Heterogeneous Temporal Hypergraph Neural Network

Huan Liu[1,2] , **Pengfei Jiao**[1,2] , **Mengzhou Gao**[1,2∗] , **Chaochao Chen**[3] and **Di Jin**[4∗]
[1]School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China
[2]Data Security Governance Zhejiang Engineering Research Center, Hangzhou, China
[3]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[4]College of Intelligence and Computing, Tianjin University, Tianjin, China
{huanliu, pjiao, mzgao}@hdu.edu.cn, zjuccc@zju.edu.cn, jindi@tju.edu.cn

Abstract

Graph representation learning (GRL) has emerged as an effective technique for modeling graph-structured data. When modeling heterogeneity and dynamics in real-world complex networks, GRL methods designed for complex heterogeneous temporal graphs (HTGs) have been proposed and have achieved successful applications in various fields. However, most existing GRL methods mainly focus on preserving the low-order topology information while ignoring higher-order group interaction relationships, which are more consistent with real-world networks. In addition, most existing hypergraph methods can only model static homogeneous graphs, limiting their ability to model high-order interactions in HTGs. Therefore, to simultaneously enable the GRL model to capture high-order interaction relationships in HTGs, we first propose a formal definition of heterogeneous temporal hypergraphs and P-uniform heterogeneous hyperedge construction algorithm that does not rely on additional information. Then, a novel Heterogeneous Temporal HyperGraph Neural network (HTHGN), is proposed to fully capture higher-order interactions in HTGs. HTHGN contains a hierarchical attention mechanism module that simultaneously performs temporal message-passing between heterogeneous nodes and hyperedges to capture rich semantics in a wider receptive field brought by hyperedges. Furthermore, HTHGN performs contrastive learning by maximizing the consistency between low-order correlated heterogeneous node pairs on HTG to avoid the low-order structural ambiguity issue. Detailed experimental results on three real-world HTG datasets verify the effectiveness of the proposed HTHGN for modeling high-order interactions in HTGs and demonstrate significant performance improvements.

1 Introduction

Graph representation learning (GRL) has emerged as an effective technique for learning real-world graph-structured data and has been widely applied in various fields [Xia *et al.*, 2021; Gao *et al.*, 2023b; Zhang *et al.*, 2023c]. In the major research of GRL, graph neural networks (GNNs) as encoders have gained universal effectiveness due to their powerful message-passing mechanism and fitting ability [Kipf and Welling, 2017; Hamilton *et al.*, 2017]. However, most methods assume that the network is homogeneous and static and only includes pairwise relationships, which often contradicts real-world systems [Antelmi *et al.*, 2023; Barros *et al.*, 2021; Wang *et al.*, 2023]. For example, in an academic network that includes multiple node types such as *author*, *paper* and *venue*, as well as evolving *co-author* and *co-citations* relationships among multiple authors and papers over time. This network contains heterogeneity and dynamics in not low-order but group interactions, which are too complex to be described by simple pairwise graphs [Antelmi *et al.*, 2023].

Considering the successful applications of GRL methods of heterogeneous graphs, dynamic graphs, and hypergraphs, we propose a formal definition of *heterogeneous temporal hypergraphs* (HTHGs) as a modeling tool to comprehensively describe high-order relationships in heterogeneous temporal graphs (HTGs). Specifically, HTHGs refer to hypergraphs that contain high-order relationships among three or more multi-type entities, and these entities and relationships could increase or delete over time. Since HTHGs involve multiple types of nodes and interaction patterns, effectively modeling the high-order correlations and semantic information inherent in HTHGs is crucial for representation learning.

However, existing GRL and GNN research usually one-sidedly simplifies HTHG in different aspects, thus losing its information integrity and seriously affecting performance. For example, on the one hand, modeling group interactions as hyperedges and performing representation learning through hypergraph GNNs has become a dominant paradigm and has achieved remarkable results [Huang and Yang, 2021; Gao *et al.*, 2023a]. However, these methods usually only focus on homogeneous hypergraphs with static structures, failing to model the heterogeneity and dynamics in HTGs. On the other hand, for heterogeneity and dynamics, GNNs designed for heterogeneous graphs [Wang *et al.*, 2023; Yang *et al.*, 2022; Ji *et al.*, 2023] and dynamic graphs [Barros *et al.*, 2021;

∗Corresponding authors.

Zhang *et al.*, 2023a; Zhang *et al.*, 2023b] have been proposed respectively and achieved encouraging results. However, this method usually performs representation learning on pairwise networks and cannot model both low-order and high-order dynamic relationships between multiple heterogeneous nodes.

Although effectively modeling semantic information, temporal dependence, and group interactions has demonstrated significant performance improvements in practice, uniformly modeling HTHGs still has the following challenges: **1) How to model group interactions without relying on expert knowledge?** Additional information or predefined structures are required to model group interactions, but their effectiveness depends on prior knowledge and is difficult to generalize. **2) How to model both low-order and high-order information on HTHGs?** The message-passing mechanism of the GNNs and cluster/star expansion are often used to model high-order interaction, but they have computational problems [Antelmi *et al.*, 2023] and cannot preserve both low-order and high-order relationships simultaneously. **3) How to perform message-passing on HTHGs?** To preserve temporal structural and semantic information in latent representation space, enhancing communication between heterogeneous nodes and hyperedges is necessary and challenging.

In this paper, to uniformly model the complex blend of heterogeneity, dynamics, and group interactions, we first provide a formal definition of HTHGs, which describes multi-scale group interaction relationships containing dynamics and heterogeneity. Furthermore, to universally model high-order interactions in HTHGs, we formally define two general heterogeneous hyperedges, k-hop, and k-ring, and P-uniform hyperedges based on high-order neighborhood sampling, which do not rely on predefined structures. Then, we propose a novel contrastive heterogeneous temporal hypergraph neural network called HTHGN to capture the high-order dynamic semantics contained in HTHGs. Specifically, HTHGN contains a hierarchical attention mechanism that simultaneously performs cross-temporal message passing between heterogeneous nodes and hyperedges to capture rich semantic information in a wider receptive field brought by hyperedges. Finally, to avoid the problem of low-order structural ambiguity, a heterogeneous low-order structure-preserving contrastive learning objective function is used to optimize the overall HTHGN. Detailed experimental results on 3 real-world datasets demonstrate that group interaction significantly gains representation learning and verifies the effectiveness of the proposed contrastive learning method HTHGN.

