Enhancing Drug Recommendations Via Heterogeneous Graph Representation Learning in EHR Networks

Haijun Zhang, Xian Yang, Liang Bai, and Jiye Liang, *Senior Member, IEEE*

***Abstract*—Electronic health records (EHRs) contain vast medical information like diagnosis, medication, and procedures, enabling personalized drug recommendations and treatment adjustments. However, current drug recommendation methods only model patients' health conditions from EHR data, neglecting the rich relationships within the data. This paper seeks to utilize a heterogeneous information network (HIN) to represent EHR and develop a graph representation learning method for medication recommendation. However, three critical issues need to be investigated: (1) co-occurrence of diagnosis and drug for the same patient does not imply their relevance; (2) patients' directly associated information may not be sufficient to reflect their health conditions; and (3) the cold start problem exists when patients have no historical EHRs. To tackle these challenges, we develop a bi-channel heterogeneous local structural encoder to decouple and extract the diverse information in HIN. Additionally, a global information capture and fusion module, aggregating meta-paths to form a global representation, is introduced to fill the information gaps in records. A longitudinal model using rich structural information available in EHR data is proposed for drug recommendations to new patients. Experimental results on real-world EHR data demonstrate significant improvements over existing approaches.**

***Index Terms*—Drug recommendation, electronic health records, heterogeneous graph representation learning.**

I. Introduction

IN THE medical field, longitudinal electronic health record (EHR) data provides a wealth of information regarding the entire process of patient diagnosis and treatment. Leveraging this data can offer valuable insights into patients' conditions and facilitate recommending prescription drugs, particularly for patients with multimorbidity. Therefore, analyzing and modeling complex information within EHR data to support drug recommendations has become a crucial challenge in the healthcare industry. One notable limitation of conventional drug recommendation methods, including GAMENet [1] and COGNet [2], is their inability to effectively leverage the complex and heterogeneous relationships inherent in electronic health record (EHR) data. These methods primarily focus on learning representations of patients' health conditions by relying on medical codes extracted from EHRs. However, they fall short of fully harnessing the rich and diverse relationships that exist within the EHR data, thereby limiting their effectiveness in capturing the intricate connections between different aspects of patient's health.

The multimodal relationships among various entities in EHR data can be effectively represented using a heterogeneous information network (HIN). To capture comprehensive information from the HIN, various methods have been developed in the field of Heterogeneous Graph Representation (HGR). Meta-path, a commonly used approach, allows for modeling complex heterogeneous relationships between nodes in HIN and can be flexibly applied to different fields. However, most meta-path-based HGR approaches generate representations using homogenous graphs or node sequences that include only one type of node, such as HERec [3] and HAN [4]. Unfortunately, this approach lacks heterogeneity in node representations and fails to integrate important properties from various node types, which goes against our goal of leveraging the diverse information present in the HIN.

To overcome this limitation, several recent methods have been proposed, including HetGNN [5], MAGNN [6], and DisenHAN [7]. These methods address the issue by encoding the local neighborhoods based on meta-paths and then fusing the representations from different views. By doing so, they preserve the heterogeneity of HIN and integrate critical properties from different node types during the representation learning process. However, it is important to note that these methods primarily focus on aggregating local neighborhood information and may not capture a global perspective that encompasses semantic information. In the context of EHR data in the medical field, this limitation can result in a partial assessment of a patient's condition and potentially impact treatment decisions by missing key information.

The primary objective of this paper is to address the difficulties associated with applying graph representation learning techniques to drug recommendations. As illustrated in Fig. 1(a),

Manuscript received 6 April 2023; revised 28 August 2023; accepted 26 October 2023. Date of publication 31 October 2023; date of current version 10 June 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2021ZD0113303 and in part by the National Natural Science Foundation of China under Grants 62022052 and 62276159. Recommended for acceptance by T. Weninger. *(Corresponding author: Liang Bai.)*

Haijun Zhang, Liang Bai, and Jiye Liang are with the Institute of Intelligent Information Processing, Shanxi University, Taiyuan 030006, China (e-mail: zhang_haijun2021@163.com; bailiang@sxu.edu.cn; ljy@sxu.edu.cn).

Xian Yang is with the Alliance Manchester Business School, The University of Manchester, M13 9PL Manchester, U.K. (e-mail: xian.yang@manchester.ac.uk).

Digital Object Identifier 10.1109/TKDE.2023.3329025

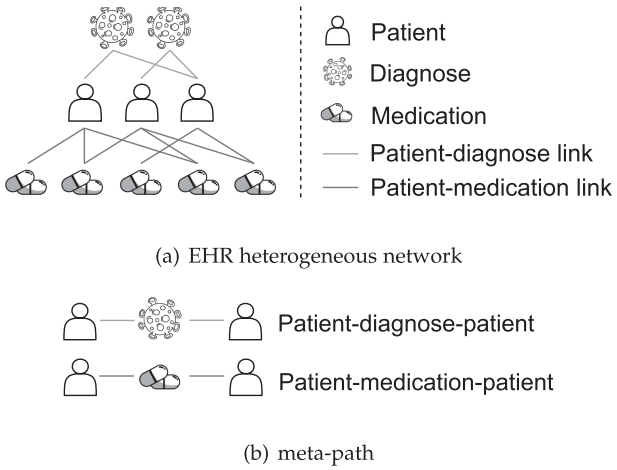


Fig. 1. Illustration of EHR Heterogeneous network.

we propose an EHR heterogeneous network that captures the structural information in EHR data, focusing specifically on medication and diagnosis information relevant to patients. The network comprises nodes representing patients, diagnoses, and medications, with links representing the relationships between patients and their corresponding diagnoses or medications. Our main focus is on addressing three key challenges.

First, we address the challenge posed by the correlational nature of EHR data, where the co-occurrence of diagnosis nodes and drug nodes for the same patient does not imply a real relationship. To overcome this, we decouple the heterogeneous EHR network and analyze patient-diagnosis and patient-medication interactions separately. This separation allows us to prevent conflicting and disparate information from converging, thereby enhancing the accuracy of our analysis. Additionally, as noted in [8], the neighboring nodes' information could be used for node representation learning, enabling multi-hop message passing to capture semantic relations beyond local neighborhoods.

Second, we tackle the challenge that patients' directly associated information in EHR data is incomplete in describing their health condition, which can hinder accurate drug recommendations and compromise patient safety. To overcome this limitation, we propose leveraging information from similar patients to fill in the gaps in the underlying disease information. Our approach involves utilizing meta-paths, powerful tools in heterogeneous network representation learning. By designing specific meta-paths, as illustrated in Fig. 1(b), we capture diverse semantic relations and incorporate domain-specific knowledge. These meta-paths connect patients through nodes of other types, enabling us to explore valuable complementary information and develop a more comprehensive understanding of the data for drug recommendation purposes. Moreover, we recognize the importance of considering the potential interactions among different medications when prescribing drugs for patient treatment. Understanding the interplay between drugs is crucial in ensuring patient health and safety. To address this challenge, we incorporate knowledge about drug-drug interactions (DDIs), allowing us to gain insights into the comprehensive medication status of patients and make well-informed decisions regarding drug recommendations. This approach emulates the process by which doctors accumulate experience from the medical histories of other patients and apply that knowledge to the diagnostic process for new patients.

Lastly, we address the cold start problem encountered in the medical field when dealing with patients who have no previous treatment history. This scenario poses a challenge for healthcare professionals who must develop new treatment plans while considering drug safety and symptom adaptation. To overcome this challenge, we propose leveraging heterogeneous graph representation learning methods to capture the relationships between medical entities and extract unique clinical visit patterns from EHR data. This enables physicians to identify potential parallels and quickly understand the conditions of new patients, allowing them to offer initial diagnostic and therapeutic options. By utilizing the information embedded in the EHR heterogeneous network, our approach assists healthcare professionals in overcoming the limitations of the cold start problem and supports them in making more informed decisions for new patients' diagnoses and treatments.

In summary, this paper introduces a novel **D**rug **Rec**ommendation Model via **H**eterogeneous **G**raph **R**epresentation Learning (**DRecHGR**) that effectively integrates both local and global information to learn graph representations, thus enabling more accurate drug recommendations in the EHR network. We propose techniques such as decoupling the network, utilizing meta-paths, incorporating knowledge about DDIs, and developing a longitudinal model that leverages the rich structural information from the EHR network. These techniques improve the representation and analysis of patient data, facilitating better drug recommendations in healthcare settings. The primary contributions of our work are outlined below.

