TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 16/31 pp215–233
DOI: 10.26599/TST.2023.9010149
Volume 30, Number 1, February 2025

Synergistic Multi-Drug Combination Prediction Based on Heterogeneous Network Representation Learning with Contrastive Learning

Xin Xi, Jinhui Yuan, Shan Lu, and Jieyue He*

Abstract: The combination of multiple drugs is a significant therapeutic strategy that can enhance treatment effectiveness and reduce medication side effects. However, identifying effective synergistic drug combinations in a vast search space remains challenging. Current methods for predicting synergistic drug combinations primarily rely on calculating drug similarity based on the drug heterogeneous network or drug information, enabling the prediction of pairwise synergistic drug combinations. However, these methods not only fail to fully study the rich information in drug heterogeneous networks, but also can only predict pairwise drug combinations. To address these limitations, we present a novel Synergistic Multi-drug Combination prediction method of Western medicine based on Heterogeneous Network representation learning with Contrastive Learning, called SMC-HNCL. Specifically, two drug features are learnt from different perspectives using the drug heterogeneous network and anatomical therapeutic chemical (ATC) codes, and fused by attention mechanism. Furthermore, a group representation method based on multi-head self-attention is employed to learn representations of drug combinations, innovatively realizing the prediction of synergistic multi-drug combinations. Experimental results demonstrate that SMC-HNCL outperforms the state-of-the-art baseline methods in predicting synergistic drug pairs on two synergistic drug combination datasets and can also effectively predict synergistic multi-drug combinations.

Key words: western medicine; synergistic drug combination; heterogeneous network; contrastive learning

1 Introduction

Synergistic drug combinations refer to the simultaneous use of two or more drugs in order to achieve enhanced therapeutic effects, reduced side effects, and prevention of drug resistance[1].

Synergistic drug combinations are an increasingly important pharmacological treatment strategy and widely used in clinical practice. For example, in cancer treatment, the combination of multiple chemotherapy drugs can improve cancer treatment outcomes and reduce the development of drug resistance. The synergistic effects of drug combinations can occur between drugs with different mechanisms, targets, and categories, which are closely related to factors such as the chemical structure and pharmacological properties of the drugs[2]. Therefore, studying synergistic drug combinations can help researchers better understand the synergistic mechanisms of drugs, facilitate the development of more effective and safer drugs, and optimize drug treatment regimens[3]. Currently, the study of synergistic drug combinations has become an

• Xin Xi, Jinhui Yuan, and Jieyue He are with School of Computer Science and Engineering, Southeast University, Nanjing 211189, China. E-mail: xixin@seu.edu.cn; jinhuiyuan@seu.edu.cn; jieyuehe@seu.edu.cn.
• Shan Lu is with Nanjing FiberHome Tiandi Co., Ltd., Nanjing 211161, China. E-mail: bfact.cn@gmail.com.
* To whom correspondence should be addressed.
Manuscript received: 2023-08-31; revised: 2023-10-28; accepted: 2023-12-06

increasingly significant research area in pharmacology. With advancements in high-throughput screening and big data technologies, research on synergistic drug combinations has been greatly promoted and accelerated. However, using traditional wet-lab methods to identify drug combinations from a large number of potential drug pairs is expensive and time-consuming. In recent years, research on the prediction of synergistic drug combinations has gained attention. The pharmacological and network topological features of drugs have been expanded in a large number of effective synergistic drug combinations. These features are used to establish computational models that calculate synergistic scores, enabling efficient screening of potential synergistic drug combinations. Relying on these computational models, available drug combination databases such as DDInter[4] and CDCDB[5] have been accumulated. These databases enrich the data foundation for the identification of synergistic drug combinations through computational methods, accelerating the process of discovering synergistic drug combinations.

Currently, computational methods for predicting synergistic drug combinations can be mainly categorized into three types[1]: feature-based methods, network-based methods, and hybrid methods. Most feature-based methods[6–9] rely on various drug information to calculate features based on drug similarity. However, many drugs lack comprehensive information, leading to suboptimal prediction performance. Network-based methods[10–14] integrate diverse data to construct drug heterogeneous networks and use a series of classical network topology similarity calculation methods to compute drug similarity as features for synergy prediction. However, the current learning of drug networks is still limited to simple topological analysis. Hybrid methods[15–20] combine the aforementioned two types of features for synergy prediction, but they also primarily utilize low-level features. In summary, the above-mentioned methods for predicting synergistic drug combinations primarily rely on drug heterogeneous networks and various drug information to compute drug similarity, so as to predict synergistic effects for drug pairs. They mostly utilize simple low-level features (such as drug molecular fingerprints[15]) and calculate drug similarity based on basic network topology (such as path distance[17]) when using drug-target heterogeneous networks. They fail to fully exploit the rich semantic and associative information within the drug-target heterogeneous networks. These factors can impact prediction accuracy, and these methods can only predict pairwise drug combinations. Currently, predicting synergistic drug combinations still faces numerous difficulties and challenges. It is of great research significance and practical value to leverage existing biomedical data resources and develop more efficient methods for predicting synergistic drug combinations to facilitate the development of novel combination drugs.

To address the aforementioned problems, we propose a novel synergistic multi-drug combination prediction method of western medicine based on heterogeneous network representation learning with contrastive learning, called SMC-HNCL. The main contributions of this paper are as follows:

(1) Heterogeneous network representation learning with contrastive learning is used to learn local and global information from each metapath in the network, resulting in high-quality network-based drug features. Additionally, another drug features based on the similarity of anatomical therapeutic chemical (ATC) codes are computed by Jaccard coefficient. An attention mechanism is employed to differentiate the importance of different drug features and fuse them to obtain enhanced drug feature.

(2) A group representation method based on multi-head self-attention is adopted to learn the importance of different drugs in drug combination, leading to effective representation of drug combination for predicting their synergy, innovatively realizing the prediction of synergistic multi-drug combinations.

(3) Extensive experiments are conducted on two real synergistic drug combination datasets. The results validate the effectiveness of SMC-HNCL and demonstrate its superiority over other state-of-the-art methods.

The rest is organized as follows. Section 2 reviews related work in the fields of computational synergistic drug combination prediction and heterogeneous network representation learning. Section 3 provides a detailed description of SMC-HNCL. Section 4 reports and analyzes the performance of SMC-HNCL. Finally, Section 5 concludes the paper.

2 Related Work

The SMC-HNCL method utilizes a heterogeneous network representation learning algorithm to obtain

drug network-based feature for synergy prediction. Therefore, the related work is divided into two parts: the prediction of synergistic drug combinations and the heterogeneous network representation learning.

2.1 Synergistic drug combination prediction

The development of combination drugs for complex diseases treatment is gaining prominence over single drugs due to their significant potential. The main challenge in developing combination drugs lies in the vast search space of possible drug combinations. Therefore, it is crucial to predict the efficacy of drug combinations and rank potential combinations using computational methods. The initial methods for identifying synergistic drug combinations include the Loewe additivity model and the Bliss independence model[21]. However, numerous computational methods have been proposed for predicting pairwise synergistic drug combinations recently. The computational methods for predicting synergistic drug combinations of western medicine can be broadly classified into three categories[1]: feature-based methods, network-based methods, and hybrid methods.