In summary, the contributions of this paper are as follows:

- We study the complex properties prevalent in real-world complex networks. To the best of our knowledge, we are the first to define HTHGs, which are used to model complex networks containing dynamics, heterogeneity, and group interactions.
- We propose a general hyperedges construction algorithm to model high-order semantic information without relying on additional information and prior knowledge.
- We propose a novel contrastive heterogeneous temporal hypergraph neural network, HTHGN, which simultaneously models low-order and high-order interactions, and extensive experimental results on 3 real-world datasets verify its superior performance.

2 Background and Preliminaries

The overall architecture of our proposed HTHGN model is shown in Figure 1. We design a pipeline to learn node representations, which can capture both low-order and high-order information within a HTG. The basic definitions in this paper are shown below:

Definition 1 (**Heterogeneous Graph**). *A **Heterogeneous Graph** can be defined as $G = (V, E, X)$, where V and E denote the node set and the edge set, respectively; $X \in \mathbb{R}^{|V| \times D}$ is the D-dimensional attribute matrix of nodes. Each node $v \in V$ and link $e \in E$ is associated with their mapping functions $\phi(v) : V \to \mathcal{A}$ and $\psi(e) : E \to \mathcal{R}$, where $\mathcal{A}$ and $\mathcal{R}$ denote the node types and link types, and $|\mathcal{A}| + |\mathcal{R}| > 2$ due to heterogeneity.*

Definition 2 (**Heterogeneous Temporal Graph**). *A **Heterogeneous Temporal Graph** is a list of observed heterogeneous snapshots $\mathcal{G} = \left\{G^1, G^2, \ldots, G^T\right\}$ ordered by timestamps, where T is the size of time window and $G^t = (V^t, E^t, X^t)$ represents the t-th snapshot. The node set V^t and edge set E^t can differ between snapshots, representing dynamic addition and removal of nodes and edges.*

Definition 3 (**Link Prediction**). *Given a heterogeneous temporal graph $\mathcal{G} = \{G^t\}_{t=1}^T$ and the learned node representations $Z \in \mathbb{R}^{|V^t| \times d}$, the **link prediction** is the problem of predicting the probability $p\left((i,j) \in E^\tau \mid z_i, z_j\right)$ where $\tau > T$. Besides, the **new link prediction** predicts the probability $p\left((i,j) \in E^\tau \mid z_i, z_j, (i,j) \notin E^T\right)$ where $\tau > T$.*

3 Heterogeneous Temporal Hypergraph Contrastive Learning: HTHGN

3.1 Hypergraph Construction

Given a HTG $\mathcal{G} = \{G^t\}_{t=1}^T$, the hypergraph construction module is leveraged to construct heterogeneous collective relations based on the heterogeneous snapshots. For this purpose, it is necessary to employ a graph structure capable of modeling interactions that encompass both multiple node types and collective behavior. Formally, we define as follows:

Definition 4 (**Heterogeneous hypergraph**). *A **heterogeneous hypergraph** can be defined as $H = (V, \mathcal{E})$, where V and $\mathcal{E}$ denote the node set and the hyperedges set, respectively; each hyperedge $e \in \mathcal{E}$ is a subset of V, i.e., $e \subseteq V$, and each node $v \in V$ is contained in at least one hyperedge $e \in E$, i.e., $V = \bigcup_{e \in \mathcal{E}} e$. Each node $v \in V$ and link $e \in \mathcal{E}$ is associated with their mapping functions $\phi_h(v) : V \to \mathcal{A}_h$ and $\psi_h(e) : E \to \mathcal{R}_h$, where $\mathcal{A}_h$ and $\mathcal{R}_h$ denote the node types and hyperedge types, and $|\mathcal{A}_h| + |\mathcal{R}_h| > 2$ due to heterogeneity. When $|\mathcal{A}_h| + |\mathcal{R}_h| = 2$, the heterogeneous hypergraph degenerates into a homogeneous hypergraph, represented as H^-.*

Due to heterogeneous hyperedges, H has a higher level of expressive power that can encompass many entity types

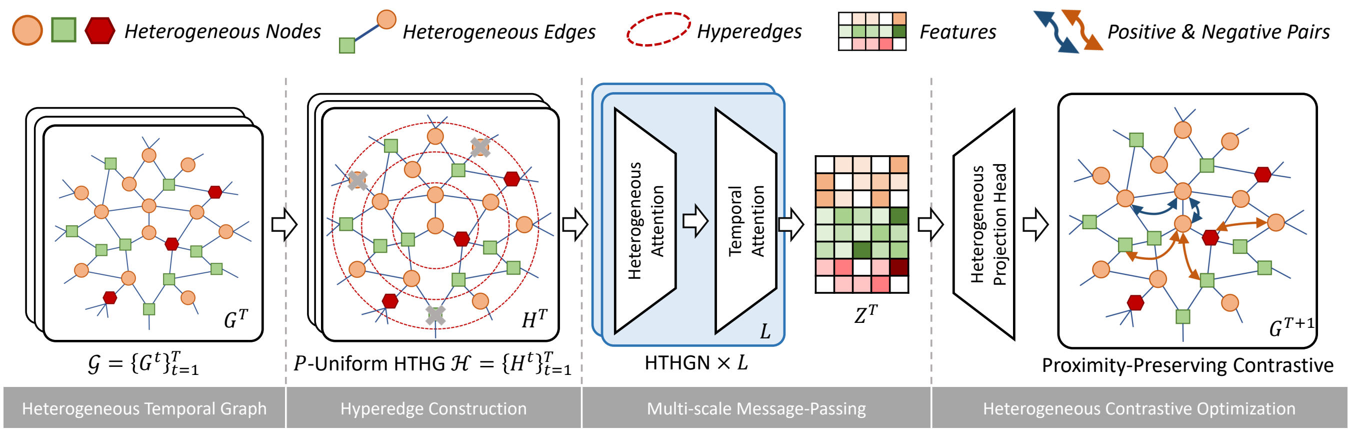


Figure 1: Overall architecture of the proposed HTHGN model.

between subtle relationships. However, hyperedges are typically defined by original data and predefined structures, such as meta-paths, which often introduce noise due to insufficient modeling and rely on specific scenarios, limiting their generalizability. Consequently, we propose heterogeneous network-structure-based hyperedge construction methods for constructing hyperedges, i.e., k-hop and k-ring heterogeneous hyperedge.

Definition 5 (**k-hop heterogeneous hyperedge**). *A **k-hop heterogeneous hyperedge** around a node $v \in V$ is defined as the set of all nodes that are **within** cumulative topological distance k edges from the node v, regardless of the types of edges and nodes. We denote the set of nodes in the receptive field of v by $e(v)_{k\text{-hop}}$, which is defined recursively as:*

$$e(v)_{k\text{-hop}} := e(v)_{(k-1)\text{-hop}} \cup \{u \in V \mid (v,u) \in E \wedge u \in \mathcal{N}_v^{k-1}\}, \tag{1}$$

where $\mathcal{N}_v^k$ represents the k-hop neighborhoods set of node v and $e(v)_{1\text{-hop}} := \mathcal{N}_v$.