- We propose a novel approach to model EHR as a heterogeneous graph, incorporating three types of nodes and two types of edges. This modeling technique allows for the integration of diverse patient information and enhances the representation of patient features within the graph.
- To capture the heterogeneous information present in the EHR graph, we introduce a bi-channel local heterogeneous information extraction module. This module effectively encodes diverse information and is complemented by a global information capture and fusion module that fuses the information, providing a global view. This approach enhances the capability to capture information from both the local and global structure of the EHR graph, resulting in more accurate drug recommendations.
- Using the valuable structural information obtained from our model, we leverage a Recurrent Neural Network (RNN) approach to suggest drugs for new patients, taking into account their longitudinal information. This enables us to capture temporal dependencies within the patient's medical history, leading to enhanced accuracy in drug recommendations.

II. Related Work

A. Heterogeneous Graph Representation Learning

Heterogeneous graph representation learning methods can model intricate heterogeneous network structures and rich attribute values to learn low-dimensional node and edge representations for HIN, which is popular in recommendation systems [9], [10], text mining [11], [12] and other applications. For example, T-ContextGGAN [13] is a method used in the medical field that formalizes EHR data as a HIN and utilizes time-aware meta-path and graph attention networks to extract the intrinsic relationship between temporal semantic information and EHR data, leading to improved accuracy in clinical risk prediction. There are four main types of existing heterogeneous graph representation learning methods.

The first category is graph-based neural network methods. The graph representation methods, GCN [14] and its variants [8], [15], [16], apply the convolutional operation to non-euclidean graph-structured data, which are quite effective in modeling complex relationships. In the heterogeneous domain, HAN [4] explores the similarity beyond local by combining the graph attention mechanism and meta-path-based homologous graphs. HetGNN [5] also samples and aggregates local neighbor information according to the type of nodes to extract features. MAGNN [6] generates node embeddings by utilizing intra-meta-path and inter-meta-path aggregation. DisenHAN [7] aggregates neighbors of the same aspect from the views divided by meta-relations. CKD [17] is a collaborative method that improves representation learning through combining regional and global knowledge and employing mutual information-guided distillation. This sort of technique has strong adaptability but may face challenges in handling particular complex relations and node types.

The second category is meta-path-based methods. Metapath2vec [18] captures node context in a heterogeneous graph via meta-paths and employs the skip-gram model [19] to learn representations while considering their proximity. HERec [3] fuses co-occurrence relationships of users and items resulting from meta-path-based random walks to learn node representations for recommendations. HHNE [20] also utilizes meta-path techniques and random walks to get heterogeneous embeddings. The meta-path is also used in different domains. HSGNN [21] uses the adjacency and the encoded representation of subgraphs built based on meta-paths to update embeddings of nodes for disease diagnosis. mSHINE [22] integrates long-term node information based on encoding meta-path-specific representation. Meta-path can also be committed to dynamic HIN, as described in [23], DyHNE augments the adjacency matrix using first and second-order proximity based on evolving meta-path and combines perturbation theory to learn node representations. The meta-path-based techniques offer flexibility, which can be designed with domain knowledge to capture complex semantic relations for networks with specific patterns.

The third category of popular approaches is Random walk-based methods. One such method is DeepWalk [24], which first extracts vertex sequences using the random walk algorithm, then applies word2vec [25] to consider the generated fixed-point sequences as sentences composed of words. Node2vec [26] extends DeepWalk by allowing the generated random walks to reflect both depth-first and breadth-first sampling characteristics. It should be noted that the methods in this category are conventional methods that are not specifically designed for heterogeneous networks. SILK [27] uses a customized random walk to capture the context information of a HIN. The process is guided by a trainable node-conditioned attention matrix that explores various semantics. These methods use random sampling to find pathways in HIN, but they face challenges of sample bias and information loss in exploring context.

Additionally, there are matrix factorization-based approaches that focus on learning embeddings of HIN. HeteroMF [28] and HeteRec [29] utilize the matrix factorization technique to model the contextual dependencies among different types of nodes and learn representations. These methods are mathematically interpretable but face storage and computing challenges and can be affected by noise and missing data.

Among all these categories, meta-path-based techniques offer unparalleled flexibility by integrating domain knowledge to capture intricate semantic relations within networks exhibiting specific patterns. This paper aims to use HIN to represent the medical EHR data and enhance drug recommendations with the help of semantic relationships captured by meta-paths. However, existing meta-path-based methods use homogenous graphs for representation learning, which underutilizes heterogeneous data. Additionally, these techniques concentrate more on collecting neighbor node information from a local perspective, which lacks the ability to regulate the overall global semantics of HIN. For EHR data that contains a wealth of patient health information, this situation may lead to insufficient comprehensive analysis of the patient's overall situation and failure to identify some critical health problems.

B. Drug Recommendation

Drug recommendation methods can be categorized into three types: instance-based, longitudinal-based, and graph-based. The instance-based approach in drug recommendation focuses on extracting features from a patient's current medical information to make recommendations, disregarding the diagnostic information from their past visits. LEAP [30] formalizes drug recommendation as a continuous decision-making process by combining label-dependent circular decoders, attention mechanisms for label instance information, and reinforcement learning. Another instance-based method is PPC [31], which integrates multi-source information such as demographics, diagnostics, and prescriptions data using three linear approaches.

Longitudinal-based approaches leverage recurrent neural networks (RNNs) to model a patient's medical history over time to make more accurate predictions, such as RETAIN [32], which can identify relevant past clinical visits and key diagnostic information in the current visit through a two-level neural attention model. GRU is another example of a longitudinal-based approach that uses patient clinical data and diagnostic codes to predict medication [33]. DMNC [34] enhances the RNNs with

memory-enhancing neural networks and uses two drug prediction modules to thoroughly simulate interactions of different types of views. However, this sort of method faces challenges related to data sparsity and data quality.

In contrast to the aforementioned methods that model individual medical codes separately, GAMENet [1] and SafeDrug [35] incorporate additional information to improve drug recommendation. GAMENet employs GCN [14] to encode drug-drug interactions (DDI) knowledge [36] and drug co-occurrence relationships, along with attention-based historical memory modules, to generate medication combinations. SafeDrug encodes the chemical molecular structure of the drug by utilizing a message-passing neural network to promote better drug recommendations. This approach mainly focuses on drug interactions as nodes for information gathering, neglecting intricate relationships among other medical entities. In [37], a causal graph is used to express the relationships between several modal entities, which simulates the prescribing process of a doctor and facilitates drug recommendations. This paper focuses on developing a graph-based approach specifically designed to extract structural information from heterogeneous EHR networks.

III. Preliminaries

This section introduces the fundamental concepts used in our heterogeneous graph representation learning model for drug recommendation in EHR networks. First, we introduce a heterogeneous EHR network that is constructed from EHR data and consists of nodes of various types, including patients, medications, and diagnoses. Next, we explain how meta-paths can be generated from the EHR network and used to extract global information. We utilize different types of meta-paths to generate two patient-patient interaction networks and one drug co-occurrence network. Additionally, we introduce interaction matrices from the EHR network that allow for the separate analysis of patient nodes' interactions with other node types, which facilitates the extraction of local heterogeneous information.

A. Heterogeneous EHR Network

This paper aims to leverage EHR data for drug recommendation. EHR data is a longitudinal collection of information regarding patients' visits to healthcare providers. In order to model patient concepts explicitly, we utilize diagnoses and medication codes from EHR data. The patient's health status at each hospital visit is represented by binary vectors containing 1 s or 0 s, indicating the presence or absence of diseases and medications, respectively. We denote the Heterogeneous EHR network as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of nodes and $\mathcal{E}$ represents the set of two edge types. The link types R_d and R_m represent patient-diagnose and patient-medication links, respectively. The node types are V_p for patients, V_d for diagnoses, and V_m for medications, and the corresponding node sets are denoted by $\mathcal{V}_p$, $\mathcal{V}_d$, and $\mathcal{V}_m$, respectively. The number of patient, diagnosis, and medication nodes are denoted as N_p, N_d, and N_m.

