ECML PKDD 2021

Dynamic Heterogeneous Graph Embedding via Heterogeneous Hawkes Process

Yugang Ji[1], Tianrui Jia[1], Yuan Fang[2], Chuan Shi[1]

[1]Beijing University of Posts and Telecommunications, Beijing, China

[2]Singapore Management University, Singapore

北京郵電大學 Beijing University of Posts and Telecommunications

SMU SINGAPORE MANAGEMENT UNIVERSITY

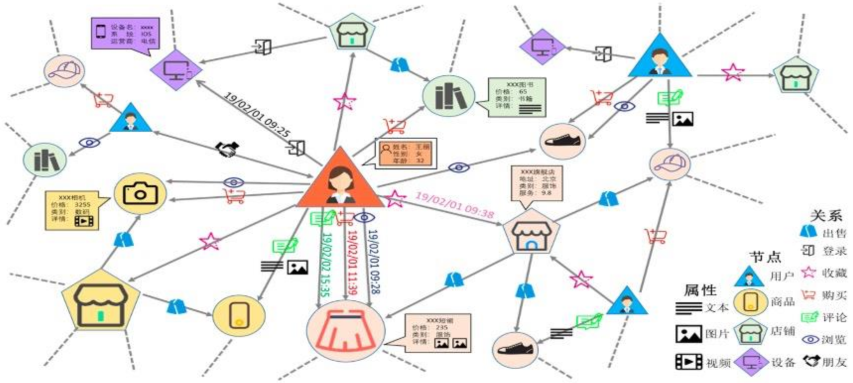
CONTENTS

CONTENTS

1 Motivation

- Graphs are universal in real-world scenarios
- Graph mining can improve users’ experiences

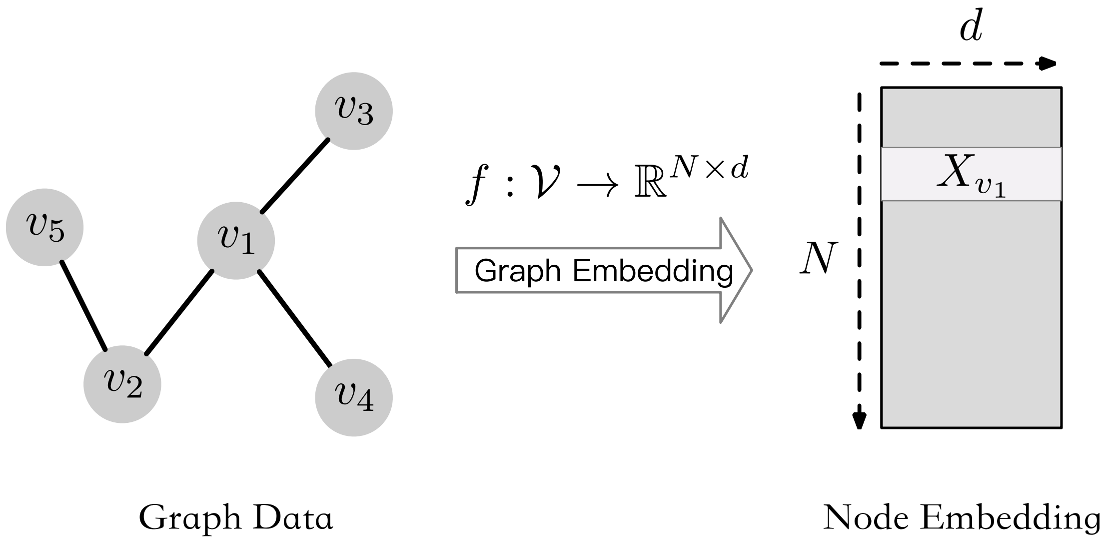
Facebook Social Network



Alibaba E-commerce Graph

Images from the Internet.

