Towards OOD Generalization in Dynamic Graphs via Causal Invariant Learning

Xinxun Zhang[1,2,3], **Pengfei Jiao**[1,2], **Mengzhou Gao**[1*], **Tianpeng Li**[4], **Xuan Guo**[4]
[1]School of Cyberspace, Hangzhou Dianzi University, China
[2]State Key Laboratory of AI Safety, Beijing
[3]School of Computer Science, Wuhan University, China
[4]College of Intelligence and Computing, Tianjin University, China
xxzhangstu@whu.edu.cn, {pjiao, mzgao}@hdu.edu.cn, {ltpnimeia, guoxuan}@tju.edu.cn,

Abstract

Although dynamic graph neural networks (DyGNNs) have demonstrated promising capabilities, most existing methods ignore out-of-distribution (OOD) shifts that commonly exist in dynamic graphs. Dynamic graph OOD generalization is non-trivial due to the following challenges: 1) Identifying invariant and variant patterns amid complex graph evolution, 2) Capturing the intrinsic evolution rationale from these patterns, and 3) Ensuring model generalization across diverse OOD shifts despite limited data distribution observations. Although several attempts have been made to tackle these challenges, none has successfully addressed all three simultaneously, and they face various limitations in complex OOD scenarios. To solve these issues, we propose a **Dy**namic graph **C**ausal **I**nvariant **L**earning (DyCIL) model for OOD generalization via exploiting invariant spatio-temporal patterns from a causal view. Specifically, we first develop a dynamic causal subgraph generator to identify causal dynamic subgraphs explicitly. Next, we design a causal-aware spatio-temporal attention module to extract the intrinsic evolution rationale behind invariant patterns. Finally, we further introduce an adaptive environment generator to capture the underlying dynamics of distributional shifts. Extensive experiments on both real-world and synthetic dynamic graph datasets demonstrate the superiority of our model over state-of-the-art baselines in handling OOD shifts.

Introduction

Dynamic graphs are ubiquitous across various domains (Yang et al. 2020a; Zhao et al. 2019; Liu et al. 2025). Many dynamic graph neural networks (DyGNNs) methods have been proposed to learn meaningful node embedding, achieving remarkable success in various downstream tasks (Poursafaei et al. 2022; Huang et al. 2023; Wu, Fang, and Liao 2024; Zhang et al. 2024). Despite their great progress, most existing methods (Seo et al. 2018; Pareja et al. 2020; Sankar et al. 2020; Zhong et al. 2024) are based on the data assumption of independent identically distributed (IID), where the training and test data share the same distribution. However, in real-world dynamic graphs, distribution shifts are common due to the confounding environmental biases in their generation process. Most DyGNNs don't consider the effect of confounding biases from unstable and variant environments, limiting their ability to capture invariant spatio-temporal patterns and generalize to dynamic graphs with out-of-distribution (OOD) shifts.

To tackle this issue, and inspired by recent research in invariant learning (Arjovsky et al. 2019; Krueger et al. 2021; Wu et al. 2022b), we argue that the causal patterns in dynamic graphs to the labels remain stable and invariant across different environments. Therefore, we study generalized dynamic graph representation learning under OOD shifts by effectively capturing invariant spatio-temporal patterns from a causal view. However, it is non-trivial due to the following key challenges: (1) Identification: How can we distinguish between invariant and variant patterns in dynamic graphs amid complex evolution? (2) Extraction: How can we extract the intrinsic stable evolution rationale from the identified patterns? (3) Generalization: How can we enable the model to generalize to a wide range of unseen data with distribution shifts in the presence of limited observable data?

Several recent attempts have been made to address these challenges (Zhang et al. 2022, 2023b; Yuan et al. 2023; Yang et al. 2024; Tieu et al. 2025). However, they suffer from various issues and limitations. First, DIDA (Zhang et al. 2022), SILD (Zhang et al. 2023b), and EAGLE (Yuan et al. 2023) fail to identify causal dynamic subgraphs or explicitly learn invariant patterns, which are crucial for understanding the mechanisms behind OOD generalization in dynamic graphs and providing better interpretability. In addition, DIDA and SILD extract evolution rationales using the entire dynamic graph, which makes the evolution rationales that they extract unstable since the dynamic graph contains environmental biases. On the other hand, EAGLE and EpoD (Yang et al. 2024) overlook the evolutionary rationales of invariant patterns. Lastly, DIDA, SILD, and EpoD generate discrete and limited environment instances by using observable graph snapshots for interventions, while OOD-linker (Tieu et al. 2025) overlooks the role of environmental factors in dynamic graphs. Both limitations weaken their generalization ability in more challenging dynamic OOD scenarios. EAGLE enhances its generalization capability by inferring environment distributions to generate sufficient instances, but it requires additional labels (a mixture of time index and environment index), which limits its scalability in various

*Corresponding author

dynamic OOD scenarios.

To address the limitations of these methods while simultaneously tackling the three challenges as a whole, we first construct a structural causal model (SCM) based on the generation process of dynamic graphs to explore the causal effects among various factors. Then, we instantiate a novel **Dy**namic graph **C**ausal **I**nvariant **L**earning (DyCIL) model based on the causal relationships to handle OOD shifts in dynamic graphs via exploiting invariant spatio-temporal patterns from a causal view. DyCIL can effectively capture the invariant spatio-temporal patterns to label remain stable across various latent environments through jointly optimizing three mutually promoting modules. Specifically, we propose a dynamic causal subgraph generator to explicitly identify causal dynamic subgraphs, offering deeper insights and enhanced interpretability into the underlying mechanisms of dynamic graphs OOD generalization. Then, we design a causal-aware spatio-temporal attention module for learning representations capable of OOD generalization under distribution shifts by exploiting the intrinsic evolution rationale of causal dynamic subgraphs. This endows DyCIL with a higher capability to learn invariant spatio-temporal patterns. Lastly, we introduce an adaptive environment generator that adaptively infers environment distribution without requiring additional information. This enables DyCIL to generate ample continuous environment instances to capture the underlying dynamics of distributional shifts, thereby improving the model's generalization ability across complex OOD scenarios. Contributions of our work are summarized as follows:

- We introduce DyCIL, a novel model designed for OOD generalization in dynamic graphs from a causal perspective. DyCIL can learn invariant spatio-temporal patterns and generalize well in dynamic scenarios with various OOD shifts.
- We develop multiple effective modules enabling DyCIL to handle OOD shifts, including a dynamic causal subgraph generator that identifies causal dynamic subgraphs, a causal-aware spatio-temporal attention module, and an adaptive environment generator to enable the model to learn generalized invariant representations under various OOD shifts.
- Extensive experiments on both real-world datasets and synthetic datasets demonstrate the superiority of our proposed model over state-of-the-art methods on different dynamic graph prediction tasks with OOD shifts.