B. Meta-Path in Heterogeneous EHR Network

As a widely used tool in heterogeneous information networks, the meta-path has a general expression of $V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} V_{l+1}$, representing the reachable path from V_1 to V_{l+1} through the composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$. By applying a known type of meta-path, the neighbors of each node in the graph can be identified, and these nodes and their connections can be used to construct a meta-path-based graph.

In this paper, we generate two patient-patient interaction graphs, namely $G_{pdp} = (\mathcal{V}_p, \mathcal{E}_{pdp})$ and $G_{pmp} = (\mathcal{V}_p, \mathcal{E}_{pmp})$, using the meta-paths $V_p \xrightarrow{R_d} V_d \xrightarrow{R_d} V_p$ and $V_p \xrightarrow{R_m} V_m \xrightarrow{R_m} V_p$, respectively. Here, we define $\mathcal{V}_p$ as the set of patient nodes, and $\mathcal{E}_{pmp}$ and $\mathcal{E}_{pdp}$ refer to the links between patient nodes resulting from the respective meta-paths. We denote the adjacent matrices of G_{pdp} and G_{pmp} as $\mathbf{A}_{pdp}$ and $\mathbf{A}_{pmp}$ with the size of $N_p \times N_p$. Additionally, we utilize a drug occurrence network generated from EHR, which is represented as $G_{mpm} = (\mathcal{V}_m, \mathcal{E}_{mpm})$ with an adjacency matrix $\mathbf{A}_{mpm}$ based on the meta-path $V_m \xrightarrow{R_m} V_p \xrightarrow{R_m} V_m$.

C. Patient-Centric Interaction Matrices From Heterogeneous EHR Network

To facilitate the extraction of local heterogeneous information, this paper splits the EHR heterogeneous network into two networks based on link type: $G_{pd} = (\mathcal{V}_{pd}, \mathcal{E}_{pd})$ for $V_p \xrightarrow{R_d} V_d$ and $G_{pm} = (\mathcal{V}_{pm}, \mathcal{E}_{pm})$ for $V_p \xrightarrow{R_m} V_m$. Here, $\mathcal{V}_{pd} = \{\mathcal{V}_p, \mathcal{V}_d\}$ and $\mathcal{V}_{pm} = \{\mathcal{V}_p, \mathcal{V}_m\}$. $\mathcal{E}_{pd}$ and $\mathcal{E}_{pm}$ are the corresponding sets of links, with patients' interaction matrices $\mathbf{A}_{pd} \in \mathbb{R}^{N_p \times N_d}$ and $\mathbf{A}_{pm} \in \mathbb{R}^{Np \times N_m}$, respectively.

IV. Our Method

A. The Overview of Our DRecHGR Model

In this section, we present a detailed explanation of the DRecHGR model, which was developed to tackle the challenges associated with drug recommendation in EHR networks. The model comprises two key modules, as shown in Fig. 2. The first module is the bi-channel local heterogeneous information extraction module, which utilizes various types of node information and structural information to enrich the node representation. This module plays a crucial role in preventing the convergence of conflicting and disparate information by analyzing patient-diagnosis and patient-medication interactions separately. The second module is the global information capture and fusion module, which is designed to combine meta-paths to aggregate high-order similar patient information to alleviate the issue of incomplete information. This module is critical for capturing complex relationships at the global level that cannot be captured by local information alone. To generate drug recommendations for new patients using their longitudinal information, we leverage the rich structural information learned from our model and apply an RNN-based approach. The captured complex relationship information between medical entities could offer initial diagnostic and therapeutic options for physicians so as to confront

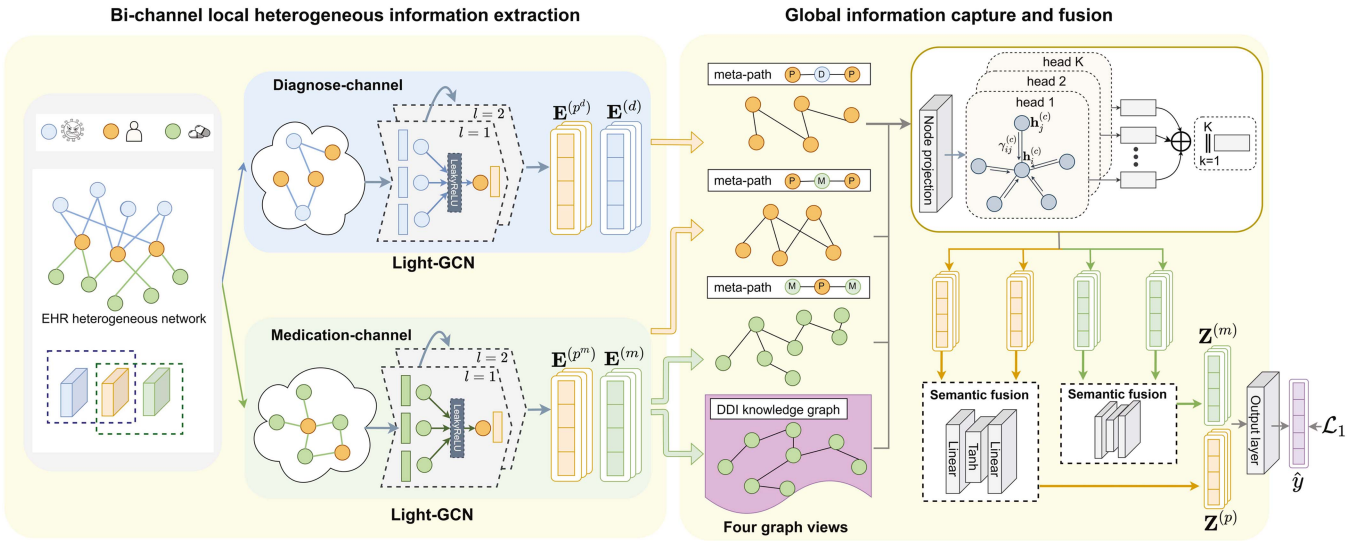


Fig. 2. Overall framework of DRecHGR. The EHR heterogeneous network is first split into two interaction graphs and encoded with the bi-channel local heterogeneous information extraction module to leverage diverse information to enhance node representation. Then, the global information capture and fusion module amalgamates meta-paths to aggregate view-specific high-order structural information in the global perspective, complementing local encoding. The patient and medication representation are redefined by semantic fusion later and used for prediction.

the cold start problem for patients with no treatment history. The following sections elaborate on the different components of our model.

B. Bi-Channel Local Heterogeneous Information Extraction

To encode the local structural information between nodes of various types in G_{pd} and G_{pm}, we utilize a two-layer lightweight graph convolutional network that takes advantage of the relationship between multiple types of nodes to enrich node embeddings, which simplifies graph convolution networks, making them efficient for recommendations on large-scale graphs, and excels in cold start scenarios by leveraging connection information [38]. We denote $\hat{\mathbf{A}}_{pd}$ and $\hat{\mathbf{A}}_{pm}$ as the normalized version of $\mathbf{A}_{pd}$ and $\mathbf{A}_{pm}$. For example, $\hat{\mathbf{A}}_{pd} = \mathbf{D}_{(p)}^{-1/2}\mathbf{A}_{pd}\mathbf{D}_{(d)}^{-1/2}$ where $\mathbf{D}_{(p)} \in \mathbb{R}^{N_p \times N_p}$, $\mathbf{D}_{(d)} \in \mathbb{R}^{N_d \times N_d}$ are diagonal degree matrices.

From G_{pd}, we can get the following representations of patients and diseases:

$$\mathbf{E}_l^{(p^d)} = \text{LeakyReLU}\left(\hat{\mathbf{A}}_{pd}\mathbf{E}_{l-1}^{(d)}\right), \quad \text{for } l \in \{1,2\}, \tag{1}$$

$$\mathbf{E}_l^{(d)} = \text{LeakyReLU}\left(\hat{\mathbf{A}}_{pd}^T\mathbf{E}_{l-1}^{(p^d)}\right), \quad \text{for } l \in \{1,2\}, \tag{2}$$

where the index l denotes the layer of the GCN, and the superscripts p^d indicates that the patient representations are from G_{pd}. Here, $\mathbf{E}_l^{(p^d)} \in \mathbb{R}^{N_p \times D}$ and $\mathbf{E}_l^{(d)} \in \mathbb{R}^{N_d \times D}$ are representations of patients and diseases from lth layer of GCN (where D is the embedding size). The initial input matrices of GCN are randomly initialized and learnable. We use the LeakyReLU activation function with a slope of 0.5 to facilitate function convergence and resolve the issue of a zero gradient in the case of negative input when applying gradient descent.

