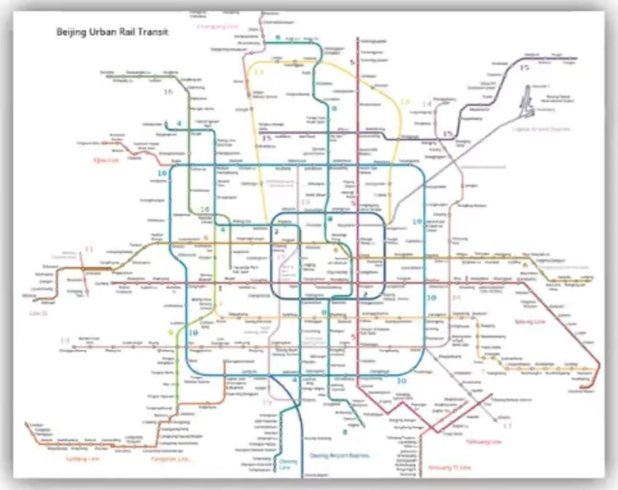
DyTed : Disentangled Representation Learning for Discrete-time Dynamic Graph

张文涛

Motivation

Motivtion

- Graphs 是我们日常生活中所面临各项任务的基础。

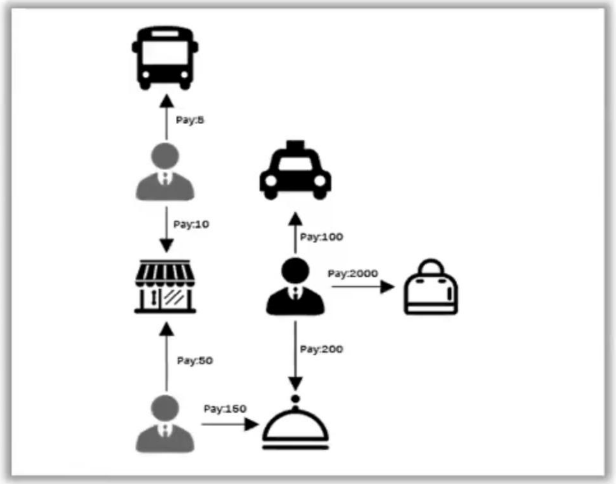


Traffic

1. Traffic Prediction
2. Traffic Planning
3. ……

Social network

1. Community Discovery
2. Social Recommendation
3. ……

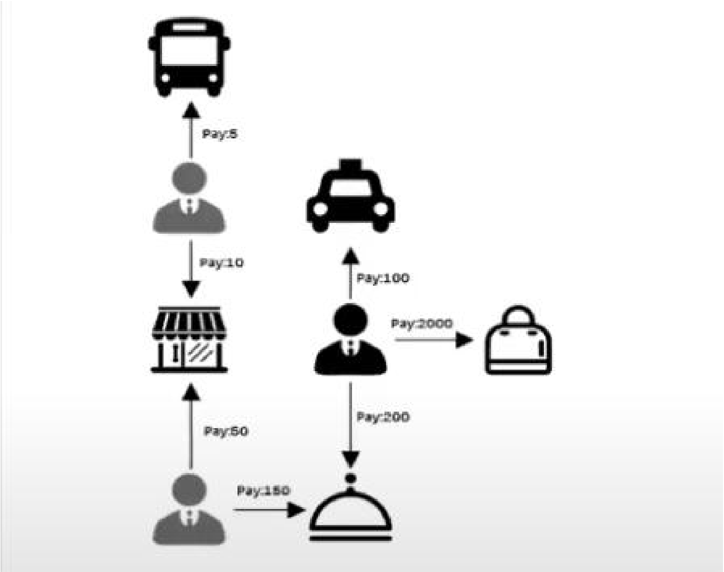


Transaction network

1. Transaction Rec
2. User Categorization
3. ……

Motivation

- 多种类型的任务需要多个端到端的训练，这将会消耗巨大的计算资源



Transaction network

End-to-end Training → **Task 1**: Transaction Rec

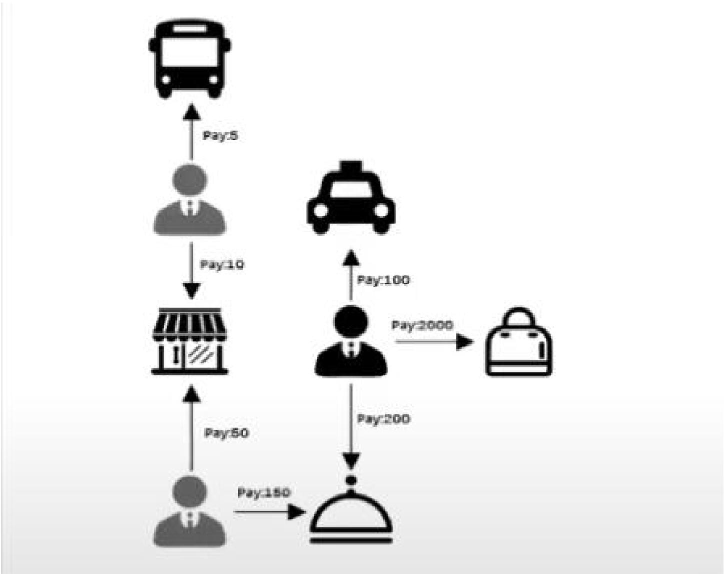
End-to-end Training → **Task 2**: User Categorization

