meta-learner

- Base Model $f_\theta = (g_\phi, h_\omega)$ parameterized by $\theta = \{\phi, \omega\}$
- Co-adaptation
  - semantic-wise adaptation

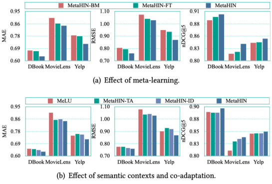
$$\phi_u^p = \phi - \alpha \frac{\partial \mathcal{L}_{\mathcal{T}_u}\left(\omega, \mathbf{x}_u^p, \mathcal{S}_u^{\mathcal{R}}\right)}{\partial \phi} = \phi - \alpha \frac{\partial \mathcal{L}_{\mathcal{T}_u}\left(\omega, \mathbf{x}_u^p, \mathcal{S}_u^{\mathcal{R}}\right)}{\partial \mathbf{x}_u^p} \frac{\partial \mathbf{x}_u^p}{\partial \phi}$$

  - task-wise adaptation

$$\omega_u^p = \omega^p - \beta \frac{\partial \mathcal{L}_{\mathcal{T}_u}\left(\omega^p, \mathbf{x}_u^{p\langle S\rangle}, \mathbf{S}_u^{\mathcal{R}}\right)}{\partial \omega^p}$$

MetaHIN—Experiments

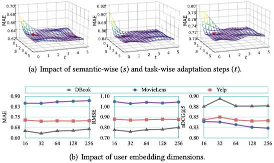
Scenario	Model	DBook MAE ↓	DBook RMSE ↓	DBook nDCG@5 ↑	MovieLens MAE ↓	MovieLens RMSE ↓	MovieLens nDCG@5 ↑	Yelp MAE ↓	Yelp RMSE↓	Yelp nDCG@5 ↑
Existing items for new users (User Cold-start or UC)	FM	0.7027	0.9158	0.8032	1.0421	1.3236	0.7303	0.9581	1.2177	0.8075
	NeuMF	0.6541	0.8058	0.8225	0.8569	1.0508	0.7708	0.9413	1.1546	0.7689
	GC-MC	0.9061	0.9767	0.7821	1.1513	1.3742	0.7213	0.9321	1.1104	0.8034
	mp2vec	0.6669	0.8391	0.8144	0.8793	1.0968	0.8233	0.8972	1.1613	0.8235
	HERec	0.6518	0.8192	0.8233	0.8691	0.9916	0.8389	0.8894	1.0998	0.8265
	DropoutNet	0.8311	0.9016	0.8114	0.9291	1.1721	0.7705	0.8557	1.0369	0.7959
	MeteEmb	0.6782	0.8553	0.8527	0.8261	1.0308	0.7795	0.8988	1.0496	0.7875
	MeLU	0.6353	0.7733	0.8793	0.8104	0.9756	0.8415	0.8341	1.0017	0.8275
	MetaHIN	**0.6019**	**0.7261**	**0.8893**	**0.7869**	**0.9593**	**0.8492**	**0.7915**	**0.9445**	**0.8385**
New items for existing users (Item Cold-start or IC)	FM	0.7186	0.9211	0.8342	1.3488	1.8503	0.7218	0.8293	1.1032	0.8122
	NeuMF	0.7063	0.8188	0.7396	0.9822	1.2042	0.6063	0.9273	1.1009	0.7722
	GC-MC	0.9081	0.9702	0.7634	1.0433	1.2753	0.7062	0.8998	1.1043	0.8023
	mp2vec	0.7371	0.9294	0.8231	1.0615	1.3004	0.6367	0.7979	1.0304	0.8337
	HERec	0.7481	0.9412	0.7827	0.9959	1.1782	0.7312	0.8107	1.0476	0.8291
	DropoutNet	0.7122	0.8021	0.8229	0.9604	1.1755	0.7547	0.8116	1.0301	0.7943
	MeteEmb	0.6741	0.7993	0.8537	0.9084	1.0874	0.8133	0.8055	0.9407	0.8092
	MeLU	0.6518	0.7738	0.8882	0.9196	1.0941	0.8041	0.7567	0.9169	0.8451
	MetaHIN	**0.6252**	**0.7469**	**0.8902**	**0.8675**	**1.0462**	**0.8341**	**0.7174**	**0.8696**	**0.8551**
New items for new users (User-Item Cold-start or UIC)	FM	0.8326	0.9587	0.8201	1.3001	1.7351	0.7015	0.8363	1.1176	0.8278
	NeuMF	0.6949	0.8217	0.8566	0.9686	1.2832	0.8063	0.9860	1.1402	0.7836
	GC-MC	0.7813	0.8908	0.8003	1.0295	1.2635	0.7302	0.8894	1.1109	0.7923
	mp2vec	0.7987	1.0135	0.8527	1.0548	1.2895	0.6687	0.8381	1.0993	0.8137
	HERec	0.7859	0.9813	0.8545	0.9974	1.1012	0.7389	0.8274	0.9887	0.8034
	DropoutNet	0.8316	0.8489	0.8012	0.9635	1.1791	0.7617	0.8225	0.9736	0.8059
	MeteEmb	0.7733	0.9901	0.8541	0.9122	1.1088	0.8087	0.8285	0.9476	0.8188
	MeLU	0.6517	0.7752	0.8891	0.9091	1.0792	0.8106	0.7358	0.8921	0.8452
	MetaHIN	**0.6318**	**0.7589**	**0.8934**	**0.8586**	**1.0286**	**0.8374**	**0.7195**	**0.8695**	**0.8521**
Existing items for existing users (Non-cold-start)	FM	0.7358	0.9763	0.8086	1.0043	1.1628	0.6493	0.8642	1.0655	0.7986
	NeuMF	0.6904	0.8373	0.7924	0.9249	1.1388	0.7335	0.7611	0.9731	0.8069
	GC-MC	0.8056	0.9249	0.8032	0.9863	1.2238	0.7147	0.8518	1.0327	0.8023
	mp2vec	0.6897	0.8471	0.8342	0.8788	1.1006	0.7091	0.7924	1.0191	0.8005
	HERec	0.6794	0.8409	0.8411	0.8652	1.0007	0.7182	0.7911	0.9897	0.8101
	DropoutNet	0.7108	0.7991	0.8268	0.9595	1.1731	0.7231	0.8219	1.0333	0.7394
	MeteEmb	0.7095	0.8218	0.7967	0.8086	1.0149	0.8077	0.7677	0.9789	0.7740
	MeLU	0.6519	0.7834	0.8697	0.8084	0.9978	0.8433	0.7382	0.9028	0.8356
	MetaHIN	**0.6393**	**0.7704**	**0.8859**	**0.7997**	**0.9491**	**0.8499**	**0.6952**	**0.8445**	**0.8477**