Similarly, from G_{pm}, we get the following representations of patients and medications:

$$\mathbf{E}_l^{(p^m)} = \text{LeakyReLU}\left(\hat{\mathbf{A}}_{pm}\mathbf{E}_{l-1}^{(m)}\right), \quad \text{for } l \in \{1,2\}, \tag{3}$$

$$\mathbf{E}_l^{(m)} = \text{LeakyReLU}\left(\hat{\mathbf{A}}_{pm}^T\mathbf{E}_{l-1}^{(p^m)}\right), \quad \text{for } l \in \{1,2\}, \tag{4}$$

where the superscripts p^m indicates that the patient representations are from G_{pm}, and $\mathbf{E}_l^{(m)} \in \mathbb{R}^{N_m \times D}$ are representations of medications from the lth layer of GCN.

As observed in [38], individual layers of embedding capture unique semantic meanings. To address the issue of over-smoothing, we aggregate embeddings from all layers to ensure the effective encoding of local structural information without causing over-smoothing. The resulting representations are $\mathbf{E}^{(p^d)} = \sum_0^2 \mathbf{E}_l^{(p^d)}$, $\mathbf{E}^{(p^m)} = \sum_0^2 \mathbf{E}_l^{(p^m)}$, $\mathbf{E}^{(d)} = \sum_0^2 \mathbf{E}_l^{(d)}$, and $\mathbf{E}^{(m)} = \sum_0^2 \mathbf{E}_l^{(m)}$.

C. Global Information Capture and Fusion

1) Learning Representations From Different Graph Views: Local aggregation of heterogeneous information may not be enough to improve drug recommendation performance. To overcome this, we constructed patient-patient interaction graphs G_{pdp} and G_{pmp}, employing threshold values δ_{pdp} and δ_{pmp} to ensure the graph's sparsity and avoid irrelevant or redundant links between patients. We established links between two patients only if they share more than δ_{pdp}/δ_{pmp} common diseases/drugs, respectively.

We then utilized an effective attention mechanism to learn representations of patient and drug nodes. Our approach aims to supplement node representation with global semantic information guided by the DDI network G_{mm} and the EHR medication

co-occurrence network G_{mpm}. We use the local node representations $\mathbf{E}^{(p^d)}$ and $\mathbf{E}^{(p^m)}$ from the previous step for G_{pdp} and G_{pmp} respectively, while $\mathbf{E}^{(m)}$ is used for both G_{mm} and G_{mpm}. To simplify notation, we use the general identifier c to indicate different graph views, where $c \in \{pdp, pmp, mpm, mm\}$.

In each graph G_c, we represent the ith row of its associated $\mathbf{E}^{(c)}$ as $\mathbf{e}_i^{(c)}$. To project these representations, we perform the following step:

$$\mathbf{h}_i^{(c)} = \mathbf{W}^{(c)}\mathbf{e}_i^{(c)}, \tag{5}$$

where $\mathbf{W}^{(c)}$ is the learnable mapping matrix. The attention mechanism is utilized to compute the attention coefficient between node i and j using the softmax function as follows:

$$\gamma_{ij}^{(c)} = \frac{\exp\left(g_{ij}^{(c)}\right)}{\sum_{k\in\mathcal{N}_i^{(c)}} \exp\left(g_{ik}^{(c)}\right)}, \tag{6}$$

where

$$g_{ij}^{(c)} = \text{LeakyReLU}\left(\boldsymbol{\beta}^{(c)^T}(\mathbf{h}_i^{(c)} \| \mathbf{h}_j^{(c)})\right). \tag{7}$$

Here, $\mathcal{N}_i^{(c)}$ is the set of neighboring nodes of node i in graph G_c, $\boldsymbol{\beta}^{(c)}$ is the learnable attention vector, and $\|$ is the concatenation operator.

The representation of node i is then obtained by aggregating the neighbor representations using attention weights, as shown below:

$$\mathbf{s}_i^{(c)} = \text{ELU}\left(\sum_{j\in\mathcal{N}_i^{(c)}} \gamma_{ij}^{(c)}\mathbf{h}_j^{(c)}\right). \tag{8}$$

where ELU stands for the exponential linear unit, which is a type of nonlinear activation layer used in neural networks.

To gather more diverse information and avoid bias towards a single perspective, we use a multi-head attention method. This involves projecting the node representation into multiple subspaces using linear layers and then aggregating the representations under each subspace. This is illustrated as follows:

$$\mathbf{s}_i^{(c)} = \|_{k=1}^{K} \text{ELU}\left(\sum_{j\in\mathcal{N}_i^{(c)(k)}} \gamma_{ij}^{(c)(k)}\mathbf{h}_j^{(c)(k)}\right), \tag{9}$$

where K is the number of heads and superscript (k) indicate the index of head, the operator $\|$ refers to concatenation. Concatenating the representation from each subspace is thought to add new information and increase the robustness of the node representation, leading to a more stable training process.

In this step, the resulting patient representations are $\mathbf{s}_i^{(pdp)}$ and $\mathbf{s}_i^{(pmp)}$ for $i \in \{1, \ldots, N_p\}$, obtained from G_{pdp} and G_{pmp} respectively. Similarly, the resulting medication representations are $\mathbf{s}_j^{(mm)}$ and $\mathbf{s}_j^{(mpm)}$ from G_{mm} and G_{mpm} respectively for $j \in \{1, \ldots, N_m\}$.

2) Semantic Fusion: Based on the above component, each patient and medication is associated with two representations that contain different semantics. To integrate this diversity, we use a learnable weighted fusion mechanism. The weight $\alpha^{(c)}$ of graph G_c for the final representation is computed as follows:

$$\alpha^{(c)} = \frac{1}{|\mathcal{V}_c|}\sum_{i\in\mathcal{V}_c} \mathbf{q}^{\mathrm{T}} \tanh\left(\mathbf{W}\mathbf{s}_i^{(c)} + \mathbf{b}\right), \tag{10}$$

where the learnable parameters shared across different graphs c are the projection matrix $\mathbf{W}$, the bias vector $\mathbf{b}$, and the attention vector $\mathbf{q}$. $\mathcal{V}_c$ represents the node set in graph G_c, and $|\mathcal{V}_c|$ is the size of the set. To improve the distinguishability of representations in different views, we normalize the weight as follows:

$$\eta^{(c)} = \frac{\exp\left(\alpha^{(c)}\right)}{\sum_{c\in\{pdp,pmp,mpm,mm\}} \exp\left(\alpha^{(c)}\right)}, \tag{11}$$

Then, the fused representations of each patient i and each medication j are obtained as follows:

$$\mathbf{z}_i^{(p)} = \eta^{(pdp)}\mathbf{s}_i^{(pdp)} + \eta^{(pmp)}\mathbf{s}_i^{(pmp)}, \tag{12}$$

$$\mathbf{z}_j^{(m)} = \eta^{(mm)}\mathbf{s}_j^{(mm)} + \eta^{(mpm)}\mathbf{s}_j^{(mpm)}. \tag{13}$$

The resulting fused representations for patients and medications can be expressed in matrix form as $\mathbf{Z}^{(p)} \in \mathbb{R}^{N_p\times d}$ and $\mathbf{Z}^{(m)} \in \mathbb{R}^{N_m\times d}$ respectively.