End-to-end Training → **Task 3**: ……

Motivation

- 基于预训练的表征学习方法是应对该问题的一种解决方案

Fine-training



Transaction network

Pre-training

Representation r

Task 1: Transaction Rec

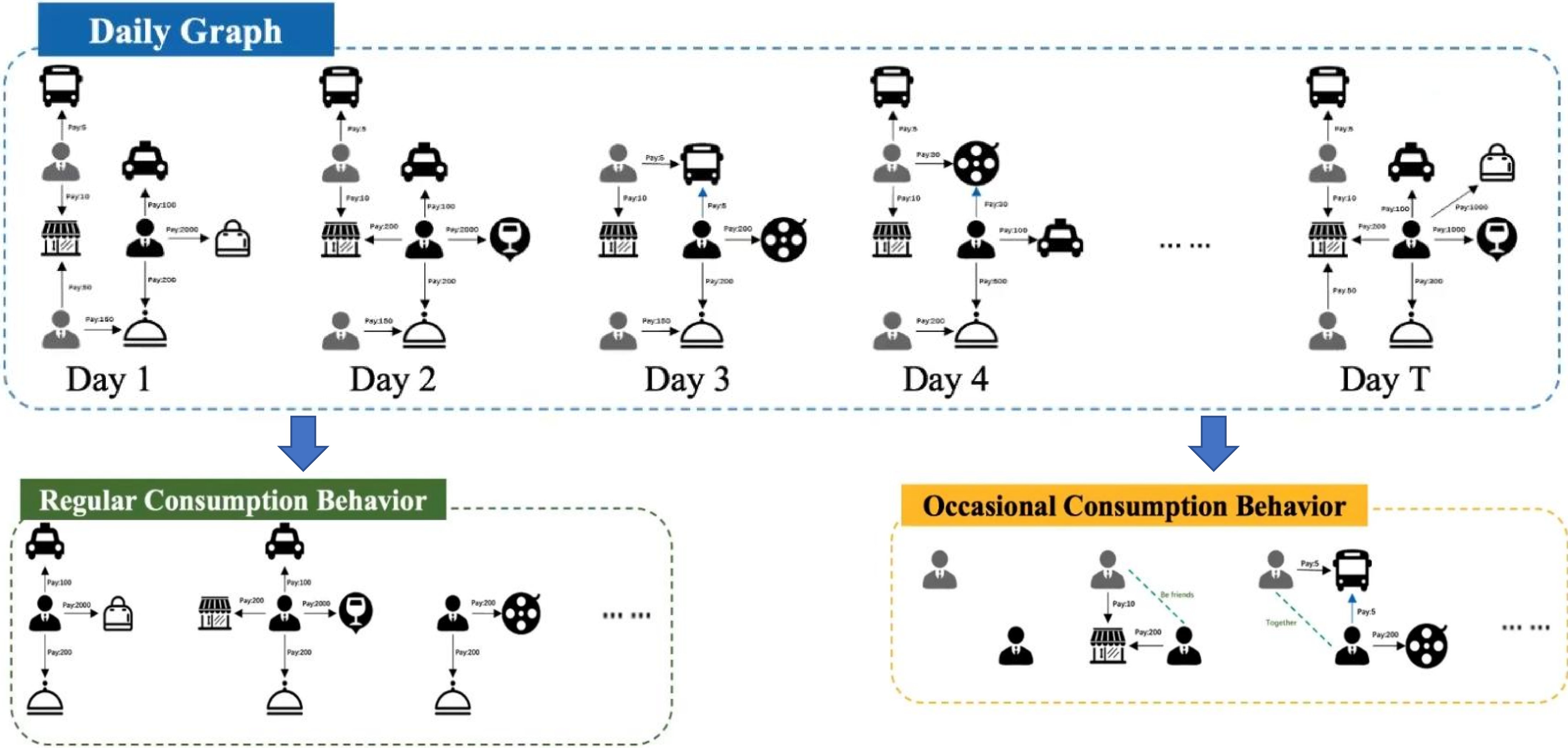
Task 2: User Categorization

Task 3: ……

表征学习可以显著的减少计算资源的需求

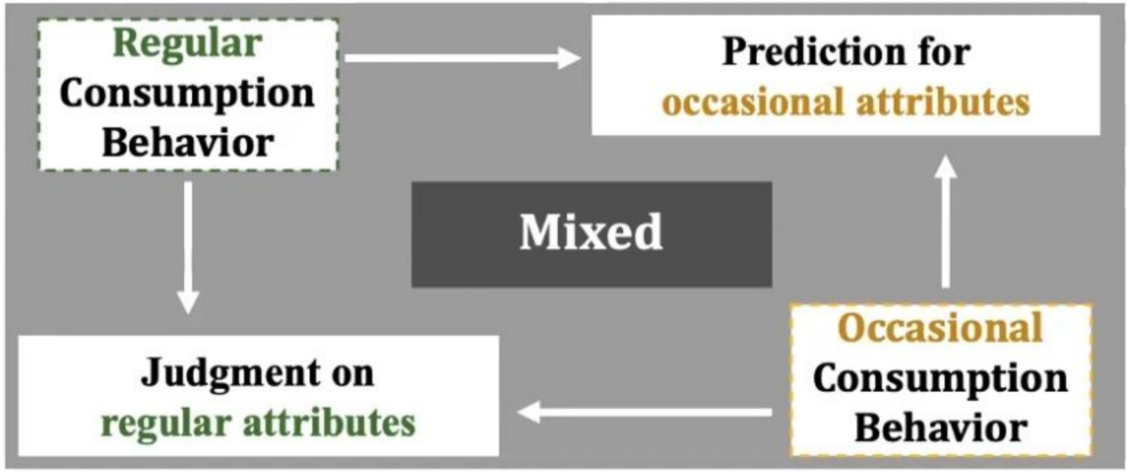
Motivation

- 另一方面，图数据在预测等任务中通常会随着时间发生变化。



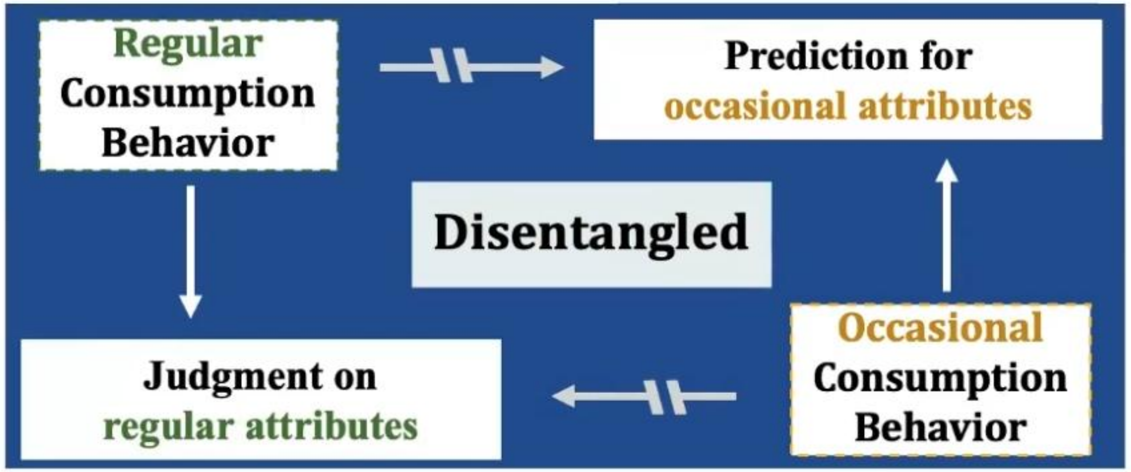
Motivation

现有方法



Mix time-invariant features with time-varying features

该论文提出的方法

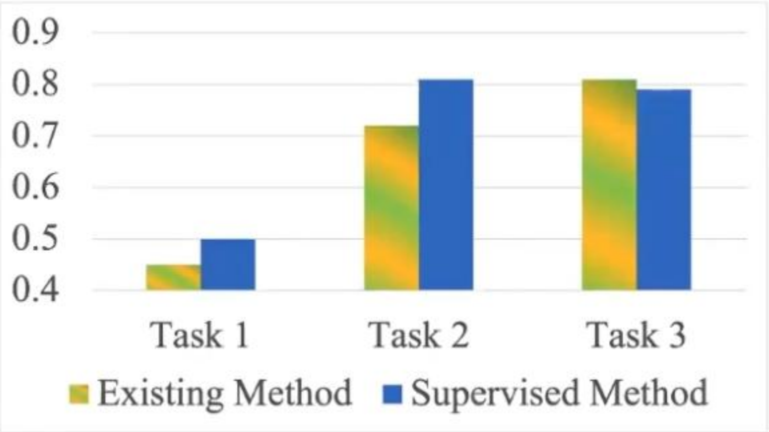


Disentangle time-invariant features with time-varying features

Motivation

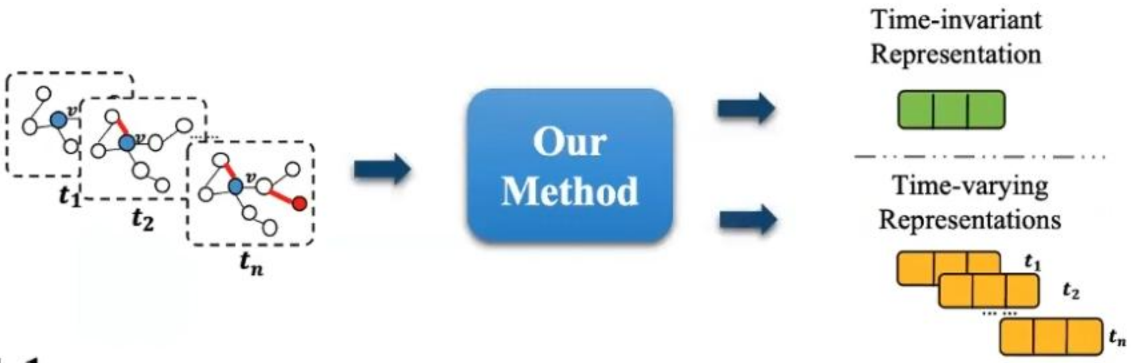
现有方法

Mix time-invariant features with time-varying features



该论文提出的方法

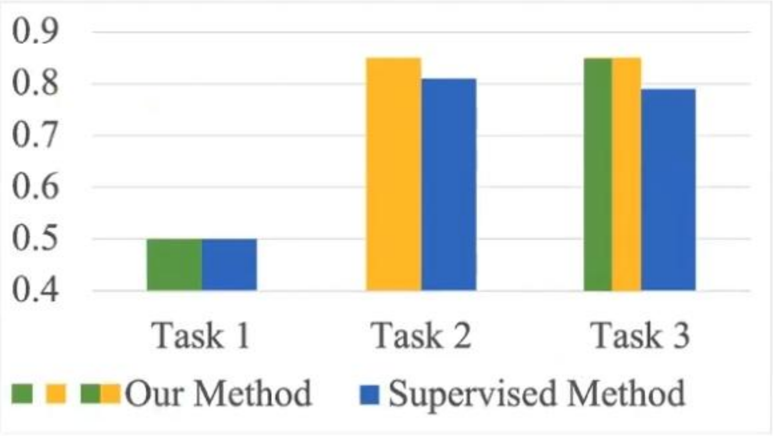
Disentangle time-invariant features with time-varying features



Task 1:
Predict User Annual Income
(user intrinsic stable characteristics)

Task 2:
Predict Consumption Fluctuation
(time-related dynamic preference)

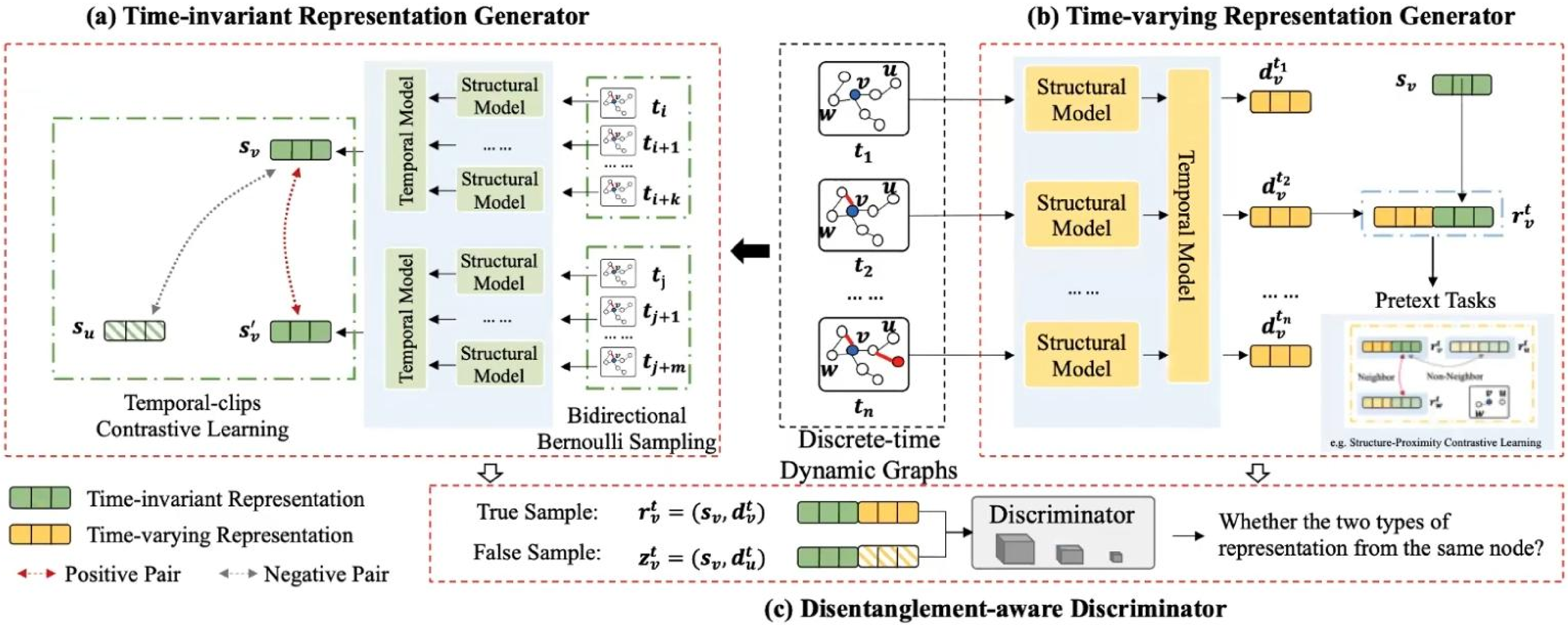
Task 3:
Predict Next Transaction
(traditional most adopted tasks)