Feature-based methods are early and common computational approaches for predicting synergistic drug combinations. These methods utilized different available information to learn drug features and employ machine learning algorithms to predict synergistic drug combinations. For instance, Zhao et al.[6] integrated molecular and pharmacological information of drugs to predict synergistic drug pairs and predict novel drug pairs. Additionally, Chen et al.[7] integrated a set of drug properties to study drug pairs' feature, and then used random forests to predict synergistic drug pairs. Sun et al.[8] utilized the Hadoop MapReduce programming model to discover synergistic drug pairs， which constructs feature of drug pairs using drug gene expression data and drug chemical information, then a support vector machine and naive bayes classifiers were employed to predict synergistic drug pairs. Xu et al.[9] simultaneously considered the chemical, biological, and pharmacological properties of drugs, and used the minimum redundancy with maximum relevance algorithm to select features. Chen et al.[22] introduced a decision stump-based solution in drug combination prediction to generate a decision tree for evaluating nodes in global view and improve generalization ability. Feature-based methods are commonly used in the early stages of synergistic drug combination research. While integrating various drug information, they overlook drug network information which may yield suboptimal results due to missing information for many drugs.

Network-based methods are widely used recently by utilizing drug networks to learn drug similarity or features for predicting synergistic drug combinations. For example, Wang et al.[10] proposed a method based on the protein-protein interaction (PPI) network to identify potential synergistic drug pairs. They simulated the effects of various drug combinations on the PPI network and selected drug combinations with the highest synergistic potential. Zou et al.[11] built a large-scale drug combination network and used a neighbor community-based approach to explore the drug synergy relationships, demonstrating that the topological and functional characteristics of neighboring communities trait effectively predict drug combinations, providing a fresh perspective for drug research. Yin et al.[12] simulated the interaction of drug combinations with targets in the PPI network and found that the effectiveness of drug combinations largely depends on the target network topology, indicating that discovering new synergistic drug combinations based on network topology is promising. Chen et al.[13] utilized a drug heterogeneous network and employed Laplacian regularized least squares to rank potential synergistic drug combinations. This method integrated known synergistic drug combinations, drug-target interaction, and drug chemical structures, enabling the discovery of synergy between main drugs and auxiliary drugs without requiring negative samples. However, this method lacks important information such as ATC codes information and PPI networks. Cheng et al.[14] proposed a prediction method for synergistic drug pairs targeting specific diseases based on a drug-target-disease heterogeneous network. They quantified the relationships between drug-targets and disease-targets in the PPI network. While network-based methods utilize drug network information, they mainly focus on topological distance, often missing crucial drug information, which affects prediction accuracy. Only utilizing different drug information or drug networks to learn drug features is insufficient to fully explore the correlation between drugs. Therefore, hybrid methods that integrated various drug information and drug heterogeneous networks have been proposed for the prediction of synergistic drug combinations recently.

For instance, EPSDC[15] introduces a synergistic drug combination prediction framework. It effectively predicts drug combinations by integrating information from multiple sources, including biology, chemistry, pharmacology, and networks. EPSDC builds feature vectors for drug pairs using various drug similarities, which are then used in a predictor to assign scores. Transductive learning is applied to a drug-target network to get network-based scores. Ultimately, two integration rules combines both scores to prioritize potential drug combinations. DSGCR[16] proposes a graph-based collaborative regularization calculation method to predict drug synergistic effects. It effectively combines drug-target networks, pharmacology information, and prior knowledge of drug combinations for synergy prediction. ISDCSMP[17] constructs a new heterogeneous network based on known drug combinations, anatomical therapeutic similarity, drug-target protein associations and protein-protein interactions, then calculates the correlation between drug pairs to obtain the combination coefficients of metapaths by ridge regression. Finally, the synergy scores can be calculated based on combination coefficients and correlations of metapaths. Unlike EPSDC, ISDCSMP can discover potential synergistic drug combinations without the negative samples. DSML[18] utilizes multitask learning, integrating drug targets, protein-protein interaction, ATC codes and known synergistic drug combinations to predict synergistic drug combinations. By reconstructing drug-target interactions with multiple knowledge sources, prediction accuracy is notably enhanced. NEWMIN[19] introduces a graph representation learning-based algorithm for predicting synergistic drug pairs. It calculates drug similarity based on six relevant drug information, to construct a multi-layer drug similarity network. A random walk algorithm samples information from the multi-layer networks and assesses each network's importance. A random forest then predicted synergistic drug pairs. Zhang et al. proposed MGAE-DC[20] to learn drug embeddings considering both synergistic, additive and antagonistic combinations through multiple channels, which make the drug embeddings become more discriminative. While hybrid methods consider various drug information and drug networks, it often miss vital information, which can introduce noise. Moreover, the learning of drug networks is still limited to topological analysis, which compromises predictive accuracy.

In summary, existing methods primarily rely on various drug information or drug networks to calculate drug similarity for predicting pairwise synergistic drug combinations. However, current methods only use simple low-level features, such as drug molecular fingerprints and drug ATC codes information, when calculating drug similarity. Similarly, when utilizing drug networks, they only consider simple network topological structures, such as path distance, without fully exploiting the rich semantic and associative information within the drug network. These limitations significantly impact the accuracy. Additionally, the main challenge in developing combination drugs lies in the vast search space of potential drug combinations, which results in existing research mainly focusing on predicting pairwise synergistic drug combinations.

2.2 Heterogeneous network representation learning

Formally, a heterogeneous network[23] is defined as a directed graph $G=\{V,E,A,R\}$, with a node type mapping function $\varphi: V \rightarrow A$ and an edge type mapping function $\Psi: E \rightarrow R$, where V, E, A, and R represent the sets of nodes, edges, node types, and edge types, respectively. Each node $v \in V$ and each edge $e \in E$ belong to a specific type in A and R, i.e., $\varphi(v) \in A$ and $\Psi(e) \in E$. Heterogeneous networks have multiple node or edge types, i.e., $|A|+|R|>2$. Specifically, if $|A|=1$ and $|R|=1$, it represents a homogeneous network with nodes and edges of the same type.

Heterogeneous network representation learning (HNRL), also called heterogeneous network embedding (HNE), aims to learn a mapping function $\Phi: V \rightarrow \mathrm{R}^{|V| \times d}$. This mapping Φ maps each node $v \in V$ to a low-dimensional space, obtaining a low-dimensional vector representation. It learns the rich latent information between different nodes and edges. Currently, mainstream methods for heterogeneous network representation learning can be categorized into shallow models and deep models[23].

