

Revealing Concept Shift in Spatio-Temporal Graphs via State Learning

Kuo Yang^{1,2}, Yunhe Guo^{1,2}, Qihe Huang^{1,2}, Zhengyang Zhou^{1,2,*}, Yang Wang^{1,2,*}

¹University of Science and Technology of China (USTC), Hefei, China

²Suzhou Institute for Advanced Research, USTC, Suzhou, China

{yangkuo, guoyunhe, hqh}@mail.ustc.edu.cn, {zzy0929, angyan}@ustc.edu.cn.

Abstract

Dynamic graphs are ubiquitous in the real world, presenting the temporal evolution of individuals within spatial associations. Recently, dynamic graph learning research is flourishing, striving to more effectively capture evolutionary patterns and spatial correlations. However, existing methods still fail to address the issue of concept shift in dynamic graphs. Concept shift manifests as a distribution shift in the mapping pattern between historical observations and future evolution. The reason is that some environment variables in dynamic graphs exert varying effects on evolution patterns, but these variables are not effectively captured by the models, leading to the intractable concept shift issue. To tackle this issue, we propose a **State-driven environment inference framework (Samen)** to achieve a dynamic graph learning framework equipped with concept generalization ability. Firstly, we propose a two-stage environment inference and compression strategy. From the perspective of state space, we introduce a prefix-suffix collaborative state learning mechanism to bidirectionally model the spatio-temporal states. A hierarchical state compressor is further designed to refine the state information resulting in concept shift. Secondly, we propose a skip-connection spatio-temporal prediction module, which effectively utilizes the inferred environments to improve the model’s generalization capability. Finally, we select seven datasets from different domains to validate the effectiveness of our model. By comparing the performance of different models on samples with concept shift, we verify that our **Samen** gains generalization capacity that existing methods fail to capture.

1 Introduction

Dynamic graphs are ubiquitous in the real world, encompassing social networks [Wang *et al.*, 2019], web-based platforms [Gao *et al.*, 2022], and traffic networks [Zhou *et al.*, 2020; Liu

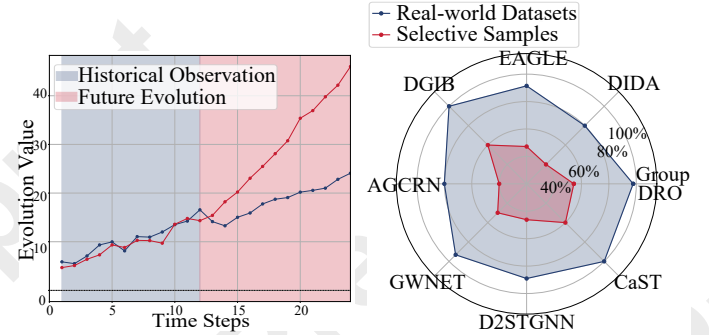


Figure 1: Our research motivation. Left Panel: a sample selected from the SD2019 dataset exhibiting concept shift. Right Panel: the performance discount comparison of existing models on the samples with a concept shift issue.

et al., 2023; Liu *et al.*, 2024]. Unlike static graphs, dynamic graphs involve the changes in nodes and edges over time, presenting a significant challenge in simultaneously modeling temporal evolution and spatial correlations [Zhou *et al.*, 2023; Cao *et al.*, 2020]. Recently, the spatio-temporal learning of coupled Graph Neural Networks and Time Series Models has started to flourish [Xia *et al.*, 2023; Yuan *et al.*, 2023; Liu *et al.*, 2025a; Liu *et al.*, 2025b; Miao *et al.*, 2024; Miao *et al.*, 2025], gaining increasing improvement for its ability to effectively model both spatial dependencies and temporal dynamics.

Although existing dynamic graph learning frameworks are theoretically proven to effectively capture spatio-temporal correlations, modeling real-world dynamic graphs is challenging due to the constantly changing evolutionary patterns influenced by various complex factors [Zhang *et al.*, 2022; Yuan *et al.*, 2023]. These factors include human noise, irregular patterns, and unpredictable external conditions, all of which can significantly disturb the evolution patterns [Zhou *et al.*, 2023; Wu *et al.*, 2022; Wang *et al.*, 2023].

One of the major challenges arising from changes in evolution patterns is the issue of concept shift. This issue manifests as similar or identical historical observations leading to different future evolutions as shown in the right panel of Fig. 1. Actually, this phenomenon is widely present in real-world datasets [Shao *et al.*, 2022a]. Empirically, we observe that

*Zhengyang Zhou and Yang Wang are corresponding authors.

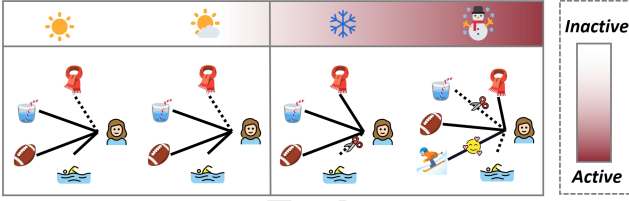


Figure 2: Human interests are influenced by the environments. Weather variables remain inactive in historical observations but become active in future observations.

the proportion of samples exhibiting concept shift in the Yelp [Sankar *et al.*, 2020] and SD-2019 [Liu *et al.*, 2023] datasets reached as high as 24.3% and 32.9%, respectively. The widespread existence of this phenomenon poses significant challenges for dynamic graph learning methods. As shown in the left panel of Fig. 1, we filter the samples that exhibit concept shift, and utilize them to evaluate existing dynamic graph learning frameworks. This distinct contrast highlights that concept shift is indeed a significant factor contributing to the lack of generalization in current dynamic graph learning models. Therefore, enhancing the model’s ability to address the concept shift issue has become an urgent challenge to resolve.

Unlike data distribution shift, which appears as discrepancies between the distributions of the training and test datasets [Chen *et al.*, 2024; Sun *et al.*, 2021; Li *et al.*, 2022; Chen *et al.*, 2022], concept shift is characterized by the shifts in the mapping relationship from data to its label. Some environment variables in dynamic graphs exert a varying effect on evolution patterns, and the failure of models to capture those information exacerbates this issue. In other words, some key environment variables often remain inactive in historical observations, but their emergence in future observations leads to significant changes in evolutionary trends. As shown in Fig. 2, clear weather often does not alter human interest patterns, resulting in their inactivity. However, as winter comes, previously inactive weather variables become activated, leading to significant shifts in human interests. Therefore, identifying environment variables with both inactive and active properties is a crucial step in addressing the concept shift.