Method

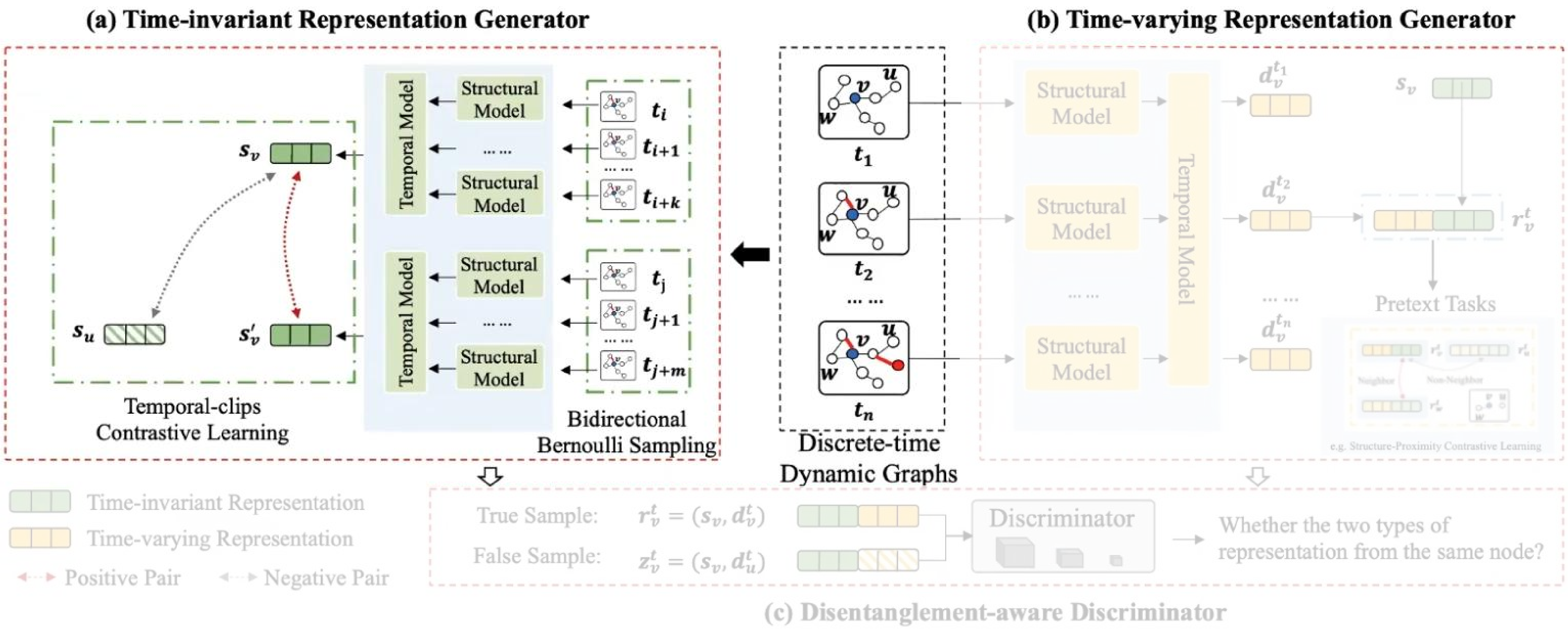
Method

Overview of DyTed



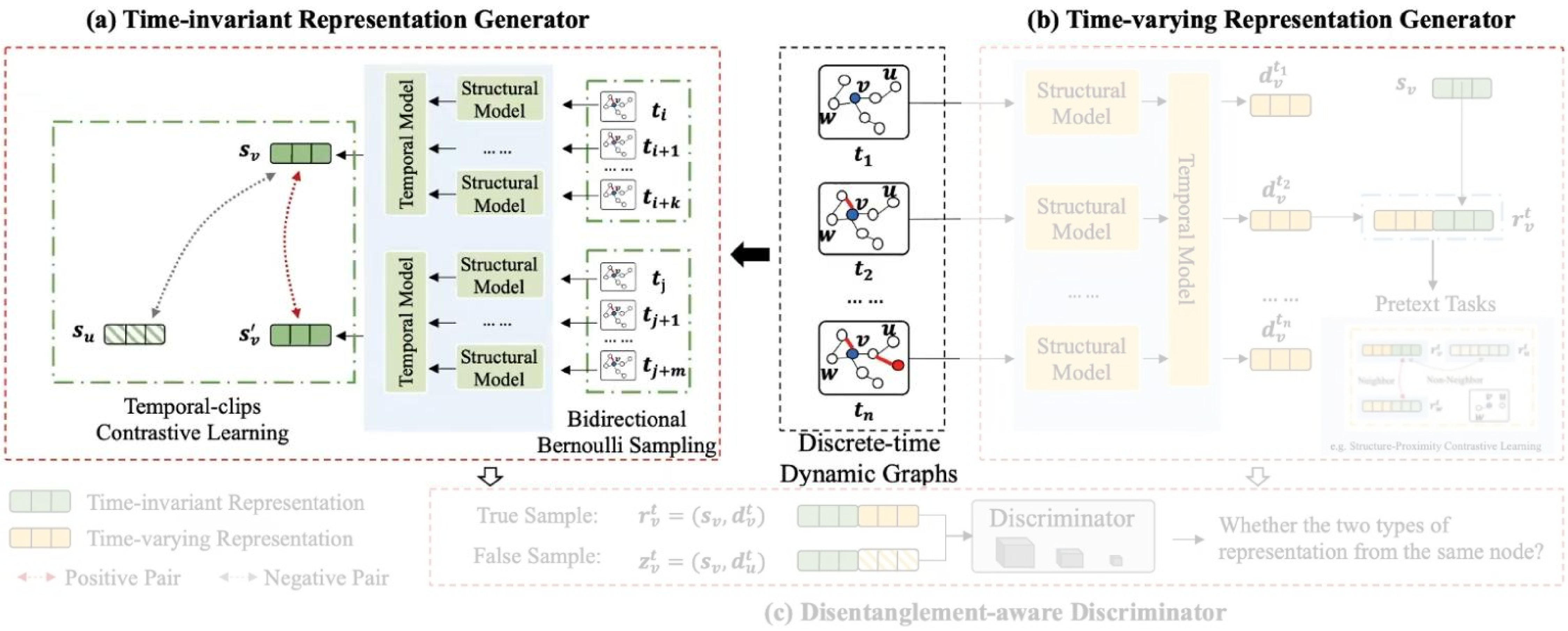
Method

Time-invariant Representation Generator



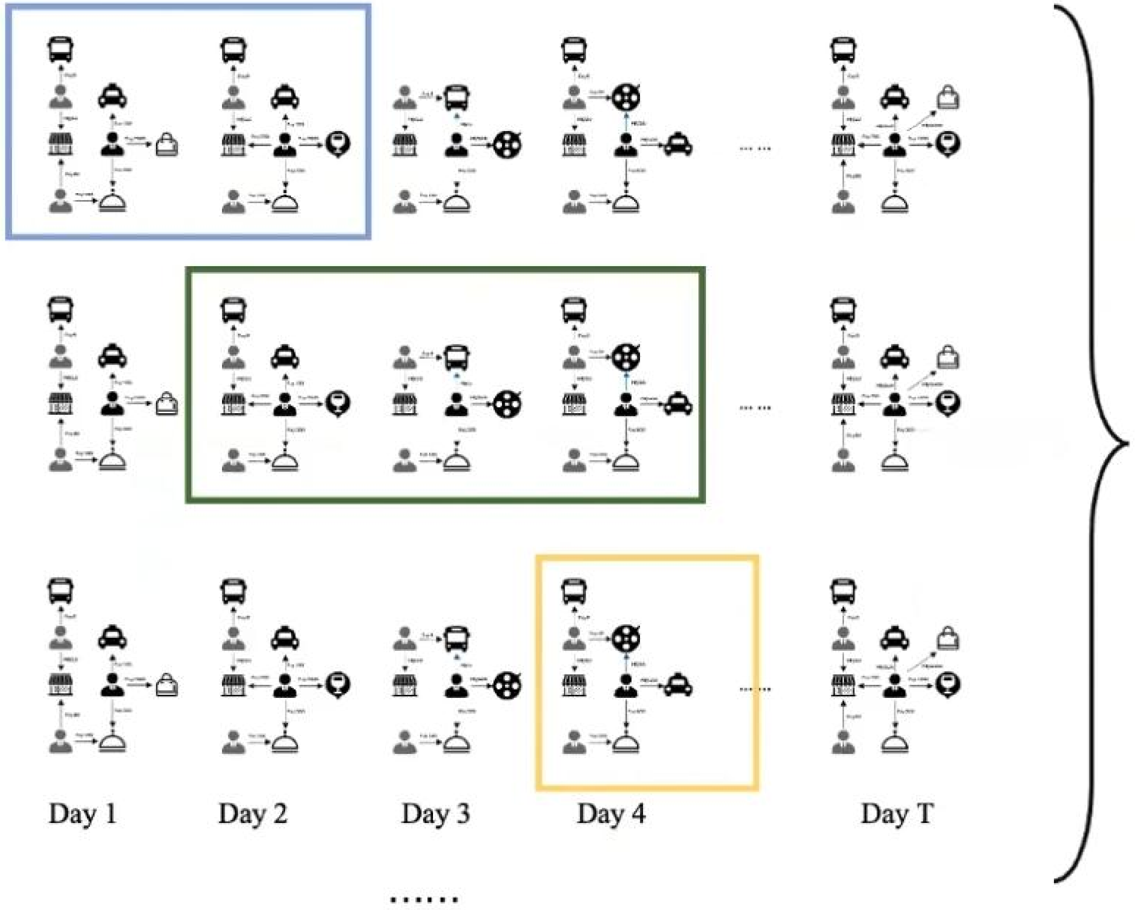
Method

Time-invariant Representation Generator



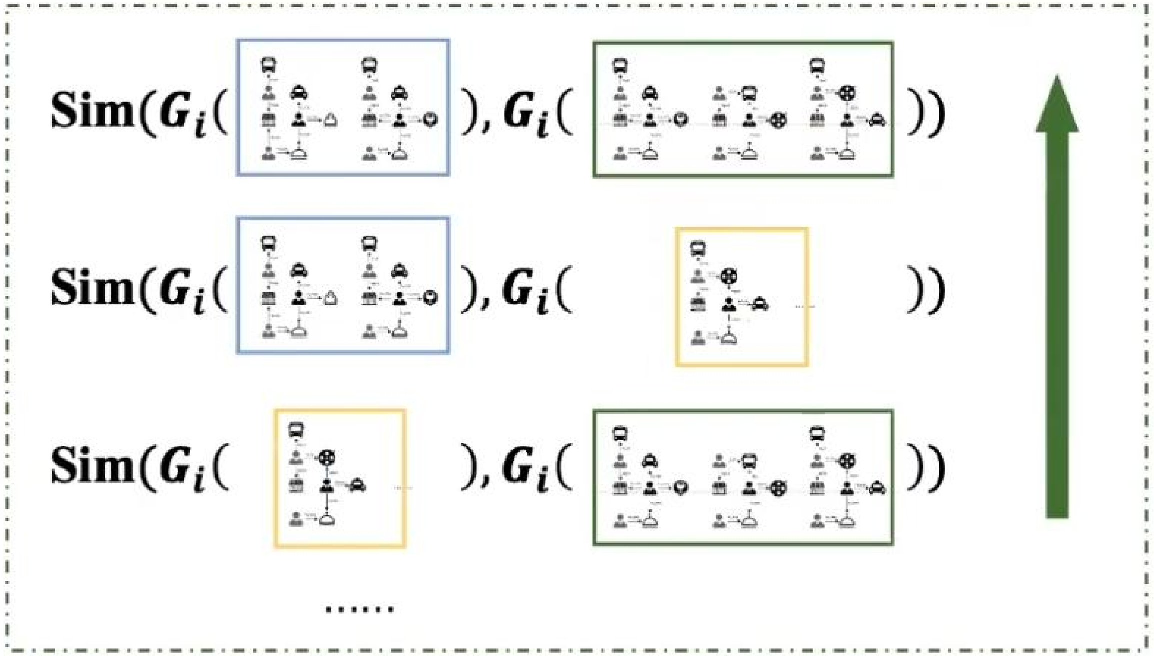
Method

How to extract **time-invariant** representations?



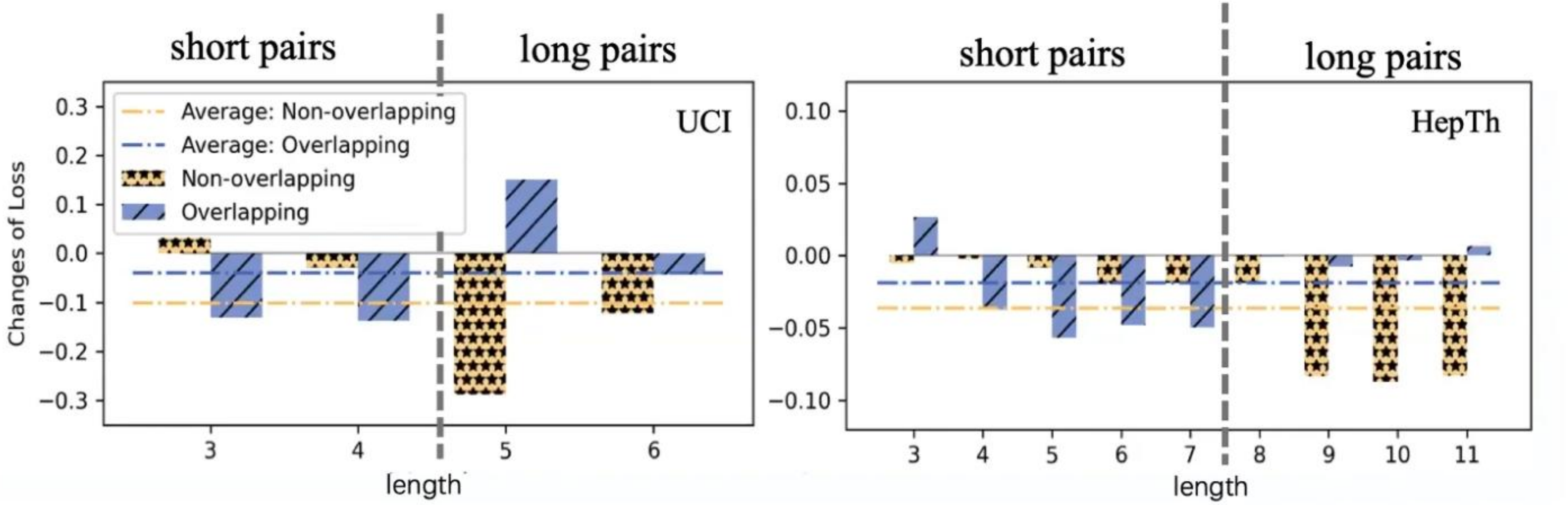
Time-invariant features:
The same properties **manifested in any temporal clip**.

Given a time-invariant representations generator G_i:



Method

Optimize Strategy



Three Rules

- Optimizing **non-overlapping pairs** benefits more for the overall loss reduction than **overlapping pairs**.
- For non-overlapping pairs, optimizing **long pairs** are more effective than **short pairs**.
- For overlapping pairs, optimizing **short pairs** are more effective than **long pairs**.

Method

Optimize Strategy

Three Rules

- Optimizing **non-overlapping pairs** ≥ overlapping pairs.
- For non-overlapping pairs, optimizing **long pairs** ≥ short pairs.
- For overlapping pairs, optimizing **short pairs** ≥ long pairs.

Long & Short → The sampling length needs to be controllable.

Non-overlapping & Overlapping → The two samplings need to be dependent.

Method

Optimize Strategy