Graph Embedding



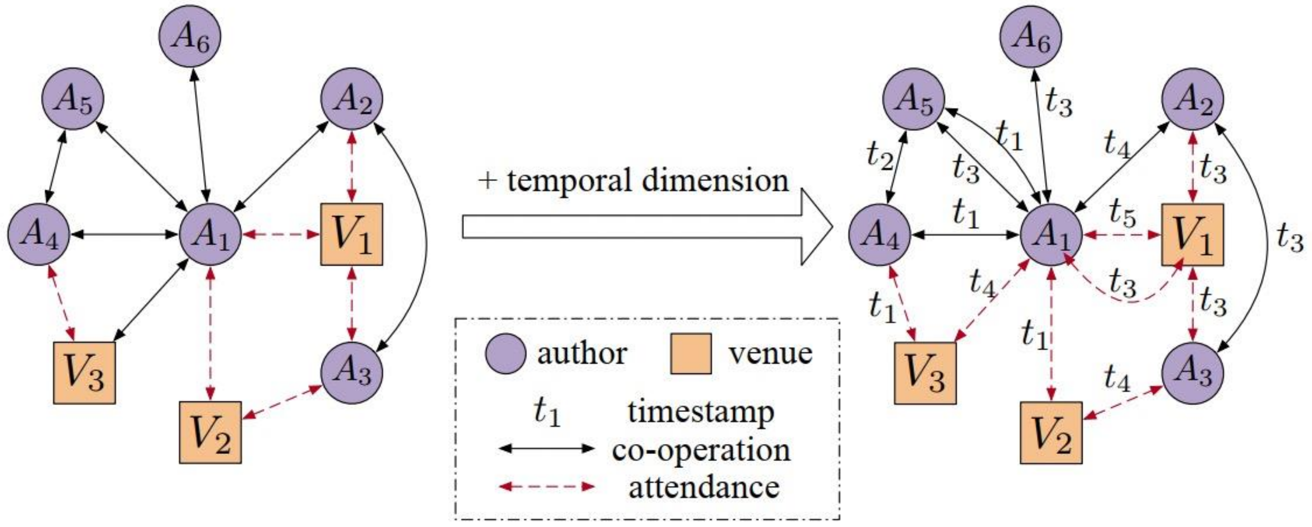
Drawbacks of existing models

- cannot preserve the heterogeneous semantics
- cannot preserve the dynamic evolutions

Dynamic Heterogeneous Graph (DHG)

- Multiple dynamic events
- historical events excite current interaction

we focus on the problem of dynamic heterogeneous graph embedding



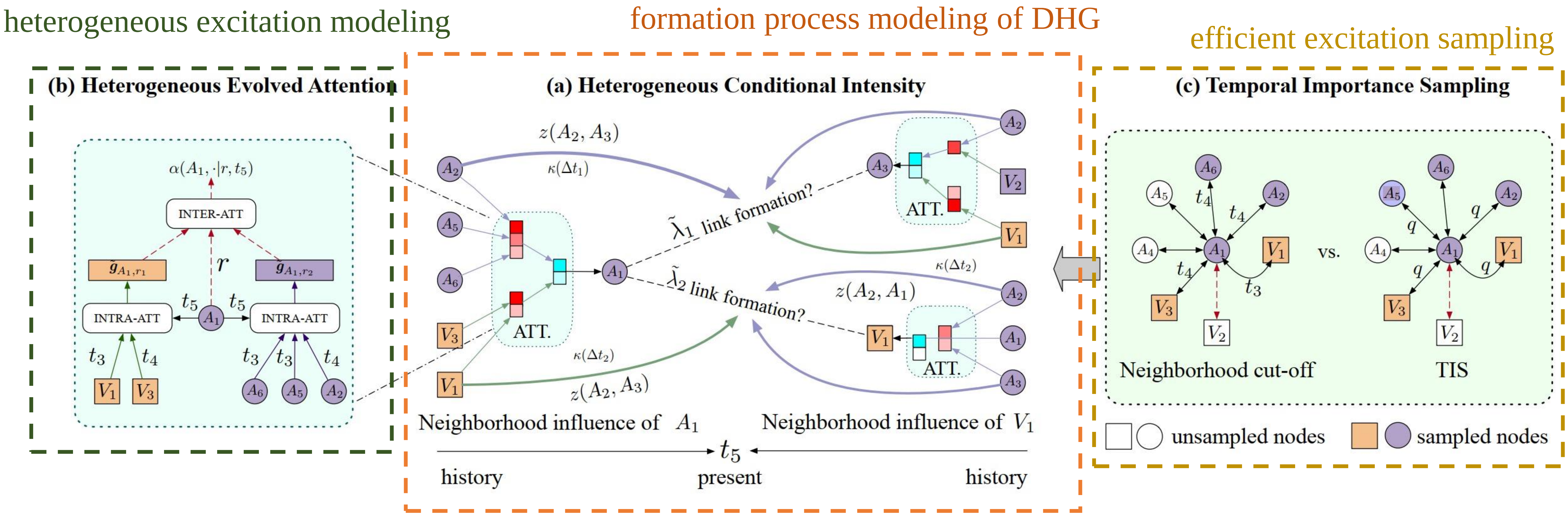
(a) static heterogeneous graph (b) dynamic heterogeneous graph