In this work, we propose a state-driven dynamic graph learning framework to address the issue of concept shift. **Firstly**, from the perspective of information theory [Jaynes, 1957], we theoretically investigate how to infer the environment variables that cause concept shift issue. Further, we summarize two practical guidelines to model the environment variables. **Secondly**, we introduce the insight of spatio-temporal state to uniformly model all potential environment variables in dynamic graphs. By designing a prefix-suffix state learning mechanism, we can bidirectionally capture environment variables. **Thirdly**, to refine the environments leading to concept shift, we propose a hierarchical state compressor to infer those pivotal environments with both inactive and active properties. By constructing hierarchical state learning tasks, we can maximize the extraction of variables from historical observations that may alter evolutionary patterns. **Finally**, we design a skip-connection state prediction framework to alleviate the limitations faced by parameterized

models in addressing concept shift. **Empirically**, we select seven datasets from different domains to validate the effectiveness of our model. **Samen** achieves better results than all baselines on most datasets. Moreover, we filter all samples exhibiting concept shift from two cross-domain datasets. By comparing the performance of different methods on those samples, we corroborate the excellent generalization in addressing the concept shift issue. Our contributions are as follows:

- We conduct a theoretical analysis of the variational upper bounds of parameterized estimation methods to address the concept shift issue. Based on this, we introduce two primary practical guidelines to address this challenge.
- We model environment variables from the perspective of state space, and propose a two-stage state modeling strategy to infer the key states that contribute to concept shift. Technically, we introduce a novel prefix-suffix collaborative state learning mechanism, and a hierarchical state compressor to infer the key states.
- We design extensive experiments to validate the superior generalization in concept shift scenarios.

2 Preliminaries and Backgrounds

2.1 Preliminaries

Notation. We define a temporal graph with N nodes over T steps as $\mathbf{TG} = \{\mathcal{G}^t\}_{t=1}^T$. Each graph snapshot at time step t is indicated as $\mathcal{G}^t = (X^t, A^t)$, where $X^t \in \mathbb{R}^{N \times d}$ denotes the node-wise features, and $A^t \in \mathbb{R}^{N \times N}$ indicates the connections between nodes, where d uniformly represents features and representations dimensions.

Problem definition. Temporal graphs learning focuses on summarizing the evolutionary patterns from historical snapshots and predicting the future trends at subsequent K steps. Specific tasks include future node properties forecast and link prediction. With the historical observations $\mathbf{X} = \mathcal{G}^{1:T}$ as input, the temporal graph neural network learns future evolution patterns $\mathbf{Y} = \mathcal{G}^{T+1:T+K}$, aiming to predict $\hat{\mathbf{Y}} \leftarrow \mathbf{X}$.

2.2 Backgrounds

Concept shift. Different from data distribution shift issue which has been widely studied [Yuan *et al.*, 2023; Xia *et al.*, 2023; Wu *et al.*, 2022; Yu *et al.*, 2023; Chen *et al.*, 2024; Li *et al.*, 2022; Chen *et al.*, 2022], the concept shift has different characteristics and reasons. Data distribution shift issue denotes the differences between the training and test data distributions, i.e., $\mathbb{P}(\mathbf{TG}_{train}) \neq \mathbb{P}(\mathbf{TG}_{te})$ [Chen *et al.*, 2024; Chen *et al.*, 2022; Li *et al.*, 2022]. However, concept shift is manifested in changes to the mapping distribution between the data and its labels,

$$\mathbb{P}(\kappa) \neq \mathbb{P}(\kappa'), \quad (1)$$

where $\mathbf{X} \xrightarrow{\kappa} \mathbf{Y}$ and $\mathbf{X} \xrightarrow{\kappa'} \mathbf{Y}'$. In other words, concept shift is an intra-sample phenomenon, whereas data distribution shift reflects inter-sample differences. Therefore, the differing causes of distribution shifts lead to distinct technical approaches for addressing these issues. The key to resolving

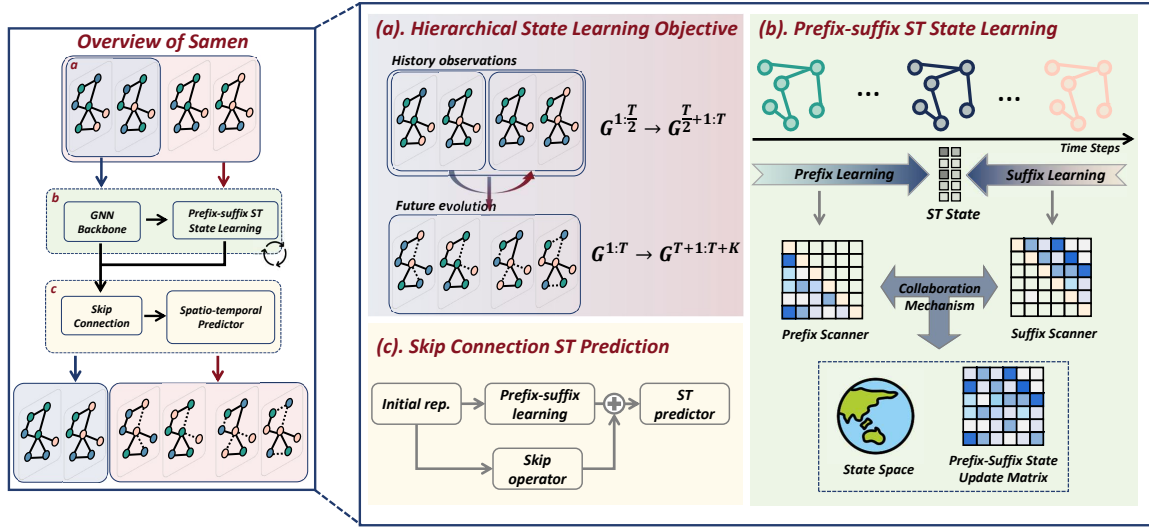


Figure 3: Left Panel: the overview of Samen. Right Panel: the detailed pipeline of Samen. Based on a hierarchical state learning objective, Samen mainly includes a prefix-suffix collaborative state learning mechanism and a skip-connection spatio-temporal prediction module.

data distribution shift lies in discovering the common feature among samples (invariant learning technique) [Zhou *et al.*, 2023; Wu *et al.*, 2022], while addressing concept shift involves uncovering additional features within the samples to bridge the gaps in the shifted mapping relationships (environment inference technique). In this work, we focus on a theoretical and practical investigation of the environment variables that contribute to concept shift from the perspective of state space.