Definition 1: Truncated geometric distribution. *Let* $p \in (0,1)$ *be the probability of success on each Bernoulli trial,* $m \in \mathbb{N}^+$ *be the number of Bernoulli trials when we get the first success. Given* $m \in [1, L]$ *it follows the truncated geometric distribution as follows:*

$$Geo(p, L) \sim f(m; p; L) = p(1-p)^{(m-1)}\Phi(L)^{-1}$$

where $\Phi(\mathrm{L}) = \sum_{m=1}^{L} p(1-p)^{(m-1)}$. *We have* $\boxed{\mathbb{E}(m) = \frac{1}{p} - \frac{L(1-p)^L}{1-(1-p)^L}.}$

$\downarrow$

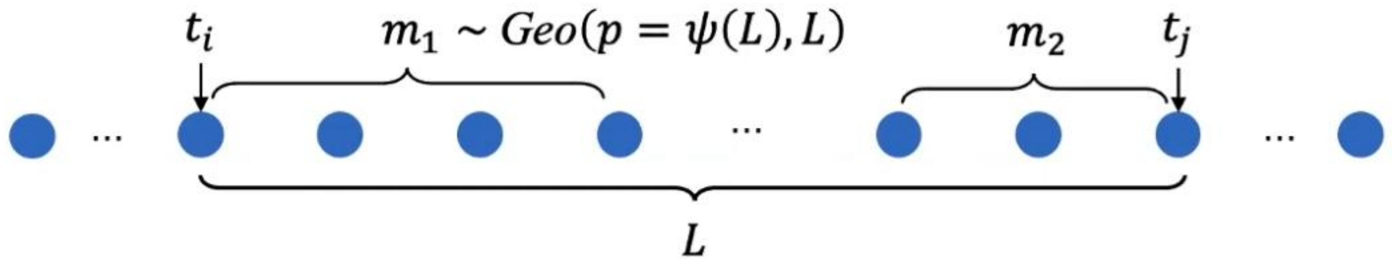
Control the sampling length

Method

Optimize Strategy

Definition 2: Bidirectional Bernoulli Sampling. *Let $t_i \in \mathbb{N}^+$ be the start timestamp for the first temporal clips and $t_j \in \mathbb{N}^+$ be the end timestamp for the second temporal clips. Let $L = t_j - t_i + 1$ denote the clips range from t_i to t_j. We sample L from uniform distribution $U(1, T)$ and t_i from $U(1, T - L)$, where T is the total number of snapshots. Then the length of the two temporal clips $m_1, m_2 \in \mathbb{N}$ is drawn i.i.d. from truncated geometric distribution $Geo(p = \psi(L), L)$, where $\psi(L)$ is the decreasing function related to L. Following the above sampling process, two temporal clips are sampled as:*

$$\mathcal{C}_1 = [\mathcal{G}^{t_i}, \mathcal{G}^{t_{i+1}}, \dots \mathcal{G}^{t_i + m_1 - 1}] \qquad \mathcal{C}_2 = [\mathcal{G}^{t_j - m_2 + 1}, \mathcal{G}^{t_j - m_2 + 2}, \dots \mathcal{G}^{t_j}]$$