Related Work

DyGNNs model evolving structures and temporal dependencies in real-world systems (Poursafaei et al. 2022; Huang et al. 2023; Jiao et al. 2025). Existing methods include continuous-time models (Nguyen et al. 2018; Jin, Li, and Pan 2022; Rossi et al. 2020) that process event streams with time encodings or memory modules, and discrete-time models (Sankar et al. 2020; Pareja et al. 2020; Zhang et al. 2023a) that represent dynamic graphs as timestamped snapshots using GNNs and sequence models. While effective for temporal representation learning, they overlook OOD shift. To enhance robustness, graph OOD generalization has been explored through disentanglement-based methods (Yang et al. 2020b; Liu et al. 2020; Li et al. 2022a) that separate invariant and variant factors, and causality-based methods (Wu et al. 2022b; Li et al. 2022b; Jia et al. 2024) that exploit stable causal mechanisms across environments. Motivated by these advances, recent studies extend OOD generalization to dynamic graphs. DIDA (Zhang et al. 2022) and SILD (Zhang et al. 2023b) address OOD shifts from the perspective of decoupled learning, while EAGLE (Yuan et al. 2023) and EpoD (Yang et al. 2024) enhance generalization by modeling latent environments. OOD-Linker (Tieu et al. 2025) further improves invariance by leveraging the information bottleneck principle.

Preliminary Analysis

A Causal View on Dynamic Graphs

We first take a causal look at the generating process of the dynamic graph and construct SCM (Pearl et al. 2000) in Figure 1. It shows the causalities among six variables, where each directed link from one variable to another denotes a cause-effect relationship between variables. The key explanations regarding SCM are as follows:

- $C_t \rightarrow \mathcal{G}_t \leftarrow E_t$: The observed dynamic graph data is generated by two unobserved latent variables: the causal factor C_t and the environment factor E_t. While C_t denotes the genuine causal relationship, E_t is the unstable variant patterns arising from changes in external variables over time. For example, living and workplace will affect a node's future status in social networks.
- $S_t \rightarrow C_t \leftarrow T_{t'}$: For a dynamic graph, the causal factor C_t consist of spatial factor S_t and temporal factor $T_{t'}$, where S_t denotes the topological information at timestamp t and $T_{t'}(t' < t)$ represents the temporal evolution information from historical snapshots.
- $C_t \rightarrow \mathcal{G}_t \rightarrow Y_t$: This link indicates that the causal factor C_t is the only endogenous parent to determine the generation of label Y_t of $\mathcal{G}_t$ in dynamic graphs.
- $E_t \rightarrow C_t$: This link means that a spurious correlation exists between causal factor C_t and environment factor E_t, which usually changes over time in dynamic graphs.

According to the SCM of dynamic graphs, we recognize a backdoor path between C_t and Y_t, *i.e.*, $C_t \leftarrow E_t \rightarrow \mathcal{G}_t \rightarrow Y_t$, which will cause a misleading correlation between C_t and Y_t. Furthermore, dynamic graphs collected from different environments exhibit various confounding biases. The spatio-temporal patterns extracted using confounding biases are unstable, leading to the failure of most DyGNNs to generalize to OOD shift scenarios.

Backdoor Adjustment

To ensure that predictions in dynamic graphs are based solely on causal factor C_t, it is imperative to tackle the confounding bias effect from the environment factor E_t. To achieve it, we utilize causal tools *do-calculus* on the variable C_t to eliminate the backdoor path between C_t and Y_t, a process commonly referred to as backdoor adjustment (Pearl

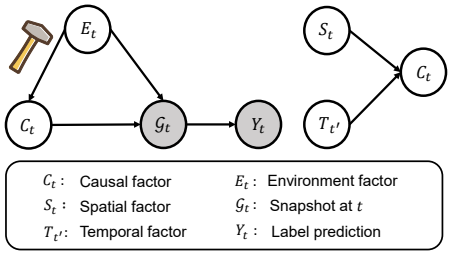

Figure 1: Structural causal model for the generating process of the dynamic graph, where grey and white variables represent observed variables and unobserved, respectively.

2014). By doing so, the probability distribution $P(Y_t|do(C_t)$ can be estimated without interference from E_t. Specifically, we can estimate $P(Y_t|do(C_t))$ by stratifying E_t between C_t and Y_t,

$$P(Y_t|do(C_t)) = \sum_{e_i \in E_t} P(Y_t|S_t, T_{t'}, E_t = e_i)P(E_t = e_i) \tag{1}$$

where e_i denotes the discrete component of environment factor E_t. Detailed derivations are provided in Appendix B.

Model Instantiations

A dynamic graph can be denoted as a series of snapshots $\mathcal{G}_{1:T} = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_T\}$, where T is the total number of snapshots. $\mathcal{G}_t$ denotes the snapshot at timestamp t, which is a graph with a node-set $\mathcal{V}_t$, an edge-set $\mathcal{E}_t$ and feature matrix $\mathbf{X}_t$, *i.e.*, $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{X}_t)$. We denote $\mathbf{A}_t$ as the adjacency matrix of $\mathcal{G}_t$.In this paper, we mainly focus on dynamic prediction tasks, which aim to utilize previous snapshot sequences to predict future states. *i.e.*, $P(Y_t|\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_t)$, where Y_t denote the node properties or interactive relationship at timestamp $t+1$. To tackle the OOD generalization challenge in dynamic graphs, we propose a model named DyCIL, which implements the aforementioned backdoor adjustment mechanism. The overall architecture of DyCIL is illustrated in Figure 2.

Dynamic Causal Subgraph Generator

Based on the above analysis, we know that dynamic graphs are generated from both causal factors and environmental factors. Therefore, we assume that each dynamic graph $\mathcal{G}_{1:T}$ has a causal dynamic subgraph $\mathcal{G}^c_{1:T} = \{\mathcal{G}^c_1, \mathcal{G}^c_2, \cdots, \mathcal{G}^c_T\}$ whose correlation with the label is invariant across different environments. We denote the complement of the causal dynamic subgraph as the variant dynamic subgraph $\mathcal{G}^e_{1:T} = \mathcal{G}_{1:t} \setminus \mathcal{G}^c_{1:t} = \{\mathcal{G}^e_1, \mathcal{G}^e_2, \cdots, \mathcal{G}^e_T\}$, whose relationship with the label is variant across different environments.