(a) Effect of meta-learning.

(b) Effect of semantic contexts and co-adaptation.

Figure 5: Analysis of MetaHIN using various ablated models.

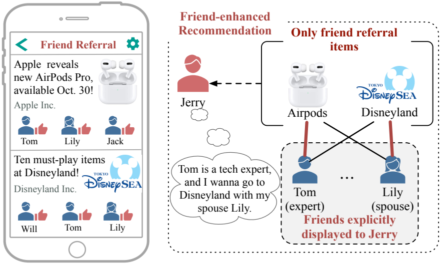


(a) Impact of semantic-wise (s) and task-wise adaptation steps (t).

(b) Impact of user embedding dimensions.

Figure 6: Analysis of parameters in MetaHIN.

Application—Friend-enhanced Rec.



Interests ?
Expert ?
Friends ?

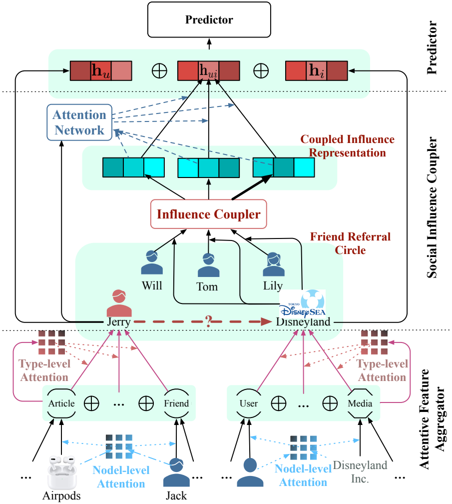
- only recommends to the user what his/her friends have interacted with
  - friends are high-quality information filters to provide more high-quality items
- all friends who have interacted with the item are explicitly displayed to the user attached to the recommended item
  - explicit social factors and the interpretability for user behaviors.