D. Model Training

Next, we concatenate the patient and medication representations to predict the probability of an edge between patient i and medication j in the EHR network:

$$\hat{y}_{ij} = fc_1\left(\left[\mathbf{z}_i^{(p)} \| \mathbf{z}_j^{(m)}\right]\right). \tag{14}$$

Here, $fc_1(\cdot)$ represents a fully connected network consisting of an activation layer and a linear layer. During training, we utilize the commonly used binary cross entropy loss function for link prediction to optimize the model parameters of this module via:

$$\begin{aligned}\mathcal{L}_1 = &-\frac{1}{N_p \times N_m}\sum_{i=1}^{N_p}\sum_{j=1}^{N_m}\left(y_{ij}\log\sigma\left(\hat{y}_{ij}\right)\right.\\ &\left.+\left(1-y_{ij}\right)\log\left(1-\sigma\left(\hat{y}_{ij}\right)\right)\right),\end{aligned} \tag{15}$$

where $\sigma(\cdot)$ is the sigmoid function and y_{ij} is the ground-truth label reflecting the link between patient i and medication j in EHRs. The complete procedure of DRecHGR is demonstrated in Algorithm 1.

E. Drug Recommendations for New Patients With Longitudinal Information

Traditional drug recommendation systems encode medical codes using recurrent neural networks, which produce representations but do not exploit structural information among different medical codes. Inspired by graph neural networks, we use the representations generated by the previous module, followed by a longitudinal model to fine-tune and adapt them to the task of drug recommendation at the visit level. As shown in Fig. 3, the inputs of this module are $\mathbf{E}^{(d)} \in \mathbb{R}^{N_d\times d}$ and $\mathbf{Z}^{(m)} \in \mathbb{R}^{N_m\times d}$ generated by our DRecHGR model. Let $\mathbf{d}_{t,i} \in \mathbb{R}^{N_d}$ indicate

Algorithm 1: The Algorithm of DRecHGR.

Input: EHR heterogeneous Information Network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

Output: Predicted link probability $\hat{y}_{ij}$ between patient node i and medication node j for all i and j in the EHR network $\mathcal{G}$.

Generate two patient-centric interaction graphs G_{pd}, G_{pm} from $\mathcal{G}$.
The graph set $\{G_{pdp}, G_{pmp}, G_{mpm}, G_{mm}\}$ of homogeneous graphs is generated from $\mathcal{G}$ based on different types of meta-paths.
while not converge **do**
  Generate local heterogeneous latent representations by (1)–(4) from G_{pd} and G_{pm};
  for $c \in \{G_{pdp}, G_{pmp}, G_{mpm}, G_{mm}\}$ **do**
    Type-specific Node transformation by (5);
    Calculate the attention weight by (6) and (7);
    Aggregate node representation by (8);
    Calculate and concatenate multi-head embedding to generate embedding in c view by (9);
    Calculate semantic weights for the c view by (10) and (11);
  end for
  Generate fused represention of patients and medications by (12) and (13);
  Predict the probability of patient-medication link by (14);
  Optimize the model by (15);
end while
return the DRecHGR model, and embeddings of nodes in EHR network.

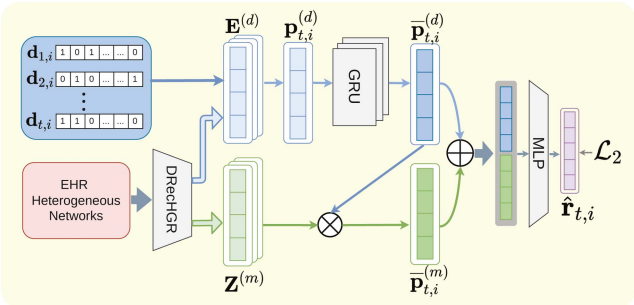


Fig. 3. Longitudinal model for drug recommendation tasks that leverages the representations obtained from DRecHGR for new patients with longitudinal information.

the presence of diseases at the current visit t for patient i. Then the aggregated disease information $\mathbf{p}_{t,i}^{(d)} \in \mathbb{R}^d$ for this patient is:

$$\mathbf{p}_{t,i}^{(d)} = \mathbf{d}_{t,i}^T \mathbf{E}^{(d)}. \tag{16}$$

Next, we utilize the parameters-shared and memorized gated recurrent neural network (GRU) to encode the longitudinal information and learn the representation as below:

$$\overline{\mathbf{p}}_{t,i}^{(d)} = f_{GRU}\left(\mathbf{p}_{1,i}^{(d)}, \mathbf{p}_{2,i}^{(d)}, \ldots, \mathbf{p}_{t,i}^{(d)}\right). \tag{17}$$

Given $\overline{\mathbf{p}}_{t,i}^{(d)} \in \mathbb{R}^d$, we compute the attention score between patient i and all N_m medications. This allows us to obtain the relevant drug information $\overline{\mathbf{p}}_{t,i}^{(m)} \in \mathbb{R}^d$ associated with this patient, as follows:

$$\overline{\mathbf{p}}_{t,i}^{(m)} = \mathbf{Z}^{(m)^\top} \text{Softmax}\left(\mathbf{Z}^{(m)} \overline{\mathbf{p}}_{t,i}^{(d)}\right). \tag{18}$$

By utilizing $\overline{\mathbf{p}}_{t,i}^{(d)}$ and $\overline{\mathbf{p}}_{t,i}^{(m)}$, we can obtain an integrated representation that contains disease and medication information related to this patient, which is used to predict drug labels as follows:

$$\hat{\mathbf{r}}_{t,i} = fc_2\left(\left[\overline{\mathbf{p}}_{t,i}^{(d)} \| \overline{\mathbf{p}}_{t,i}^{(m)}\right]\right), \tag{19}$$

where we define $\hat{\mathbf{r}}_{t,i} \in \mathbb{R}^{N_m}$ as the probability values for recommending different medications to patient i at time t, where its jth value $\hat{r}_{t,i}^j$ indicates the probability of recommending medication j. $fc_2(\cdot)$ represents a fully connected neural network. To learn drug recommendations for new patients, we use the following loss function:

$$\mathcal{L}_2 = -\frac{1}{N_p \times N_m} \sum_i^{N_p} \frac{1}{T_i} \sum_t^{T_i} \sum_j^{N_m} \left(r_{t,i}^j \log \sigma\left(\hat{r}_{t,i}^j\right) + \left(1 - r_{t,i}^j\right) \log\left(1 - \sigma\left(\hat{r}_{t,i}^j\right)\right)\right), \tag{20}$$

where $r_{t,i}^j$ is the ground-truth label, and T_i is the number of hospital visits for patient i. By combining the temporal information from multiple hospital visits with the structural data found in the EHR network, we can obtain a more complete understanding of a patient's health. This enables us to offer more precise drug recommendations based on a more complete picture of the patient's health status.

TABLE I
SUMMARY OF OUR EXPERIMENTAL DATASETS

	MIMIC	MIMIC-HM
# patients	6350	1378
# disease	1958	1576
# drug	112	136
avg. # of visits	2.37	1.28
avg. # of diseases per visit	10.51	9.01
avg. # of drugs per visit	11.65	13.80
# of patient-disease links	75300	8091
# of patient-drug links	87650	12376

V. EXPERIMENTS

A. Dataset and Evaluation Metrics

To assess our model's performance, we utilized the MIMIC Dataset [39] and its subset, which only includes patients with a high likelihood of mortality labeled as MIMIC-HM. Table I

displays the dataset statistics, including the number of patients, prescriptions, and diagnoses (i.e., diseases) in the data. The datasets are publicly available and do not contain any personal information. We performed various data processing operations on both datasets, such as removing missing and duplicate values and sorting the prescription, procedure, and diagnosis files. We also filter out drugs that appeared less frequently in each dataset. In our model's global information capture and fusion module, we incorporated drug-drug interaction knowledge from the TWOSIDES dataset [36], which contains information on over 1,000 drug pairs that are known to cause adverse effects when taken together. This allowed us to better understand drug interactions.

To evaluate the performance of our model, we used multiple metrics, including AUC, AP scores, and TopK measures. The AUC evaluates the ability of our model to differentiate between positive and negative drug recommendations, while the AP score considers both precision and recall. The TopK measures the percentage of the top K drugs recommended by our model that are actually taken by the patient. Specifically, we evaluated our model using top5, top10, and top20. These metrics allowed us to assess the effectiveness of our approach in predicting the most appropriate drugs for individual patients and provided insight into the potential clinical relevance of our findings.