We employ a generator to identify the causal dynamic subgraph $\mathcal{G}^c_{1:T} = \Psi(\mathcal{G}_{1:T})$. According to the invariant learning theory (Rojas-Carulla et al. 2018), we make the following assumption about $\Psi(\mathcal{G}_{1:T})$,

Assumption 1. *Given a dynamic graph, there exists a causal dynamic subgraph generator $\Psi(\mathcal{G}_{1:t})$ such that: (1) $Y_t = f(\Psi(\mathcal{G}_{1:t})) + \epsilon$, i.e., $Y_t \perp (\mathcal{G}_{1:t} \setminus \Psi(\mathcal{G}_{1:t})) | \Psi(\mathcal{G}_{1:t})$, where f is the predictor function, $\perp$ denotes probabilistic independence, and ϵ is random noise; (2) $\forall \mathbf{e}_i, \mathbf{e}_j \in \mathbf{E}$, $p(Y_t|\Psi(\mathcal{G}_{1:t}), \mathbf{e}_i) = p(Y_t|\Psi(\mathcal{G}_{1:t}), \mathbf{e}_j)$, where $\mathbf{e}_i, \mathbf{e}_j$ denote instances randomly sampled from environment $\mathbf{E}$.*

Assumption 1 shows that we can develop a subgraph generator to generate causal dynamic subgraphs with sufficient and stable predictive capability for dynamic node-level prediction tasks across different environments. To characterize the optimality of causal dynamic subgraph generator, we derive the following result from Assumption 1, which formalizes the mutual information maximization property of the optimal generator.

Theorem 1. *Let Ψ be a subgraph generator mapping from $\mathcal{G}_{1:t}$ to a subgraph. Under Assumption 1, the following results hold:*

- *The optimal causal subgraph generator Ψ^* satisfies:*

$$\Psi^* = \arg\max_{\Psi} I(\Psi(\mathcal{G}_{1:t}); Y_t). \tag{2}$$

  However, directly optimizing Eq. 2 is challenging, so we address this issue by approximating it with a variational lower bound.
- *For any generator Ψ, the mutual information can be lower bounded by a variational approximation:*

$$I(\Psi(\mathcal{G}_{1:t}); Y_t) \geq \mathbb{E}_{p(\mathcal{G}_{1:t}, Y_t)} \left[\log q_\phi(Y_t|\Psi(\mathcal{G}_{1:t}))\right] + H(Y_t) \tag{3}$$

  where q_ϕ is any variational distribution approximating $p(Y_t|\Psi(\mathcal{G}_{1:t}))$, parameterized by ϕ, and $H(Y_t)$ is the entropy of label Y_t.

We provide the detailed proof in Appendix C. In our work, the bound is indirectly optimized from a causal perspective through the invariance loss and the intervention loss. Therefore, Ψ can be instantiated with learnable parameters to generate the causal dynamic subgraph. In static graphs, existing methods commonly utilize GNNs to generate soft masks for identifying causal subgraphs (Wu et al. 2022b; Li et al. 2022b). However, such methods assume the causal structure to be fixed, which is incompatible with dynamic graphs where both the topology and underlying causal subgraphs evolve over time. To address this, we propose to quantify the spatio-temporal importance of dynamic subgraphs by jointly encoding spatial structure and temporal dynamics, thereby enabling the generation of causal dynamic subgraphs with temporal evolution.

Spatial and Temporal Encoding. We first utilize the degree centrality commonly adopted in literature (Marwick and Boyd 2011; Ying et al. 2021) to capture the spatial feature of nodes within the graph. Then we employ the temporal encoding technique (da Xu et al. 2020) to incorporate temporal dependency information. For node u in snapshot $\mathcal{G}_t$, we have,

$$\mathbf{H}^u_t = \mathbf{W}((\mathbf{X}^u_t + deg(D^u_t)) \oplus TE(t)), \tag{4}$$

where $\mathbf{W}$ is the trainable parameters, $deg(\cdot)$ is learnable embedding vectors specified by degree D^u_t of node u in times-

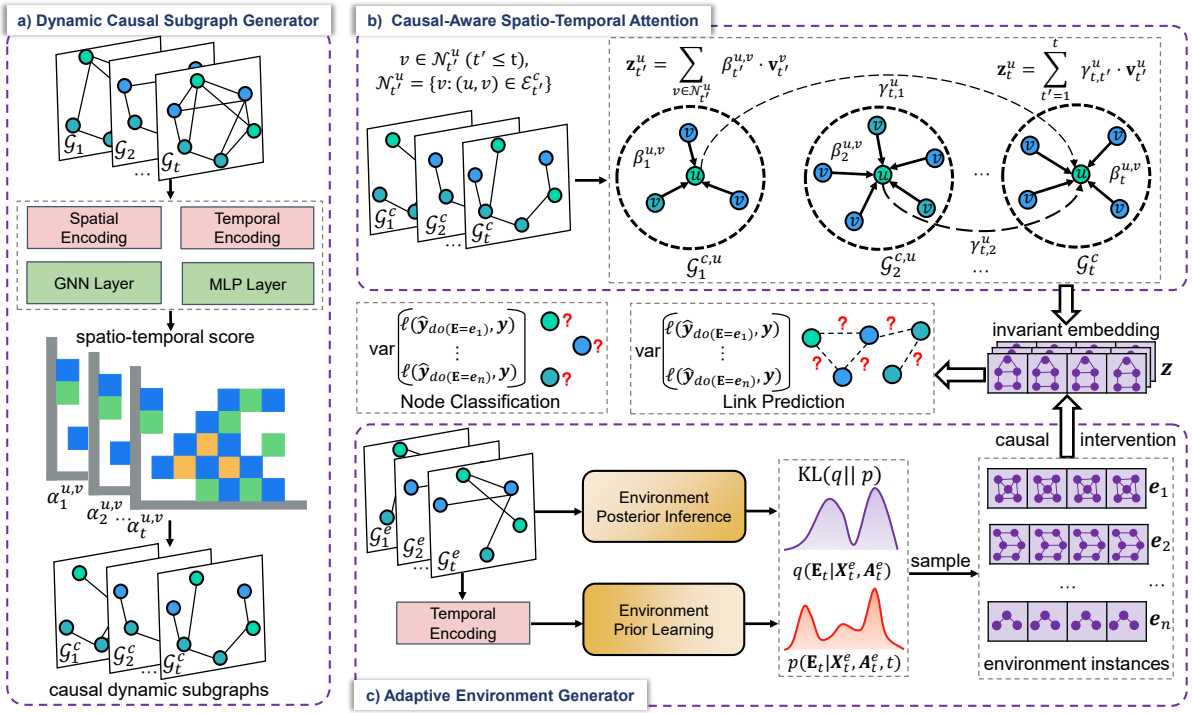

Figure 2: The framework of DyCIL.

tamp t as spatial encoding function, and $\oplus$ is the element-wise addition. $TE(t)$ denotes the temporal encoding function, specifically,

$$TE(t) = \sqrt{\frac{1}{d}} \left[\cos\left(w_1 t\right), \sin\left(w_1 t\right), \cdots, \sin\left(w_d t\right)\right], \quad (5)$$

where $w_1, \cdots, w_d$ are learnable parameters, d is the encoding dimension. In this way, the spatial feature and temporal dependency information can be inherently incorporated into $\mathbf{H}_t^u$.