Friend-Enhanced Recommendation

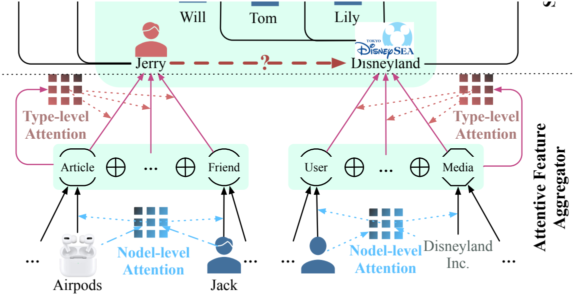
- Friend Referral Circle (FRC)
  Given an HSG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we define the friend referral circle of a user u w.r.t. a non-interacting item i (i.e., $\langle u, i \rangle \notin \mathcal{E}_R$) as $\mathcal{C}_u(i) = \{v | \langle u, v \rangle \in \mathcal{E}_F \cap \langle v, i \rangle \in \mathcal{E}_R\}$. Here v is called an influential friend of user u.
- Friend-Enhanced Recommendation (FER)
  Given an HSG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and the FRC $\mathcal{C}_u(i)$ of a user u w.r.t. a non-interacting item i, the FER aims to predict whether user u has a potential preference to item i. That is, a prediction function $\hat{y}_{ui} = \mathcal{F}(\mathcal{G}, \mathcal{C}_u(i); \Theta)$ is to be learned, where $\hat{y}_{ui}$ is the probability that user u will interact with item i, and Θ is the model parameters.

SIAN

Social Influence Attentive Neural Network for Friend-Enhanced Recommendation



SIAN—Attentive Feature Aggregator



- Node-level Attentive Aggregation

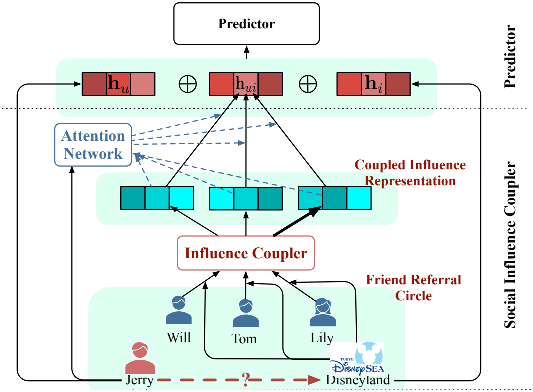
$$\mathbf{p}_u^t = \mathrm{ReLU}\left(\mathbf{W}_p\left(\sum_{k\in\mathcal{N}_u^t}\alpha_{ku}\mathbf{x}_k\right) + \mathbf{b}_p\right)$$

- Type-level Attentive Aggregation

$$\mathbf{h}_u = \mathrm{ReLU}\left(\mathbf{W}_h \sum_{t\in\mathcal{T}}\beta_{tu}\mathbf{p}_u^t + \mathbf{b}_h\right)$$

$$\beta_{tu} = \frac{\exp\left(\mathbf{a}_t^\top\left[\mathbf{p}_u^{t_1}\oplus\mathbf{p}_u^{t_2}\oplus\cdots\oplus\mathbf{p}_u^{t_1}{}^{T_1}\right]\right)}{\sum_{t'\in\mathcal{T}}\exp\left(\mathbf{a}_{t'}^\top\left[\mathbf{p}_u^{t_1}\oplus\mathbf{p}_u^{t_2}\oplus\cdots\oplus\mathbf{p}_u^{t_1}{}^{T_1}\right]\right)}$$

SIAN—Social Influence Coupler



- Coupled Inuence Representation.

  $\mathbf{c}_{\langle v,i \rangle} = \sigma \left(\mathbf{W}_c \phi \left(\mathbf{h}_v, \mathbf{h}_i \right) + \mathbf{b}_c \right)$

- Attentive Inuence Degree.