Method

Optimize Strategy

Definition 2: Bidirectional Bernoulli Sampling. *Let $t_i \in \mathbb{N}^+$ be the start timestamp for the first temporal clips and $t_j \in \mathbb{N}^+$ be the end timestamp for the second temporal clips. Let $L = t_j - t_i + 1$ denote the clips range from t_i to t_j. We sample L from uniform distribution $U(1, T)$ and t_i from $U(1, T - L)$, where T is the total number of snapshots. Then the length of the two temporal clips $m_1, m_2 \in \mathbb{N}$ is drawn i.i.d. from truncated geometric distribution $Geo(p = \psi(L), L)$, where $\psi(L)$ is the decreasing function related to L. Following the above sampling process, two temporal clips are sampled as:*

$$\mathcal{C}_1 = [\mathcal{G}^{t_i}, \mathcal{G}^{t_{i+1}}, \ldots \mathcal{G}^{t_i+m_1-1}] \qquad \mathcal{C}_2 = [\mathcal{G}^{t_j-m_2+1}, \mathcal{G}^{t_j-m_2+2}, \ldots \mathcal{G}^{t_j}]$$

As L **increases**, $\psi(L)$ **decreases** slowly. → **Long pairs** are more likely to be **non-overlapping**.

As L **decreases**, $\psi(L)$ **increases** slowly. → **Short pairs** are more likely to be **overlapping**.

Optimize Strategy

Proposition 1: *Let pairs of two temporal clips $\mathcal{C}_1$ and $\mathcal{C}_2$ be sampled from bidirectional Bernoulli sampling. Let $X = 1$ denote that $\mathcal{C}_1$ and $\mathcal{C}_2$ have overlapped snapshots, otherwise $X = 0$. When probability p in truncated geometric distribution satisfies that $\frac{2}{L+2} \leq p < 1$ and $L \geq 3$, then we have:*

$$\frac{\Pr(X=1)}{\Pr(X=0)} \leq 1, \quad \frac{\Pr(L=l+1 \mid X=0)}{\Pr((L=l \mid X=0)} \geq 1, \quad \frac{\Pr(L=l+1 \mid X=1)}{\Pr((L=l \mid X=1)} \leq 1$$

Three Rules

- Optimizing **non-overlapping pairs** ≥ overlapping pairs. $\frac{\Pr(X=1)}{\Pr(X=0)} \leq 1$
- For non-overlapping pairs, optimizing **long pairs** ≥ short pairs. $\frac{\Pr(L=l+1 \mid X=0)}{\Pr((L=l \mid X=0)} \geq 1$
- For overlapping pairs, optimizing **short pairs** ≥ long pairs. $\frac{\Pr(L=l+1 \mid X=1)}{\Pr((L=l \mid X=1)} \leq 1$

Method

Optimize Strategy

Proposition 1: *Let pairs of two temporal clips $\mathcal{C}_1$ and $\mathcal{C}_2$ be sampled from bidirectional Bernoulli sampling. Let $X = 1$ denote that $\mathcal{C}_1$ and $\mathcal{C}_2$ have overlapped snapshots, otherwise $X = 0$. When probability p in truncated geometric distribution satisfies that $\frac{2}{L+2} \leq p < 1$ and $L \geq 3$, then we have:*

$$\frac{\Pr(X=1)}{\Pr(X=0)} \leq 1, \quad \frac{\Pr(L=l+1 \mid X=0)}{\Pr((L=l \mid X=0)} \geq 1, \quad \frac{\Pr(L=l+1 \mid X=1)}{\Pr((L=l \mid X=1)} \leq 1$$

$$\frac{2}{L+2} \leq p = \psi(L) = 1 - \alpha \frac{2}{L+2} < 1$$

$$0 < \boxed{\alpha} \leq 1$$

Learnable parameter

Method

Optimize Strategy

Given two sampled temporal clips $\mathcal{C}_1$ and $\mathcal{C}_2$,

the time-invariant representation for node v is denoted as: $\boldsymbol{s}_v^1 = \boldsymbol{G}_i(\boldsymbol{\mathcal{C}}_1)_v, \boldsymbol{s}_v^2 = \boldsymbol{G}_i(\boldsymbol{\mathcal{C}}_2)_v$

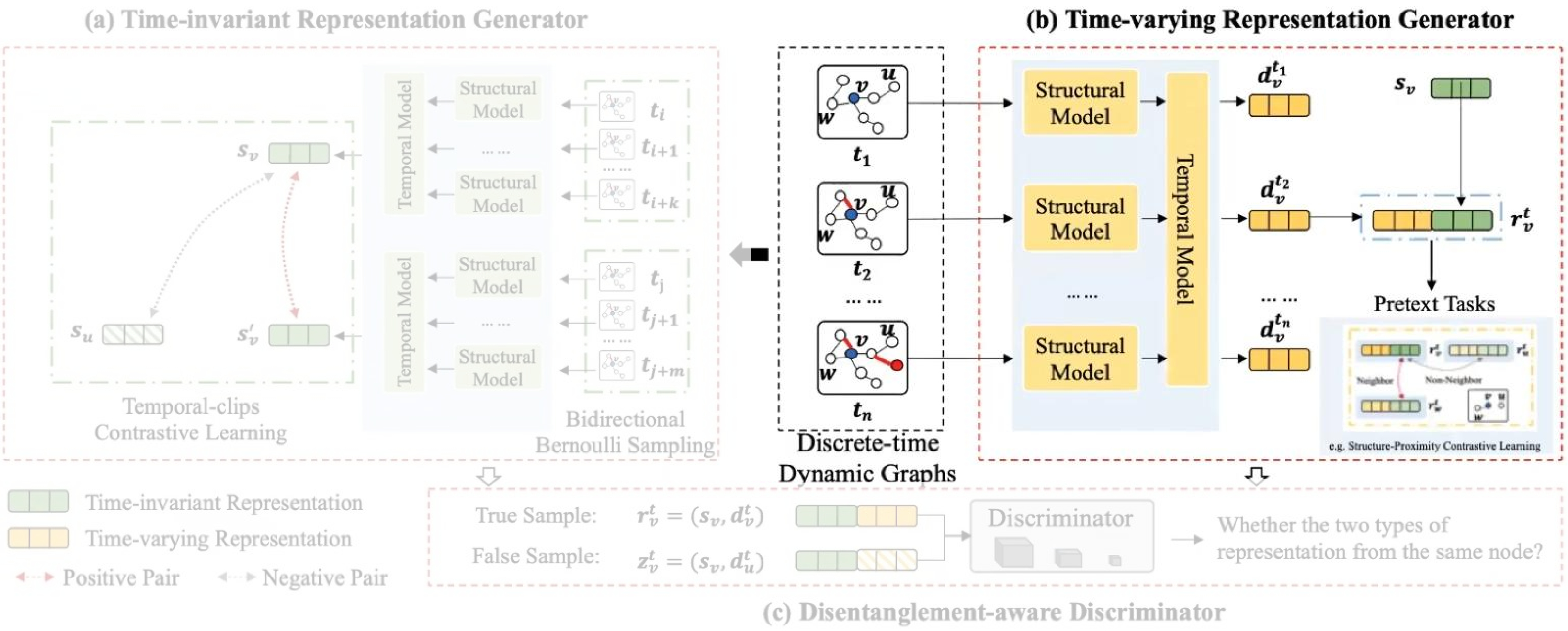
We take representations from the same node as $\underbrace{\textbf{positive pairs}}_{(s_v^1, s_v^2)}$,

and those from different nodes as $\underbrace{\textbf{negative pairs}}_{(s_v^1, s_u^1)}$.

InfoNCE[1] Loss: $$L_i(G_i) = \mathbb{E}_v\left[-\log\left(\frac{\exp\left(\mathrm{sim}(s_v^1, s_v^2)/\tau\right)}{\sum_{u\in\mathcal{V}}\exp\left(\mathrm{sim}(s_v^1, s_u^1)/\tau\right)}\right)\right]$$

Method

Time-varying Representation Generator



Method

Optimize the time-varying generator G_v

The time-varying representations for node v is denoted as: $[\boldsymbol{d}_v^1, \boldsymbol{d}_v^2, \dots, \boldsymbol{d}_v^T] = \boldsymbol{G}_v([\mathcal{G}^1, \mathcal{G}^2, \dots, \mathcal{G}^T])_v$

⬇

Considering that the graph structure and the evaluation are the **comprehensive embodiment** of both the time-invariant and time-varying representations:

$$\textbf{Combination}: \boldsymbol{r}_v^t = (\boldsymbol{s}_v,\ \boldsymbol{d}_v^t)$$