B. Baseline Models

Our model was compared to a group of graph-based learning models, with the predicted probability of a connection between patient and drug obtained by measuring the similarity of their representations. The following is a list of the graph-based learning models that were evaluated:

- *Deepwalk* [24]: DeepWalk is an unsupervised graph-based learning method that incorporates random walk and word2vec algorithms to learn network information and represent nodes as vectors.
- *Node2vec* [26]: Node2vec is a network embedding method that utilizes a biased random walk strategy to produce node sequences. It then applies the word2vec algorithm to learn the embedding vectors of nodes.
- *GCN* [14]: GCN is a representative semi-supervised graph convolutional neural network method that operates in the spectral domain. It iteratively aggregates neighbor information through message passing to update node representation.
- *GAT* [15]: GAT is a popular spatial-based graph convolutional neural network that uses attention mechanisms for iterative aggregation of neighboring nodes' representations.
- *Metapath2vec* [18]: Metapath2vec constructs heterogeneous neighbors through a meta-path-based random walk and then utilizes a heterogeneous skip-gram learning approach to model nodes that share similar structure and semantics.
- *HAN* [4]: HAN is a heterogeneous graph-based learning method that uses node attention and semantic attention mechanisms to capture a comprehensive representation of nodes, leveraging a diverse range of information.
- *CKD* [17]: CKD is a collaborative knowledge distillation approach that leverages regional and global knowledge within meta-paths and employs mutual information-guided distillation for enhanced representation learning.

To assess the efficacy of our model for drug recommendation in new patients with longitudinal information, we fine-tuned our model and compared it with several drug recommendation models specifically designed to capture longitudinal information. The list of these models is presented below:

- *LEAP* [30]: LEAP is an instance-based machine learning method that provides personalized medication combination recommendations by identifying patterns in previous instances of patients with similar medical histories.
- *RETAIN* [32]: RETAIN uses RNNs and a two-stage neural attention model to identify important information from a patient's medical history for drug recommendation.
- *DMNC* [34]: DMNC is a memory-based machine learning method that utilizes longitudinal patient data to improve the coding efficiency of patient representations through a memory-augmented neural network.
- *GAMENet* [1]: GAMENet is another popular longitudinal memory-based machine learning approach that uses both memory networks and graph neural networks to enhance medication representation.
- *COGNet* [2]: COGNet combines transformer-based networks [40] and graph convolutional networks to encode diagnoses, procedures, and medications. It uses hierarchical selection mechanisms to fuse historical medication information at the visit and medication levels.

C. Implementation Details

In order to conduct our experiments, each dataset was divided equally into two halves. The first half was used to evaluate the performance of our model compared to other graph-based learning baselines. The second half of the dataset was used to fine-tune our model and compare it with several drug recommendation models that were specifically designed to capture longitudinal information. By fine-tuning the model, we aimed to adapt it to the task of drug recommendation and capture the dynamic nature of patients' medical histories over time. Each half of the dataset was further split into training, validation, and testing sets in a ratio of 2/3 : 1/6 : 1/6. To construct the two patient-patient interaction graphs G_{pdp} and G_{pmp}, we set appropriate threshold values for δ_{pdp} and δ_{pmp} to maintain graph sparsity by retaining one percent of the edges. Further details on the threshold selection process will be provided in the sensitivity analysis section. We make our codes available at website.[1]

We implement all baseline approaches using the PyTorch version of the code with the same hyperparameter settings. To guarantee a fair comparison, we initialized the embeddings of all nodes in the baseline models randomly and without using any attribute information for initialization. We set the dimension of the embedding vectors to 64 for all models. For methods utilizing multi-head attention mechanisms such as GAT, HAN,

[1]https://github.com/HjZ1998/DRecHGR-master

TABLE II
COMPARING THE PERFORMANCE OF DRECHGR WITH GRAPH-BASED MODELS IN LINK PREDICTION

Dataset	MIMIC														
Training	67%					40%					20%				
Metrics	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20
Deepwalk	87.03	72.04	17.31	33.30	55.59	86.57	68.41	15.78	31.89	54.68	86.95	69.61	15.99	31.97	55.64
Node2vec	86.95	70.14	17.00	32.40	55.02	87.46	70.52	16.96	31.66	55.91	87.28	69.63	16.46	31.72	55.55
GCN	89.80	77.43	19.16	36.19	57.04	89.78	77.40	19.16	36.19	56.92	89.42	76.89	19.16	36.19	55.54
GAT	90.24	78.63	19.27	36.12	58.04	90.42	78.62	19.27	36.19	58.15	90.39	78.52	**19.27**	36.14	58.15
Metapath2vec	85.83	75.09	19.13	35.17	58.32	85.18	74.12	18.73	34.85	57.54	84.64	73.05	18.62	34.54	56.37
CKD	76.65	63.13	18.61	34.03	54.91	76.61	62.74	18.61	34.01	55.17	77.69	63.56	18.06	32.87	54.47
HAN	86.78	68.93	17.18	31.16	46.98	86.75	68.99	17.18	31.16	46.99	86.39	68.47	17.18	31.16	47.88
DRecHGR	**90.89**	**79.21**	**19.31**	**36.34**	**59.03**	**90.88**	**79.11**	**19.29**	**36.33**	**58.97**	**90.84**	**79.09**	19.15	**36.27**	**58.71**
Dataset	MIMIC-HM														
Training	67%					40%					20%				
Metrics	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20
Deepwalk	86.29	54.33	21.82	36.89	59.73	85.79	51.93	20.96	34.39	56.46	85.88	51.77	20.64	33.11	54.04
Node2vec	84.71	53.51	22.34	36.88	56.10	83.05	47.03	18.54	29.98	51.08	83.89	49.91	20.43	32.64	53.80
GCN	90.83	65.24	21.36	40.96	63.33	90.72	64.75	21.36	37.98	63.33	90.66	64.60	21.36	37.98	63.48
GAT	91.45	66.13	21.36	40.96	64.94	91.42	65.99	21.36	40.96	65.36	91.37	65.46	21.36	40.96	64.84
Metapath2vec	80.87	61.05	**30.81**	42.83	57.50	80.83	59.57	**29.55**	42.50	56.61	78.07	52.98	25.03	37.75	53.48
CKD	67.55	38.50	18.95	32.86	39.91	66.50	37.07	18.96	30.94	37.97	64.23	33.06	14.13	24.79	34.41
HAN	90.68	64.75	22.54	39.76	64.13	90.61	64.58	22.54	39.78	64.13	90.60	64.11	22.54	39.76	64.13
DRecHGR	**93.21**	**73.13**	29.78	**47.80**	**70.31**	**92.96**	**73.05**	29.30	**47.50**	**69.50**	**92.96**	**72.31**	**28.43**	**46.49**	**70.34**

and DRecHGR, we set the number of heads to 8. For models encoding longitudinal information, we use GRU to encode temporal visit information, and the hidden layer size is set to 128. The models are trained using the Adam optimizer with a learning rate and weight decay of 0.001, and both the pre-training and fine-tuning stages are trained for 100 epochs. To prevent overfitting, we employ the early stopping strategy with a patience setting of 100 during fine-tuning. If there was no decrease in the loss on the validation set and the AUC evaluation index does not improve for 100 consecutive rounds, the training would be terminated.

D. The Performance of Graph Representation Learning

1) Comparison With Baseline Methods: We compared DRecHGR with several graph-based learning models, and their performance is summarized in Table II. As stated in the experimental setting section, 2/3 of the data samples were used in the training process. In addition to reporting the performance when all 2/3 samples were used for training, we also evaluated the models' performance using fewer samples, specifically at the ratios of 2/5 and 1/5. From Table II, we can first observe that DRecHGR consistently outperforms the other models, achieving the highest AUC, AP, and TopK values across different training proportions. Notably, the performance of the models was robust and not greatly influenced by the size of the training dataset. Further analysis reveals that among all baseline methods, the GAT model had the largest AUC values, which were comparable to the AUC values obtained from DRecHGR. However, the performance of GAT on the MIMIC-III High Mortality Dataset was significantly lower than that of DRecHGR in terms of AP and TopK values. These findings suggest that DRecHGR may hold promise as a superior approach for graph-based learning tasks, particularly in cases where high AP and TopK values are important.