Early methods for heterogeneous network representation learning mainly focuses on shallow models. These models begin with randomly initialized nodes to get initial embeddings, which will be learned through suitable objective functions. Furthermore, shallow models can be divided into two categories: random walk-based methods and decomposition-based methods. Random walk-based methods are initially employed in homogeneous networks to generate node

sequences for capturing the local structure information of the network[24]. However, in heterogeneous networks, node sequences need to not only incorporate the complex network structural information but also capture the underlying semantic information. To address this, numerous semantic-aware random walk methods have been proposed. For example, metapath2vec[25] used metapath-guided random walks to generate heterogeneous node sequences with rich semantics and a heterogeneous skip-gram technique to preserve the similarity between the target node and its neighboring nodes. HERec[26] utilizes multiple metapaths in the heterogeneous network to guide random walk and learn node representations under multiple semantic information. Metagraph2vec[27] designed a metagraph-guided random walk, preserving more complex similarities between nodes. Decomposition-based methods aim to decompose the heterogeneous network into multiple subgraphs while maintaining the proximity of nodes within each subgraph to achieve node embedding. A classical approach is to process nodes in different metric spaces. PME[28] is a representative work that treated each edge type as a relationship and mapped nodes into different metric spaces using relation-specific matrix. This allows nodes connected by different types of edges to maintain proximity, capturing the heterogeneity of the network. HEBE[29] performs subgraph sampling on the heterogeneous network using hyperedges while preserving the similarity between the center node and the subgraph to learn node embeddings. Each hyperedge encapsulates abundant information, making HEBE robust to sparse data.

Deep models utilize neural network models to learn node representations from either node attributes or complex network structures. Compared to shallow models, deep models require more space and time during training but have relatively lower memory costs during inference. Deep models can be classified into three categories: message passing-based, encoder decoder-based, and adversarial-based models. Message passing-based methods adopt the idea of nodes receiving and aggregating messages from their neighboring nodes to update their own representations, which has been widely used in graph neural networks (GNNs). The core of message passing is designing efficient message aggregation functions to better capture the potential semantic and relational information in heterogeneous networks. For instance, HetGNN[30] consists of three parts: content aggregation, neighbor aggregation, and type aggregation. It employs three different message aggregation strategies to learn diverse node embeddings and then integrates them to better preserve the network structure and node heterogeneity. Wang et al. proposed heterogeneous attention network (HAN)[31], which introduces a hierarchical attention mechanism to learn the importance of different nodes and metapaths, capturing both complex local neighborhood structural information and underlying semantic information in heterogeneous networks. HGT[32] treates one type of node as a query to compute the importance of other types of nodes around it, capturing interactions between different types of nodes and assigning different weights for aggregation. GTN and its variant FastGTN[33] design an aggregation function that automatically generates metapaths during the message passing process, aiming to automatically discover useful metapaths in the process of learning node embeddings. HDHGR[34] investigates the effect of homophily properties on the performance of GNNs and concludes that heterogeneous GNNs perform better on datasets with strong homophily properties. SimpleHGN[35] extends the graph attention network (GAT)[36] by incorporating learnable edge type embeddings, residual connections, and L2 regularization on the output to capture heterogeneous information in the network. HINormer[37] then further introduces information about the local structure of heterogeneous graphs and encodes heterogeneous relations. RHGNN[38] decomposes the heterogeneous graph into multiple relationship subgraphs based on edge relationships. It learns vertex representations from each subgraph and connected them via cross-relation message passing mechanism. RHGNN not only learns node representations from various edge types but also captures rich semantic information under the relationships. Encoder decoder-based models typically utilize neural network models as encoders to learn embeddings from node attributes and design decoders to retain certain properties of the network. For example, HNE[39] learns embeddings from images and texts using convolutional neural networks (CNNs) and autoencoders, respectively, and then predicts whether there exists a link between images and texts using learned embeddings. Camel[40] uses GRU as the encoder to learn paper embeddings from abstracts and employed the skip-gram objective function to preserve the local structure of the graph. SHNE[41] combines

heterogeneous skip-gram and deep semantic encoding and performed deep optimization. The core of SHNE is to extract the heterogeneous structural closeness between nodes and the unstructured semantic relationships, considering them as functions of node content. Adversarial-based models leverage the adversarial training between generators and discriminators to learn high-quality node representations. In homogeneous networks, adversarial-based models only consider the heterogeneous network structure information. For example, GraphGAN[42] uses breadth-first search to generate virtual nodes. However, in heterogeneous networks, adversarial-based models design discriminators and generators that are aware of relationships to capture rich semantics on the heterogeneous network. HeGAN[43] is the first study to employ generative adversarial networks (GANs) on heterogeneous network embedding. It incorporates multiple relationships into both the generator and discriminator to consider the heterogeneity of graph. MVACM[44] uses GANs to generate complementary views by computing the similarity of nodes in different views.

In summary, shallow models are easily parallelizable but require a two-stage training, where embedding is independent of downstream tasks, and they have high memory costs. Deep models, which are now mainstream, offer end-to-end training and are more memory-efficient. Specifically, message passing-based models demonstrate proficiency in encoding both network structure and node attributes, integrating different semantic information. They are widely used recently. Compared to message passing-based models, encoder decoder-based models are weaker in information fusion due to the lack of message passing mechanisms. However, they introduce various objective functions through different decoders, providing greater flexibility. Adversarial-based methods tend to enhance the robustness of embeddings by utilizing negative samples, but the choice of negative samples greatly affects performance, resulting in higher variance. Overall, different methods for heterogeneous network representation learning have their own characteristics. Selecting an appropriate heterogeneous network representation learning approach is imperative for synergistic drug combination prediction.

3 Method

3.1 Problem definition

Given a drug combination $\text{comb} = \{d_1, d_2, \ldots, d_n\}$, consisting of n drugs, the objective is to predict whether all n drugs in the combination are synergistic. We aim to learn a prediction function $\widehat{y}_{\text{comb}} = \Gamma(d_1, d_2, \ldots, d_n|\theta)$, where $\widehat{y}_{\text{comb}}$ represents the probability that the drug combination comb is synergistic, and θ denotes the model parameters of the function Γ.

In particular, up to now, research on synergistic drug combinations has focused on combinations of two drugs. Given a combination of drugs d_i and d_j, along with the drug heterogeneity graph G and the ATC codes of drugs, the objective is to predict whether drugs d_i and d_j are synergistic, i.e., learning a prediction function $\widehat{y}_{d_i d_j} = \Gamma\left(d_i, d_j|\theta\right)$, where $\widehat{y}_{d_i d_j}$ represents the probability that drugs d_i and d_j are a synergistic combination, and θ denotes the model parameters of the function Γ.

3.2 Overall framework of SMC-HNCL

The proposed SMC-HNCL is a novel approach for predicting synergistic multi-drug combinations of western medicine. It utilizes heterogeneous network representation learning with contrastive learning to obtain network-based drug features without the need for supervised training using known synergistic drug combinations. SMC-HNCL leverages the Jaccard coefficient to calculate the drug similarity based on ATC codes, enabling the extraction of another drug feature. Moreover, we employ an attention mechanism to aggregate different drug features and learn the importance of individual drugs in the drug combinations, thereby obtaining a representation for drug combinations and enabling synergistic predictions. Figure 1 illustrates the overall framework of SMC-HNCL.