This approach to constructing hyperedges based on k-hop connectivity implicitly integrates multifaceted semantic layers and relationships between entities, offering a nuanced understanding of their interactions. As illustrated in Figure 2, which has *Author(A)*, *Paper(P)* and *Venue(V)*. when $k = 1$, the 1-hop hyperedge of node *A1* represents all direct neighbors, that is, $e(A1)_{\text{1-hop}} = \{P1, V1\}$. When $k = 2$, the 2-hop hyperedge encompasses both 1-hop and 2-hop neighbors, thus, $e(A1)_{\text{2-hop}} = \{P1, V1, A2, A3, P2\}$. This encapsulates rich semantic information that can be recognized as multiple meta-paths, such as *AP*, *AV*, *APA*, *AVA*, *AVP*, *etc*.

Definition 6 (**k-ring heterogeneous hyperedge**). *A **k-ring heterogeneous hyperedge** around a node $v \in V$ is defined as the set of all heterogeneous nodes that reside **exactly** a topological distance of k edges from v, where the types of nodes and edges along the path are not necessarily the same, denoted as $e(v)_{k\text{-ring}}$, which is defined recursively as:*

$$\begin{aligned} e(v)_{1\text{-ring}} &:= \mathcal{N}_v, \\ e(v)_{k\text{-ring}} &:= \{u \in V \mid (v,u) \in E \wedge u \in \mathcal{N}_v^{k-1}\}. \end{aligned} \tag{2}$$

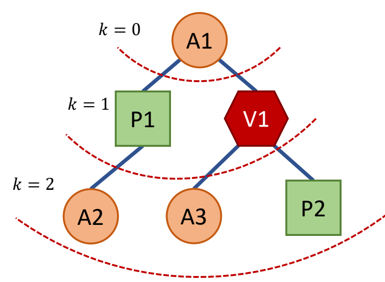


Figure 2: k-hop heterogeneous hyperedge toy example.

Different from k-hop heterogeneous hyperedge, the k-ring focuses on the heterogeneous nodes at a specific distance, which helps understanding group interactions at that exact distance. Given the substantial augmentation in the number of nodes and edges typically resulting from the k-hop/k-ring expansion in hypergraphs, we propose the concept of a P-uniform heterogeneous hypergraph.

Definition 7 (**P-uniform heterogeneous hypergraph**). *Given $k \in \mathbb{N}$ and $P \in \mathbb{N}$, a **P-uniform hypergraph**, denoted as $H_P = (V, \mathcal{E}_k)$, is a heterogeneous hypergraph that every k-hop/k-ring hyperedge $e \in \mathcal{E}_k$ connects exactly P nodes from V. That is:*

$$H_P = (V, \mathcal{E}_k) \text{ is } P\text{-uniform} \Leftrightarrow \forall e \in \mathcal{E}_k, |e| = P, \tag{3}$$

where $|e|$ denotes the cardinality, i.e., the number of heterogeneous nodes of the hyperedge e.

This P-uniform heterogeneous hypergraph addresses the computational and complexity challenges inherent to hypergraph expansions by introducing a uniform sampling of hyperedges, exhibiting invariance in heterogeneous hyperedge cardinality.

Theorem 1 (**Scalability of k-hop/k-ring structures in P-uniform heterogeneous hypergraph**). *For a P-uniform heterogeneous hypergraph $H_P = (V, \mathcal{E}_k)$, let $|V| \to \infty$ while maintaining $|e| = P, \forall e \in \mathcal{E}_k$. It follows that the number of k-hop/k-rings structures grows polynomially with the size of V, assuming a constant average degree.*

The P-uniform heterogeneous hypergraph condenses the hypergraph by uniformly sampling the incorporating nodes and thus balanced representation of the underlying heterogeneous relationships while mitigating the exponential increase in graph components. Then, we construct a HTHG based on the heterogeneous hypergraph snapshots.

Definition 8 (**Heterogeneous Temporal Hypergraph**). *A Heterogeneous Temporal Hypergraph is denoted as $\mathcal{H} = \{H^t\}_{t=1}^T$, where each $H^t = (V^t, \mathcal{E}^t)$ is a snapshot of the hypergraph at time t. Each snapshot consists of a node set V^t, hyperedges set $\mathcal{E}^t$, type-assignment functions for nodes $\phi_h : V^t \to \mathcal{A}_h$ and for hyperedges $\psi_h : \mathcal{E}^t \to \mathcal{R}_h$ at time t. The evolution of this hypergraph over time T captures the dynamic interactions and relationships among entities.*

The hyperedges of HTHGs can dynamically expand to encompass new relationships or contracts to exclude obsolete ones, thereby accurately reflecting the temporal modifications of the system. Then, to harness the analytical power of graph-based algorithms on HTHGs, we propose a novel heterogeneous star expansion strategy that preserves the essential lower-order structures and the directness of connectivity while aptly encapsulating the heterogeneous group interactions. Thus, we offer a refined and effective approach for analyzing complex interaction networks.

Given a P-uniform heterogeneous hypergraph $H_P = (V, \mathcal{E}_k)$, the heterogeneous star expansion constructs a heterogeneous graph $G_* = (V_*, E_*, X)$ from H_P by introducing a new node for every heterogeneous hyperedge $e \in \mathcal{E}_k$, thus $V_* = V \cup \mathcal{E}_k$. The new nodes $v \in V_*$ connect each node in the hyperedge e, i.e., $E_* = E \cup \{(v, e) \mid v \in e\}$.

By introducing a distinct node for each heterogeneous hyperedge $e \in \mathcal{E}$ and connecting it to nodes within e, HTHGN meticulously maintains the heterogeneous nature of the original hypergraph while retaining the connectivity encapsulated in the higher-order relationships of the original hypergraph.

3.2 Multi-scale Message-passing

By constructing a k-hop/k-ring HTHG and conducting the heterogeneous expansion, we obtain the expanded graph $\mathcal{G}_* = \{G_*^t\}_{t=0}^T$. We then utilize a heterogeneous attention message-passing mechanism to aggregate information among heterogeneous nodes as well as between nodes and hyperedges. Specifically, we initialize the features of hyperedge nodes to all zero and execute a single-stage message-passing of the nodes and hyperedges concurrently. This encompasses heterogeneous attention aggregation tailored for heterogeneous relationships within snapshots and temporal attention aggregation designed for dynamics across snapshots.