Pretext Tasks:

$$L_v(G = \{G_i, G_v\}) = \sum_{t=1}^{T} \mathbb{E}_{(u,v)\in\mathcal{E}_t}\left[-\log\left(\frac{\exp\left(\text{sim}(r_v^t, r_u^t)/\tau\right)}{\sum_{w\in\mathcal{U}} \exp\left(\text{sim}(r_v^t, r_w^t)/\tau\right)}\right)\right]$$

or

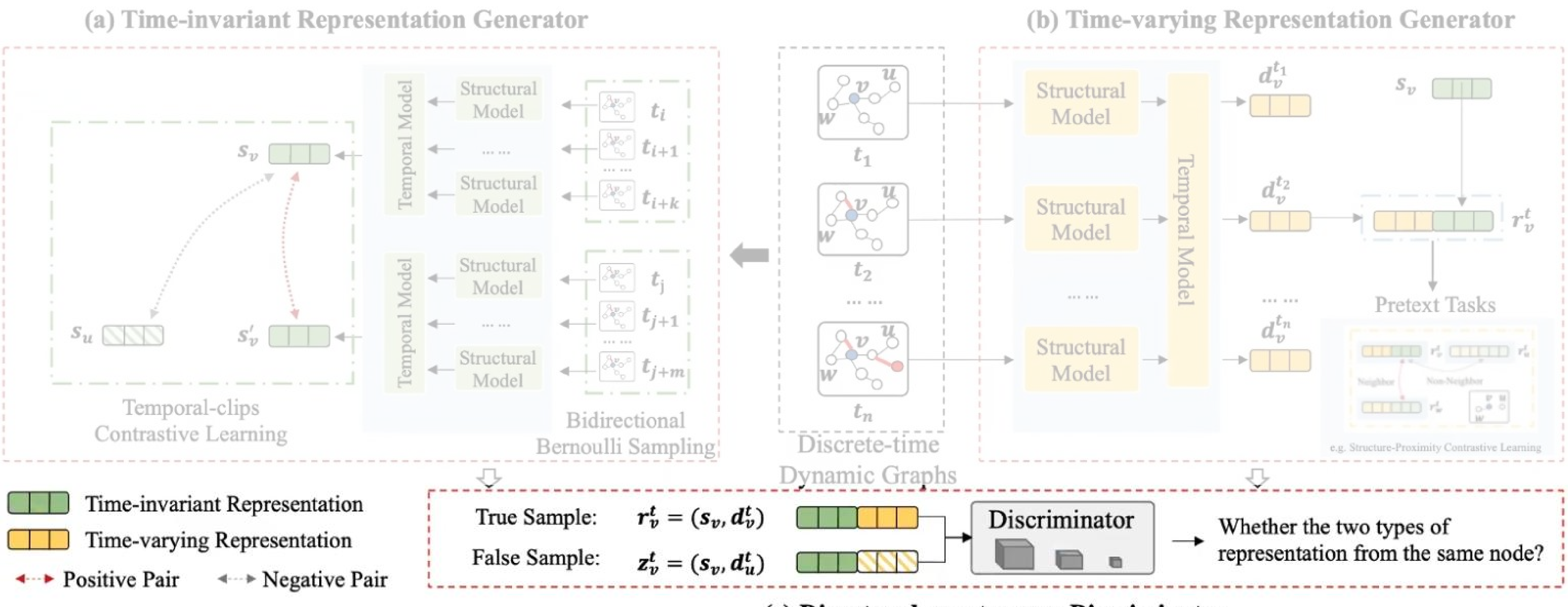
$$L_v(G) = -\sum_{t=1}^{T-1}\left[\mathbb{E}_{(u,v)\in\mathcal{E}_{t+1}} \log\left(r_v^{t\top} r_u^t\right) + \mathbb{E}_{(u,w)\notin\mathcal{E}_{t+1}} \log\left(1 - (r_v^{t\top} r_w^t)\right)\right]$$

or

…

Method

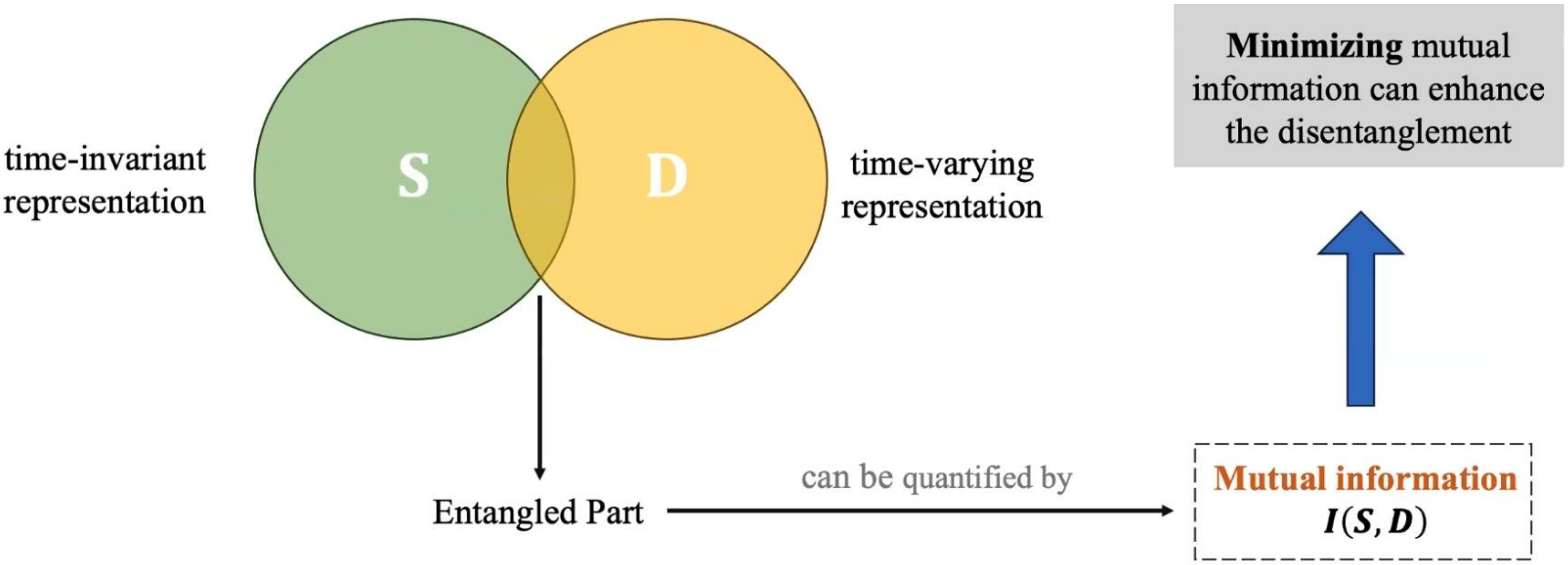
Disentanglement-aware Discriminator



(c) Disentanglement-aware Discriminator

Method

解耦表示学习



Method

解耦表示学习

Based on the definition of mutual information:

$$I(\boldsymbol{S},\boldsymbol{D}) = \sum_{s\in S}\sum_{d\in D} p(\boldsymbol{s},\boldsymbol{d})\log\frac{p(\boldsymbol{s},\boldsymbol{d})}{p(\boldsymbol{s})p(\boldsymbol{d})} = \mathcal{D}_{kl}(p(\boldsymbol{S},\boldsymbol{D})||p(\boldsymbol{S})p(\boldsymbol{D}))$$

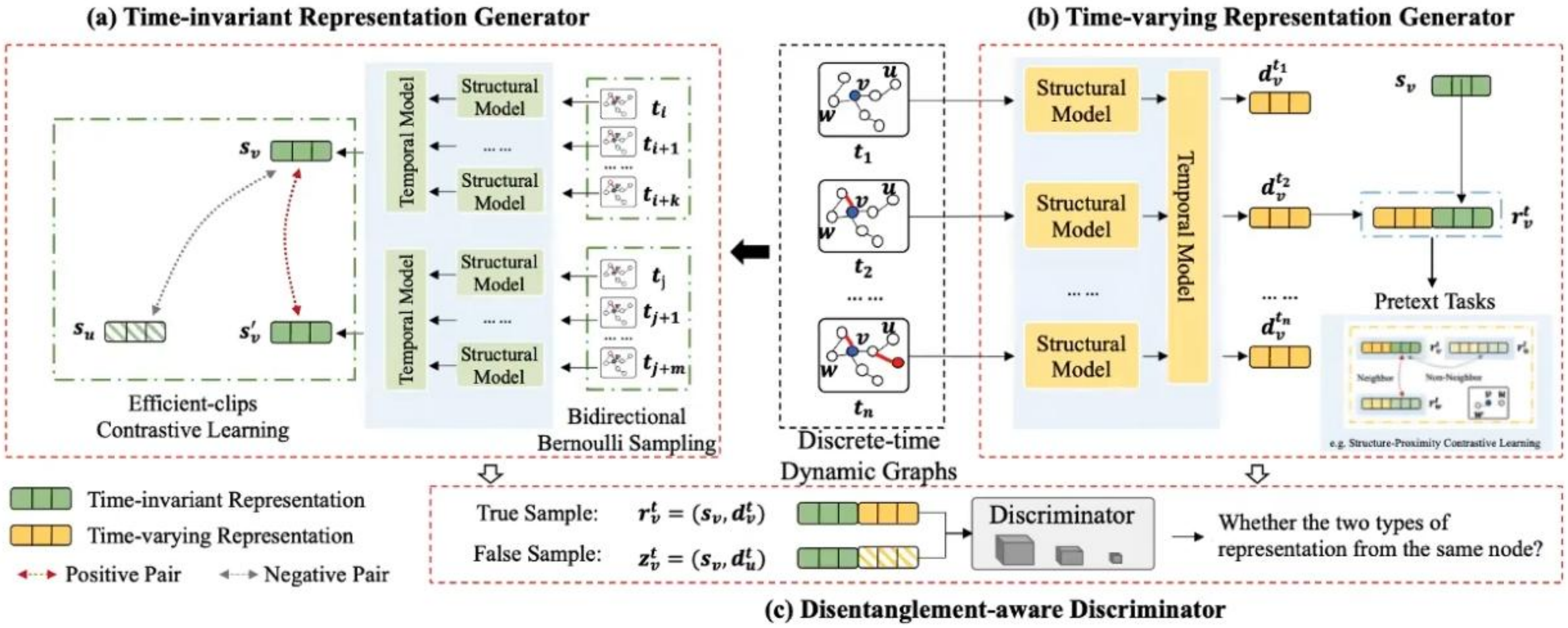
Sample $r_i^t = (s_v, d_v^t) \sim p(\boldsymbol{S},\boldsymbol{D})$ and $z_i^t = (s_v, d_u^t) \sim p(\boldsymbol{S})p(\boldsymbol{D})$.