The overall framework of the SMC-HNCL method consists of three sub-modules:

(1) The drug feature extraction module based on heterogeneous network representation learning with contrastive learning: this module first constructs a heterogeneous network of drugs and then utilizes a heterogeneous network representation learning algorithm based on contrastive learning to get drug feature representations. This process includes drug

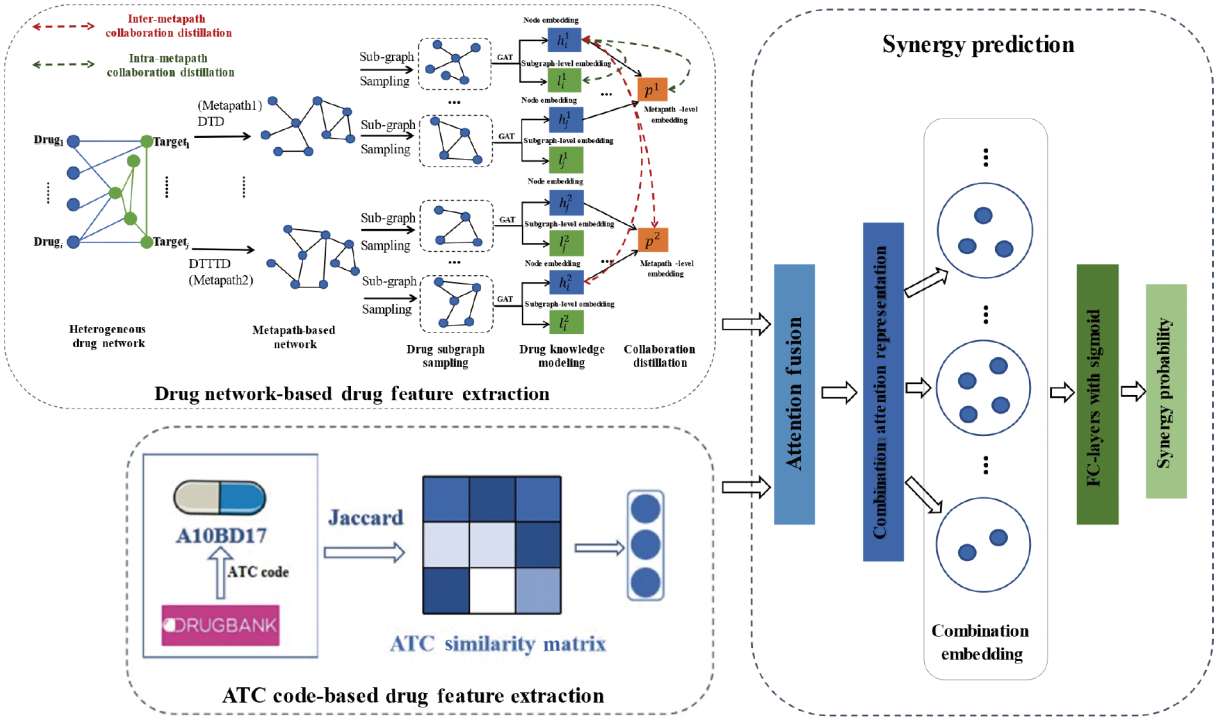

Fig. 1 Overall framework.

subgraph sampling based on metapaths, drug knowledge modeling and contrastive learning.

(2) The drug feature extraction module based on drug ATC codes: this module calculates drug similarity based on ATC codes using the Jaccard coefficient. It constructs a drug similarity matrix where each row represents the drug features used for prediction. We opt to utilize the ATC information for extracting drug feature, primarily due to its distinctive application in cataloging western pharmaceuticals. It is not only readily accessible and has been proven to be suitable for synergistic prediction[15−20].

(3) The synergistic prediction module based on attentional group representation: this module processes drug features from previous two modules, using an attention mechanism to combines them and obtain a singular drug feature. Finally, it employs an attention-based group representation method to obtain the representation of drug combinations for synergistic multi-drug combination prediction.

The detailed description of each module in the framework will be provided next.

3.3 Drug feature extraction module based on HNRL with contrastive learning

In recent years, there have been numerous studies in the field of synergistic drug combinations that have utilized the drug heterogeneous network. For instance, Wang et al.[10] reported that 62% of drug combinations are synergistic when the number of targets jointly affected by drug combinations is equal to or less than three in the target-target interaction network. Zou et al.[11] demonstrated that the topological structure of neighboring network communities can be effectively utilized to predict synergistic drug combinations. Yin et al.[12] found that the effectiveness of drug combinations largely depends on the network's topological structure. Cheng et al.[14] also revealed that for a drug combination to achieve a synergistic therapeutic effect beyond individual drug impacts, the combination's targets probable jointly affect the set of disease-related targets. Therefore, the constructed drug heterogeneous network in this study mainly includes two types of edges: drug-target and target-target. Relevant data for constructing the drug-target heterogeneous network are obtained from different databases. Figure 2 illustrates the drug-target heterogeneous network constructed in this study.

Based on the constructed drug-target heterogeneous network, inspired by heterogeneous network contrastive learning[45], this model utilizes a contrastive learning-based heterogeneous graph neural network to learn the embedding of drugs. Specifically, the GAT[36] is employed to learn therepresentations of target drug

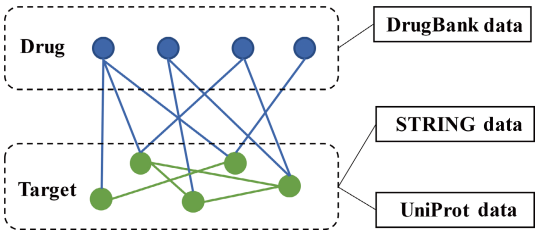

Fig. 2 Drug-target heterogeneous network.

nodes in subgraphs under different metapaths. Simultaneously, contrastive learning is used to explicitly model the correlations between different metapaths, as illustrated in Fig. 3. This process consists of three main parts: metapath-based drug subgraph sampling, drug knowledge modeling, and contrastive learning.

（1）Metapath-based drug subgraph sampling

The success of GNNs has demonstrated the effectiveness of aggregating information from neighboring nodes to learn node embeddings. However, applying GNNs directly for node representation learning on different metapaths is insufficient for heterogeneous networks due to the issues of sparsity and redundancy. The sparsity issue in GNNs occurs when valuable node relationships are sparse, limiting information aggregation. While the redundancy issue arises when the relationships between nodes are rich but redundant, causing GNNs to aggregate noise information.

To address these issues, we employ graph diffusion techniques[46] to smooth the neighbors of graphs corresponding to different metapaths. Then, we sample a fixed-size drug subgraph that contains sufficient structural information for metapath-based embedding learning. Specifically, given the drug heterogeneous network $G=\{V,E,A,R\}$ and the metapath set M, for each metapath $m\in M$, we first map the drug heterogeneous network to a metapath-based drug homogeneous network G^m (if drug nodes d_i and d_j are connected via metapath m, $G^m_{ij}=1$, otherwise $G^m_{ij}=0$). Then, we use generalized graph diffusion, such as personalized pagerank (PPR)[47], to measure the connectivity between nodes in G^m, as shown in Eq. (1):

$$S^m=\alpha\left(I_n-(1-\alpha)D_m^{-1/2}A^mD_m^{-1/2}\right)^{-1} \tag{1}$$

where $S^m\in\boldsymbol{R}^{N\times N}$ is the diffusion matrix, N is the number of target nodes, i.e., the number of drug nodes, $A^m\in\boldsymbol{R}^{N\times N}$ is the adjacency matrix of the metapath m-based homogeneous network G^m, D_m is a diagonal matrix with $D_m(i,i)=\sum_j A^m(i,j)$, and α is a fixed parameter set to 0.15. The diffusion matrix S^m has been shown to effectively recover meaningful neighbors from noisy graphs[46], which is crucial for overcoming the sparsity and redundancy issues in metapath-based homogeneous networks.