Heterogeneous Attention Aggregation. Heterogeneous attention aggregation is utilized to accomplish message-passing within snapshots of the HTHG.

For each heterogeneous node $i \in V_*^t$ on snapshot $G_*^t = (V_*^t, E_*^t, X^t)$, type-preserving attribute projection is performed through the heterogeneous input layer:

$$z_i^t = \sigma\left(W_{\phi(i)} x_i^t + b_{\phi(i)}\right), \tag{4}$$

where $x_i^t \in \mathbb{R}^D$ and $z_i^t \in \mathbb{R}^d$ are the original attributes and hidden representation vectors of node i; $W_{\phi(i)} \in \mathbb{R}^{d \times D}$ and $b_{\phi(i)}$ are type-dependent learnable transfer matrices and bias vectors; $\sigma(\cdot)$ represents a activation function such as ReLU.

Subsequently, we introduce a relationship type-dependent graph attention mechanism to model distinct semantic relationships within the expanded graph G_*^t. Specifically, for neighbor nodes $\mathcal{N}_i^a$ connected to node i under a certain type of relationship $a \in \mathcal{A}_i$, we execute the following K-head attention aggregation:

$$\begin{aligned} z_{i,a}^t &= \|_{k=1}^K \left[\sum_{j \in \mathcal{N}_i^a} \alpha_{ij}^k W_a^k z_i^t \right], \\ \alpha_{ij}^k &= \text{Softmax}\left(c \cdot \sigma(W_{\phi(i)}^k z_i^t + W_{\phi(j)}^k z_j^t)\right), \end{aligned} \tag{5}$$

where $z_{i,a}^t \in \mathbb{R}^d$ is the attention aggregated representation of all neighbors $j \in \mathcal{N}_i^a$ with respect to relation type a; $\alpha_{ij}^k \in \mathbb{R}$ is the normalized mutual attention coefficient of node i with j; W_a^k and $W_{\phi(i)}^k \in \mathbb{R}^{d \times d}$ are the learnable key and value vectors transfer matrices; $c \in \mathbb{R}^d$ is the learnable weight vector for calculating the attention coefficient; $\sigma(\cdot)$ represents a nonlinear activation function such as LeakyReLU.

Following the attention aggregation for neighbor nodes under specific relationship types, we further introduce a self-attention mechanism to aggregate the hidden representations of node i concerning neighbors of different relationship types:

$$\begin{aligned} z_i^t &= \sigma\left(\sum_{a \in \mathcal{A}_i} \beta_a z_{i,a}^t\right), \\ \beta_a &= \text{Softmax}\left(\frac{1}{|V_*^t|} \sum_{i=1}^{|V_*^t|} q \cdot \tanh(W_a z_{i,a}^t)\right), \end{aligned} \tag{6}$$

where z_i^t is the representation of node i under the t-th snapshot of expanded HTHG; $\beta_a \in \mathbb{R}$ is the normalized attention score with respect to relationship type a; $W_a \in \mathbb{R}^{d \times d}$ and $q \in \mathbb{R}^d$ are learnable attention transformation matrices respectively; $\sigma(\cdot)$ represents the activation function ReLU.

By executing the above heterogeneous attention aggregation within each snapshot, lower-order heterogeneous semantic information is attentively captured, and higher-order interaction and complex semantics are also aggregated into the hyperedge nodes, which ensures a comprehensive integration of both low- and high-order relation, enhancing the overall representation capacity of the hypergraph by encapsulating a broad of interactions and semantics within its structure.

Temporal Attention Aggregation. This module aggregates node representations calculated under different snapshots and generates dynamic node representations.

Since the temporal attention mechanism will uniformly calculate the node representation under all snapshots, we first add temporal position encoding to each snapshot:

$$z_{i,p}^t = z_i^t + p^t, p_j^t = \begin{cases} \sin(t/10000^{2j/d}), & (\text{if } j \text{ is even}) \\ \cos(t/10000^{2j/d}), & (\text{if } j \text{ is odd}) \end{cases} \tag{7}$$

where $z_{i,p}^t \in \mathbb{R}^d$ is the hidden representation of node i with position encoding about time t; $p^t \in \mathbb{R}^d$ is a deterministic position code with respect to time t.

Then, a temporal attention aggregation module is used to aggregate the representations under different snapshots:

$$\begin{aligned}\bar{z}_i^t &= \sigma\left(\mathrm{FC}\left(\sum_{t'=1}^{T}(\gamma_i^{t,t'} \cdot W_{\mathrm{V}} z_{i,p}^t)\right)\right), \\ \gamma_i^{t,t'} &= \mathrm{Softmax}\left(\frac{1}{\sqrt{d}}[W_{\mathrm{K}} z_{i,p}^t] \cdot [W_{\mathrm{Q}} z_{i,p}^{t'}]\right),\end{aligned} \tag{8}$$

where $\bar{z}_i^t \in \mathbb{R}^d$ is the representation vector obtained by attentionally synthesizing each snapshot of node i; $\gamma_i^{t,t'} \in \mathbb{R}$ is the normalized mutual attention coefficient between the t-th and t'-th snapshots of node i; $W_{\mathrm{K}} \in \mathbb{R}^{d\times d}$, $W_{\mathrm{Q}} \in \mathbb{R}^{d\times d}$ and $W_{\mathrm{V}} \in \mathbb{R}^{d\times d}$ are learnable transformation matrices of Key, Query and Value vectors used to calculate attention weights; $\mathrm{FC}(\cdot)$ is a trainable fully connected layer. The above parameters are not shared among different target node types $\phi(i)$.

Then, a heterogeneous gated residual connection mechanism is used to connect with the previous layer input and calculate the final node representation by summing all snapshots.

$$\hat{z}_i = \sum_{t=1}^{T}\left(r_{\phi(i)}^t \cdot \bar{z}_i^t + (1 - r_{\phi(i)}^t)\,\mathrm{FC}(z_i^t)\right), \tag{9}$$

where z_i^t and $\hat{z}_i$ are the residual node attribute vector obtained by Equation (6) and the updated node representation vector updated by the whole message-passing mechanism; $r_{\phi(i)} \in \mathbb{R}$ a is a trainable variable used to control the update strength of node i; $\mathrm{FC}(\cdot)$ is a trainable fully connected layer.

Stacking two or more layers of the above attention modules enables a single-stage message-passing process: simultaneously from heterogeneous nodes to hyperedge nodes and back to the heterogeneous nodes. This layered approach enhances the depth of information integration, allowing for the iterative refinement of node representations and facilitates a comprehensive bidirectional flow of information. This bidirectional message passing harnesses the strengths of both direct and higher-order interactions, enhancing the ability to capture and understand the complex, multi-faceted relationships present within the hypergraph structure.