Generate Causal Dynamic Subgraph. Subsequently, we can obtain the spatio-temporal score of each edge in $\mathcal{G}_t$ with a GNN-based generator,

$$\alpha_t^{u,v} = \text{MLP}(\mathbf{h}_t^u || \mathbf{h}_t^v), (u,v) \in \mathcal{E}_t, \quad \mathbf{h}_t = \text{GNN}(\mathcal{G}_t, \mathbf{H}_t), \quad (6)$$

where $||$ denotes concatenation operation, $\mathbf{h}_t \in \mathbb{R}^{|\mathcal{V}_t| \times |d|}$ summarizes the d-dimensional spatio-temporal matrix of all nodes in $\mathcal{G}_t$, $\mathbf{h}_t^u$ and $\mathbf{h}_t^v$ denotes the vector of node u and v in $\mathbf{h}_t$, and $\alpha_t^{u,v}$ denotes the spatio-temporal score of edge between node u and node v in $\mathcal{G}_t$. Then we will select the edges with the highest spatio-temporal score to construct a causal dynamic subgraph $\mathcal{G}_t^c$ and collect the complement of $\mathcal{G}_t^c$ as variant dynamic subgraph $\mathcal{G}_t^e$, particularly,

$$\mathcal{E}_t^c = \text{Top}_r(\alpha_t), \quad \mathcal{E}_t^e = \text{Top}_{1-r}(1-\alpha_t), \quad (7)$$

where α_t denotes the spatio-temporal scores of all edges in $\mathcal{G}_t$, and $\text{Top}_r(\cdot)$ selects the top$-K$ edges with $K = |\mathcal{E}_t| \times r$, and r is the causal ratio.

Causal-Aware Spatio-Temporal Attention

In dynamic graphs, the state of node u at timestamp t is collaboratively determined from its current neighbors and past neighbors, *i.e.*, dynamic ego-graph. Therefore, inspired by some literature (Sankar et al. 2020; Zhang et al. 2022), we propose causal-aware spatio-temporal attention to learn the invariant spatio-temporal pattern of each node with its causal dynamic ego-graph, further to project the intrinsic evolution rationale behind the causal dynamic subgraphs into node embedding.

Specifically, we first construct the causal dynamic ego-graph of node u at timestamp t based on the identified causal dynamic subgraphs, denoted as $\mathcal{G}_{1:t}^{c,u} = \{\mathcal{G}_1^{c,u}, \mathcal{G}_2^{c,u}, \cdots, \mathcal{G}_t^{c,u}\}$. For node u at timestamp $t'(t' \leq t)$, we utilize spatial self-attention integrated with spatial-temporal scores $\alpha_{t'}^{u,v}$ to aggregate information from the neighbors of $\mathcal{G}_{t'}^{c,u}$. On the one hand, $\alpha_{t'}^{u,v}$ helps learn causal information with diversity, thereby aiding the model in effectively extracting invariant patterns. On the other hand, involving the dynamic causal subgraph generator in the optimization process allows for end-to-end training. This process is formalized as follows,

$$\mathbf{z}_{t'}^u = \sum_{v \in \mathcal{N}_{t'}^u} \beta_{t'}^{u,v} \mathbf{v}_{t'}^v, \beta_{t'}^{u,v} = \frac{exp\left(\frac{\mathbf{q}_{t'}^u (\mathbf{k}_{t'}^v)^T}{\sqrt{d'}} \alpha_{t'}^{u,v}\right)}{\sum_{j \in \mathcal{N}_{t'}^u} exp\left(\frac{\mathbf{q}_{t'}^u (\mathbf{k}_{t'}^j)^T}{\sqrt{d'}} \alpha_{t'}^{u,j}\right)} \quad (8)$$

$$\mathbf{q}_{t'}^u = \mathbf{W}^q \mathbf{X}_{t'}^u, \quad \mathbf{k}_{t'}^v = \mathbf{W}^k \mathbf{X}_{t'}^v, \quad \mathbf{v}_{t'}^v = \mathbf{W}^v \mathbf{X}_{t'}^v, \quad (9)$$

where $\mathbf{z}_{t'}^u$ summarizes the causal patterns at timestamp t'

$\mathcal{N}_{t'}^u = \{v : (u, v) \in \mathcal{E}_{t'}^c\}$ is the L-hop causal neighbors of node u at timestamp t', and d' denotes the feature dimension. $\mathbf{q}_{t'}^u$, $\mathbf{k}_{t'}^v$, $\mathbf{v}_{t'}^v$ represents the query vector of node u, key and value vector of node v in timestamp t', respectively.

After getting causal patterns $\mathbf{z}_{1:t'}^u$ of node u across various snapshots, we employ a temporal self-attention mechanism to aggregate information from its causal dynamic ego-graph $\mathcal{G}_{1:t}^{c,u}$ to obtain the final invariant spatio-temporal node embedding $\mathbf{z}_t^u$ of node u at timestamp t. Specially,

$$\mathbf{z}_t^u = \sum_{t'=1}^{t} \gamma_{t,t'}^u \mathbf{v}_{t'}^u, \quad \gamma_{t,t'}^u = \frac{exp\left(\frac{\mathbf{q}_t^u (\mathbf{k}_{t'}^u)^T}{\sqrt{d'}}\right)}{\sum_{i=1}^{t} exp\left(\frac{\mathbf{q}_t^u (\mathbf{k}_i^u)^T}{\sqrt{d'}}\right)} \tag{10}$$

$$\begin{gathered} \mathbf{q}_t^u = \mathbf{W}^q \left(\mathbf{z}_t^u || TE(t)\right), \mathbf{k}_{t'}^u = \mathbf{W}^k \left(\mathbf{z}_{t'}^u || TE(t')\right), \\ \mathbf{v}_{t'}^u = \mathbf{W}^v \left(\mathbf{z}_{t'}^u || TE(t')\right) \end{gathered} \tag{11}$$

where $\mathbf{q}_t^u$, $\mathbf{k}_{t'}^u$, $\mathbf{v}_{t'}^u$ represents the query, key and value vector of node u in timestamp t and t', respectively. $TE(t)$ is utilized to account for the inherent temporal dependencies across different snapshots.