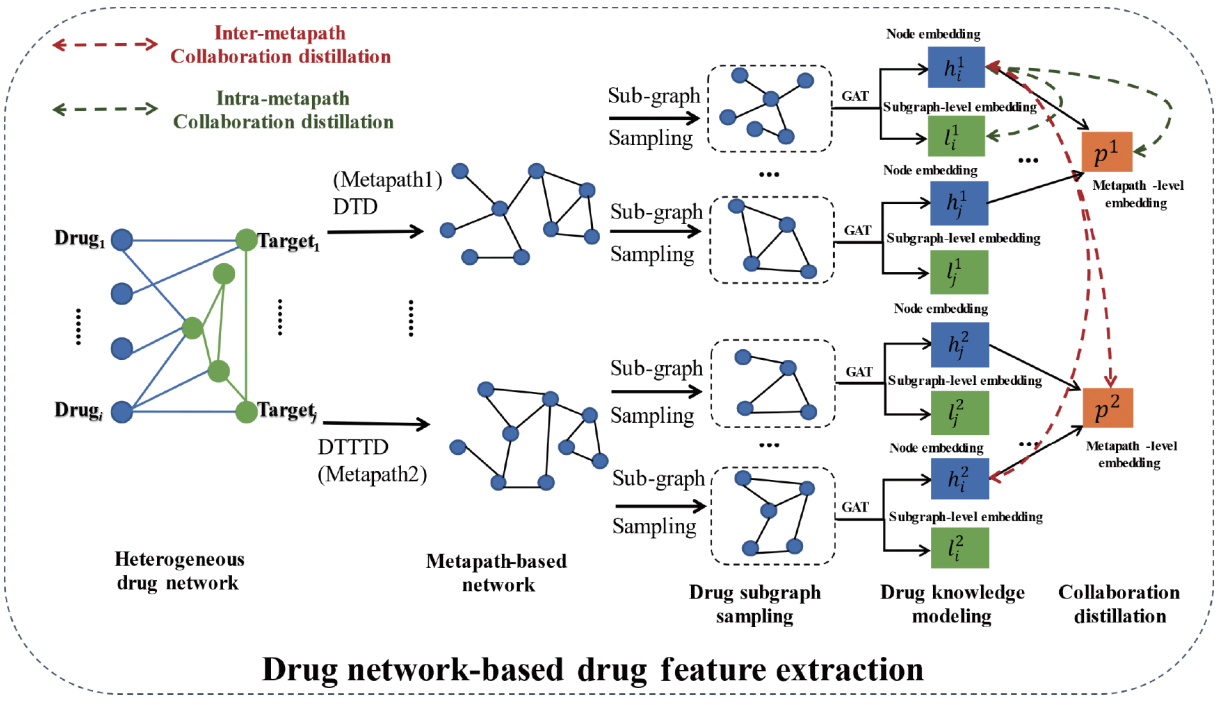

Fig. 3 Heterogeneous network representation learning based on contrastive learning.

Generally, nodes are more related to their local neighborhood, and the information provided by distant nodes for node embedding is limited. Therefore, to further enhance efficiency, this study applies a semantic context subgraph sampling strategy. Based on the diffusion matrix S^m, the top K important neighbors (including the node itself) are sampled to construct the meta-path-based drug subgraph G_i^m, which can be expressed as

$$G_i^m = \text{top}_{\text{similarity}(S^m(i,:),K)} \tag{2}$$

where K is the size of the drug subgraph, and the $\text{top}_{\text{similarity}}$ function returns the neighboring nodes sorted by similarity.

（2）Drug knowledge modeling

Given the drug heterogeneous network $G = \{V, E, A, R\}$ and metapath m, the heterogeneous network can be mapped to a semantically homogeneous network G^m based on the metapath m. This study explicitly extracts drug heterogeneous network embeddings by modeling both local and global knowledge within each metapath's semantic space.

For a drug subgraph G_i^m centered around drug node d_i based on the metapath m, node embedding h_i^m of the drug d_i is learned by aggregating the information from neighboring nodes in the drug subgraph G_i^m using GAT[36]:

$$\alpha_{i,j} = \frac{\exp\left(\text{LeaklyReLU}\left(a^{\mathrm{T}}\left[Wh_i \| Wh_j\right]\right)\right)}{\sum_{k \in N_{(i)}^m} \exp\left(\text{LeaklyReLU}\left(a^{\mathrm{T}}\left[Wh_i \| Wh_k\right]\right)\right)}$$

$$h_i^m = \sigma\left(\sum_{j \in N_{(i)}^m} \alpha_{i,j} W h_j\right) \tag{3}$$

$$h_i^m = \|_{x=1}^{X} h_i^m$$

where W is the learnable weight matrix for linear transformation of node features, σ is the activation function, a is a learnable parameter, $N_{(i)}^m$ represents metapath m-based neighbors set of node i, $\|$ denotes the vector concatenation operation, and X represents the number of attention heads.

In each drug subgraph G_i^m, we learn the node embeddings of the central drug node d_i and the neighboring nodes $d_{j,j\neq i}$. Only the embedding h_i^m of the central drug node learned from drug subgraph G_i^m will be used for optimization and downstream tasks.

For each drug node in G, they exhibit different local connectivity patterns with their neighboring nodes, reflecting the surrounding region nodes. However, in the semantic space of different metapaths, these connectivity patterns may vary, which is insufficient for metapath-based embedding learning. Therefore, we model the local knowledge in each semantic space as the drug subgraph-level embedding. Given a sampled drug subgraph G_i^m with the center drug node d_i and the metapath m, corresponding drug subgraph-level embedding l_i^m is obtained by summing and averaging drug node embeddings learned within each drug subgraph using a local readout function $\mathcal{R}_l : \boldsymbol{R}^{(K)\times d} \rightarrow \boldsymbol{R}^d$:

$$l_i^m = \mathcal{R}_l\left(G_i^m\right) = \sigma\left(\frac{1}{K}\sum_{j=1}^{K} h_j^m\right) \tag{4}$$

where h_j^m represents the embedding of each drug node in the drug subgraph G_i^m, including the central drug node and other neighboring drug nodes, σ is the Sigmoid function, and K denotes the number of drug nodes in the drug subgraph G_i^m.

Although the subgraph-level local knowledge provides the local connectivity patterns for each node, it fails to capture global connectivity patterns of the heterogeneous network shared across all locations. To address this, we define the global knowledge as metapath-level embedding p_m, which is obtained by averaging the central drug node embeddings in each drug subgraph G_i^m using a global readout function $\mathcal{R}_g : \boldsymbol{R}^{N\times d} \rightarrow \boldsymbol{R}^d$:

$$p^m = \mathcal{R}_g\left(H^m\right) = \sigma\left(\frac{1}{N}\sum_{i=1}^{N} h_i^m\right) \tag{5}$$

where h_i^m represents the embedding of the central drug node in each drug subgraph G_i^m, σ is a non-linear activation function, and N denotes the number of drug nodes in the homogenous graph G^m based on the metapath m.