  $d'_{u \leftarrow \langle v,i \rangle} = \sigma \left(\mathbf{W}_2 \left(\sigma \left(\mathbf{W}_1 \phi \left(\mathbf{c}_{v,i}, \mathbf{h}_u \right) + \mathbf{b}_1 \right) \right) + b_2 \right)$

  $\mathbf{h}_{ui} = \sum_{v \in \mathcal{C}_u(i)} d_{u \leftarrow \langle v,i \rangle} \mathbf{c}_{\langle v,i \rangle}$

SIAN—Behavior Prediction and Model Learning

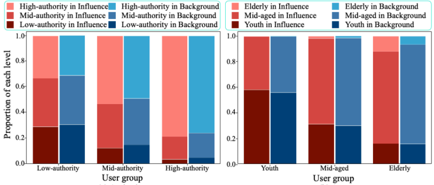
$$\mathbf{h}_o = \sigma\left(\mathbf{W}_{o_2}\left(\sigma\left(\mathbf{W}_{o_1}\left([\mathbf{h}_u \oplus \mathbf{h}_{ui} \oplus \mathbf{h}_i]\right) + \mathbf{b}_{o_1}\right) + \mathbf{b}_{o_2}\right)\right)$$

$$\hat{y}_{ui} = \text{sigmoid}\left(\mathbf{w}_y^\top \mathbf{h}_o + b_y\right)$$

$$-\sum_{\langle u,i\rangle \in \mathcal{E}_R} \left(y_{ui} \log \hat{y}_{ui} + (1 - y_{ui}) \log (1 - \hat{y}_{ui})\right) + \lambda \|\Theta\|_2^2$$

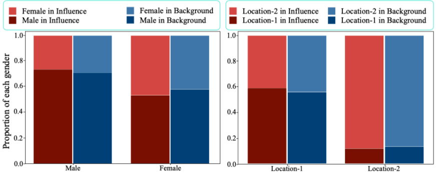
SIAN—Experiments

Dataset	Model	AUC		F1		Accuracy	
		d=32	d=64	d=32	d=64	d=32	d=64
Yelp	MLP	0.6704	0.6876	0.6001	0.6209	0.6589	0.6795
	DeepWalk	0.7693	0.7964	0.6024	0.6393	0.7001	0.7264
	node2vec	0.7903	0.8026	0.6287	0.6531	0.7102	0.7342
	metapath2vec	0.8194	0.8346	0.6309	0.6539	0.7076	0.7399
	DeepWalk+fea	0.7899	0.8067	0.6096	0.6391	0.7493	0.7629
	node2vec+fea	0.8011	0.8116	0.6634	0.6871	0.7215	0.7442
	metapath2vec+fea	0.8301	0.8427	0.6621	0.6804	0.7611	0.7856
	GCN	0.8022	0.8251	0.6779	0.6922	0.7602	0.7882
	GAT	0.8076	0.8456	0.6735	0.6945	0.7783	0.7934
	HAN	0.8218	0.8476	0.7003	0.7312	0.7893	0.8102
	TrustMF	0.8183	0.8301	0.6823	0.7093	0.7931	0.8027
	DiffNet	0.8793	0.8929	0.8724	0.8923	0.8698	0.8905
	SIAN	**0.9486***	**0.9571***	**0.8976***	**0.9128***	**0.9096***	**0.9295***
Douban	MLP	0.7689	0.7945	0.7567	0.7732	0.7641	0.7894
	DeepWalk	0.8084	0.8301	0.7995	0.8054	0.8295	0.8464
	node2vec	0.8545	0.8623	0.8304	0.8416	0.8578	0.8594
	metapath2vec	0.8709	0.8901	0.8593	0.8648	0.8609	0.8783
	DeepWalk+fea	0.8535	0.8795	0.8347	0.8578	0.8548	0.8693
	node2vec+fea	0.8994	0.9045	0.8732	0.8958	0.8896	0.8935
	metapath2vec+fea	0.9248	0.9309	0.8998	0.9134	0.8975	0.9104
	GCN	0.9032	0.9098	0.8934	0.9123	0.9032	0.9112
	GAT	0.9214	0.9385	0.8987	0.9103	0.8998	0.9145
	HAN	0.9321	0.9523	0.9096	0.9221	0.9098	0.9205
	TrustMF	0.9034	0.9342	0.8798	0.9054	0.9002	0.9145
	DiffNet	0.9509	0.9634	0.9005	0.9259	0.9024	0.9301
	SIAN	**0.9742***	**0.9873***	**0.9139***	**0.9429***	**0.9171***	**0.9457***
FWD	MLP	0.5094	0.5182	0.1883	0.1932	0.2205	0.2302
	DeepWalk	0.5587	0.5636	0.2673	0.2781	0.1997	0.2056
	node2vec	0.5632	0.5712	0.2674	0.2715	0.2699	0.2767
	metapath2vec	0.5744	0.5834	0.2651	0.2724	0.4152	0.4244
	DeepWalk+fea	0.5301	0.5433	0.2689	0.2799	0.2377	0.2495
	node2vec+fea	0.5672	0.5715	0.2691	0.2744	0.3547	0.3603
	metapath2vec+fea	0.5685	0.5871	0.2511	0.2635	0.4698	0.4935
	GCN	0.5875	0.5986	0.2607	0.2789	0.4782	0.4853
	GAT	0.5944	0.6006	0.2867	0.2912	0.4812	0.4936
	HAN	0.5913	0.6025	0.2932	0.3011	0.4807	0.4937
	TrustMF	0.6001	0.6023	0.3013	0.3154	0.5298	0.5404
	DiffNet	0.6418	0.6594	0.3228	0.3379	0.6493	0.6576
	SIAN	**0.6845***	**0.6928***	**0.3517***	**0.3651***	**0.6933***	**0.7018***