State Space Model (SSM). SSM is a mathematical framework used to describe the dynamic evolution of a system [Gu *et al.*, 2020; Smith *et al.*, 2022; Zhang *et al.*, 2023; Gu and Dao, 2023; Gu *et al.*, 2021a]. It is widely applied in areas such as signal processing, control systems, and time series analysis. By utilizing a set of state variables, SSM represents the internal state of a system at any given point, allowing for the prediction of the system’s future behavior or observations [Kantas *et al.*, 2015]. Most practices of SSM consist of two main components: the state transition equation (Eq. 2), which updates state based on the system evolves, and the output equation (Eq. 3), which relates the updated states to the final output.

$$h'(t) = Ah(t) + Bx(t), \quad (2)$$

$$y(t) = Ch'(t) + Dx(t), \quad (3)$$

where A, B, C, D are learnable parameters most practices [Gu *et al.*, 2021a], $h(t)$ and $x(t)$ denotes the state and input at step t respectively.

One of the most significant advantages of SSM is that it can effectively captures long-term correlation [Gu and Dao, 2023]. This success typically stems from two main objectives: the first is to efficiently learn prefix states, and the second is to adaptively adjust the update weight based on the actual sequence information. Formally, given a binary associative prefix operator \circ (i.e., $(a \circ b) \circ c = a \circ (b \circ c)$) and a sequence of T graphs $[G^1, G^2, \dots, G^T]$, the prefix scanner

of SSM is fed with historical observations to return the state sequences,

$$[G^1, (G^1 \circ G^2), \dots, (G^1 \circ G^2 \circ \dots \circ G^T)]. \quad (4)$$

Our proposed state-driven environment inference framework is built upon the foundation of SSM, but it is strictly different from existing works. Specifically, we observe that the state in dynamic graphs exhibit a unique duality. This requires that comprehensive state modeling must incorporate the cooperative learning mechanism of both prefix and suffix states.

3 Related Work

Generalization of Dynamic Graph Models. The issue of generalization is a significant challenge currently faced by dynamic graph learning methods. Numerous studies have focused on developing dynamic graph learning approaches with stronger generalization capabilities. Existing work primarily focuses on addressing out-of-distribution (OOD) generalization, i.e., $\mathbb{P}(\mathbf{TG}_{train}) \neq \mathbb{P}(\mathbf{TG}_{te})$ [Chen *et al.*, 2024; Chen *et al.*, 2022; Li *et al.*, 2022; Yuan *et al.*, 2023; Xia *et al.*, 2023]. Unlike existing works, we focus on concept shift in dynamic graphs. Concept shift is manifested in changes to the mapping distribution between the data and its labels $\mathbb{P}(\kappa) \neq \mathbb{P}(\kappa')$, where $\mathbf{X} \xrightarrow{\kappa} \mathbf{Y}$ and $\mathbf{X} \xrightarrow{\kappa'} \mathbf{Y}'$. In this work, we focus on solving the concept shift problem on dynamic graphs.

State Space Model. The State Space Model (SSM) is a mathematical model used to describe dynamic systems, which has a long history of development [Welch, 1995]. With the development of deep learning, SSM has been widely used in many time series scenarios [Rangapuram *et al.*, 2018; Gu *et al.*, 2020; Gu *et al.*, 2021b; Gu *et al.*, 2021a; Smith *et al.*, 2022]. Among these studies, Mamba [Gu and Dao, 2023] has made a significant impact across various research

fields, due to its adaptive selectivity and parallel scanning capabilities, which deliver superior efficiency and performance. Actually, all variants of SSM solely focus on learning prefix states. Based on a thorough analysis and understanding of dynamic graphs in the real world, we observe the duality of spatio-temporal states. This insight inspires us to propose a prefix-suffix collaborative state learning model.

4 Upper Bound for Tackling Concept Shift

The issue of concept shift in dynamic graphs fundamentally arises from the presence of unseen environment variables. Therefore, understanding the upper bound of the model in addressing concept shift is primary for tackling this challenge.

We first define the ground-truth distribution of historical observations \mathbf{X} and future evolution \mathbf{Y} as $\mathbb{Q}_{\mathbf{X}}$ and $\mathbb{Q}_{\mathbf{Y}}$. Our goal is to obtain an approximate $\hat{\mathbb{Q}}$, which can align the evolution patterns of $\mathbb{Q}_{\mathbf{X}}$ and $\mathbb{Q}_{\mathbf{Y}}$. This objective can be formalized as,

$$\hat{\mathbb{Q}} = \arg \min_{\mathbb{Q}} -I(\mathbf{X}, \mathbf{E}; \mathbf{Y}). \quad (5)$$

To obtain this estimate, we need to understand the role of the environment variable \mathbf{E} in historical observations \mathbf{X} and future evolution \mathbf{Y} . Therefore, the objective of Eq. 5 can be updated as,

$$\hat{\mathbb{Q}} = \arg \min_{\mathbb{Q}} -I(\mathbf{X}, \mathbf{E}; \mathbf{Y}) - I(\mathbf{E}; \mathbf{Y}) + I(\mathbf{X}; \mathbf{E}). \quad (6)$$

For the first term $I(\mathbf{X}, \mathbf{E}; \mathbf{Y})$ of Eq. 6, we can obtain the lower bound by considering the parameterized variational estimations,

$$\begin{aligned} I(\mathbf{X}, \mathbf{E}; \mathbf{Y}) &= \mathbb{E}_{\mathbf{X}, \mathbf{Y}, \mathbf{E}} \left[\log \frac{\mathbb{P}(\mathbf{Y}|\mathbf{E}, \mathbf{X})}{\mathbb{P}(\mathbf{Y})} \right] \\ &\geq \mathbb{E}_{\mathbf{X}, \mathbf{E}} [\log \mathbb{P}(\mathbf{Y}|\mathbf{X}, \mathbf{E})] + H(\mathbf{Y}). \end{aligned} \quad (7)$$

For the second term $I(\mathbf{E}; \mathbf{Y})$ of Eq. 6, there exists,

$$\begin{aligned} I(\mathbf{E}; \mathbf{Y}) &= \mathbb{E}_{\mathbf{E}, \mathbf{Y}} \left[\log \frac{\mathbb{P}(\mathbf{Y}|\mathbf{E})}{\mathbb{P}(\mathbf{Y})} \right] \\ &\geq \mathbb{E}_{\mathbf{E}, \mathbf{Y}} [\log \mathbb{P}_{\theta}(\mathbf{Y}|\mathbf{E})] + H(\mathbf{Y}). \end{aligned} \quad (8)$$