（3）Contrastive learning

Based on the aforementioned modeling of drug knowledge, it is expected to extract both the local and global knowledge of drugs from all metapaths $m \in M$, which is a typical contrastive learning problem. This paper utilizes mutual information (MI) as the measure for contrast, which has been widely used to capture nonlinear statistical dependencies between variables. For convenience, this paper adopts the mutual information estimation based on Jensen–Shannon divergence (JSD)[48], which is formulated as follows:

$$\text{MI}(X, Y) = \boldsymbol{E}_{\boldsymbol{P}}\left[-\text{sp}\left(-f(x,y)\right)\right] - \boldsymbol{E}_{\boldsymbol{P}\times\tilde{\boldsymbol{P}}}\left[\text{sp}\left(f(x,\tilde{y})\right)\right] \tag{6}$$

where $\mathrm{sp}(x)=\log(1+\mathrm{e}^x)$ is the softplus function, $\tilde{\boldsymbol{P}}$ represents the positive sample distribution, and $\tilde{\boldsymbol{P}}$ represents the negative sample distribution.

Intra-metapath contrastive learning enhances topological embedding optimization by extracting both local and global knowledge within each metapath. For a specified metapath m, it measures the mutual information between the drug node embedding h_i^m, the drug subgraph-level embedding l_i^m and metapath-level embedding p^m. The objective function is defined as

$$\mathcal{L}_{\text{intra}}=-\sum_{m\in M}\left(\sum_{i}^{|N|}\left(\mathrm{MI}\left(h_i^m,l_i^m\right)+\mathrm{MI}\left(h_i^m,p^m\right)\right)\right) \tag{7}$$

where N represents the number of drug nodes in the homogeneous graph G^m based on the metapath m. For local knowledge, subgraph-level embeddings l_i^m derived from all metapaths $m\in M$ are treated as positive samples, while other randomly sampled nodes serve as negative samples.

Inter-metapath contrastive learning aims to extract both local and global knowledge simultaneously across different metapaths. A approach similar to intra-metapath contrastive learning is employed, but the embeddings come from different metapaths. The objective of inter-metapath contrastive learning is defined as

$$\mathcal{L}_{\text{inter}}=-\sum_{i}^{|N|}\left(\sum_{m\in M}\sum_{n\in M,n\neq m}\mathrm{MI}\left(h_i^m,l_i^n\right)+\mathrm{MI}\left(h_i^m,p^n\right)\right) \tag{8}$$

For global knowledge, the metapath-level embedding p^n is used as a positive sample. While treat the metapath-level embeddings as negative samples after disrupting node attributes.

The drug feature extraction module does not require known data on drug combinations during training. Finally, this module combines both intra-metapath and inter-metapath contrastive learning objectives, expressed as

$$L=\mathcal{L}_{\text{intra}}+\mathcal{L}_{\text{inter}} \tag{9}$$

We utilize the Adam optimizer to minimize the objective and learn the optimal drug embeddings in a self-supervised way. As the objective is optimized, we obtain the drug embeddings h_i^m for each metapath m. The final embedding f_i^{network} for drug i is learned by simply summing up all the metapath-based drug node embeddings:

$$f_i^{\text{network}}=\sum_{m\in M}h_i^m \tag{10}$$

3.4 Drug feature extraction module based on drug ATC codes

The ATC classification system codes, abbreviated as ATC codes, is the official classification system for pharmaceutical drugs by the World Health Organization[49]. Previous studies have shown that in the ATC system, drugs in a drug combination often belong to the same anatomical group and therapeutic group[50]. Therefore, the ATC codes can be used to differentiate effective synergistic drug combinations from a large search space. Furthermore, drug ATC codes reflect the similarity of drugs, and this concept has been widely used in the discovery of new drug targets and drug combinations. The ATC classification system categorizes drugs into different levels based on their target organs and chemical, therapeutic, and pharmacological properties, which involves five different levels of drugs, it is represented by a letter, two digits, a letter, a letter, and two digits, respectively. These levels represent the main anatomical group, main therapeutic group, therapeutic/pharmacological subgroup, chemical/therapeutic/pharmacological subgroup, and chemical subgroup. Similar to previous studies[12−17], the anatomical, therapeutic and chemical similarity of drug pairs is typically calculated using the Jaccard similarity. Since the similarity of most drug ATC codes at the fourth and fifth levels in the dataset is zero, this study calculates the anatomical therapeutic chemical similarity of drug pairs based on the first three levels of the drug ATC codes from the DrugBank database.

Assuming $\mathrm{ATC}_k(d_i)$ and $\mathrm{ATC}_k\left(d_j\right)$ represent the ATC code of drug d_i and d_j at the k-th level, the drug ATC similarity $S_k(d_i,d_j)$ at the k-th level between drug d_i and d_j is calculated using the Jaccard coefficient as shown in Eq. (11):

$$S_k\left(d_i,d_j\right)=\frac{\left|\mathrm{ATC}_k(d_i)\cap\mathrm{ATC}_k\left(d_j\right)\right|}{\left|\mathrm{ATC}_k(d_i)\cup\mathrm{ATC}_k\left(d_j\right)\right|},k=1,2,3 \tag{11}$$

where $|\cdot|$ denotes the cardinality of the ATC code. Then, the anatomical therapeutic chemical similarity S^{ATC} between drugs d_i and d_j is computed by taking the average of the ATC similarities at the first three levels. The calculation is given by Eq. (12):

$$S^{\mathrm{ATC}}\left(d_i,d_j\right)=\frac{\sum_{k=1}^{3}S_k(d_i,d_j)}{3} \tag{12}$$

It is worth noting that multiple different ATC codes may represent the same drug. For example, D03BA02, M09AB02, and D03BA52 all represent the drug collagenase clostridium histolyticum, while D04AA01, R06AC06, and R01AC06 all represent the drug thonzylamine. Therefore, the ATC similarity between collagenase clostridium histolyticum and thonzylamine at the first level is

$$S_1(\text{collagenase clostridium histolyticum, thonzylamine}) = \frac{|D|}{|D,M,R|} = \frac{1}{3} \tag{13}$$

By calculating the similarity of drug pairs based on ATC codes, we can obtain the drug similarity matrix M^{ATC}, the i-th row vector of M^{ATC} is used as the ATC-based feature representation f_i^{ATC} for drug d_i:

$$f_i^{\text{ATC}} = M_{i,:}^{\text{ATC}} \tag{14}$$

3.5 Synergistic prediction module based on attentional group representation

After obtaining the drug features f_i^{network} based on heterogeneous network and the drug features f_i^{ATC} based on drug ATC codes, this module aims to measure the importance of different drug features and predict synergistic multi-drug combinations. Firstly, an attention mechanism is employed to fuse the two drug features. Then, inspired by GBERT[51], a group representation method based on multi-head self-attention is used to learn the representation of each drug combination for synergy prediction.