2) Case Studies: In addition to quantitative evaluations, we conducted case studies to further illustrate the effectiveness of our proposed model. For comparison, we chose GAT as a representative baseline. We randomly selected three patients from each dataset to visualize the representations of the patients and their associated drugs using t-SNE. We then plotted the patient and drug representations on a scatter plot, as shown in Table III. In the scatter plot, blue dots represent patients, green dots represent the drugs that were accurately predicted in the Top20 prediction results, and orange dots represent drugs that were not correctly identified. From Table III, we can find that our model outperformed the GAT baseline, with a higher proportion of accurately predicted drugs represented by green dots and a lower proportion of misidentified drugs represented by orange dots. These qualitative experiments provide further evidence of the effectiveness of our proposed method.

3) Sensitivity Analysis: We set δ_{pmp} and δ_{pdp} to 29 and 9 for the MIMIC and 20 and 5 for the MIMIC-HM dataset. By adjusting the values of δ_{pmp} and δ_{pdp}, different edge ratios in G_{pmp} and G_{pdp} can be obtained, resulting in patient-patient interaction graphs of varying sparsity. In this sensitivity analysis subsection, our study investigated how sparsity would influence our model's overall performance on the MIMIC dataset. Fig. 4 illustrates how varying the edge ratio in G_{pmp} (Fig. 4(a)–(e)) and G_{pdp} (Fig. 4(f)–(j)) affects the model performance. These figures reveal a general downward trend in model performance as the proportion of edges in the graph increases, highlighting

TABLE III
CASE STUDIES TO ILLUSTRATE THE PERFORMANCE OF DRECHGR BY EXAMINING DRUG RECOMMENDATION RESULTS FOR RANDOMLY SELECTED PATIENTS

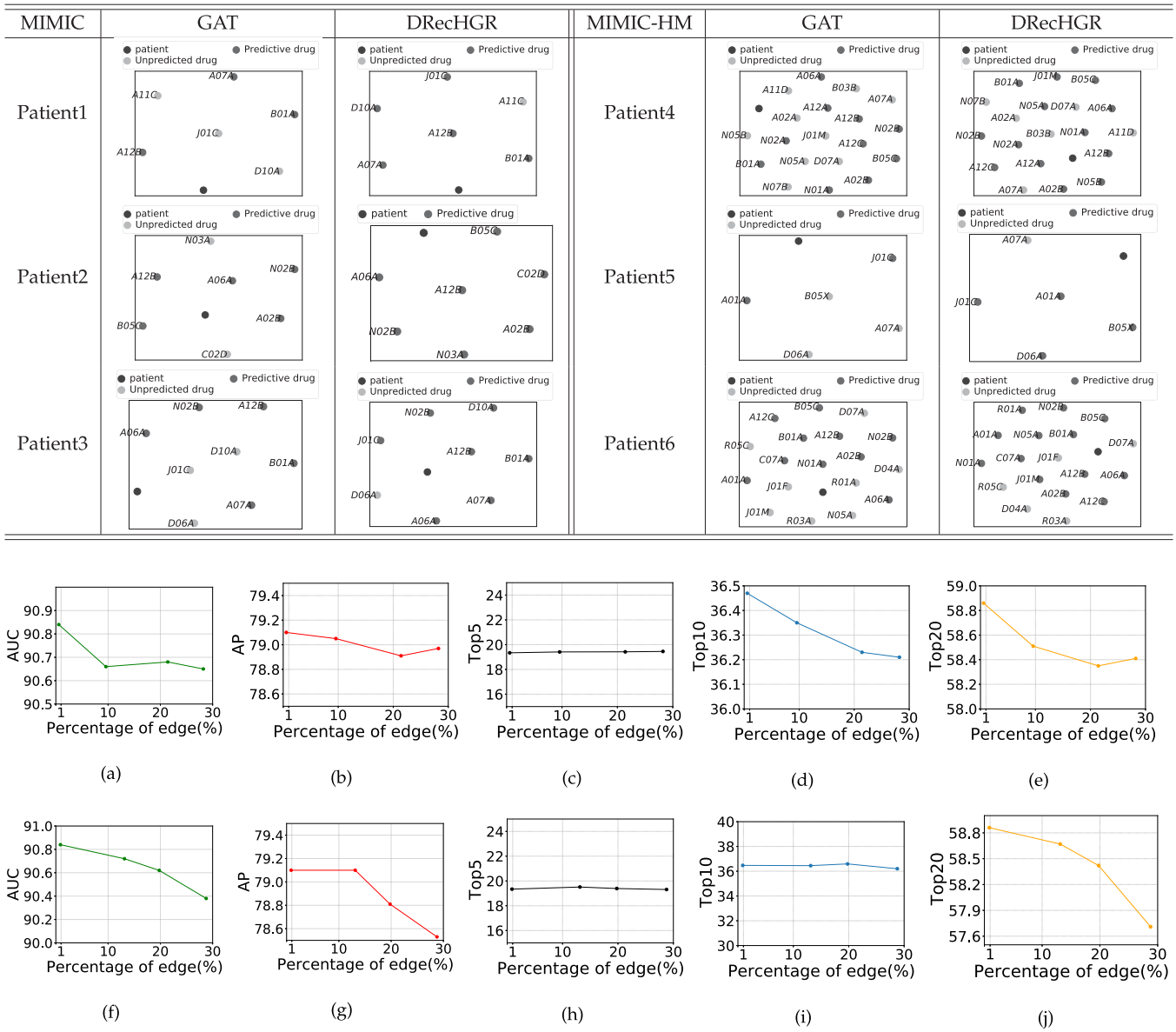


Fig. 4. Sensitivity analysis. (a)–(e) and (f)–(j) illustrates the impact of varying the edge ratio in G_{pmp} and G_{pdp} on the model performance, respectively.

the effect of graph sparsity on representation learning. However, the reduction in performance is not significant, indicating the robustness of our model. Conducting the ablation study on G_{pmp} and G_{pdp} is detailed in Appendix A, available online. G_{mm} and G_{mpm} are also considered.

E. The Performance of Drug Recommendation for New Patients

To evaluate the effectiveness of our proposed model for drug recommendation in patients with longitudinal information, we fine-tuned our model and compared it with several drug recommendation models specifically designed to capture longitudinal information. Same as described in Section V-D1, here we also evaluated the models' performance at different sample sizes. The results are presented in Table IV. It can be observed that our proposed model achieves the best performance under different data settings compared to all the baselines. Among all baselines, GAMENet shows the best performance for most datasets, which is attributed to its combination of memory network module and graph representation learning techniques to suggest safe and effective medications while also utilizing previous visit data. However, DRecHGR-task outperforms GAMENet, demonstrating the power of graph representation learning. The improvement of our method results from the full exploitation of medication information by global information capture and fusion and the integration of various types of medical code information into the representation by encoding the EHR network structure. Besides,

TABLE IV
COMPARING THE PERFORMANCE OF DRECHGR WITH BASELINE APPROACHES IN DRUG RECOMMENDATION FOR NEW PATIENTS WITH LONGITUDINAL INFORMATION

Dataset	MIMIC														
Training	67%					40%					20%				
Metrics	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20
DMNC	86.31	60.65	18.42	35.20	58.84	85.58	58.72	17.36	34.54	58.14	84.92	58.34	17.44	34.26	57.47
Retain	91.65	75.36	23.69	42.34	66.46	91.08	74.14	23.53	41.84	65.83	89.97	71.75	22.79	41.12	63.20
LEAP	89.06	65.59	19.68	36.35	62.73	88.56	64.13	19.19	35.54	61.68	87.61	62.41	18.66	34.53	60.20
GAMENet	91.32	74.51	23.80	42.60	67.21	90.51	72.30	22.94	41.54	65.28	89.35	70.01	22.70	40.40	63.27
COGNet	90.86	**76.20**	22.73	36.36	**68.18**	87.37	71.95	22.73	40.91	54.55	90.10	**74.77**	22.73	40.91	54.55
DRecHGR-Task	**91.84**	75.55	**24.15**	**43.36**	67.11	**91.37**	**74.72**	**24.15**	**43.17**	**66.19**	**90.15**	72.17	**23.81**	**41.96**	**64.05**
Dataset	MIMIC-HM														
Training	67%					40%					20%				
Metrics	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20	AUC	AP	Top5	Top10	Top20
DMNC	88.92	57.74	24.86	42.14	64.51	87.73	57.60	24.88	40.09	62.13	87.83	54.41	25.51	40.24	61.26
Retain	89.65	62.70	19.12	36.57	57.10	88.60	62.10	21.45	34.64	57.73	86.85	60.99	20.12	34.98	57.54
LEAP	91.03	61.49	26.65	42.78	67.70	90.13	59.91	25.13	40.82	65.16	89.94	61.21	26.66	42.58	64.09
GAMENet	92.14	68.07	29.17	48.36	70.55	91.47	66.82	30.27	47.14	69.61	90.93	67.03	30.52	46.94	**69.00**
COGNet	87.12	60.12	28.57	42.86	64.29	86.65	55.26	28.57	42.86	57.14	85.48	57.85	28.57	**50.00**	57.14
DRecHGR-Task	**93.02**	**71.31**	**31.36**	**50.56**	**72.28**	**93.02**	**71.31**	**31.36**	**50.56**	**72.28**	**91.82**	**69.18**	**31.31**	49.72	68.98

the analysis of the impacts of using historical records is provided in Appendix B, available online.