For the third term $I(\mathbf{X}; \mathbf{E})$ of Eq. 6, there exists,

$$\begin{aligned} I(\mathbf{X}; \mathbf{E}) &= \mathbb{E}_{\mathbf{X}, \mathbf{E}} \left[\log \frac{\mathbb{P}(\mathbf{E}|\mathbf{X})}{\mathbb{P}(\mathbf{E})} \right] \\ &\leq \mathbb{E}_{\mathcal{G}} [\text{KL}(\mathbb{P}_{\phi}(\mathbf{E}|\mathcal{G})||\mathbb{P}(\mathbf{E}))]. \end{aligned} \quad (9)$$

Given the above derivations, we can further update the objective of Eq. 5 by formalize the variational lower bound,

$$\begin{aligned} \hat{\mathbb{Q}} &= \arg \min_{\mathbb{Q}} \mathbb{E}[-\log \mathbb{P}(\mathbf{Y}|\mathbf{X}, \mathbf{E}) + \log \mathbb{P}_{\phi}(\mathbf{E}|\mathcal{G}) \\ &\quad - \log \mathbb{P}_{\theta}(\mathbf{Y}|\mathbf{E}) + \text{KL}(\mathbb{P}_{\phi}(\mathbf{E}|\mathcal{G})||\mathbb{P}(\mathbf{E}))]. \end{aligned} \quad (10)$$

The parameterized variational estimates $\mathbb{P}_{\theta}(\mathbf{Y}|\mathbf{E})$ and $\mathbb{P}_{\phi}(\mathbf{E}|\mathbf{X})$ establish the inference line $\langle \mathbf{X} \rightarrow \mathbf{E} \rightarrow \mathbf{Y} \rangle$. The $\mathbb{P}(\mathbf{Y}|\mathbf{X}, \mathbf{E})$ highlights the inference process $\langle \mathbf{X} \rightarrow \mathbf{Y} \leftarrow \mathbf{E} \rangle$. The inductive bias $\mathbb{P}(\mathbf{E})$ highlights the extraction process of the environment variables $\langle \mathbf{X} \rightarrow \mathbf{E} \rangle$. Therefore, current parameterized models that bridge the gap between \mathbf{X} and \mathbf{Y} are inherently constrained by modeling $\mathbb{P}(\mathbf{E})$ and $\mathbb{P}(\mathbf{Y}|\mathbf{X}, \mathbf{E})$.

Practical Guidance. The theoretical analysis above provides us with practical insights in tackling concept shift issue.

- Understanding the inductive bias $\mathbb{P}(\mathbf{E})$ of concept shifts is essential for achieving accurate inference of environment variables from history observations $\langle \mathbf{X} \rightarrow \mathbf{E} \rangle$.
- Adopting the collaborative inference lines of $\langle \mathbf{X} \rightarrow \mathbf{E} \rightarrow \mathbf{Y} \rangle$ and $\langle \mathbf{X} \rightarrow \mathbf{Y} \leftarrow \mathbf{E} \rangle$ to model the unseen environment variables that cause concept shift provides theoretical insights.

In the following research, we build on these two practical guidances to enhanced generalization capability in addressing concept shift issue.

5 Environment Inference from the Perspective of Spatio-Temporal State

In this section, we aim to infer the environment variables \mathbf{E} that cause concept shift issue, i.e., the inference path $\langle \mathbf{X} \rightarrow \mathbf{E} \rangle$. This corresponds to the first practical guidance derived from our theoretical analysis. Specifically, we adopt a two-stage modeling strategy, initially extracting environment variables comprehensively and subsequently refining the information that causes concept shift issue.

5.1 Prefix-Suffix Collaborative State Learning Mechanism

In this subsection, we adopt a two-stage modeling strategy, initially extracting environment variables comprehensively and subsequently refining the information that causes concept shift issue. We first introduce the spatio-temporal state (ST state) to construct a unified perspective for capturing environment information \mathbf{E} in dynamic graphs. This design is inspired by the exciting performance achieved in the studies of sequence data from the perspective of state space. In dynamic graphs, the ST state exhibit a *duality*, characterized by both *continuity* and *hysteresis*.

- The *continuity* of ST state signifies that each individual state is often reflected by the accumulation of past state patterns. For example, the timestamp state is typically the cumulative result of historical evolutionary processes.
- The *hysteresis* of ST state means that certain individual state information is retroactively extrapolated from future evolution patterns. For instance, unrecorded weather conditions or individual affinities can only be inferred through future evolutionary processes.

Given the *duality* of ST state, we propose a prefix-suffix collaborative state learning method. This design incorporates a state learning operator that effectively models prefix information while also enabling the inference of suffix information, distinguishing it from traditional State Space Models (SSMs) that focus exclusively on learning prefix information. Additionally, existing works have confirmed that achieving data-adaptive state updates is equally important [Gu and Dao, 2023]. This means the model should dynamically adjust the weights of state updates based on the characteristics of sequence data. Consequently, we aim to implement a ST state learning framework that satisfies the following two requirements.

- **Bidirectional state learning.** The ability to capture both prefix and suffix information in the state learning process.
- **Data-adaptive state updating.** A mechanism that adjusts the state updates dynamically based on the characteristics of the input sequence.

To achieve this target, we propose a Bidirectional Mixing Mechanism (BMM), which encompasses a bidirectional mixing aggregation process and a adaptive state update operator. We first design a bidirectional aggregation kernel to achieve prefix-suffix state aggregation.

Bidirectional feature mixing kernel. Given the parameterized graph neural network backbone GNN_α , we first conduct spatial encoding on the graph, resulting in a temporal static representation $Z^t = \text{GNN}_\alpha(\mathcal{G}^t) \in \mathbb{R}^{N \times d}$ at step t . Next, we achieve bidirectional mixing of local steps through a message-passing kernel,

$$\hat{Z}_{pre}^t = [Z^t || Z^{t-1}] \cdot [I || W_\psi]^T, \quad (11)$$

$$\hat{Z}_{suf}^t = [Z^t || Z^{t+1}] \cdot [I || W_\xi]^T, \quad (12)$$

where $\hat{Z}_{pre}^1 = Z^1$ and $\hat{Z}_{suf}^T = Z^T$, $[\cdot || \cdot]$ indicates the concatenation operation. $W_\xi \in \mathbb{R}^{d \times d}$ and $W_\psi \in \mathbb{R}^{d \times d}$ are formalized as learnable aggregation kernel. Although the aggregation between the local sequences does not capture the long-term evolutionary patterns, it distinctly determines the direction of information compression. Building on this certain direction, we achieve state updates for long-term observations by amplifying the dependencies of the short sequences

