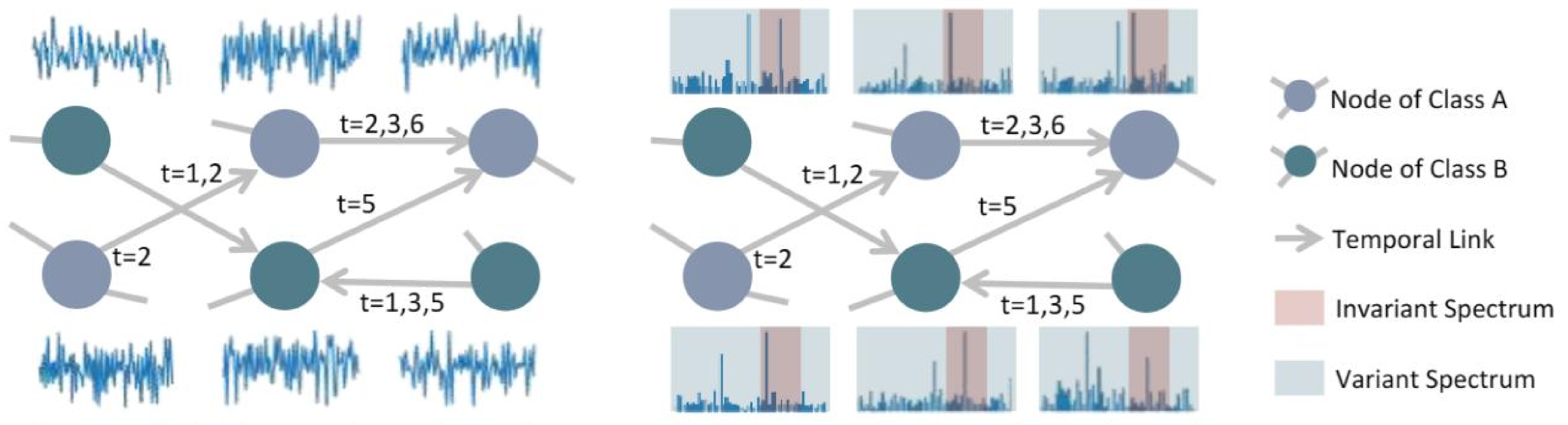
Spectral Invariant Learning for Dynamic Graphs under Distribution Shifts

Zeyang Zhang[1], Xin Wang[1], Ziwei Zhang[1], Zhou Qin[2], Weigao Wen[2], Hui Xue[2], Haoyang Li[1], Wenwu Zhu[1]
[1]Tsinghua University, [2]Alibaba Group

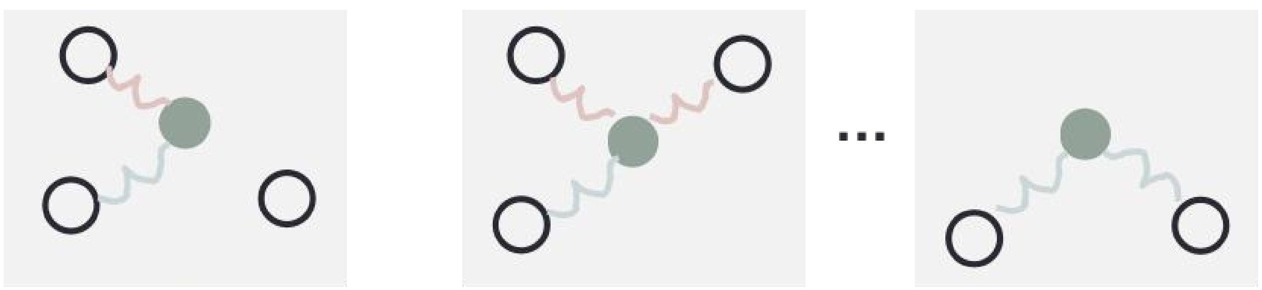
➢ Motivations

- **Spatio-temporal distribution shift naturally exist on dynamic graphs**
  - Due to Survivorship bias/Selection bias/Trending/"Black Swan"/……
  - Data distributions may shift in train and test stages along both spatial and temporal dimensions.
- **Existing DyGNNs handle distribution shifts in time domain**
  - assume that in the time domain, the distribution shift is observable and the invariant and variant patterns can be easily disentangled
- **However, there exist cases that the distribution shift is unobservable in the time domain while observable in the spectral domain**