3.3 Heterogeneous Contrastive Optimization

To enable the HTHGN to learn heterogeneous semantic information from network data adaptively and to circumvent the issue of lower-order structural information loss caused by the introduction of hyperedges, we optimize the entire model through a self-supervised contrastive learning objective. For the given HTHG $\mathcal{G}_* = \{G_*^t\}_{t=1}^T$ and the target node $i \in V$, we select its heterogeneous neighbors at following snapshot as positive sample set $\mathbb{P}_i^{T+1}$ and uniformly sample Q non-neighbors as negative sample set $\mathbb{N}_i^{T+1}$:

$$\begin{aligned}\mathbb{P}_i^{T+1} &= \{u \mid u \in V^{T+1} \wedge u \in \mathcal{N}_i^{T+1}\}, \\ \mathbb{N}_i^{T+1} &= \{v \mid v \in V^{T+1} \wedge v \notin \mathcal{N}_i^{T+1} \wedge v \neq i\},\end{aligned} \tag{10}$$

Subsequently, we employ a projection head as a discriminator to assess the likelihood of the existence of lower-order relationships between node pairs in the $T+1$-th snapshot:

$$\mathcal{D}(\hat{z}_i, \hat{z}_j) = \mathrm{FC}\left(\sigma(\mathrm{FC}\left(\hat{z}_i \| \hat{z}_j\right))\right). \tag{11}$$

We draw inspiration from the Deep InfoMax for the objective function, adopting a noise-contrastive estimation framework paired with a binary cross-entropy (BCE) loss. This loss function discriminates between pairs of samples originating from the joint distribution of nodes and their corresponding heterogeneous neighbors (positive examples) and those from the marginal distributions (negative examples):

$$\begin{aligned}\mathcal{L} = \sum_{i\in V^t}\Bigg(&\sum_{j\in\mathbb{P}_i^{T+1}} \mathbb{E}\left[\log \mathcal{D}(\hat{z}_i, \hat{z}_j)\right] \\ &+ \sum_{j\in\mathbb{N}_i^{T+1}} \mathbb{E}\left[\log\left(1 - \mathcal{D}(\hat{z}_i, \hat{z}_j)\right)\right]\Bigg).\end{aligned} \tag{12}$$

The model is effectively trained to distinguish between authentic neighboring node relationships and unrelated node pairs across snapshots, thereby enhancing its ability to infer and preserve lower-order connections.

4 Experiments

4.1 Datasets and Baselines

This section evaluates the proposed HTHGN and baselines on three real-world datasets: Yelp, DBLP, and AMiner. We compare static homogeneous methods **VGAE** [Kipf and Welling, 2016], **GATv2** [Brody *et al.*, 2022], **DGI** [Veličković *et al.*, 2019]; dynamic homogeneous GNNs **EvolveGCN** [Pareja *et al.*, 2020], **DySAT** [Sankar *et al.*, 2020]; hypergraph methods **HyperGCN** [Yadati *et al.*, 2019], **UniGCN** and **UniGAT** [Huang and Yang, 2021] , **HGNNP** [Gao *et al.*, 2023a]; static heterogeneous methods **metapath2vec** [Dong *et al.*, 2017], **R-GCN** [Schlichtkrull *et al.*, 2018], **HGT** [Hu *et al.*, 2020], **H-GVAE** [Dalvi *et al.*, 2022], **HPN** [Ji *et al.*, 2023]; dynamic heterogeneous GNNs **DyHATR** [Xue *et al.*, 2020] and **HTGNN** [Fan *et al.*, 2022].

4.2 Experiment Setup

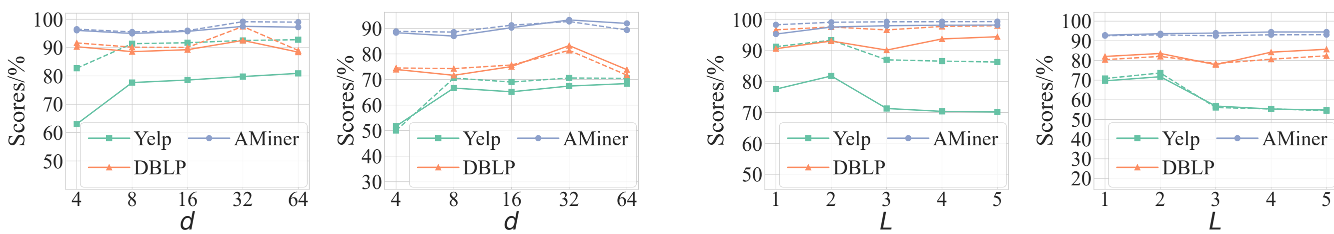
We conducted dynamic link prediction and new link prediction experiments to verify the gain of higher-order interactions on representation learning performance. We held out the last 3 snapshots for testing and trained the model on the remaining snapshots. The link prediction uses all edges in $T+1$-th snapshot as positive edges, while the new link prediction only evaluates edges that have not appeared. We performed 5 repeated randomized experiments for all methods and reported their means and standard deviations. More setup and implementation details see Appendix B.

4.3 Experiment Results

Link Prediction Our comparative experimental results are summarized in Table 1, and more results are in Appendix E. The results show that HTHGN achieves excellent performance on both AUC and AP metrics in all datasets. In particular, we note that methods designed for heterogeneous graphs generally outperform homogeneous graph methods, demonstrating the clear gains of introducing higher-order heterogeneous type information into representation learning. We