State update operation. After each bidirectional feature mixing layer, we perform a global state update based on the local mixed information. Inspired by the successful practices of existing state space models, we introduce learnable state transition matrix $A \in \mathbb{R}^{N \times d \times d}$ and input control matrix $B \in \mathbb{R}^{N \times d \times d}$ to update state,

$$S_{pre}^t(n) = A_n S_{pre}^{t-1}(n) + B_n \hat{Z}_{pre}^{t-1}(n), \quad (13)$$

$$S_{suf}^t(n) = A_n S_{suf}^{t+1}(n) + B_n \hat{Z}_{suf}^{t+1}(n), \quad (14)$$

where $S_{pre}^1 = \hat{Z}_{suf}^1$, $S_{suf}^T = \hat{Z}_{pre}^T$ and $n \in [1, \dots, N]$ denotes the node. Thus, we can obtain the states derived from both prefix learning and suffix learning, i.e., $S_{pre} \in \mathbb{R}^{T \times N \times D}$ and $S_{suf} \in \mathbb{R}^{T \times N \times D}$. We then couple them into a unified state space representation,

$$S = S_{pre} \odot S_{suf}, \quad (15)$$

where \odot is the element-wise product. Guided by the *duality* of ST state in dynamic graphs, we comprehensively model the ST state $S = \{S^1, S^2, \dots, S^T\} \in \mathbb{R}^{T \times N \times d}$. The next challenge lie in refining the information that contributes to conceptual distribution shift.

5.2 Hierarchical Compressor for Inferring Activable States

We start by investigating the characteristics of those states that contribute to the concept shift issue. We argue that the concept shift is controlled by those ST states, which remain inactive in historical observations but become active in future

evolution. Therefore, we propose a hierarchical compressor for inferring activable states. Specifically, we conduct a fine-grained partition of the historical data $\mathcal{G}^{1:T}$ to create a hierarchical history-future data form $\mathcal{G}^{1:T} = (\mathcal{G}^H, \mathcal{G}^F)$, where $H \cup F = \{1, \dots, T\}$. In the implementation, H and F are divided equally over the set $\{1, \dots, T\}$. Therefore, we construct a hierarchical learning tasks,

$$\mathcal{G}^H \rightarrow \mathcal{G}^F, \quad \mathcal{G}^{1:T} \rightarrow \mathcal{G}^{T+1:T+K}. \quad (16)$$

We then obtain corresponding training objectives,

$$\mathcal{L}_{obs} := \mathcal{L}(\mathcal{G}^H \rightarrow \mathcal{G}^F), \quad \mathcal{L}_{fut} := \mathcal{L}(\mathcal{G}^{1:T} \rightarrow \mathcal{G}^{T+1:T+K}). \quad (17)$$

In this section, we undertake a comprehensive investigation of environment variables \mathbf{E} from the perspective of state space. We innovatively summarize two pivotal inductive biases: the *duality* of states and the *activatable property* of states that contribute to concept shift. We then propose an effective modeling framework for the spatio-temporal states leading to the concept shift issue.

6 Skip-connection Dynamic Graph Prediction

In this section, we aim to develop a dynamic graph prediction module that leverages the inferred states to address concept shift issue, i.e., the collaborative inference paths of $\langle \mathbf{X} \rightarrow \mathbf{E} \rightarrow \mathbf{Y} \rangle$ and $\langle \mathbf{X} \rightarrow \mathbf{Y} \leftarrow \mathbf{E} \rangle$. This corresponds to the second practical guidance derived from our theoretical analysis. Based on the definition of conditional probability, we can derive it by marginalizing over \mathbf{X} ,

$$\mathbb{P}(\mathbf{Y}|\mathbf{E}) = \sum_{\mathbf{X}} \mathbb{P}(\mathbf{Y}|\mathbf{X}, \mathbf{E}) \mathbb{P}(\mathbf{X}|\mathbf{E}). \quad (18)$$

Thus, $\mathbb{P}(\mathbf{Y}|\mathbf{E})$ can be obtained by marginalizing over \mathbf{X} . This involves integrating the probabilities of \mathbf{X} to derive the probability distribution of \mathbf{Y} given \mathbf{E} . We can further utilize a unified parameterized framework $\mathbb{P}(\mathbf{Y}|\hat{\mathbf{E}})$ to align the collaborative mechanisms ($\langle \mathbf{X} \rightarrow \mathbf{E} \rightarrow \mathbf{Y} \rangle$ & $\langle \mathbf{X} \rightarrow \mathbf{Y} \leftarrow \mathbf{E} \rangle$) for predicting future evolution, where $\hat{\mathbf{E}}$ is condensed from \mathbf{X} and \mathbf{E} .

To achieve this target, we propose a skip-connection dynamic graph prediction framework. Given the initial representation $Z \in \mathbb{R}^{T \times N \times d}$ obtained from the GNN encoder and the ST state $S \in \mathbb{R}^{T \times N \times d}$ learned from the prefix-suffix modeling mechanism, we implement a connection based on skip manner,

$$\hat{S} = S \oplus Z \cdot D, \quad (19)$$

where $D \in \mathbb{R}^{d \times d}$ represents learnable parameters demoting the skip weights and \oplus is the element-wise addition. Then, skip-connected output $\hat{S} \sim \hat{\mathbf{E}}$ will be fed into a spatio-temporal prediction layer \mathbb{P}_θ to predict future evolution. The learning objective \mathcal{L}_{fut} for predicting future evolution \mathbf{Y} based on historical observations \mathbf{X} is,

$$\mathcal{L}_{fut} = -\mathbb{E}[\log \mathbb{P}_\theta(\mathbf{Y}|\hat{S})] + \mathbb{E}[\log(\mathbb{P}_\phi(\hat{S}|\mathbf{X}))], \quad (20)$$

where $\phi = \{\alpha, W_\psi, W_\xi, A, B, D\}$ denotes the parameterized prefix-suffix state learning mechanism, θ indicates the spatio-temporal prediction process. Therefore, the learning objective of the auxiliary task is,

$$\mathcal{L}_{obs} = -\mathbb{E}[\log \mathbb{P}_\theta(\mathcal{G}^U|\hat{S}^U)] + \mathbb{E}[\log(\mathbb{P}_\phi(\hat{S}^H|\mathcal{G}^H))]. \quad (21)$$

Model	COLLAB	Yelp	ACT
DySAT	88.77±0.23	78.87±0.57	78.52±0.40
VREx	88.31±0.32	79.04±0.16	83.11±0.29
GroupDRO	88.76±0.12	79.38±0.42	85.19±0.53
DIDA	91.97±0.05	78.22±0.40	89.84±0.82
EAGLE	92.45±0.21	78.97±0.31	92.37±0.53
DGIB	92.68±0.20	79.53±0.20	94.89±0.20
Samen	92.71±0.15	79.96±0.21*	94.41±0.19

Table 1: AUC score (%) of future link prediction task on real-world social relationship datasets. The best results are shown in bold and the second best results are underlined. * indicates statistical significance against the second-best results.