Minimize $\boldsymbol{\mathcal{D}_{kl}(p(S,D)||p(S)p(D))}$ by GAN[2]:

$$\min_{G=\{G_i,G_v\}}\max_{D} V(G,D) = \min_{G=\{G_i,G_v\}}\max_{D} \mathbb{E}_T\left(\log\left(D\left(r_i^t\right)\right) + \log\left(1 - D\left(z_i^t\right)\right)\right)$$

Method

Loss function



Loss function of generators:

$$Loss(G = \{G_i, G_v\}) = L_v(G) + \lambda_1 L_i(G_i) + \lambda_2 V(G,D) + \lambda_3 \|w\|_2^2$$

Loss function of discriminator:

$$Loss(D) = -V(G,D)$$

Experiments

Experiments

Setup

We conducted experimental tests on **five public datasets** and **two real transaction datasets**.

DATASET	#Nodes	#Edges	#Snapshots
UCI	1,809	56,459	13
Bitcoin	3,782	483,700	20
AS733	4,648	532,230	30
HepTh	7,576	196,463	23
HepPh	10,404	339,556	20
Tencent-alpha	11,623	102,464	30
Tencent-beta	115,190	6,680,276	30

We apply the proposed framework DyTed to the following **five baselines**.

- LSTMGCN
- EvolveGCN (AAAI 20)
- ROLAND (KDD 22)
- DySAT (WSDM 20)
- HTGN (KDD 21)

Experiments

- RQ1: Whether the DyTed can improve the performance of existing methods in various downstream tasks
  - Task 1: Node Classification with time-invariant label
  - Task 2: Node Classification with time-varying label
  - Task 3: Link Prediction
- RQ2: What does each component of DyTed bring?
  - Ablation study
  - Disentanglement degree analysis
- RQ3: Is there any additional benefit of disentanglement?
  - Requirement of downstream training resources
  - Robustness against noise

Experiments

RQ1: Performance compare for Task 1

Model	Annual Income		Age		Assets		Financing Risk		Consumption Fluctuation	
	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1
LSTMGCN	50.38 ± 0.38	27.12 ± 0.32	34.86 ± 0.48	17.98 ± 1.08	24.06 ± 0.46	16.37 ± 0.39	49.13 ± 0.32	29.56 ± 0.64	93.70 ± 0.06	93.65 ± 0.06
LSTMGCN-DyTed	(+7.23% ↑)	(+10.55% ↑)	(+1.55% ↑)	(+23.30% ↑)	(+6.77% ↑)	(+20.95% ↑)	(+2.89% ↑)	(+6.76% ↑)	(+3.18% ↑)	(+2.36% ↑)
-Combine	51.14 ± 0.46	28.21 ± 0.49	33.64 ± 0.56	21.13 ± 0.43	25.51 ± 0.54	19.14 ± 0.55	48.29 ± 0.55	**31.56 ± 0.21**	95.91 ± 0.33	95.83 ± 0.32
-Time-invariant	**54.02 ± 0.35**	**29.98 ± 0.72**	**35.40 ± 0.20**	**22.17 ± 0.21**	**25.69 ± 0.26**	**19.80 ± 0.41**	**50.55 ± 0.50**	31.42 ± 0.98	74.05 ± 0.99	73.32 ± 1.02
-Time-varying	52.45 ± 0.18	23.82 ± 0.21	34.04 ± 0.55	15.12 ± 0.47	25.39 ± 0.18	15.02 ± 0.31	49.03 ± 0.35	24.98 ± 1.09	**96.68 ± 0.21**	**95.86 ± 0.20**
DySAT	46.65 ± 3.12	25.34 ± 0.97	28.92 ± 3.18	13.54 ± 1.28	25.50 ± 0.58	13.06 ± 0.46	42.95 ± 2.46	25.03 ± 0.74	72.73 ± 2.06	71.21 ± 2.18
DySAT-DyTed	(+13.48% ↑)	(+10.22% ↑)	(+15.08% ↑)	(+38.63% ↑)	(+9.73% ↑)	(+26.26% ↑)	(+15.32% ↑)	(+14.74% ↑)	(+16.72% ↑)	(+19.15% ↑)
-Combine	49.39 ± 3.31	27.82 ± 1.76	28.07 ± 4.55	13.93 ± 2.04	25.88 ± 0.52	13.12 ± 0.62	45.66 ± 4.65	**33.13 ± 3.09**	80.11 ± 1.96	79.98 ± 2.05
-Time-invariant	**52.94 ± 0.48**	**27.93 ± 0.50**	**33.28 ± 0.73**	**18.77 ± 0.94**	**27.98 ± 0.71**	**16.49 ± 0.82**	**49.53 ± 0.42**	28.72 ± 1.32	71.48 ± 2.05	71.11 ± 2.07
-Time-varying	40.41 ± 4.87	25.16 ± 1.96	23.87 ± 1.42	11.74 ± 1.08	23.06 ± 1.22	11.25 ± 0.87	42.01 ± 3.10	26.05 ± 0.88	**84.89 ± 1.66**	**84.85 ± 1.66**
EvolveGCN	48.26 ± 1.01	23.89 ± 1.07	31.02 ± 0.88	12.64 ± 0.74	24.39 ± 1.02	11.62 ± 0.17	43.76 ± 1.00	25.76 ± 0.40	92.00 ± 0.39	92.14 ± 0.39
EvolveGCN-DyTed	(+11.33% ↑)	(+3.47% ↑)	(+15.73% ↑)	(+19.86% ↑)	(+11.19% ↑)	(+30.12% ↑)	(+15.68% ↑)	(+1.28% ↑)	(+4.62% ↑)	(+5.18% ↑)
-Combine	53.00 ± 0.84	**24.72 ± 1.19**	35.06 ± 0.64	13.43 ± 0.54	26.86 ± 0.35	12.39 ± 0.18	47.96 ± 1.50	26.08 ± 0.29	**96.25 ± 0.08**	96.20 ± 0.06
-Time-invariant	**53.73 ± 0.40**	23.38 ± 0.08	**35.90 ± 0.31**	**15.15 ± 0.45**	**27.12 ± 0.30**	**15.12 ± 0.64**	**50.62 ± 0.58**	**26.09 ± 0.92**	76.45 ± 2.11	79.71 ± 2.11
-Time-varying	52.53 ± 1.03	24.63 ± 1.12	34.78 ± 0.58	12.11 ± 0.61	26.37 ± 0.89	10.87 ± 0.17	47.82 ± 1.65	24.05 ± 0.55	96.02 ± 0.08	**96.91 ± 0.08**
HTGN	54.11 ± 0.46	23.41 ± 0.13	35.74 ± 0.36	10.53 ± 0.08	26.70 ± 0.39	11.25 ± 0.71	50.13 ± 0.19	22.26 ± 0.06	92.25 ± 1.54	92.14 ± 1.57
HTGN-DyTed	(+1.81% ↑)	(+1.11% ↑)	(+3.47% ↑)	(+9.78% ↑)	(+6.55% ↑)	(+29.69% ↑)	(+3.97% ↑)	(+4.49% ↑)	(+4.80% ↑)	(+4.87% ↑)
-Combine	53.26 ± 0.55	23.17 ± 0.16	36.50 ± 0.39	10.50 ± 0.08	27.43 ± 0.38	12.96 ± 0.77	49.86 ± 0.36	22.18 ± 0.11	96.57 ± 0.09	96.52 ± 0.09
-Time-invariant	**55.09 ± 0.46**	**23.67 ± 0.13**	**36.98 ± 0.34**	**11.56 ± 0.07**	**28.45 ± 0.29**	**14.59 ± 0.07**	**52.12 ± 0.25**	**23.26 ± 0.08**	93.30 ± 0.29	93.21 ± 0.29
-Time-varying	53.06 ± 0.37	23.11 ± 0.10	35.66 ± 0.46	10.51 ± 0.10	26.99 ± 0.37	8.97 ± 0.21	49.08 ± 0.30	21.95 ± 0.09	**96.68 ± 0.11**	**96.63 ± 0.12**
ROLAND	50.62 ± 0.40	20.27 ± 0.11	31.21 ± 0.34	10.41 ± 0.08	28.08 ± 0.43	8.77 ± 0.10	49.34 ± 0.42	22.02 ± 0.12	83.27 ± 0.16	75.18 ± 0.07
ROLAND-DyTed	(+9.40% ↑)	(+24.32% ↑)	(+15.54% ↑)	(+20.94% ↑)	(+3.74% ↑)	(+11.97% ↑)	(+5.78% ↑)	(+11.58% ↑)	(+8.47% ↑)	(+7.37% ↑)
-Combine	54.01 ± 0.51	23.38 ± 0.14	36.04 ± 0.23	11.60 ± 0.05	28.50 ± 0.32	8.87 ± 0.08	50.11 ± 0.24	22.25 ± 0.07	85.45 ± 0.21	75.67 ± 0.09
-Time-invariant	**55.38 ± 0.36**	**25.20 ± 0.10**	**36.06 ± 0.47**	**12.59 ± 0.10**	**29.13 ± 0.37**	**9.82 ± 0.17**	**52.19 ± 0.36**	**24.57 ± 0.11**	84.25 ± 0.20	75.17 ± 0.09
-Time-varying	53.76 ± 0.32	23.01 ± 0.09	35.13 ± 0.34	10.40 ± 0.08	27.63 ± 0.47	8.67 ± 0.12	49.15 ± 0.46	21.97 ± 0.14	**90.32 ± 0.27**	**80.72 ± 0.12**