Adaptive Environment Generator

Inferring environment instances directly from observed data suffers from two key drawbacks: it yields only a limited number of discrete environments and fails to reflect the continuous nature of dynamic OOD shifts. For example, in citation networks, authors gradually transition between research fields. To overcome this, we infer the environment distribution and sample a large number of semantically similar yet distinct instances. These continuous environment variations help the model capture the underlying dynamics of distributional shifts, thereby improving its generalization ability across complex OOD scenarios. Specifically,

$$q(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e) = \prod_{i=1}^{n} q(\mathbf{E}_t^i | \mathbf{X}_t^e, \mathbf{A}_t^e), \tag{12}$$

$$q(\mathbf{E}_t^i | \mathbf{X}_t^e, \mathbf{A}_t^t) = \mathcal{N}(\mathbf{E}_t^i | \boldsymbol{\mu}_t^i, diag((\boldsymbol{\sigma}_t^i)^2)), \tag{13}$$

where $\mathbf{X}_t^e$ and $\mathbf{A}_t^e$ denote the feature matrix and adjacency matrix of variant subgraph $\mathcal{G}_t^e$, respectively, $\mathbf{E}_t^i$ represents the i-th vector of $\mathbf{E}_t$, $\boldsymbol{\mu}_t^i$ and $(\boldsymbol{\sigma}_t^i)^2$ are the i-th values of environment mean vectors $\boldsymbol{\mu}_t$ and environment variance vector $(\boldsymbol{\sigma}_t)^2$. Then we sample sufficient environment instances $\mathbf{e}_i$ from the approximate posterior distribution $q(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e)$ of latent environment to conduct causal intervention, which can endow the model with stronger generalization capabilities across various complex dynamic OOD scenarios.

Additionally, to accurately characterize the environmental distribution, we take a new prior distribution on the environment by allowing for prior parameters to be modeled by information of variant dynamic subgraphs. In particular, we can write the construction of the prior distribution adopted in our work as follows,

$$p(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e, t) = \mathcal{N}(\boldsymbol{\mu}_t^p, diag((\boldsymbol{\sigma}_t^p)^2)), \tag{14}$$

$$(\boldsymbol{\mu}_t^p, \boldsymbol{\sigma}_t^p) = \mathrm{MLP}(\mathrm{GNN}(\mathbf{X}_t^e, \mathbf{A}_t^e) || TE(t)), \tag{15}$$

where $\boldsymbol{\mu}_t^p$ and $(\boldsymbol{\sigma}_t^p)^2)$ are the learned prior mean vector and variance vector, respectively. Here, the temporal encoding function is employed to preserve the temporal dependency information in prior distribution. We can enforce the approximate posterior $q(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e)$ to be close to the prior $p(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e, t)$ by minimizing their KL divergence to optimize the process,

$$\mathcal{L}_E = \mathrm{KL}[q(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e) || p(\mathbf{E}_t | \mathbf{X}_t^e, \mathbf{A}_t^e, t)]. \tag{16}$$

Optimization with Causal Intervention

Based on the learned invariant spatio-temporal embedding $\mathbf{z}$ and generated environment instance $\mathbf{e}_i$, we further instantiate the training objective of OOD generalization with causal intervention. We first utilize two classifiers to project $\mathbf{z}$ and $\mathbf{e}_i$ into a probability distribution over labels $\mathbf{y}$, where $\mathbf{z}$ and $\mathbf{y}$ are the summarized invariant node embeddings and labels. Inspired by (Wu et al. 2022b; Cadene et al. 2019), then we formulate the joint prediction $\hat{\mathbf{y}}_{do(E=e_i)}$ under the intervention $do(\mathbf{E} = \mathbf{e}_i)$ as $\hat{\mathbf{y}}_c$ masked by $\hat{\mathbf{y}}_{e_i}$,

$$\hat{\mathbf{y}}_{do(\mathbf{E}=\mathbf{e}_i)} = \hat{\mathbf{y}}_c \odot \sigma(\hat{\mathbf{y}}_{e_i}), \quad \hat{\mathbf{y}}_c = \Phi_c(\mathbf{z}), \quad \hat{\mathbf{y}}_{e_i} = \Phi_e(\mathbf{e_i}), \tag{17}$$

where $\hat{\mathbf{y}}_c$ is the predicted label obtained solely from the causal dynamic subgraphs, $\hat{\mathbf{y}}_{e_i}$ measures the predictive ability of the variant dynamic subgraphs form environment, and $\odot$ denotes element-wise multiplication, σ is the sigmoid function used to help discover the causal subgraph via adjusting the output logits of $\mathbf{z}$. Then we calculate the the intervention loss and task loss as follows,

$$\mathcal{L}_{do} = \mathrm{Var}_{\mathbf{e}_i \in \mathbf{E}} \left\{\ell(\hat{\mathbf{y}}_{do(\mathbf{E}=\mathbf{e}_i)}, y)\right\}, \quad \mathcal{L}_{inv} = \ell(\hat{\mathbf{y}}_c, y). \tag{18}$$

The final loss function consists of three parts: invariance loss, intervention loss, and the KL divergence of the environmental distribution,

$$\mathcal{L} = \mathcal{L}_{inv} + \lambda(\mathcal{L}_{do} + \mathcal{L}_E), \tag{19}$$

where λ is the trade-off weight. We provide the overall training algorithm, complexity analysis, and time cost of DyCIL in Appendix A, D, and G.

Experiments

Datasets and Baselines

We utilize three real-world datasets, namely Collab (Tang et al. 2012), ACT (Kumar, Zhang, and Leskovec 2019), and Aminer (Tang et al. 2008). Additionally, we synthesize two datasets by introducing manually designed distribution shifts, Temporal-Motif and Synthetic-Collab. Table 1 summarizes the statistics of all datasets. We compare our method with representative approaches from DyGNNs (GCRN (Seo et al. 2018), EvolveGCN (Pareja et al. 2020), DySAT (Sankar et al. 2020)), OOD generalization (IRM (Arjovsky et al. 2019), V-REx (Krueger et al. 2021), GraphDRO (Sagawa* et al. 2020)), graph OOD generalization (DIR (Wu et al. 2022b), EERM (Wu et al. 2022a), and dynamic graph OOD generalization (DIDA (Zhang et al. 2022), SILD (Zhang et al. 2023b), EAGLE (Yuan et al. 2023), OOD-Linker (Tieu et al. 2025)). More details about the datasets, baselines, reproducibility, and experiment results are in Appendix E, F, and G.