The overall training objectives of the proposed **Samen** is,

$$\mathcal{L}_{ST} = \mathcal{L}_{fut} + \mathcal{L}_{obs}. \quad (22)$$

7 Experiment

We evaluate our **Samen** by answering the following questions. **Q1**. Does our **Samen** demonstrate competitiveness compared to existing state-of-the-art methods on real-world datasets? **Q2**. Does our approach effectively tackle the issues that remain unresolved in existing works? **Q3**. Does each component in our **Samen** effectively enhance the generalization capacity? **Q4**. Does our framework operate with high efficiency?

7.1 Experiment Setup

In this subsection, we introduce datasets, baselines and experiment settings briefly.

Datasets. We employ seven cross-domain real-world dynamic graph datasets to evaluate our **Samen**. COLLAB [Tang *et al.*, 2012] is an academic collaboration dataset consisting of papers published over a span of 16 years. Yelp [Sankar *et al.*, 2020] is a dataset that contains business reviews, while ACT [Kumar *et al.*, 2019] tracks students’ activities on a MOOC platform over a 30-day period. PEMS08 and PEMS04 [Song *et al.*, 2020] are classic traffic network datasets from California, featuring 5-minute intervals. SD-2019 and GBA-2019 [Liu *et al.*, 2023] are newly introduced large-scale traffic network datasets.

Baselines. We compare **Samen** against two categories of baselines: six social link forecasting methods and six traffic flow prediction frameworks. Social link forecasting methods consist of DySAT [Sankar *et al.*, 2020], VREx [Krueger *et al.*, 2021], GroupDRO [Sagawa *et al.*, 2019], DIDA [Yuan *et al.*, 2023], EAGLE [Yuan *et al.*, 2023], DGIB [Yuan *et al.*, 2024]. Traffic flow prediction baselines include AGCRN [Bai *et al.*, 2020], GWNET [Wu *et al.*, 2019], Z-GCNETs [Chen *et al.*, 2021], D2STGNN [Shao *et al.*, 2022b], STGNCDE [Choi *et al.*, 2022], CaST [Xia *et al.*, 2023].

Experiment Settings. We filter samples exhibiting concept shift and investigate the effectiveness of our **Samen** on concept shift scenario.

- The task of social relationship analysis is to exploit past graphs to make link prediction in the next time step. The principle for selecting samples that exist concept shift is as

follows. For a given sample (X, Y) , if there exists another sample (X', Y') satisfying $X = X'$ and $Y \neq Y'$, the current sample will be identified as exhibiting concept shift.

- The traffic flow forecast task is to predict the next 12 steps based on historical 12 steps observations (12 \rightarrow 12). The principle for selecting samples that exist concept shift: given a sample (X, Y) , if there exists another sample (X', Y') satisfying $MAD(X, X') \leq \sigma$ and $MAD(Y, Y') \geq 10\sigma$, the current sample will be identified as exhibiting concept shift¹.

7.2 Performance Analysis on Real-world Datasets

We evaluate the performance of **Samen** across seven real-world datasets from different domains to answer **Q1**.

Social link forecasting. Tab. 1 presents the performance of **Samen** on social link prediction tasks. Our model outperforms all baselines on two datasets. We also observe that DGIB [Yuan *et al.*, 2024] achieves one best performance on ACT, and DGIB [Yuan *et al.*, 2023] achieve many sub-optimal results. It proves that the approaches of tackling data out-of-distribution issue can also bring some insights into solving concept shift.

Traffic flow prediction. Tab. ?? shows the results of **Samen** on traffic flow dataset. We can observe that **Samen** outperforms all baselines on two datasets. We note that CaST obtains suboptimal results on most datasets and even optimal results on PEMS04. This indicates that practices centered on inferring environment variables may offer potential solutions for addressing both data distribution shift and concept shift. Based on the comparison of different methods across datasets from two domains, we observe that **Samen** is competitive with existing state-of-the-art frameworks. More importantly, we emphasize the performance of our method in addressing challenges that have not been tackled by previous works.

7.3 Generalization Analysis

To answer **Q2**, we compare the performance degradation of **Samen** with existing methods on samples affected by concept shift. We adopt the concept shift samples filtered from the Yelp and SD-2019 datasets, as used in Fig. 1. As shown in Fig. 4, we provide the performance comparison of the models on the raw dataset and the concept-shifted sample set. We can observe that **Samen** exhibits the least performance degradation on the samples with concept shift, and significantly outperforms all existing methods. We can conclude that our framework demonstrates excellent generalization ability in countering concept shift issue.

7.4 Ablation Study

To answer **Q3**, we investigate each component of **Samen**. Specifically, we conduct ablation studies to explore the effectiveness of prefix-suffix cooperative state learning mechanism, hierarchical compressor and skip-connection dynamic graph predictor. We design three variants of **Samen** for each of the three main modules, i.e. *w.o. pre-suf*, *w.o.hicom* and

¹MAD denotes Mean Absolute Distance, which follows the same calculation method as MAE (Mean Absolute Error). In practice, σ is set to 0.5.

Model	PEMS08		PEMS04		SD-2019		GBA-2019	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
GWNET	16.67±0.9	26.77±1.1	20.69±0.8	33.02±1.3	17.74±1.1	29.62±2.9	20.91±2.6	33.41±2.3
AGCRN	15.95±0.2	25.75±0.6	19.83±0.3	32.26±0.2	18.09±1.2	32.01±1.3	21.01±1.6	34.25±2.0
Z-GCNets	15.76±0.3	26.31±0.2	19.50±0.5	31.91±0.7	18.21±1.7	32.76±1.8	21.84±1.2	35.12±2.2
D2STGNN	15.69±0.3	26.41±0.4	19.55±0.7	31.99±0.5	17.85±1.5	29.51±1.7	20.71±2.0	33.65±2.7
STGNCDE	15.32±0.6	26.38±0.3	18.94±0.6	32.39±0.4	19.55±0.7	33.57±1.5	21.79±1.5	35.37±2.4
CaST	15.08±0.3	25.16±0.4	18.56±0.9	31.44±0.3	17.61±1.3	29.12±1.4	21.30±1.5	34.67±1.5
Samen	14.70±0.2*	24.26±0.6*	17.91±0.4	<u>31.62±0.3</u>	17.18±1.2	28.32±1.4	20.26±1.5	<u>33.54±1.8</u>

Table 2: The performance of traffic prediction tasks (12 → 12) on four real-world datasets. The best results are shown in bold and the second best results are underlined. * indicates statistical significance against the second-best results.