Dataset	Yelp		DBLP		AMiner		Avg. Rank
Metrics	AUC	AP	AUC	AP	AUC	AP	
VGAE	58.62 ± 4.79	59.71 ± 4.48	77.40 ± 1.41	80.55 ± 1.53	84.56 ± 2.68	87.13 ± 2.53	9.67
GATv2	59.87 ± 1.31	57.20 ± 1.76	83.28 ± 0.13	84.71 ± 0.27	89.12 ± 0.30	90.60 ± 0.40	6.17
DGI	55.68 ± 3.11	57.36 ± 2.66	73.36 ± 2.93	77.65 ± 2.47	80.71 ± 5.59	84.04 ± 4.60	12.33
EvolveGCN	54.85 ± 5.51	54.79 ± 4.07	71.26 ± 6.87	75.33 ± 5.70	74.90 ± 7.87	78.97 ± 6.28	14.67
DySAT	61.88 ± 2.68	58.57 ± 2.71	78.61 ± 1.54	80.56 ± 1.42	83.76 ± 0.98	85.31 ± 1.18	9.50
HyperGCN	59.18 ± 2.05	55.64 ± 2.03	72.60 ± 1.04	73.96 ± 0.68	75.07 ± 1.24	75.64 ± 2.31	13.67
UniGCN	57.47 ± 5.37	54.99 ± 4.13	75.14 ± 0.76	74.45 ± 3.26	82.35 ± 0.62	81.50 ± 1.45	12.67
UniGAT	55.47 ± 0.57	52.01 ± 0.69	83.88 ± 0.31	86.54 ± 0.34	89.13 ± 0.51	91.19 ± 0.62	7.00
HGNNP	62.16 ± 3.80	60.14 ± 3.54	80.39 ± 0.20	83.39 ± 0.38	85.59 ± 0.82	88.19 ± 0.71	7.17
metapath2vec	63.79 ± 0.41	59.28 ± 0.46	64.28 ± 2.15	60.28 ± 2.56	71.05 ± 1.61	68.95 ± 1.87	13.00
R-GCN	52.72 ± 2.25	51.66 ± 1.48	82.28 ± 3.79	84.18 ± 3.95	88.16 ± 1.55	89.66 ± 1.47	9.83
HGT	55.86 ± 1.97	54.07 ± 2.00	82.32 ± 0.46	85.32 ± 0.55	87.27 ± 0.63	89.94 ± 0.53	8.33
HetSANN-GVAE	59.28 ± 2.41	57.43 ± 2.34	81.51 ± 1.53	85.08 ± 2.44	87.52 ± 0.55	90.19 ± 0.80	6.83
HPN	62.02 ± 1.04	60.24 ± 1.53	81.30 ± 1.12	82.64 ± 1.37	84.39 ± 1.06	87.97 ± 0.77	7.67
DyHATR	63.58 ± 1.37	63.60 ± 1.29	69.61 ± 1.64	69.82 ± 1.72	75.90 ± 2.51	76.80 ± 2.43	11.33
HTGNN	70.43 ± 3.36	67.45 ± 4.18	85.94 ± 3.47	87.17 ± 3.30	90.50 ± 3.33	90.81 ± 3.42	2.17
HTHGN	**74.04 ± 4.82**	**89.56 ± 3.01**	**91.33 ± 1.61**	**96.97 ± 0.56**	**96.58 ± 1.14**	**98.80 ± 0.40**	**1.00**

Table 1: AUC and AP scores of link prediction tasks between HTHGN and baselines in three datasets.



(a) Link Prediction (b) New Link Prediction

Figure 3: Impact of dimension.

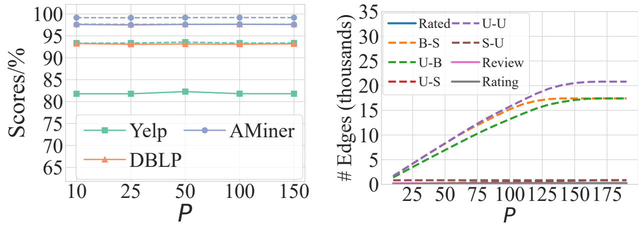
(a) Link Prediction (b) New Link Prediction

Figure 4: Impact of layer.

believe this is because more semantic relatedness can be captured through type information, where homogeneous graph methods are limited. Furthermore, it should be noted that although models designed for homogeneous hypergraphs, such as UniGAT and HGNNP, enlarge the receptive field, their performance suffers due to more noise and the inability to model semantic information. We further observe that compared to static heterogeneous graph methods, dynamic heterogeneous graph methods achieve more competitive performance, thus validating the advantages of modeling network temporal evolution. Compared with the novel HTGNN, our method achieves better results by modeling heterogeneous high-order interaction. We believe this can be attributed to the sparsity of the network structure, and HTHGN rescues this through heterogeneous hyperedges and significantly improves the attention encoder performance. Similar and more significant experimental conclusions can be drawn from the more challenging new link prediction experiment, which is reported in Appendix E.1.

Impact of Hypergraph Construction For the k-hop/k-ring hyperedge construction method proposed in this paper, we set different values for k and compare its performance on 3 datasets under the HTHGN model. From Table 4 in Appendix E.2, we can observe that when $k = 1$, the performance gap between 1-hop and 1-ring is not obvious. This may be because when $k = 1$, the hyperedge is equivalent to its direct neighbors, so there is no functional difference between 1-hop and 1-ring hyperedges. In addition, we should also notice that in all 3 datasets, the model performance gradually increases when k increases. This reflects that increasing the group interaction receptive field for HTHGN can enrich the heterogeneous relationship semantic and thereby improve performance. It should be noted that in all 3 datasets, the HTHGN of 3-ring hyperedge is significantly better than that of 3-hop hyperedge. We believe that due to the 2-layer design of the HTHGN encoder, 2-order neighbors can complete message-passing through low-order interactions. Therefore, the 3-ring hyperedge focuses on heterogeneous nodes with a distance of exactly 3, which can bring more pure and inspiring high-order interactive information.

Impact of P-uniform To explore the impact of P-uniform on HTHGN representation learning performance, we analyzed the hyperedge numbers and model performance under different values of P, which are reported in Figure 5. It can be seen from the results that as P increases, the size of the hyperedge in HTHG increases sharply. This is because the expansion of the hypergraph often leads to a significant increase in the number of nodes and edges. On the contrary,



(a) Link Prediction (b) Yelp Dataset

Figure 5: Impact of P-uniform on HTHGN. (a) shows AUC and AP scores; (b) shows the number of low- and high-order hyperedges on the Yelp dataset. Results indicate that increasing P enlarges the hypergraph but HTHGN maintains stable performance.

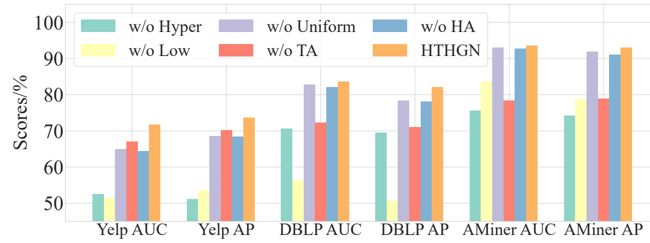


Figure 6: Ablation results of HTHGN and its five ablated variants on three datasets (Yelp, DBLP, and AMiner) with AUC and AP as evaluation metrics. The results demonstrate the performance contribution of each component in the model.

P-uniform HTHG usually has fewer nodes and edges in the converted ordinary graph, which can balance the contradiction between model performance and computing resources and improve computing efficiency. In addition, it is worth noting that the model performance of the proposed HTHGN is relatively stable under different values of P. This result shows the proposed HTHGN can also perform constantly well under a smaller hyperedge scale, verifying that high-order relationships are important for improving the general validity of learning performance. Besides, this result demonstrates that compared with the naive cluster/star expansion algorithm, HTHGN can significantly improve computational efficiency by controlling the P-uniform hyperedges.