Dataset	# Nodes	# Edges	# Snapshots	# Features	# Classes	Train/Val/Test	Temporal Granularity	Distribution Shift
Collab	23,035	151,790	16	32	-	10/1/5	year	Cross-Domain Transfers
ACT	20,408	202,339	30	32	-	20/2/8	day	Cross-Domain Transfers
Synthetic-Collab	23,035	151,790	16	64	-	10/1/5	-	Feature Evolution
Temporal-Motif	5,000	298,709	30	8	3	24/3/3	-	Motif Structure Transfers
Aminer	43,141	851,527	17	128	20	11/3/3	year	Temporal Evolution

Table 1: The summary of the dataset statistics.

	Dataset	Collab		ACT		Synthetic-Collab		
Method Type	Model	w/o OOD	w/ OOD	w/o OOD	w/ OOD	$\bar{p} = 0.4$	$\bar{p} = 0.6$	$\bar{p} = 0.8$
DyGNNs	GCRN	82.78±0.54	69.72±0.45	76.28±0.51	64.35±1.24	72.57±0.72	72.29±0.47	67.26±0.22
	EvolveGCN	86.62±0.95	76.15±0.91	74.55±0.33	63.17±1.05	69.00±0.53	62.70±1.14	60.13±0.89
	DySAT	88.77±0.23	76.59±0.20	78.52±0.40	66.55±1.21	70.24±1.26	64.01±0.19	62.19±0.39
OOD	IRM	87.96±0.90	75.42±0.87	80.02±0.57	69.19±1.35	69.40±0.09	63.97±0.37	62.66±0.33
	VREx	88.31±0.32	76.24±0.77	83.11±0.29	70.15±1.09	70.44±1.08	63.99±0.21	62.21±0.40
	GroupDRO	88.76±0.12	76.33±0.29	85.19±0.53	74.35±1.62	70.30±1.23	64.05±0.21	62.13±0.35
Graph OOD	EERM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	DIR	88.18±0.12	76.07±0.30	88.24±0.46	75.71±2.15	76.81±1.11	67.97±0.35	66.53±0.60
Dynamic Graph OOD	DIDA	91.97±0.05	81.87±0.40	89.84±0.82	78.64±0.97	85.20±0.84	82.89±0.23	72.59±3.31
	SILD	92.36±0.19	84.14±0.31	89.28±0.41	79.91±0.65	85.95±0.18	84.69±1.18	78.01±0.71
	EAGLE	92.34±0.13	84.24±0.19	92.43±0.71	82.99±0.83	88.12±0.47	86.97±0.50	82.08±0.96
	OOD-Linker	—	85.30±0.31	—	**85.98±1.00**	85.58±1.54	83.09±1.82	79.83±1.69
	DyCIL	**95.00±0.09**	**86.54±0.17**	**93.41±0.13**	85.66±0.21	**90.98±0.09**	**87.69±0.23**	**83.31±0.19**

Table 2: AUC score (%) of link prediction. The best are bolded and the second best are underlined.

Results on Link Prediction Task

We show the results in Table 2. DyGNNs methods deteriorate significantly under distribution shifts, while DyCIL achieves more progress in OOD scenarios. Specifically, compared to DySAT, DyCIL achieves improvements of 7.01% and 12.9%, and 18.9% and 28.7% for the w/o OOD and w/ OOD scenarios on the Collab and ACT datasets, respectively. The general OOD generalization methods only have limited improvements, as the environment labels they rely on are not available in real-world dynamic graphs. Graph OOD methods focus only on the causal rationale of the structure without considering the temporal dynamics of dynamic graphs, hindering their ability to handle OOD shifts in dynamic graphs. DIDA and SILD rely solely on observable data to perform causal interventions, which weakens their generalization ability in scenarios with severe distribution shifts. Specifically, in the Synthetic-Collab dataset with $\bar{p} = 0.8$, the performance of DIDA and SILD decrease by 10.3% and 6.68% compared to $\bar{p} = 0.6$, while our model only drops by 4.41%. This demonstrates that DyCIL has better generalization capabilities in dynamic graphs with severe OOD shifts by capturing the underlying dynamics of distributional shifts adaptively. EAGLE achieves further improvements by focusing on latent environment modeling of dynamic graphs. However, it requires additional labels to infer the environment and overlooks the intrinsic evolution rationale of invariant patterns that are commonly present in real dynamic graphs, which hinders its scalability and performance. Therefore, DyCIL shows a more significant improvement in the Collab and ACT real-world datasets compared to EAGLE. Due to its neglect of environmental influences, OOD-Linker performs significantly worse than EAGLE and DyCIL on three synthetic datasets characterized by severe OOD distribution shifts.

Results on Node Classification Task

The results are shown in Table 3. DyCIL outperforms all baseline methods on both node classification datasets by a significantly large margin. Specifically, DyCIL improves by 4.77%, 8.43%, and 12.23% over the best baseline, SILD, on the Temporal-Motif dataset. As time evolves, the improvement of DyCIL becomes more significant, mainly because the emergence of more unseen variant motif structures leads to more severe distribution shifts as time goes on. SILD struggles to handle cases with severe OOD shifts, resulting in a noticeable performance decline over time. In contrast, DyCIL can continuously capture causal motif structures under different degrees of distribution shifts, thus significantly outperforming various baselines. On the Aminer dataset, DyCIL handles OOD shifts better than all the baselines, which validates that DyCIL can capture invariant spatio-temporal patterns under distribution shifts. In addition, DyCIL exhibits smaller variance in most cases, showing it's less sensitive to spurious correlations under different distribution shifts.

Ablation Study

To further validate the contribution of each component in our model, we conduct ablation experiments. We name the DyCIL variants as follows, **w/o SG**: DyCIL without the dy-

	Dataset	Temporal-Motif			Aminer		
Method Type	Model	T-M28	T-M29	T-M30	Aminer15	Aminer16	Aminer17
DyGNNs	GCRN	51.93±1.68	51.55±1.26	51.31±1.28	47.96±1.12	51.33±0.62	42.93±0.71
	EvolveGCN	45.63±0.33	45.50±0.28	45.29±0.42	44.14±1.12	46.28±1.84	37.71±1.84
	DySAT	48.44±0.56	48.20±0.45	47.99±0.45	48.41±0.81	49.76±0.96	42.39±0.62
OOD	IRM	50.17±0.71	50.01±0.52	49.84±0.67	48.44±0.13	50.18±0.73	42.40±0.27
	VREx	51.32±0.77	52.64±0.63	50.79±0.46	48.70±0.73	49.24±0.27	42.59±0.37
	GroupDRO	50.89±0.38	50.95±0.87	50.18±0.53	48.73±0.61	49.74±0.26	42.80±0.36
Graph OOD	EERM	45.23±0.47	45.05±0.40	44.68±0.41	OOM	OOM	OOM
	DIR	47.07±0.45	47.42±0.68	45.74±0.56	46.85±1.84	49.69±1.71	41.73±1.41
Dynamic Graph OOD	DIDA	64.26±0.31	63.71±0.41	61.94±0.33	50.34±0.81	51.43±0.27	44.69±0.06
	SILD	<u>77.89±0.74</u>	<u>72.70±0.84</u>	<u>66.72±0.93</u>	<u>51.61±1.12</u>	<u>52.85±1.05</u>	<u>44.47±1.18</u>
	EAGLE	67.26±1.17	66.98±1.06	65.46±1.05	OOM	OOM	OOM
	DyCIL	**82.66±0.34**	**81.13±0.20**	**78.95±0.19**	**53.04±0.75**	**54.30±0.21**	**45.89±0.45**

Table 3: ACC score (%) of node classification. The best are bolded and the second best are underlined.