- *How to model the* ***continuous dynamics of heterogeneous interactions****?*
  - heterogeneous formation process

- *How to model the* ***complex influence of different semantics****?*
  - historical events are heterogeneous
  - current interactions are heterogeneous
  - excitations continuously decay

CONTENTS

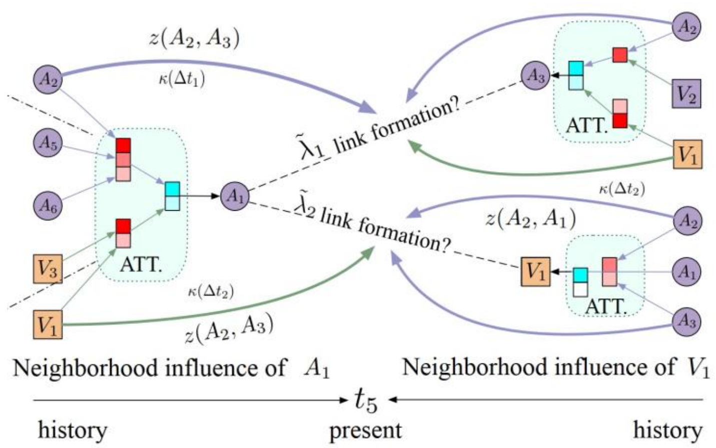
Overall framework



Overall framework of HPGE.

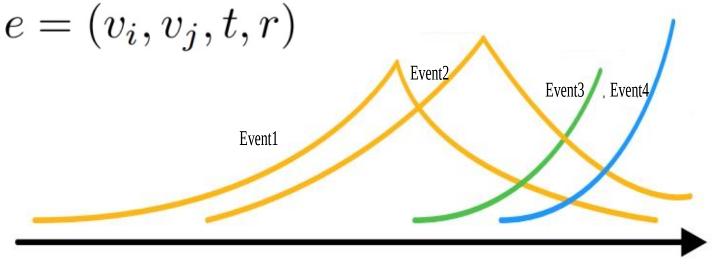
Heterogeneous Conditional Intensity

$$\mu_r(v_i, v_j) = -\sigma(f(\boldsymbol{h}_i \boldsymbol{W}_{\phi(v_i)} - \boldsymbol{h}_j \boldsymbol{W}_{\phi(v_j)})\boldsymbol{W}_r + b_r)$$



$$\tilde{\lambda}(e) = \underbrace{\mu_r(v_i, v_j)}_{\text{base rate}} + \gamma_1 \underbrace{\sum_{r' \in \mathcal{R}} \sum_{p \in \mathcal{N}_{i,r',<t}} \alpha(p,e) z(v_p, v_j) \kappa_i(t - t_p)}_{\text{neighborhood influence on source } v_i} + \gamma_2 \underbrace{\sum_{r'' \in \mathcal{R}} \sum_{q \in \mathcal{N}_{j,r'',<t}} \alpha(q,e) z(v_q, v_i) \kappa_j(t - t_q)}_{\text{neighborhood influence on target } v_j};$$