Specifically, for drug i, after the two drug feature extraction modules aforementioned, we obtain the two drug features f_i^{network} and f_i^{ATC}. The attention mechanism is used to fuse the two drug features. First, we project f_i^{network} and f_i^{ATC} into K and Q, and then calculate the attention weight α between them. Finally we fuse them to obtain the fused drug feature f_i^{fusion} based on α. The process is described by Eqs. (15)–(18):

$$K = K - \text{Linear}\left(f_i^{\text{network}}\right) \tag{15}$$

$$Q = Q - \text{Linear}\left(f_i^{\text{ATC}}\right) \tag{16}$$

$$\alpha = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{17}$$

$$f_i^{\text{fusion}} = \alpha \cdot \text{Linear}_{\text{ATC}}\left(f_i^{\text{ATC}}\right) + (1-\alpha) \cdot f_i^{\text{network}} \tag{18}$$

where $K-\text{Linear}$ and $Q-\text{Linear}$ are used to transform the two different drug features into the same space, with d_k representing the dimension of K. $\text{Linear}_{\text{ATC}}$ is used to transform f_i^{ATC} into the same dimension as f_i^{network}.

To learn a representation for a multi-drug combination containing n drugs $\{d_1, d_2, \ldots, d_n\}$, we utilizes a multi-head self-attention in the Transformer[52] encoder to measure the importance of different drugs and obtain the representation of the drug combination f^{comb}:

$$f^{\text{comb}} = \text{Self}-\text{Attention}_{\text{Multi-Head}}\left(\Big\|_{i\in[1,n]} f_i^{\text{fusion}}\right) \tag{19}$$

where $\|$ denotes vector concatenation, $\text{Self}-\text{Attention}_{\text{Multi-Head}}$ represents the multi-head self-attention in the Transformer encoder. The representation of the drug combination f^{comb} is then passed through a fully connected layer and a sigmoid activation function to output the probability of synergy prediction. This process is described by the following Eq. (20):

$$\widehat{y}_{\text{comb}} = \sigma\left(\text{FC}-\text{Layers}\left(f^{\text{comb}}\right)\right) \tag{20}$$

where $\widehat{y}_{\text{comb}}$ represents the probability of synergy for the drug combination, σ is the sigmoid activation function. The final output of SMC-HNCL is the probability of a drug combination being synergistic. The drug combination representation is passed through a fully connected layer and a sigmoid activation function to output the synergy prediction probability. Therefore, the binary cross-entropy loss, BCELoss, is chosen as the loss function. The model parameters are updated by minimizing the loss using the Adam optimizer. The calculation of the cross-entropy loss is described by the following equation:

$$\text{BCELoss}(\hat{y}_{\text{comb}}, y_{\text{comb}}) = -\frac{1}{N}\sum_{(\text{comb})\in D}\left(y_{\text{comb}}\log(\hat{y}_{\text{comb}}) + (1-y_{\text{comb}})\log(1-\hat{y}_{\text{comb}})\right) \tag{21}$$

where N represents the number of samples, D represents the training set. And a dropout layer are added to prevent overfitting.

4 Experimental

In order to verify the effectiveness of the SMC-HNCL method, experiments are conducted on two synergistic drug combination datasets, DDInter[4] and CDCDB[5], including synergistic drug pairs on both datasets and synergistic multi-drug combinations on CDCDB dataset. The experiments includes the following three

parts:

(1) Comparative experiments mainly compare the performance of SMC-HNCL with other state-of-the-art baseline methods[15–19] on DDInter and CDCDB synergistic drug pair datasets, as well as the performance of SMC-HNCL on synergistic multi-drug combinations in the CDCDB dataset.

(2) The ablation experiment mainly verifies the effectiveness of two different feature extraction modules in the SMC-HNCL and the drug combination representation learning module based on attention mechanism.

(3) The parameter sensitivity experiment mainly analyzes the impact of important hyperparameters in the SMC-HNCL, including the embedding dimension of drug features and the sampling size of drug subgraphs.

4.1 Dataset

We use the synergistic drug combination data from the latest two databases to evaluate the proposed SMC-HNCL method and baseline methods. The two datasets are DDInter[4] and CDCDB[5], respectively. And, the DDInter database was launched in October 2021, in which we obtain 1551 kinds of drugs, 86 997 pairs of synergistic drugs and 10 878 pairs of antagonistic drugs (i.e., known non-synergistic drug pairs), without synergistic multi-drug combination data; The CDCDB database was launched in June 2022. We obtain 3226 kinds of drugs, 10 561 pairs of synergetic drugs and 6457 combinations of synergetic drugs. In addition, in the construction of drug-target heterogeneous network, we use the drug target data from DrugBank[53] database and screens human targets through Uniprot[54] database, and uses STRING[55] database to screen target-target interaction (PPI) data. Tables 1 and 2 summarize the details of the two datasets.

4.2 Data preprocessing and parameter setting

At present, the number of negative samples is small or missing in the existing synergistic drug combinations, so it is necessary to preprocess the two datasets, similar to the data processing method of the baseline methods[15–19], in which the synergistic drug pairs are positive samples. For balanced data samples, negative samples (i.e., non-synergistic drug pairs) make the proportion of positive and negative samples 1:1 through the random negative sampling strategy. For example, there are 10 561 positive samples (synergistic drug pairs) in the CDCDB dataset, and 10 561 negative samples (non-synergistic drug pairs) constitute balanced data samples through random negative sampling. For unbalanced data samples, negative samples are all drug pairs except positive samples. For example, there are 10 561 positive samples (synergistic drug pairs) in the CDCDB dataset while 5 191 364 drug pairs except positive samples are negative samples. The data is shown in Table 3 and Table 4.

In addition, SMC-HNCL can also be used for the prediction of synergistic multi-drug combinations. Therefore, this paper also adopts the random negative sampling strategy for the negative samples of synergistic multi-drug combinations on the CDCDB dataset, and ensures that each number of drug combinations has the same number, so as to keep a 1:1

Table 1 Data statistics of DDInter.

Data type	Instance	Data source
Drug	1 551	DDInter
Target	1 855	DrugBank, Uniprot
Synergistic drug pair	86 997	DDInter
Antagonists drug pair	10 878	DDInter
Drug-target	14 188	DrugBank
Target-target (PPI)	38 313	STRING

Table 2 Data statistics of CDCDB.

Data type	Instance	Data source
Drug	3 226	CDCDB
Target	13 960	DrugBank, Uniprot
Synergistic drug pair	10 561	CDCDB
Synergistic multi-drug combination	6 457	CDCDB
Drug-target	16 771	DrugBank
Target-target (PPI)	231 016	STRING

Table 3 Positive and negative samples of drug pairs on DDInter.

Sampling strategy	Positive sample	Negative sample
DDInter balanced data samples	86 997	86 997
DDInter unbalanced data samples	86 997	1 115 028

Table 4 Positive and negative samples of drug pairs on CDCDB.

Sampling strategy	Positive sample	Negative sample
CDCDB balanced data samples	10 561	10 561
CDCDB unbalanced data samples	10 561	5 191 364

ratio of positive and negative samples to balance the multi-drug combination data. The statistics is shown in Table 5.

For the hyperparameters of SMC-HNCL, the important parameters are mainly concentrated in the drug feature extraction module of heterogeneous network representation learning based on contrastive learning. Specifically, the learning rate is set to 0.001, dropout is set to 0.2, and other important parameters are grid searched. The size of the drug subgraph, that is, the number of nodes sampled by the subgraph, is searched in {5, 10, 15, 20, 25, 30}, and the embedding dimension of the drug is adjusted in {8, 16, 32, 64, 128, 256}. In addition, it should be noted that the metapath used by SMC-HNCL is {DTD, DTTD, DTTTD, DTDTD}, where D represents the drug node, and T represents the target node.