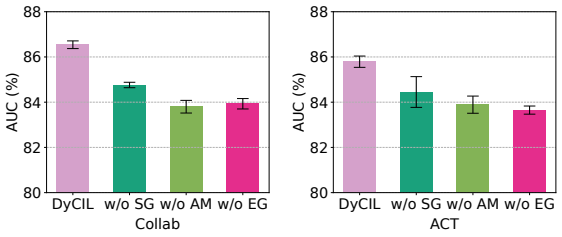

Figure 3: The ablation experiment results.

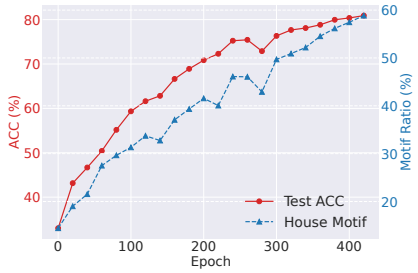

Figure 4: A case study on Temporal-Motif dataset.

namic causal subgraph generator. **w/o AM**: DyCIL without the causal-aware spatio-temporal attention module. **w/o EG**: DyCIL without adaptive environment generator. We present the ablation experiment results in Figure 3. Overall, DyCIL outperforms all its variants across various datasets. Removing any of the three components leads to a significant performance drop, further validating the effectiveness of each component of our model. In summary, DyCIL jointly optimizes three mutually promoting modules to effectively capture invariant spatio-temporal patterns and achieve OOD generalization under distribution shifts in dynamic graphs. More ablation analyses are provided in Appendix G.

Case Study

To verify whether DyCIL truly learns the causal dynamic subgraphs and captures the evolutionary rationale in dynamic graphs, we conduct a case study on the Temporal-Motif dataset. Since the node labels are determined solely by the evolving 'house' motif that constitutes the causal rationale, the causal dynamic subgraphs generated by the model should contain more 'house' motifs. Therefore, we calculate the ratio of 'house' motifs in the causal dynamic subgraphs to the total number of 'house' motifs and show its relationship with performance in Figure 4. We can notice that as the number of training epochs increases, the identified causal dynamic subgraphs contain more 'house' motifs, enabling successive improvements in performance. In other words, the performance improvement of DyCIL comes from capturing more 'house' motifs. It demonstrates that DyCIL captures genuine causal patterns to make stable predictions across various distribution shifts and explains why DyCIL has good OOD generalization ability. This serves as additional evidence that DyCIL enhances interpretability by identifying causal subgraphs in dynamic graph OOD generalization.

Conclusion

In this paper, we propose DyCIL, a novel model to handle OOD shifts in dynamic graphs by exploiting spatio-temporal invariant patterns from a causal view. DyCIL first generates causal dynamic subgraphs via a dynamic causal subgraph generator. Then DyCIL learns invariant node representations by capturing the evolutionary rationale behind its causal dynamic ego-graph. Finally, DyCIL instantiates latent environments through inferred environment distribution and performs causal interventions. Experiment results demonstrate that DyCIL can better handle OOD shifts compared to state-of-the-art baselines. The ablation study further validates the contribution of each module. Additionally, we conduct a case study to demonstrate DyCIL's capability to identify causal dynamic subgraphs. One limitation is that we mainly consider discrete dynamic homogeneous graphs. Future work can explore OOD generalization in dynamic heterogeneous graphs and continuous dynamic graphs, which are more common and challenging in the real world.

Acknowledgments

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LDT23F01015F01 and LMS25F030011, in part by the Key Technology Research and Development Program of the Zhejiang Province under Grant No. 2025C1023, in part by the National Natural Science Foundation of China under Grants 62372146, and in part by the Open Research Fund of The State Key Laboratory of AI Safety, Chinese Academy of Sciences (2025-02).

References

Arjovsky, M.; Bottou, L.; Gulrajani, I.; and Lopez-Paz, D. 2019. Invariant Risk Minimization. *arXiv preprint arXiv:1907.02893*.

Cadene, R.; Dancette, C.; Cord, M.; Parikh, D.; et al. 2019. RUBi: Reducing Unimodal Biases in Visual Question Answering. In *Advances in neural information processing systems*, volume 32.

da Xu; chuanwei ruan; evren korpeoglu; sushant kumar; and kannan achan. 2020. Inductive Representation Learning on Temporal Graphs. In *International Conference on Learning Representations*.

Huang, S.; Poursafaei, F.; Danovitch, J.; Fey, M.; Hu, W.; Rossi, E.; Leskovec, J.; Bronstein, M.; Rabusseau, G.; and Rabbany, R. 2023. Temporal Graph Benchmark for Machine Learning on Temporal Graphs. In *Advances in Neural Information Processing Systems*, volume 36, 2056–2073.

Jia, T.; Li, H.; Yang, C.; Tao, T.; and Shi, C. 2024. Graph Invariant Learning with Subgraph Co-mixup for Out-Of-Distribution Generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8562–8570.

Jiao, P.; Chen, H.; Guo, X.; Zhao, Z.; He, D.; and Jin, D. 2025. A Survey on Temporal Interaction Graph Representation Learning: Progress, Challenges, and Opportunities. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 10499–10507.

Jin, M.; Li, Y.-F.; and Pan, S. 2022. Neural temporal walks: Motif-aware representation learning on continuous-time dynamic graphs. In *Advances in Neural Information Processing Systems*, volume 35, 19874–19886.

Krueger, D.; Caballero, E.; Jacobsen, J.-H.; Zhang, A.; Binas, J.; Zhang, D.; Le Priol, R.; and Courville, A. 2021. Out-of-Distribution Generalization via Risk Extrapolation (REx). In *International conference on machine learning*, 5815–5826. PMLR.