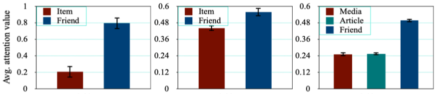
User group
(a) Authority

User group
(b) Age



User group
(C) Gender

User group
(d) Location



Neighbor Type
(a) Yelp

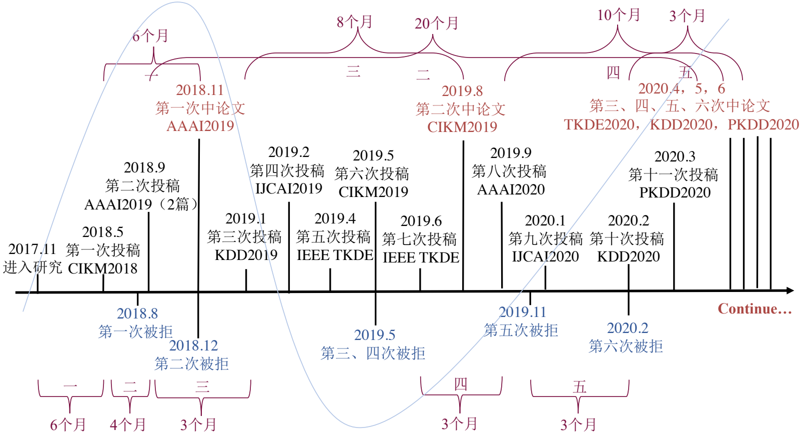
Neighbor Type
(b) Douban

Neighbor Type
(c) FWD

Outline

- Dynamic Homogeneous Information Network (CIKM)
- Static Heterogeneous Information Network (AAAI, TKDE)
- Dynamic Heterogeneous Information Network (TKDE)
- Application (KDD, ECML-PKDD)
- Conclusion

Conclusion



不抛弃，不放弃

Conclusion

- 认真思考问题
  - 不限于动脑
  - 不限于眼下
- 认真读/写论文
  - 笔记、总结、分类
  - 英语、积累、模仿
- 认真看/写代码
  - 看流程、写注释、理解即可
  - 想清楚、写注释、规范统一

References

- **Lu Yuanfu**, Fang Yuan, Shi Chuan. Meta-learning on Heterogeneous Information Networks for Cold-start Recommendation. In KDD, 2020.
- **Lu Yuanfu**, Shi Chuan, Hu Linmei, Liu Zhiyuan. Relation Structure-aware Heterogeneous Information Network Embedding. In AAAI, 2019.
- **Lu Yuanfu**, Wang Xiao, Shi Chuan, Philip S. Yu, Yanfang Ye. Temporal Network Embedding with Micro-and Macro-dynamics. In CIKM, 2019.
- **Lu Yuanfu**, Xie Ruobing, Shi Chuan, Yuan Fang, Wei Wang, Xu Zhang, and Leyu Lin. Social Influence Attentive Neural Network for Friend-Enhanced Recommendation. In ECML-PKDD, 2020.
- Shi Chuan, **Lu Yuanfu**, Hu Linmei, Liu Zhiyuan, Ma Huadong. RHINE: Relation Structure-Aware Heterogeneous Information Network Embedding. IEEE TKDE, 2020.
- Wang Xiao, **Lu Yuanfu**, Shi Chuan, Ruijia Wang, Peng Cui, Shuai Mou. Dynamic Heterogeneous Information Network Embedding with Meta-path based Proximity. IEEE TKDE, 2020.

Thank You

Q&A

http://shichuan.org
https://github.com/rootlu