$e = (v_i, v_j, t, r)$



$$z(v_p, v_j) = -\|\boldsymbol{h}_p \boldsymbol{W}_{\phi(p)} - \boldsymbol{h}_j \boldsymbol{W}_{\phi(j)}\|_2^2;$$

Type-aware influence

Heterogeneous evolved attention mechanism

Event-level excitation via intra-attention

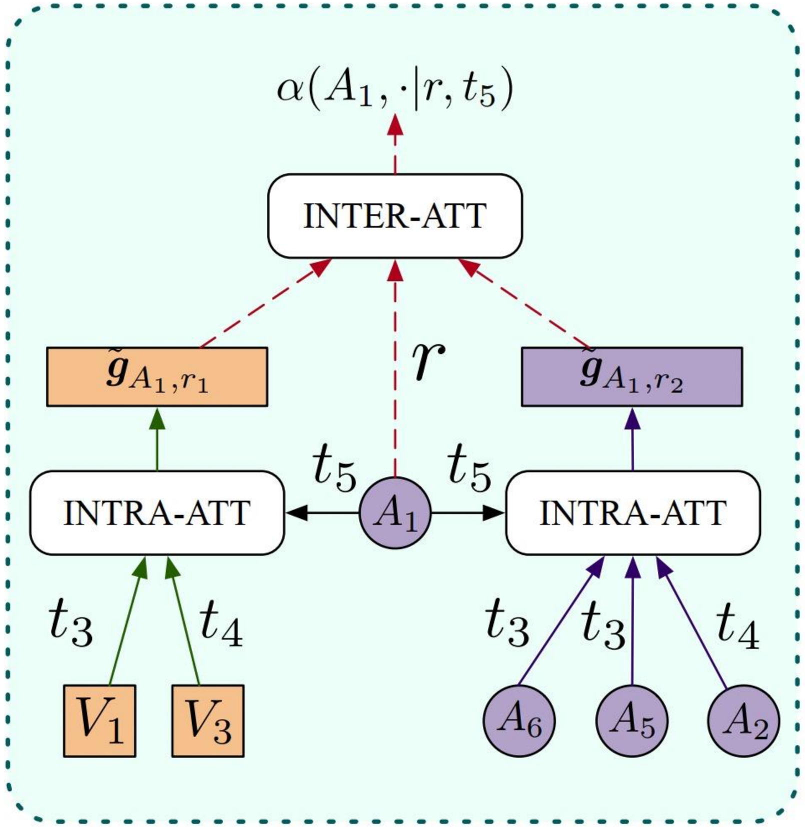
$$\xi(v_p, t_p | r', v_i, t) = \mathrm{softmax}(\sigma(\kappa_i(t - t_p)[\boldsymbol{h}_i \boldsymbol{W}_{\phi(v_i)} \oplus \boldsymbol{h}_j \boldsymbol{W}_{\phi(v_j)}]\boldsymbol{W}_\xi))$$