Kumar, S.; Zhang, X.; and Leskovec, J. 2019. Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 1269–1278.

Li, H.; Zhang, Z.; Wang, X.; and Zhu, W. 2022a. Disentangled graph contrastive learning with independence promotion. *IEEE Transactions on Knowledge and Data Engineering*, 35(8): 7856–7869.

Li, H.; Zhang, Z.; Wang, X.; and Zhu, W. 2022b. Learning Invariant Graph Representations for Out-of-Distribution Generalization. In *Advances in Neural Information Processing Systems*, volume 35, 11828–11841.

Liu, H.; Jiao, P.; Gao, M.; Chen, C.; and Jin, D. 2025. Heterogeneous Temporal Hypergraph Neural Network. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 3117–3125.

Liu, Y.; Wang, X.; Wu, S.; and Xiao, Z. 2020. Independence promoted graph disentangled networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 4916–4923.

Marwick, A.; and Boyd, D. 2011. To See and Be Seen: Celebrity Practice on Twitter. *Convergence*, 17(2): 139–158.

Nguyen, G. H.; Lee, J. B.; Rossi, R. A.; Ahmed, N. K.; Koh, E.; and Kim, S. 2018. Continuous-time dynamic network embeddings. In *Companion proceedings of the the web conference 2018*, 969–976.

Pareja, A.; Domeniconi, G.; Chen, J.; Ma, T.; Suzumura, T.; Kanezashi, H.; Kaler, T.; Schardl, T.; and Leiserson, C. 2020. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 5363–5370.

Pearl, J. 2014. Interpretation and Identification of Causal Mediation. *Psychological methods*, 19(4): 459.

Pearl, J.; et al. 2000. Models, Reasoning and Inference. *Cambridge, UK: CambridgeUniversityPress*, 19(2): 3.

Poursafaei, F.; Huang, S.; Pelrine, K.; and Rabbany, R. 2022. Towards Better Evaluation for Dynamic Link Prediction. In *Advances in Neural Information Processing Systems*, volume 35, 32928–32941.

Rojas-Carulla, M.; Schölkopf, B.; Turner, R.; and Peters, J. 2018. Invariant Models for Causal Transfer Learning. *Journal of Machine Learning Research*, 19(36): 1–34.

Rossi, E.; Chamberlain, B.; Frasca, F.; Eynard, D.; Monti, F.; and Bronstein, M. 2020. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*.

Sagawa*, S.; Koh*, P. W.; Hashimoto, T. B.; and Liang, P. 2020. Distributionally Robust Neural Networks. In *International Conference on Learning Representations*.

Sankar, A.; Wu, Y.; Gou, L.; Zhang, W.; and Yang, H. 2020. 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*, 519–527.

Seo, Y.; Defferrard, M.; Vandergheynst, P.; and Bresson, X. 2018. Structured Sequence Modeling with Graph Convolutional Recurrent Networks. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part I 25*, 362–373. Springer.

Tang, J.; Wu, S.; Sun, J.; and Su, H. 2012. Cross-domain Collaboration Recommendation. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1285–1293.

Tang, J.; Zhang, J.; Yao, L.; Li, J.; Zhang, L.; and Su, Z. 2008. ArnetMiner: Extraction and Mining of Academic Social Networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 990–998.

Tieu, K.; Fu, D.; Wu, J.; and He, J. 2025. Invariant Link Selector for Spatial-Temporal Out-of-Distribution Problem. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258, 4753–4761.

Wu, Q.; Zhang, H.; Yan, J.; and Wipf, D. 2022a. Handling Distribution Shifts on Graphs: An Invariance Perspective. In *International Conference on Learning Representations*.

Wu, Y.; Fang, Y.; and Liao, L. 2024. On the Feasibility of Simple Transformer for Dynamic Graph Modeling. In *Proceedings of the ACM on Web Conference 2024*, 870–880.

Wu, Y.; Wang, X.; Zhang, A.; He, X.; and Chua, T.-S. 2022b. Discovering Invariant Rationales for Graph Neural Networks. In *International Conference on Learning Representations*.

Yang, C.; Zhang, J.; Wang, H.; Li, S.; Kim, M.; Walker, M.; Xiao, Y.; and Han, J. 2020a. Relation Learning on Social Networks with Multi-Modal Graph Edge Variational Autoencoders. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 699–707.

Yang, K.; Zhou, Z.; Huang, Q.; Li, L.; Liang, Y.; and Wang, Y. 2024. Improving generalization of dynamic graph learning via environment prompt. *Advances in Neural Information Processing Systems*, 37: 70048–70075.

Yang, Y.; Feng, Z.; Song, M.; and Wang, X. 2020b. Factorizable graph convolutional networks. *Advances in Neural Information Processing Systems*, 33: 20286–20296.

Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do Transformers Really Perform Badly for Graph Representation? In *Advances in Neural Information Processing Systems*, volume 34, 28877–28888.

Yuan, H.; Sun, Q.; Fu, X.; Zhang, Z.; Ji, C.; Peng, H.; and Li, J. 2023. Environment-Aware Dynamic Graph Learning for Out-of-Distribution Generalization. In *Advances in Neural Information Processing Systems*, volume 36, 49715–49747.

Zhang, K.; Cao, Q.; Fang, G.; Xu, B.; Zou, H.; Shen, H.; and Cheng, X. 2023a. DyTed: Disentangled Representation Learning for Discrete-time Dynamic Graph. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3309–3320.

Zhang, X.; Jiao, P.; Gao, M.; Li, T.; Wu, Y.; Wu, H.; and Zhao, Z. 2024. VGGM: Variational Graph Gaussian Mixture Model for Unsupervised Change Point Detection in Dynamic Networks. *IEEE Transactions on Information Forensics and Security*, 19: 4272–4284.

Zhang, Z.; Wang, X.; Zhang, Z.; Li, H.; Qin, Z.; and Zhu, W. 2022. Dynamic Graph Neural Networks Under Spatio-Temporal Distribution Shift. In *Advances in neural information processing systems*, volume 35, 6074–6089.

Zhang, Z.; Wang, X.; Zhang, Z.; Qin, Z.; Wen, W.; Xue', H.; Li, H.; and Zhu, W. 2023b. Spectral Invariant Learning for Dynamic Graphs under Distribution Shifts. In *Advances in Neural Information Processing Systems*, volume 36, 6619–6633.

Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; and Li, H. 2019. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE transactions on intelligent transportation systems*, 21(9): 3848–3858.

Zhong, Y.; Vu, H.; Yang, T.; and Adhikari, B. 2024. Efficient and Effective Implicit Dynamic Graph Neural Network. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4595–4606.