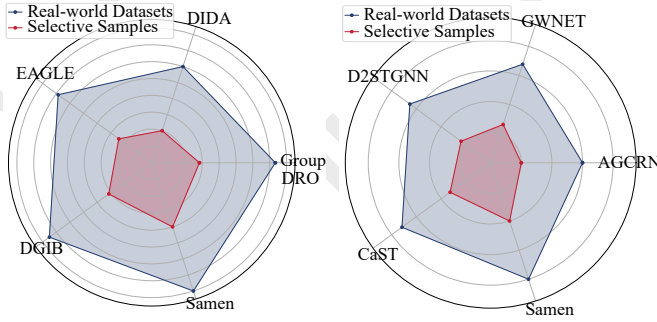


Figure 4: The performance discount of different methods on both the raw dataset and shifted dataset.

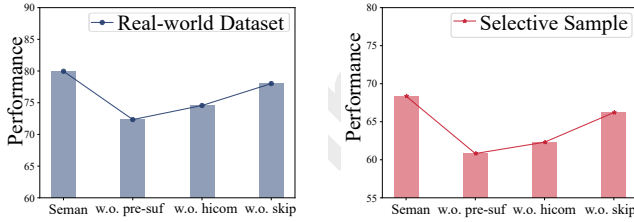


Figure 5: Ablation Study. We design three variants of Samen and compare their performance on raw Yelp (left) and sample set with concept shift (right).

w.o. skip. They respectively represent the variant that removes suffix state learning, the variant that removes the hierarchical compressor and focuses solely on single-prediction mode, and the variant that remove skip connections. As shown in Fig. 5, we observe significant performance degradation in each variant, with the degradation being particularly pronounced on samples affected by concept shift. By comparing the degree of performance degradation, we find that prefix-suffix state learning module plays the most critical role, followed by the hierarchical state compressor. Therefore, we infer that learning suffix state in dynamic graph scenarios helps the model to comprehensively capturing the spatio-temporal evolution patterns.

7.5 Efficiency Analysis

The time complexity of **Samen** is $\mathcal{O}(T \cdot L \cdot N(2d^2 + 1) + T \cdot L \cdot |\mathcal{E}|)$, where N represents the number of nodes, $|\mathcal{E}|$ denotes

the number of edges, d is the feature dimension, T represents the step of history observation, L denotes number of layers of the prefix-suffix state learning mechanism. The time cost of our **Samen** primarily lies in the ST state learning process with $\mathcal{O}(2 \cdot T \cdot L \cdot N \cdot d^2)$, and the time complexity for graph learning with $\mathcal{O}(L \cdot (N + \mathcal{E}))$. Our method has linear time complexity with high training efficiency. As shown in Fig. 3, we empirically conduct efficiency comparisons of **Samen**, DIDA, and EAGLE, measuring the time taken per epoch. We observe that the training efficiency of our method is competitive with existing approaches.

Models	DIDA	EAGLE	Samen
COLLAB	11.27	12.23	10.83
Yelp	7.92	6.89	7.12
ACT	10.27	9.34	9.23

Table 3: The training efficiency of Samen with other baselines on COLLAB, Yelp and ACT (s/epoch).

8 Conclusion

In this work, we theoretically unveil the intractable challenge posed by concept shift in dynamic graphs, and propose a state-driven environment inference framework (**Samen**) to address this issue. Building upon a comprehensive investigation of concept shift from an information theory perspective, we distill two practical guidelines to design dynamic graph learning model with enhanced generalization ability. Unifying the environment factors in dynamic graph within the state space, we propose a prefix-suffix collaborative state learning mechanism to bidirectionally capture the environment information, and introduce a hierarchical state compressor to refine the environment information leading to concept shift. Extensive experiments over seven datasets verify the effectiveness of our model.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.12227901), Natural Science Foundation of Jiangsu Province (BK.20240460, BK.20240461), the grant from State Key Laboratory of Resources and Environmental Information System.

References

- [Bai *et al.*, 2020] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33:17804–17815, 2020.
- [Cao *et al.*, 2020] Da Cao, Yawen Zeng, Meng Liu, Xiangnan He, Meng Wang, and Zheng Qin. Strong: Spatio-temporal reinforcement learning for cross-modal video moment localization. In *Proceedings of the 28th ACM international conference on multimedia*, pages 4162–4170, 2020.
- [Chen *et al.*, 2021] Yuzhou Chen, Ignacio Segovia, and Yulia R Gel. Z-gcnets: Time zigzags at graph convolutional networks for time series forecasting. In *International Conference on Machine Learning*, pages 1684–1694. PMLR, 2021.
- [Chen *et al.*, 2022] Yongqiang Chen, Yonggang Zhang, Yatao Bian, Han Yang, MA Kaili, Binghui Xie, Tongliang Liu, Bo Han, and James Cheng. Learning causally invariant representations for out-of-distribution generalization on graphs. *Advances in Neural Information Processing Systems*, 35:22131–22148, 2022.
- [Chen *et al.*, 2024] Yongqiang Chen, Yatao Bian, Kaiwen Zhou, Binghui Xie, Bo Han, and James Cheng. Does invariant graph learning via environment augmentation learn invariance? *Advances in Neural Information Processing Systems*, 36, 2024.
- [Choi *et al.*, 2022] Jeongwhan Choi, Hwangyong Choi, Jeehyun Hwang, and Noseong Park. Graph neural controlled differential equations for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6367–6374, 2022.
- [Gao *et al.*, 2022] Chen Gao, Xiang Wang, Xiangnan He, and Yong Li. Graph neural networks for recommender system. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, pages 1623–1625, 2022.
- [Gu and Dao, 2023] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
- [Gu *et al.*, 2020] Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra, and Christopher Ré. Hippo: Recurrent memory with optimal polynomial projections. *Advances in neural information processing systems*, 33:1474–1487, 2020.
- [Gu *et al.*, 2021a] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.
- [Gu *et al.*, 2021b] Albert Gu, Isys Johnson, Karan Goel, Khaled Saab, Tri Dao, Atri Rudra, and Christopher Ré. Combining recurrent, convolutional, and continuous-time models with linear state space layers. *Advances in neural information processing systems*, 34:572–585, 2021.
- [Jaynes, 1957] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [Kantas *et al.*, 2015] Nikolas Kantas, Arnaud Doucet, Sumeetpal S Singh, Jan Maciejowski, and Nicolas Chopin. On particle methods for parameter estimation in state-space models. 2015.
- [Krueger *et al.*, 2021] David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghui Zhang, Remi Le Priol, and Aaron Courville. Out-of-distribution generalization via risk extrapolation (rex). In *International Conference on Machine Learning*, pages 5815–5826. PMLR, 2021.
- [Kumar *et al.*, 2019] Srikanth Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1269–1278, 2019.
- [Li *et al.*, 2022] Haoyang Li, Ziwei Zhang, Xin Wang, and Wenwu Zhu. Learning invariant graph representations for out-of-distribution generalization. *Advances in Neural Information Processing Systems*, 35:11828–11841, 2022.
- [Liu *et al.*, 2023] Xu Liu, Yutong Xia, Yuxuan Liang, Junfeng Hu, Yiwei Wang, Lei Bai, Chao Huang, Zhenguang Liu, Bryan Hooi, and Roger Zimmermann. Largest: A benchmark dataset for large-scale traffic forecasting. *arXiv preprint arXiv:2306.08259*, 2023.
- [Liu *et al.*, 2024] Chenxi Liu, Sun Yang, Qianxiong Xu, Zhishuai Li, Cheng Long, Ziyue Li, and Rui Zhao. Spatial-temporal large language model for traffic prediction. In *25th IEEE International Conference on Mobile Data Management*, pages 31–40, 2024.
- [Liu *et al.*, 2025a] Chenxi Liu, Hao Miao, Qianxiong Xu, Shaowen Zhou, Cheng Long, Yan Zhao, Ziyue Li, and Rui Zhao. Efficient multivariate time series forecasting via calibrated language models with privileged knowledge distillation. In *41th IEEE International Conference on Data Engineering*, 2025.
- [Liu *et al.*, 2025b] Chenxi Liu, Qianxiong Xu, Hao Miao, Sun Yang, Lingzheng Zhang, Cheng Long, Ziyue Li, and Rui Zhao. TimeCMA: Towards llm-empowered multivariate time series forecasting via cross-modality alignment. In *Thirty-Ninth AAAI Conference on Artificial Intelligence*, 2025.
- [Miao *et al.*, 2024] Hao Miao, Ziqiao Liu, Yan Zhao, Chenjuan Guo, Bin Yang, Kai Zheng, and Christian S Jensen. Less is more: Efficient time series dataset condensation via two-fold modal matching. *PVLDB*, 18(2):226–238, 2024.
- [Miao *et al.*, 2025] Hao Miao, Ronghui Xu, Yan Zhao, Senzhang Wang, Jianxin Wang, Philip S Yu, and Christian S Jensen. A parameter-efficient federated framework for streaming time series anomaly detection via lightweight adaptation. *TMC*, 2025.
- [Rangapuram *et al.*, 2018] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. *Advances in neural information processing systems*, 31, 2018.