Semantic-level excitation via inter-attention

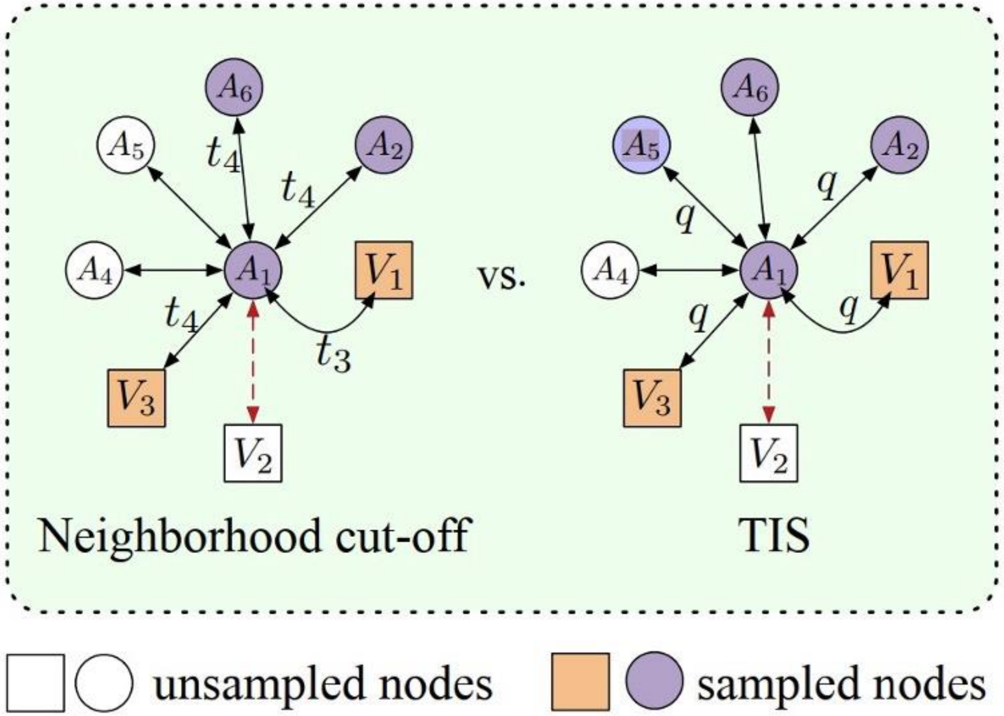
$$\beta(r | r', v_i, t) = \mathrm{softmax}(\tanh(\tilde{\boldsymbol{g}}_i \boldsymbol{W}_r)\boldsymbol{w}_r)^{\mathrm{T}},$$

Heterogeneous evolved attention

$$\alpha(p, e) = \xi(v_p, t_p | r', v_i, t)\beta(r | r', v_i, t)$$



Temporal Importance Sampling (TIS)



Samper

$$q(v_p|v_i,r',t)=\frac{\kappa_i(t-t_p)N_i(v_p)}{\sum_{v_{p'}\in N_{i,r',<t}}\kappa_i(t-t'_p)N_i(v'_p)},$$

Estimator

$$z(\hat{v}_p,v_j)=\frac{1}{n}\cdot\frac{z(\hat{v}_p,v_j)}{q(\hat{v}_p|v_i,r',t)},\quad \hat{v}_p\sim q(v_p|v_i,r',t)$$

Loss Function

$$\mathcal{L} = \mathcal{L}_{hp} + \omega_1 \mathcal{L}_{task} + \omega_2 \Omega(\Theta)$$

Regularization of parameters

$$\mathcal{L}_{hp}(e) = -\sum_{e \in \mathcal{E}} \log \sigma(\tilde{\lambda}(e)) - \sum_{k} \mathbb{E}_{j'} \log \sigma(-\tilde{\lambda}(e_{j'})) - \sum_{k} \mathbb{E}_{i'} \log \sigma(-\tilde{\lambda}(e_{i'}))$$

Heterogeneous Hawkes process loss

Link prediction
Node classification
……

CONTENTS

Datasets

Table 1: Statistics of the three public datasets.

Datasets	Node Types	#Nodes	Event Types	#Events	Time Span
Aminer	Author (A)	23,037	A-A	71,121	16 years
	Conference (C)	22	A-C	52,399	
DBLP	Author (A)	34,766	A-A	133,684	10 years
	Venue (V)	20	A-V	98,262	
Yelp	User (U)	494,524	BrU	1,145,070	60 quarters
	Business (B)	13,507	BtU	226,728	

Baselines

- M2V. & HEP & HAN & HGT
- CTDNE & E.GCN & M^2DNE
- DHNE & DyHNE & DyHATR

Tasks

- Effectiveness analysis
  - Node classification
  - Temporal Link Prediction
- Model analysis

Node classification

Table 2: Performance evaluation (with standard deviation) on node classification. The best performance is bolded and the second best is underlined.