Temporal signals on a dynamic graph   Spectrums on the same dynamic graph

- **Motivation example for tackling distribution shifts in spectral domain**



t=1   t=2   t=T

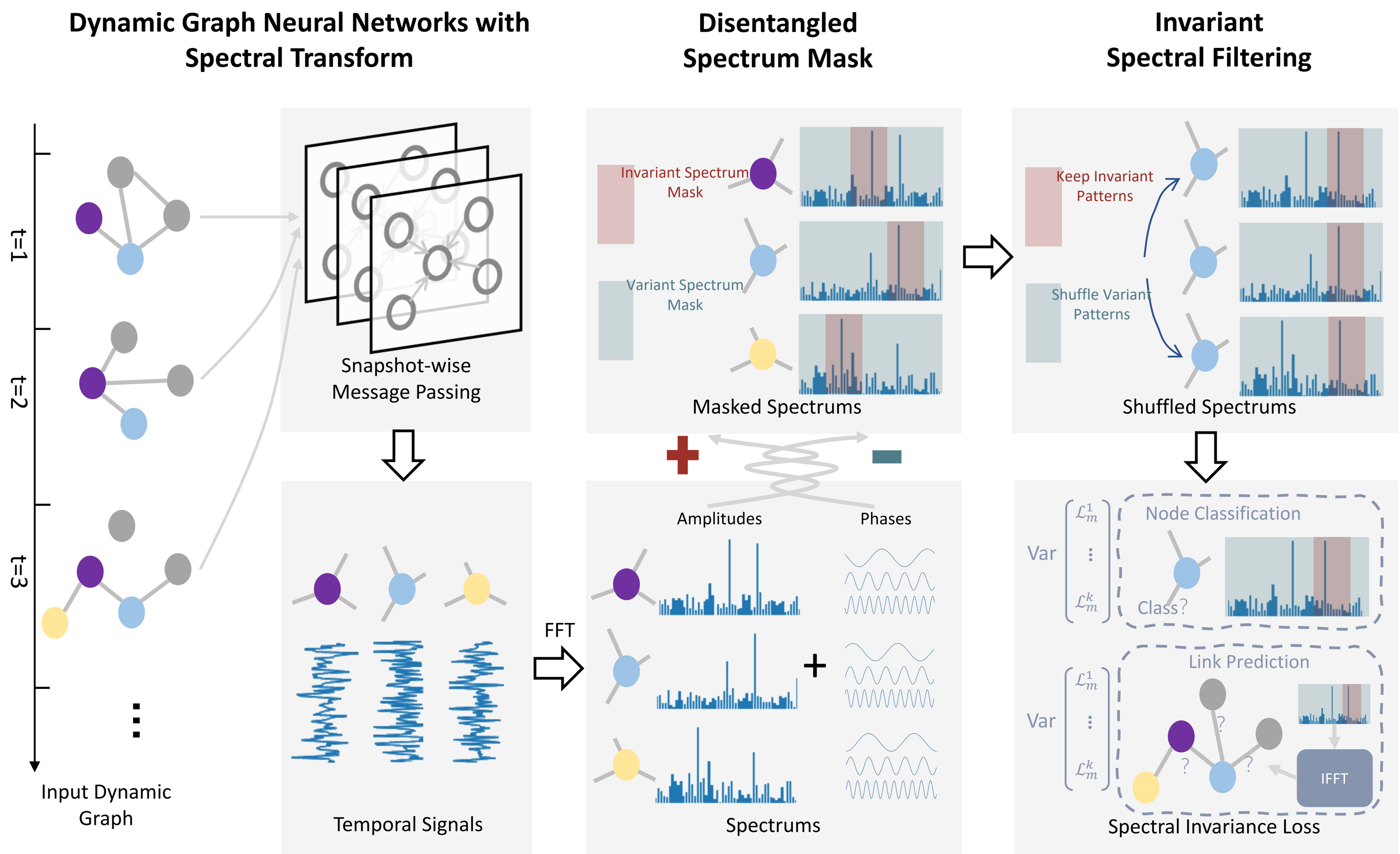
$$\mathbf{h}_v = \sum_{u \in \mathcal{N}_v} 1 = \mathbf{d}_{v,1} + \mathbf{d}_{v,2}$$
$$y_v = \mathbf{g}^\top \mathbf{d}_{v,1}$$
$$\hat{y}_v = \mathbf{w}^\top (\mathbf{m} \odot \mathbf{h}_v)$$
D_{tr} is $R_{tr}(\mathbf{w}) = \frac{1}{|D_{tr}|} \sum_{v \in D_{tr}} (\hat{y}_v - y_v)^2$

Proposition 1 *For any mask* $\mathbf{m} \in \mathbb{R}^{T \times 1}$, *for the optimal classifier in the training data* $\mathbf{w}^* = \arg\min_{\mathbf{w}} R_{tr}(\mathbf{w})$, *if* $||\mathbf{m} \odot \mathbf{w}^*||_2 \neq 0$, *there exist OOD nodes with unbounded error, i.e.,* $\exists v$ *s.t.* $\lim_{||\mathbf{d}_{v,2}|| \to \infty} (\hat{y}_v - y_v)^2 = \infty$.