VI. CONCLUSION

This paper addressed the difficulties of extracting information from heterogeneous EHR networks, specifically for drug recommendation tasks. We introduced a novel approach, the DRecHGR model, that can effectively combine both local and global information to learn graph representations, enabling more accurate drug recommendations in EHR networks. Specifically, we proposed a bi-channel heterogeneous local structural encoder that enhances node representation and captures the essence of heterogeneous graphs. In addition, we introduced a global information capture and fusion mechanism to capture global information and integrate both local and global information, resulting in a more comprehensive graph representation. Furthermore, in order to evaluate the effectiveness of our proposed approach for drug recommendation tasks for new patients with longitudinal information, we developed a longitudinal model that utilizes the rich structural information present in EHR data. This structural information is obtained through the use of our DRecHGR model. Our experimental results demonstrate our approach's effectiveness, showing our proposed DRecHGR model has the potential to enhance clinical decision-making by providing more accurate drug recommendations and improving patient outcomes.

REFERENCES

[1] J. Shang, C. Xiao, T. Ma, H. Li, and J. Sun, "GameNet: Graph augmented memory networks for recommending medication combination," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 1126–1133.

[2] R. Wu, Z. Qiu, J. Jiang, G. Qi, and X. Wu, "Conditional generation net for medication recommendation," in *Proc. ACM Web Conf.*, 2022, pp. 935–945.

[3] C. Shi, B. Hu, W. X. Zhao, and S. Y. Philip, "Heterogeneous information network embedding for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 357–370, Feb. 2019.

[4] X. Wang et al., "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.

[5] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 793–803.

[6] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proc. Web Conf.*, 2020, pp. 2331–2341.

[7] Y. Wang, S. Tang, Y. Lei, W. Song, S. Wang, and M. Zhang, "DisenHAN: Disentangled heterogeneous graph attention network for recommendation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1605–1614.

[8] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[9] W. Feng and J. Wang, "Incorporating heterogeneous information for personalized tag recommendation in social tagging systems," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2012, pp. 1276–1284.

[10] L. Wu, P. Sun, Y. Fu, R. Hong, X. Wang, and M. Wang, "A neural influence diffusion model for social recommendation," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 235–244.

[11] H. Linmei, T. Yang, C. Shi, H. Ji, and X. Li, "Heterogeneous graph attention networks for semi-supervised short text classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process.*, 2019, pp. 4821–4830.

[12] Y. Liang et al., "Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification," *Inf. Sci.*, vol. 548, pp. 295–312, 2021.

[13] Y. Xu et al., "Time-aware context-gated graph attention network for clinical risk prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 7557–7568, Jul. 2023.

[14] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

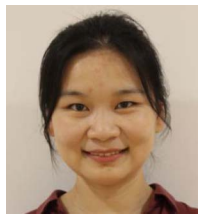
[15] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv: 1710.10903*.

[16] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6861–6871.

[17] C. Wang et al., "Collaborative knowledge distillation for heterogeneous information network embedding," in *Proc. ACM Web Conf.*, 2022, pp. 1631–1639.

[18] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks," in *Proc. 23 rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, pp. 135–144.
[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[20] X. Wang, Y. Zhang, and C. Shi, "Hyperbolic heterogeneous information network embedding," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5337–5344.
[21] Z. Liu, X. Li, H. Peng, L. He, and S. Y. Philip, "Heterogeneous similarity graph neural network on electronic health records," in *Proc. IEEE Int. Conf. Big Data*, 2020, pp. 1196–1205.
[22] X. Zhang and L. Chen, "mSHINE: A multiple-meta-paths simultaneous learning framework for heterogeneous information network embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 7, pp. 3391–3404, Jul. 2022.
[23] X. Wang, Y. Lu, C. Shi, R. Wang, P. Cui, and S. Mou, "Dynamic heterogeneous information network embedding with meta-path based proximity," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1117–1132, Mar. 2022.
[24] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 701–710.
[25] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
[26] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 855–864.
[27] Y. Hao, X. Wang, X. Wang, X. Wang, C. Chen, and M. Song, "Walking with attention: Self-guided walking for heterogeneous graph embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 6047–6060, Dec. 2022.
[28] M. Jamali and L. Lakshmanan, "HeteroMF: Recommendation in heterogeneous information networks using context dependent factor models," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 643–654.
[29] X. Yu et al., "Recommendation in heterogeneous information networks with implicit user feedback," in *Proc. 7th ACM Conf. Recommender Syst.*, 2013, pp. 347–350.
[30] Y. Zhang, R. Chen, J. Tang, W. F. Stewart, and J. Sun, "Leap: Learning to prescribe effective and safe treatment combinations for multimorbidity," in *Proc. 23 rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, pp. 1315–1324.
[31] L. Wang, W. Zhang, X. He, and H. Zha, "Personalized prescription for comorbidity," in *Proc. Database Syst. Adv. Appl.: 23 rd Int. Conf.*, 2018, pp. 3–19.
[32] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3512–3520.
[33] J. M. Bajor and T. A. Lasko, "Predicting medications from diagnostic codes with recurrent neural networks," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–19.
[34] H. Le, T. Tran, and S. Venkatesh, "Dual memory neural computer for asynchronous two-view sequential learning," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 1637–1645.
[35] C. Yang, C. Xiao, F. Ma, L. Glass, and J. Sun, "SafeDrug: Dual molecular graph encoders for recommending effective and safe drug combinations," 2021, *arXiv:2105.02711*.
[36] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions," *Sci. Transl. Med.*, vol. 4, no. 125, pp. 125ra31–125ra31, 2012.
[37] H. Sun, S. Xie, S. Li, Y. Chen, J.-R. Wen, and R. Yan, "Debiased, longitudinal and coordinated drug recommendation through multi-visit clinic records," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 27837–27849.
[38] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and powering graph convolution network for recommendation," in *Proc. 43 rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2020, pp. 639–648.
[39] A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.
[40] A. Vaswani et al., "Attention is all you need," 2017, *arXiv: 1706.03762*.

Haijun Zhang is currently working toward the postgraduation research degree in the School of Computer and Information Technology at Shanxi University. Her research interest is in the areas of recommendation systems in healthcare.



Xian Yang received the PhD degree from the Department of Computing, Imperial College London, in 2016. She is currently a lecturer (assistant professor) with Alliance Manchester Business School, the University of Manchester. Before joining AMBS, she worked as an assistant professor with the Department of Computer Science from HKBU, a researcher with Microsoft Research Asia and a research fellow, the Data Science Institute of Imperial College London. Her research interests include artificial intelligence in healthcare, natural language processing, data mining and computational epidemiology.

Liang Bai received the PhD degree in computer science from Shanxi University, in 2012. He is currently a professor with the School of Computer and Information Technology, Shanxi University. His research interest is in the areas of cluster analysis. He has published several papers in his research fields, including *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Fuzzy Systems*, ICML, KDD, and AAAI.

Jiye Liang (Senior Member, IEEE) received the PhD degree from Xi'an Jiaotong University. He is a professor in Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, the School of Computer and Information Technology, Shanxi University. His research interests include artificial intelligence, granular computing, data mining, and machine learning. He has published more than 160 international papers in his research fields, including *Artificial Intelligence*, *Journal of Machine Learning Research*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Fuzzy Systems*, *Data Mining and Knowledge Discovery*, ICML, KDD, and AAAI.