- [Sagawa *et al.*, 2019] Shiori Sagawa, Pang Wei Koh, Tatsunori B Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. *arXiv preprint arXiv:1911.08731*, 2019.
- [Sankar *et al.*, 2020] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the 13th international conference on web search and data mining*, pages 519–527, 2020.
- [Shao *et al.*, 2022a] Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 4454–4458, 2022.
- [Shao *et al.*, 2022b] Zezhi Shao, Zhao Zhang, Wei Wei, Fei Wang, Yongjun Xu, Xin Cao, and Christian S Jensen. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *arXiv preprint arXiv:2206.09112*, 2022.
- [Smith *et al.*, 2022] Jimmy TH Smith, Andrew Warrington, and Scott W Linderman. Simplified state space layers for sequence modeling. *arXiv preprint arXiv:2208.04933*, 2022.
- [Song *et al.*, 2020] Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 914–921, 2020.
- [Sun *et al.*, 2021] Qingyun Sun, Jianxin Li, Hao Peng, Jia Wu, Yuanxing Ning, Philip S Yu, and Lifang He. Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism. In *Proceedings of the Web Conference 2021*, pages 2081–2091, 2021.
- [Tang *et al.*, 2012] Jie Tang, Sen Wu, Jimeng Sun, and Hang Su. Cross-domain collaboration recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1285–1293, 2012.
- [Wang *et al.*, 2019] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019.
- [Wang *et al.*, 2023] Xu Wang, Lianliang Chen, Hongbo Zhang, Pengkun Wang, Zhengyang Zhou, and Yang Wang. A multi-graph fusion based spatiotemporal dynamic learning framework. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 294–302, 2023.
- [Welch, 1995] G Welch. An introduction to the kalman filter. 1995.
- [Wu *et al.*, 2019] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.
- [Wu *et al.*, 2022] Qitian Wu, Hengrui Zhang, Junchi Yan, and David Wipf. Handling distribution shifts on graphs: An invariance perspective. *arXiv preprint arXiv:2202.02466*, 2022.
- [Xia *et al.*, 2023] Yutong Xia, Yuxuan Liang, Haomin Wen, Xu Liu, Kun Wang, Zhengyang Zhou, and Roger Zimmermann. Deciphering spatio-temporal graph forecasting: A causal lens and treatment. *arXiv preprint arXiv:2309.13378*, 2023.
- [Yu *et al.*, 2023] Junchi Yu, Jian Liang, and Ran He. Mind the label shift of augmentation-based graph ood generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11620–11630, 2023.
- [Yuan *et al.*, 2023] Haonan Yuan, Qingyun Sun, Xingcheng Fu, Ziwei Zhang, Cheng Ji, Hao Peng, and Jianxin Li. Environment-aware dynamic graph learning for out-of-distribution generalization. *arXiv preprint arXiv:2311.11114*, 2023.
- [Yuan *et al.*, 2024] Haonan Yuan, Qingyun Sun, Xingcheng Fu, Cheng Ji, and Jianxin Li. Dynamic graph information bottleneck. In *Proceedings of the ACM on Web Conference 2024*, pages 469–480, 2024.
- [Zhang *et al.*, 2022] Zeyang Zhang, Xin Wang, Ziwei Zhang, Haoyang Li, Zhou Qin, and Wenwu Zhu. Dynamic graph neural networks under spatio-temporal distribution shift. *Advances in Neural Information Processing Systems*, 35:6074–6089, 2022.
- [Zhang *et al.*, 2023] Michael Zhang, Khaled K Saab, Michael Poli, Tri Dao, Karan Goel, and Christopher Ré. Effectively modeling time series with simple discrete state spaces. *arXiv preprint arXiv:2303.09489*, 2023.
- [Zhou *et al.*, 2020] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. Riskoracle: A minute-level citywide traffic accident forecasting framework. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1258–1265, 2020.
- [Zhou *et al.*, 2023] Zhengyang Zhou, Qihe Huang, Kuo Yang, Kun Wang, Xu Wang, Yudong Zhang, Yuxuan Liang, and Yang Wang. Maintaining the status quo: Capturing invariant relations for ood spatiotemporal learning. 2023.