Proposition 2 *If* $\left(\overline{\Phi \mathbf{d}_{v,1}} \odot \Phi \mathbf{d}_{v,1}\right) \odot \left(\overline{\Phi \mathbf{d}_{v,2}} \odot \Phi \mathbf{d}_{v,2}\right) = \mathbf{0}, \forall \mathbf{d}_{v,1}, \mathbf{d}_{v,2}$, *then* $\exists \mathbf{m} \in \mathbb{C}^{T \times 1}$ *such that the optimal spectral classifier in the training data has bounded error, i.e., for* $\mathbf{w}^* = \arg\min_{\mathbf{w}} R_{tr}(\mathbf{w})$, $\exists \epsilon > 0$, $\forall v$, $\lim_{||\mathbf{d}_{v,2}|| \to \infty} (\hat{y}_v - y_v)^2 < \epsilon$.

➢ Model (SILD)

- **DyGNN with Fourier transform** to obtain the spectrums of ego-graph trajectories
- **Disentangled spectrum mask** to obtain invariant and variant spectrum masks in the spectral domain
- **invariant spectral filtering mechanism** exploits invariant spectral patterns by minimizing the variance of predictions with exposure to various variant patterns



➢ Experiments

Table 1: Results of different methods on real-world link prediction and node classification datasets. The best results are in bold and the second-best results are underlined. The year in the Aminer dataset denotes the test split, *e.g.*, 'Aminer15' denotes the average test accuracy in 2015.

Task	Link Prediction (AUC%)		Node Classification (ACC%)		
Dataset	Collab	Yelp	Aminer15	Aminer16	Aminer17
GCRN	69.72±0.45	54.68±7.59	47.96±1.12	51.33±0.62	42.93±0.71
EGCN	76.15±0.91	53.82±2.06	44.14±1.12	46.28±1.84	37.71±1.84
DySAT	76.59±0.20	66.09±1.42	48.41±0.81	49.76±0.96	42.39±0.62
IRM	75.42±0.87	56.02±16.08	48.44±0.13	50.18±0.73	42.40±0.27
VREx	76.24±0.77	66.41±1.87	48.70±0.73	49.24±0.27	42.59±0.37
GroupDRO	76.33±0.29	66.97±0.61	48.73±0.61	49.74±0.26	42.80±0.36
DIDA	81.87±0.40	75.92±0.90	50.34±0.81	51.43±0.27	44.69±0.06
SILD	**84.09**±0.16	**78.65**±2.22	**52.35**±1.04	**54.11**±0.62	**45.54**±1.19

Real-world dynamic graph datasets

SILD achieves significant performance improvements on both link prediction and node classification tasks under distribution shifts

Synthetic dynamic graph datasets

SILD can exploit invariant patterns to consistently alleviate harmful effects of variant patterns under different distribution shift levels. SILD is especially effective under strong distribution shifts in spectral domain

Table 2: Results of different methods on synthetic link prediction and node classification datasets. The best results are in bold and the second-best results are underlined. A larger 'shift' denotes a higher distribution shift level.

Dataset	Link-Synthetic (AUC%)			Node-Synthetic (ACC%)		
Shift	0.4	0.6	0.8	0.4	0.6	0.8
GCRN	72.57±0.72	72.29±0.47	67.26±0.22	27.19±2.18	25.95±0.80	29.26±0.69
EGCN	69.00±0.53	62.70±1.14	60.13±0.89	24.01±2.29	22.75±0.96	24.98±1.32
DySAT	70.24±1.26	64.01±0.19	62.19±0.39	40.95±2.89	37.94±1.01	30.90±1.97
IRM	69.40±0.09	63.97±0.37	62.66±0.33	33.23±4.70	30.29±1.71	29.43±1.38
VREx	70.44±1.08	63.99±0.21	62.21±0.40	41.78±1.30	38.11±2.81	29.56±0.44
GroupDRO	70.30±1.23	64.05±0.21	62.13±0.35	41.35±2.19	35.74±3.93	31.03±1.24
DIDA	85.20±0.84	82.89±0.23	72.59±3.31	43.33±7.74	39.48±7.93	28.14±3.07
SILD	**85.95**±0.18	**84.69**±1.18	**78.01**±0.71	**43.62**±2.74	**39.78**±3.56	**38.64**±2.76

Contact: Zeyang Zhang, Email: zy-zhang20@mails.tsinghua.edu.cn