Parameter Sensitive Analysis Here, we analyze the impact of the main hyperparameters in HTHGN on model performance. As shown in Figure 3 and 4, HTHGN's performance fluctuates slightly under different configurations, which shows that the model is stable on most tasks and datasets.

Ablation Study To verify the effectiveness of each module, we performed ablation experiments and reported the results as shown in Figure 6 and Appendix E.4. Among them, *w/o Hyper* means removing the hypergraph structure, and *w/o Low* means removing the low-order structure. Both of them significantly affect the performance of each dataset. *w/o Uniform* uses ununiformed hyperedges. *w/o TA* and *w/o HA* respectively represent the removal of the Temporal and Heterogeneous Attention Aggregation modules. Their impact on performance verifies that both temporal dependence and heterogeneous semantics are indispensable.

5 Related Works

Graph representation learning. Real-world complex networks containing heterogeneity and dynamics are ubiquitous, and representation learning and link prediction about them are usually divided into two separate research directions. On the one hand, methods for heterogeneity modeling include meta-path-based methods [Wang *et al.*, 2019; Ji *et al.*, 2023; Yang *et al.*, 2023; Fang *et al.*, 2022; Wang *et al.*, 2022] and heterogeneous message-passing-based methods [Schlichtkrull *et al.*, 2018; Hu *et al.*, 2020; Dalvi *et al.*, 2022; Mao *et al.*, 2023; Fan *et al.*, 2022]. On the other hand, methods for dynamics include decomposition-based methods [Yu *et al.*, 2017; Ma *et al.*, 2017], temporal random walk-based methods [Liu *et al.*, 2020; Huan *et al.*, 2023], and deep learning-based methods [Pareja *et al.*, 2020; Liu *et al.*, 2024]. However, these methods cannot simultaneously effectively capture the dynamics, heterogeneity, and high-order interaction.

Hypergraph representation learning. Hypergraph representation learning aims to embed the hypergraph into a low-dimensional space and maintain the original hypergraph structural information [Antelmi *et al.*, 2023]. Early hypergraph representation learning methods were based on spectral theory [Zhou *et al.*, 2006; Saito *et al.*, 2018] and structure-preserving [Sybrandt *et al.*, 2022; Sybrandt and Safro, 2020] learning node representations. However, these methods are shallow and cannot model highly nonlinear relationships. The advent of GNN-based methods offered a breakthrough with their end-to-end learning and scalability [Antelmi *et al.*, 2023; Yadati *et al.*, 2019; Huang and Yang, 2021; Gao *et al.*, 2023a; Jiao *et al.*, 2024], which define the hypergraph Laplacian and train classic GNNs on hypergraphs. Besides, GNNs designed for heterogeneous hypergraphs usually use hyperedge type [Baytas *et al.*, 2018; Sun *et al.*, 2021; Lu *et al.*, 2023] or meta-path [Li *et al.*, 2023; Yan *et al.*, 2023] to decompose the heterogeneous hypergraph into different semantic relations. This dependency on specific prior knowledge and the static nature limits their efficiency in capturing the dynamic within hypergraphs.

6 Conclusions

In this paper, we propose a HTHGN method to construct and learn heterogeneous high-order interactions in dynamic heterogeneous graphs without additional knowledge. To better divide different receptive fields, we define and analyze two different types of heterogeneous uniform hyperedge construction methods. The effectiveness of the HTHGN proposed in this paper is verified through extensive experiments on three real-world datasets. The limitation of this work is that the process of HTHGN modeling high-order semantic information is relatively complex, and the interpretability of its effectiveness still needs to be explored.

Acknowledgments

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grants LDT23F01015F01 and LMS25F030011, in part by the Key Technology Research and Development Program of Zhejiang Province under Grant 2025C1023, and in part by the National Natural Science Foundation of China under Grant 62372146.

References

[Antelmi *et al.*, 2023] Alessia Antelmi, Gennaro Cordasco, Mirko Polato, Vittorio Scarano, Carmine Spagnuolo, and Dingqi Yang. A survey on hypergraph representation learning. *ACM Computing Surveys*, 56(1), 2023.

[Barros *et al.*, 2021] Claudio D. T. Barros, Matheus R. F. Mendonça, Alex B. Vieira, and Artur Ziviani. A survey on embedding dynamic graphs. *ACM Computing Surveys*, 55(1), 2021.

[Baytas *et al.*, 2018] Inci M. Baytas, Cao Xiao, Fei Wang, Anil K. Jain, and Jiayu Zhou. Heterogeneous hyper-network embedding. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 875–880, 2018.

[Brody *et al.*, 2022] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *ICLR*, 2022.

[Dalvi *et al.*, 2022] Abhishek Dalvi, Ayan Acharya, Jing Gao, and Vasant G Honavar. Variational graph auto-encoders for heterogeneous information network. In *NeurIPS 2022 Workshop: New Frontiers in Graph Learning*, 2022.

[Dong *et al.*, 2017] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. Metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 135–144, 2017.

[Fan *et al.*, 2022] Yujie Fan, Mingxuan Ju, Chuxu Zhang, and Yanfang Ye. Heterogeneous temporal graph neural network. In *Proceedings of the 2022 SIAM International Conference on Data Mining*, pages 657–665, 2022.

[Fang *et al.*, 2022] Yang Fang, Xiang Zhao, Peixin Huang, Weidong Xiao, and Maarten de Rijke. Scalable representation learning for dynamic heterogeneous information networks via metagraphs. *ACM Transactions on Information Systems*, 40(4), 2022.

[Gao *et al.*, 2023a] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2023.

[Gao *et al.*, 2023b] Ziqi Gao, Chenran Jiang, Jiawen Zhang, Xiaosen Jiang, Lanqing Li, Peilin Zhao, Huanming Yang, Yong Huang, and Jia Li. Hierarchical graph learning for protein–protein interaction. *Nature Communications*, 14(1):1093, 2023.

[Hamilton *et al.*, 2017] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NeurIPS*, page 1025–1035, 2017.

[Hu *et al.*, 2020] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of The Web Conference 2020*, page 2704–2710, 2020.

[Huan *et al.*, 2023] Chengying Huan, Shuaiwen Leon Song, Santosh Pandey, Hang Liu, Yongchao Liu, Baptiste Lepers, Changhua He, Kang Chen, Jinlei Jiang, and Yongwei Wu. Tea: A general-purpose temporal graph random walk engine. In *Proceedings of the Eighteenth European Conference on Computer Systems*, page 182–198, 2023.