Dataset	Aminer		DBLP		Yelp	
Metric	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
M2V	0.824(0.029)	0.853(0.032)	0.874(0.024)	0.885(0.029)	0.537(0.023)	0.642(0.017)
HEP	0.949(0.016)	0.952(0.013)	0.903(0.022)	0.913(0.018)	0.622(0.012)	0.694(0.009)
HAN	0.967(0.008)	0.970(0.009)	0.912(0.014)	0.914(0.007)	0.621(0.019)	0.691(0.025)
HGT	0.963(0.007)	0.971(0.011)	0.920(0.002)	0.927(0.001)	0.633(0.026)	0.705(0.022)
CTDNE	0.897(0.038)	0.895(0.025)	0.872(0.001)	0.892(0.005)	0.512(0.011)	0.639(0.011)
E.GCN	0.952(0.020)	0.955(0.018)	0.887(0.009)	0.881(0.010)	0.611(0.009)	0.687(0.008)
M2DNE	0.969(0.015)	0.972(0.018)	0.891(0.022)	0.909(0.027)	0.619(0.003)	0.693(0.005)
DHNE	0.901(0.010)	0.913(0.009)	0.888(0.007)	0.909(0.008)	0.578(0.001)	0.665(0.001)
DyHNE	0.970(0.008)	<u>0.978(0.007)</u>	0.922(0.003)	0.922(0.004)	0.622(0.011)	<u>0.721(0.015)</u>
DyHATR	<u>0.973(0.002)</u>	0.969(0.003)	<u>0.933(0.011)</u>	<u>0.935(0.010)</u>	<u>0.627(0.008)</u>	0.717(0.007)
HPGE	**0.988(0.002)**	**0.984(0.003)**	**0.951(0.005)**	**0.952(0.004)**	**0.649(0.010)**	**0.731(0.012)**

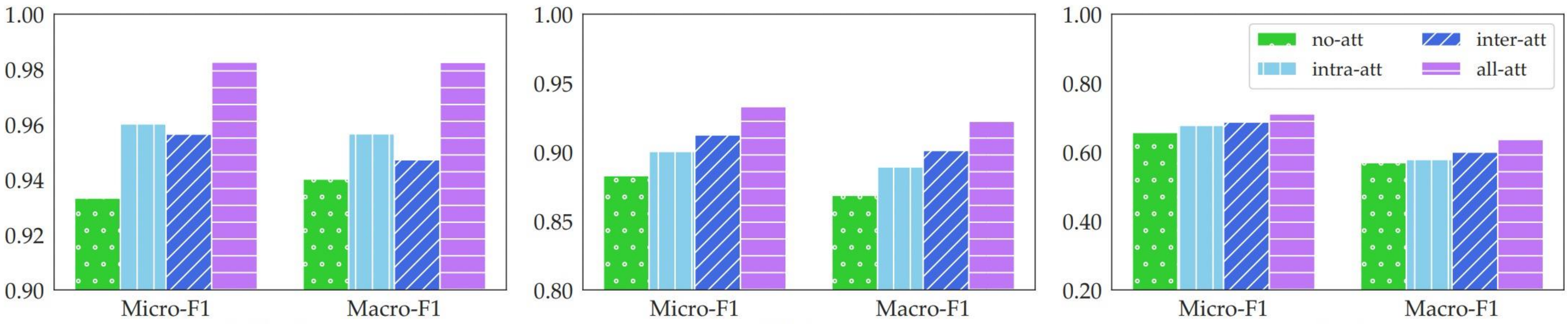
Temporal link prediction

Table 3: Performance evaluation on temporal link prediction. The best performance is bolded and the second best is underlined.