Experiments

RQ1: Performance compare for Task 1

Model	Annual Income		Age		Assets		Financing Risk	
	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1	micro-F1	macro-F1
GCN	0.3581	0.1719	0.1727	0.0760	0.0957	0.0638	0.3053	0.1785
GRU-GCN	0.1607	0.0954	0.0467	0.0220	0.1093	0.0483	0.2771	0.1929
EvolveGCN	0.1680	0.0947	0.0123	0.0072	0.0739	0.0485	0.2656	0.1412
HTGN	0.2496	0.1310	0.0056	0.0037	0.0043	0.0031	0.2629	0.1768
DySAT	0.3247	0.2044	0.1381	0.0610	0.1286	0.0606	0.3873	0.1879
DyTed-Fluctuate	0.3230	0.2397	0.2183	0.1156	0.1600	0.0917	0.3153	0.2338
DyTed	<u>0.3687</u>	<u>0.2629</u>	<u>0.2729</u>	<u>0.1363</u>	<u>0.2175</u>	<u>0.1097</u>	<u>0.3984</u>	<u>0.2730</u>
DyTed-Invariant	**0.4069**	**0.3224**	**0.2866**	**0.1839**	**0.2276**	**0.1746**	**0.3914**	**0.3503**

RQ1: Performance compare for Task 2

Table 3: Node classification with dynamic-fluctuate labels

Model	micro-F1	macro-F1
GCN	0.4725	0.3895
GRU-GCN	0.5450	0.5723
DySAT	0.5500	0.5763
HTGN	0.5761	0.4776
EvolveGCN	0.7697	0.7678
DyTed-Invariant	0.5477	0.5456
DyTed	0.7817	0.7812
DyTed-Fluctuate	**0.8197**	**0.8191**

Experiments

RQ1: Performance compare for Task 3

Model	Enron		UCI		AS733		HepTh		HepPh	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
GCN	0.7719	0.7622	0.6824	0.6800	0.8104	0.8133	0.8466	0.8521	0.8422	0.8438
GRU-GCN	0.7763	0.7751	0.7518	0.7219	0.8223	0.8244	0.8184	0.8261	0.8618	0.8528
EvolveGCN	0.7659	0.7681	0.7632	0.7828	0.9376	0.9364	0.7373	0.6651	0.9347	0.9483
HTGN	0.8018	0.8138	0.7390	0.6604	0.8768	0.8731	0.9244	0.9163	0.9473	0.9397
DySAT	0.8235	0.7760	0.7352	0.8158	**0.9499**	**0.9584**	0.8131	0.7547	0.9219	0.8779
DyTed-Invariant	0.7517	0.7112	0.7809	0.7699	0.8527	0.8183	0.8752	0.8360	0.9173	0.8940
DyTed-Fluctuate	0.7984	0.7481	0.7362	0.7103	0.8662	0.8438	0.9243	0.8976	0.9408	0.9215
DyTed	**0.8869**	**0.8766**	**0.8642**	**0.8693**	0.9365	0.9421	**0.9569**	**0.9587**	**0.9701**	**0.9700**

RQ2: Component Analysis

Ablation Study

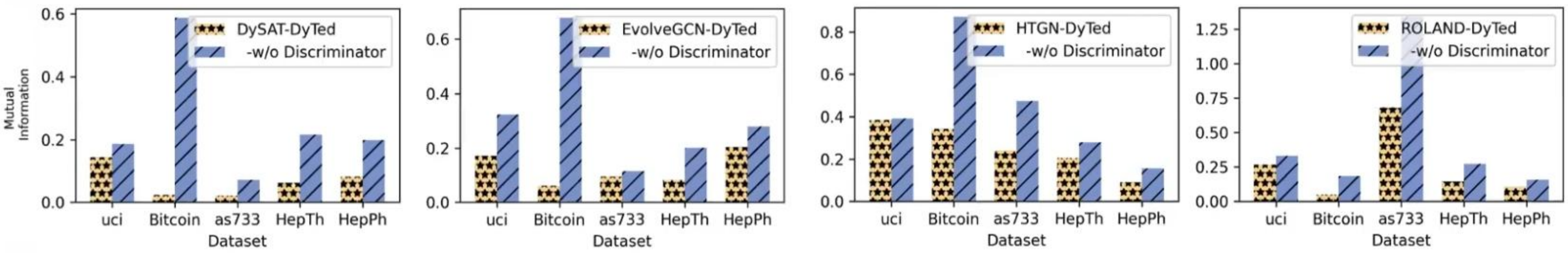
- **-DyTed-Radom Sampling**: Replace bidirectional Bernoulli sampling with random sampling.
- **-DyTed-w/o-Time-varying Generator**: Remove the time-varying representation generator.
- **-DyTed-w/o-Discriminator**: Remove the disentanglement-aware discriminator (adversarial learning).

Each component contributed to the enhancement of the DyTed framework.

Model	Uci	Bitcoin	As722	HepTh	HepPh
LSTMGCN-DyTed	**76.06**	**79.02**	**81.82**	**85.88**	**87.12**
-Random Sampling	71.09	77.72	80.57	85.79	87.03
-w/o Time-varying Generator	72.85	69.83	73.69	76.49	82.78
-w/o Discriminator	76.03	76.88	80.52	85.58	86.12
DySAT-DyTed	**86.12**	**83.32**	**84.47**	**80.76**	**78.68**
-Random Sampling	74.24	82.91	83.04	79.26	75.69
-w/o Time-varying Generator	71.26	79.61	78.7	74.87	71.66
-w/o Discriminator	78.20	80.21	84.38	77.47	76.01
EvolveGCN-DyTed	**84.15**	**85.50**	**80.97**	**87.52**	**80.21**
-Random Sampling	80.62	79.14	79.85	78.22	79.64
-w/o Time-varying Generator	82.55	78.85	78.10	67.12	73.19
-w/o Discriminator	83.75	84.15	77.65	79.79	70.54
HTGN-DyTed	**92.08**	**84.56**	**77.08**	**75.61**	**74.46**
-Random Sampling	85.09	82.21	76.00	73.05	74.26
-w/o Time-varying Generator	85.73	70.54	76.92	73.81	73.85
-w/o Discriminator	90.93	84.01	77.01	72.31	74.21
ROLAND-DyTed	**88.05**	**88.59**	**76.96**	**81.14**	**80.50**
-Random Sampling	84.11	86.32	74.89	78.11	79.19
-w/o Time-varying Generator	73.97	86.95	74.83	78.72	79.87
-w/o Discriminator	85.76	84.34	74.16	80.47	80.10

RQ2: Component Analysis

Evaluation of Disentanglement Degree

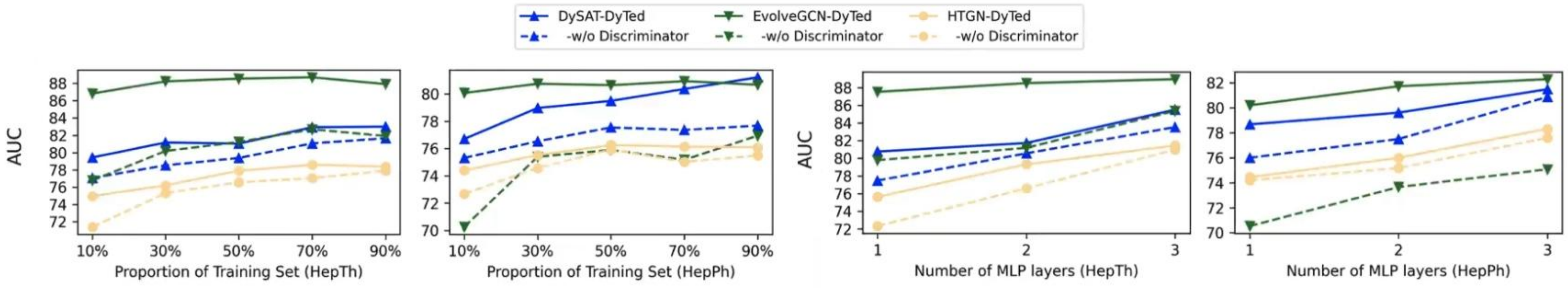


Disentanglement-aware discriminator significantly improve the disentangling degree

Experiments

RQ3: Other benefit

Requirement of Downstream Training Resources

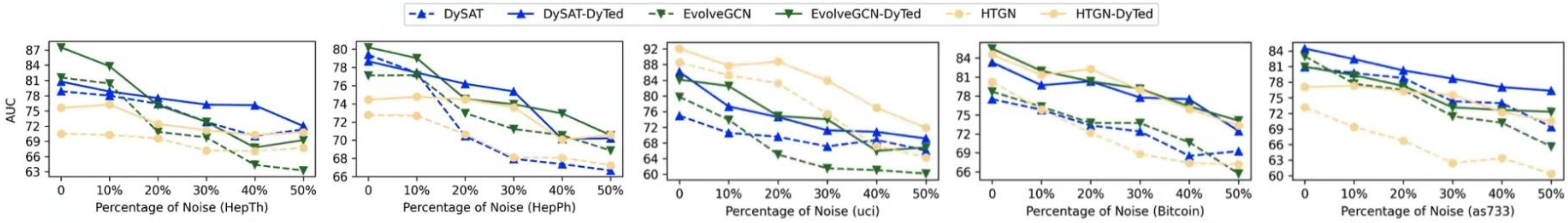


Representations with a high degree of disentanglement require fewer downstream resources

Experiments

RQ3: Other benefit

Robustness against noise



DyTed framework can significantly improve the robustness of the backbone model.

Thanks