[Huang and Yang, 2021] Jing Huang and Jie Yang. Unignn: a unified framework for graph and hypergraph neural networks. In *IJCAI*, pages 2563–2569, 2021.

[Ji *et al.*, 2023] Houye Ji, Xiao Wang, Chuan Shi, Bai Wang, and Philip S. Yu. Heterogeneous graph propagation network. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):521–532, 2023.

[Jiao *et al.*, 2024] Pengfei Jiao, Hongqian Chen, Qing Bao, Wang Zhang, and Huaming Wu. Enhancing multi-scale diffusion prediction via sequential hypergraphs and adversarial learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8):8571–8581, 2024.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Li *et al.*, 2023] Jianxin Li, Hao Peng, Yuwei Cao, Yingtong Dou, Hekai Zhang, Philip S. Yu, and Lifang He. Higher-order attribute-enhancing heterogeneous graph neural networks. *IEEE Transactions on Knowledge & Data Engineering*, 35(01):560–574, 2023.

[Liu *et al.*, 2020] Zhining Liu, Dawei Zhou, Yada Zhu, Jinjie Gu, and Jingrui He. Towards fine-grained temporal network representation via time-reinforced random walk. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):4973–4980, 2020.

[Liu *et al.*, 2024] Huan Liu, Pengfei Jiao, Xuan Guo, Huaming Wu, Mengzhou Gao, and Jilin Zhang. Hgn2t: A simple but plug-and-play framework extending hgnns on heterogeneous temporal graphs. *IEEE Transactions on Big Data*, 10(5):620–632, 2024.

[Lu *et al.*, 2023] Yifan Lu, Mengzhou Gao, Huan Liu, Zehao Liu, Wei Yu, Xiaoming Li, and Pengfei Jiao. Neighborhood overlap-aware heterogeneous hypergraph neural network for link prediction. *Pattern Recognition*, 144:109818, 2023.

[Ma *et al.*, 2017] Xiaoke Ma, Penggang Sun, and Guimin Qin. Nonnegative matrix factorization algorithms for link prediction in temporal networks using graph communicability. *Pattern Recognition*, 71:361–374, 2017.

[Mao *et al.*, 2023] Qiheng Mao, Zemin Liu, Chenghao Liu, and Jianling Sun. Hinormer: Representation learning on heterogeneous information networks with graph transformer. In *Proceedings of the ACM Web Conference*, page 599–610, 2023.

[Pareja *et al.*, 2020] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *AAAI*, volume 34, pages 5363–5370, 2020.

[Saito *et al.*, 2018] Shota Saito, Danilo P Mandic, and Hideyuki Suzuki. Hypergraph p-laplacian: a differential geometry view. In *AAAI*, 2018.

[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*, page 519–527, 2020.

[Schlichtkrull *et al.*, 2018] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web*, pages 593–607, 2018.

[Sun *et al.*, 2021] Xiangguo Sun, Hongzhi Yin, Bo Liu, Hongxu Chen, Jiuxin Cao, Yingxia Shao, and Nguyen Quoc Viet Hung. Heterogeneous hypergraph embedding for graph classification. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, page 725–733, 2021.

[Sybrandt and Safro, 2020] Justin Sybrandt and Ilya Safro. First and higher-order bipartite embeddings. In *Proceedings of the 16th International Workshop on Mining and Learning with Graphs (MLG)*, 2020.

[Sybrandt *et al.*, 2022] Justin Sybrandt, Ruslan Shaydulin, and Ilya Safro. Hypergraph partitioning with embeddings. *IEEE Transactions on Knowledge and Data Engineering*, 34(6):2771–2782, 2022.

[Veličković *et al.*, 2019] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *ICLR*, 2019.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, page 2022–2032, 2019.

[Wang *et al.*, 2022] Xiao Wang, Yuanfu Lu, Chuan Shi, Ruijia Wang, Peng Cui, and Shuai Mou. Dynamic heterogeneous information network embedding with meta-path based proximity. *IEEE Transactions on Knowledge and Data Engineering*, 34(3):1117–1132, 2022.

[Wang *et al.*, 2023] Xiao Wang, Deyu Bo, Chuan Shi, Shaohua Fan, Yanfang Ye, and Philip S. Yu. A survey on heterogeneous graph embedding: Methods, techniques, applications and sources. *IEEE Transactions on Big Data*, 9(2):415–436, 2023.

[Xia *et al.*, 2021] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(2):109–127, 2021.

[Xue *et al.*, 2020] Hansheng Xue, Luwei Yang, Wen Jiang, Yi Wei, Yi Hu, and Yu Lin. Modeling dynamic heterogeneous network for link prediction using hierarchical attention with temporal rnn. In *Machine Learning and Knowledge Discovery in Databases: European Conference*, page 282–298, 2020.

[Yadati *et al.*, 2019] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: a new method of training graph convolutional networks on hypergraphs. In *NeurIPS*, 2019.

[Yan *et al.*, 2023] Bo Yan, Cheng Yang, Chuan Shi, Jiawei Liu, and Xiaochen Wang. Abnormal event detection via hypergraph contrastive learning. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, pages 712–720, 2023.

[Yang *et al.*, 2022] Carl Yang, Yuxin Xiao, Yu Zhang, Yizhou Sun, and Jiawei Han. Heterogeneous network representation learning: A unified framework with survey and benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 34(10):4854–4873, 2022.

[Yang *et al.*, 2023] Xiaocheng Yang, Mingyu Yan, Shirui Pan, Xiaochun Ye, and Dongrui Fan. Simple and efficient heterogeneous graph neural network. *AAAI*, 37(9):10816–10824, 2023.

[Yu *et al.*, 2017] Wenchao Yu, Wei Cheng, Charu C Aggarwal, Haifeng Chen, and Wei Wang. Link prediction with spatial and temporal consistency in dynamic networks. In *IJCAI*, pages 3343–3349, 2017.

[Zhang *et al.*, 2023a] Guolin Zhang, Zehui Hu, Guoqiu Wen, Junbo Ma, and Xiaofeng Zhu. Dynamic graph convolutional networks by semi-supervised contrastive learning. *Pattern Recognition*, 139:109486, 2023.

[Zhang *et al.*, 2023b] Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu, and Liang Wang. Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4741–4753, 2023.

[Zhang *et al.*, 2023c] Yingying Zhang, Xian Wu, Quan Fang, Shengsheng Qian, and Changsheng Xu. Knowledge-enhanced attributed multi-task learning for medicine recommendation. *ACM Transactions on Information Systems*, 41(1), 2023.

[Zhou *et al.*, 2006] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *NeurIPS*, volume 19, 2006.