Dataset	Aminer			Yelp			DBLP		
Metric	ACC	F1	AUC	ACC	F1	AUC	ACC	F1	AUC
M2V	0.806	0.359	0.759	0.790	0.419	0.702	0.798	0.375	0.656
HEP	0.921	0.814	0.944	0.853	0.566	0.829	0.910	0.753	0.934
HAN	0.923	0.811	0.955	0.855	0.591	0.833	0.903	0.751	0.940
HGT	0.938	0.822	0.963	0.859	0.588	0.833	0.899	0.761	0.941
CTDNE	0.824	0.382	0.763	0.806	0.342	0.635	0.713	0.345	0.653
E.GCN	0.904	0.767	0.922	0.822	0.526	0.785	0.853	0.714	0.905
M2DNE	0.929	0.790	0.951	0.854	0.547	0.818	0.896	0.734	0.939
DHNE	0.875	0.634	0.827	0.831	0.504	0.717	0.821	0.668	0.808
DyHNE	0.928	**0.838**	0.959	0.861	0.592	0.831	0.909	0.767	0.940
DyHATR	<u>0.941</u>	0.832	<u>0.966</u>	<u>0.870</u>	<u>0.598</u>	<u>0.843</u>	<u>0.914</u>	<u>0.773</u>	0.936
HPGE	**0.953**	<u>0.835</u>	**0.976**	**0.873**	**0.603**	**0.850**	**0.938**	**0.793**	**0.957**

Effective attention mechanism

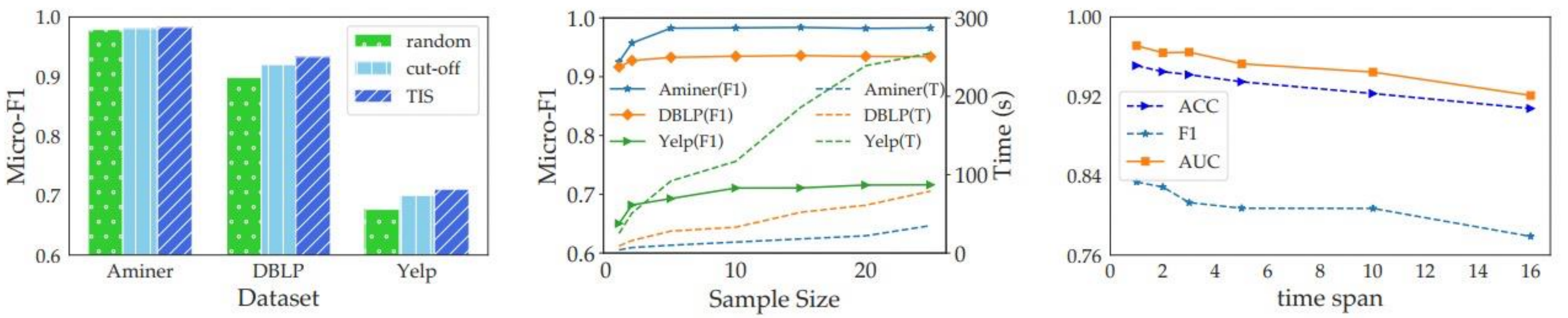
- Effectiveness of intra-attention of events: all-att vs. inter-att vs. no-att
- Effectiveness of inter-attention of semantics: all-att vs. intra-att vs. no-att



(a) Aminer (b) DBLP (C) Yelp

- Effective sampling strategy & Effective sample size

- Effective evolution modeling



(a) sampling strategies (b) effective sample size (c) varying the dynamics

CONTENTS

- We study the problem of dynamic heterogeneous graph embedding via heterogeneous Hawkes process .
- To make full use of dynamic and heterogeneous information, we propose the HPGE to model the formation process of temporal heterogeneous interactions by considering both event-level and semantic-level excitation to preserve all dynamics and semantics.
- Experimental results on three real-world datasets demonstrate the effectiveness of our proposed model.

ECML PKDD 2021

Thank you !

Q&A

北京郵電大學 Beijing University of Posts and Telecommunications

SMU SINGAPORE MANAGEMENT UNIVERSITY