4.3 Baseline methods and evaluation metric

In this paper, the SMC-HNCL method is compared with other synergistic drug combination prediction methods, and its performance is evaluated. According to the summary of synergistic drug combination prediction method in 2.1, it will be compared with the following state-of-the-art baseline methods:

EPSDC[15]: an integrated prediction framework for synergistic drug pairs, which can accurately and effectively predict synergistic drug pairs by integrating drug information from multiple sources, including biological, chemical, pharmacological, and network knowledge of drugs.

DSGCR[16]: a synergistic drug pair prediction method based on graph collaborative regularization, which can effectively combine the prior knowledge of drug target network, pharmacological information, and drug combination.

ISDCSMP[17]: a synergistic drug pair prediction method based on symmetric metapath, which can accurately predict the combination of synergistic drugs in the drug-target heterogeneous network integrating multi-source information.

DSML[18]: a synergistic drug pair prediction method based on multitask learning, which integrated drug targets, PPI, ATC codes, etc., and significantly improved the prediction performance of synergistic drug pairs through the reconstruction of drug target interaction and the combination of multisource knowledge.

NEWMIN[19]: a synergistic drug pair prediction method based on graph representation learning, which constructed a multi-layer drug similarity network according to six kinds of drug information, then used the random walk algorithm to sample the multi-layer network information and evaluated the importance of each network in the multi-layer network to fuse the representation of drugs, and finally used the random forest to predict the synergistic drug pair.

The essence of the SMC-HNCL method is a binary classification model to predict whether multiple drugs in the drug combination are synergistic. Therefore, AUC, AUPR, and F1 of the traditional binary classification task are used as evaluation metrics. AUC is calculated based on receiver operating characteristic (ROC) curve. The area of the curve is the AUC, and the AUC range is 0 to 1. The higher the value, the better the classification effect. ROC curve can help us select the best model performance under different classification thresholds, and compare different classification models. However, it should be noted that under unbalanced data samples, the results given by ROC curve may be inaccurate. Similarly, the precision recall (PR) curve is also a curve used to evaluate the performance of the binary classification model. PR curve can reflect the highest precision rate under different recall rates, so it is more appropriate to use PR curve in the classification of unbalanced data samples. Area under the PR curve (AUPR) is a metric to evaluate the performance of the binary classification model. Similar to AUC, the value range of AUPR is 0 to 1. The larger the value, the better the performance of the model. F1 is the harmonic average of recall rate and precision.

Table 5 Balanced sample of multi-drug combination on CDCDB.

Sampling strategy	Positive sample	Negative sample
Multi-drug combination balance data sample	6457	6457

4.4 Comparative experiment

To validate the effectiveness of SMC-HNCL, various experiments were conducted on drug pair data from two real drug combination datasets, and comparisons were made with relevant baseline methods. In the experiments, SMC-HNCL employed the same cross-validation approach as the baseline methods[15−19]. To ensure a fair comparison among all baseline methods, the results of 10 rounds of 5-fold cross-validation

experiments were averaged to obtain the final results. This included both balanced 5-fold cross-validation and imbalanced 5-fold cross-validation. Balanced 5-fold cross-validation was performed on balanced data samples with a 1:1 ratio of positive and negative samples, while imbalanced 5-fold cross-validation was performed on imbalanced data samples (Table 3 and Table 4). Additionally, a special 5-fold cross-validation was conducted on balanced data samples, where 1/5 of the data was used for training and 4/5 for testing, to validate the effectiveness of the model with limited training data.

Tables 6 and 7 present the results of SMC-HNCL on balanced data samples from the DDInter and CDCDB datasets using 5-fold cross-validation, respectively. Tables 8 and 9 show the results of SMC-HNCL on imbalanced data samples from the DDInter and CDCDB datasets using 5-fold cross-validation, respectively. Tables 10 and 11 display the results of SMC-HNCL on balanced data samples from the DDInter and CDCDB datasets using special 5-fold cross-validation, respectively (i.e., 1/5 for training and 4/5 for testing). For all baseline methods[15−19], the same data and partitions were used as SMC-HNCL, and the best results were selected for comparison. Bold values represent the best results among all methods, and underlined values indicate the second best results among all methods.

From Tables 6 to 11, it can be seen that the SMC-HNCL method is superior to other synergistic drug combination prediction methods in terms of AUC, AUPR and F1 for all different data samples of the two datasets, indicating that our synergistic multi-drug combination prediction method based on heterogeneous network contrastive learning is still efficient without using the known synergistic drug combination as the monitoring information. It shows that the contrastive learning of local knowledge and

Table 6 Comparative experiment results 1 on DDInter balance data samples.

Method	AUC	AUPR	F1
EPSDC	0.9064	0.8933	0.8019
ISDCSMP	0.9614	0.9524	0.8984
DSGCR	0.9729	0.9701	0.9120
DSML	0.9738	0.9711	0.9140
NEWMIN	0.9545	0.9336	0.8781
SMC-HNCL	**0.9843**	**0.9802**	**0.9587**

Table 7 Comparative experiment results 1 on CDCDB balance data samples

Method	AUC	AUPR	F1
EPSDC	0.8323	0.8174	0.7482
ISDCSMP	0.8711	0.8401	0.7544
DSGCR	0.8745	0.8588	0.7623
DSML	0.8881	0.8571	0.7641
NEWMIN	0.8632	0.8285	0.7509
SMC-HNCL	**0.9015**	**0.9021**	**0.8311**

Table 8 Comparative experiment results on DDInter unbalanced data samples.

Method	AUC	AUPR	F1
EPSDC	0.9113	0.1930	0.3207
ISDCSMP	0.9639	0.3253	0.3966
DSGCR	0.9753	0.5188	0.5526
DSML	0.9749	0.5202	0.5921
NEWMIN	0.9583	0.2305	0.4137
SMC-HNCL	**0.9882**	**0.5376**	**0.6225**

Table 9 Comparative experiment results on CDCDB unbalanced data samples.

Method	AUC	AUPR	F1
EPSDC	0.8067	0.0013	0.0068
ISDCSMP	0.8507	0.0065	0.0283
DSGCR	0.8491	0.0069	0.0322
DSML	0.8549	0.0089	0.0406
NEWMIN	0.8273	0.0051	0.0186
SMC-HNCL	**0.9114**	**0.0373**	**0.1058**

Table 10 Comparative experiment results 2 on DDInter balance data samples

Method	AUC	AUPR	F1
EPSDC	0.8843	0.8661	0.7804
ISDCSMP	0.9537	0.9499	0.8856
DSGCR	0.9540	0.9507	0.8800
DSML	0.9552	0.9532	0.8878
NEWMIN	0.9132	0.8893	0.8372
SMC-HNCL	**0.9674**	**0.9645**	**0.9178**

Table 11 Comparative experiment results 2 on CDCDB balance data samples.

Method	AUC	AUPR	F1
EPSDC	0.7691	0.7572	0.7215
ISDCSMP	0.7957	0.8031	0.7402
DSGCR	0.8039	0.7921	0.7318
DSML	0.8022	0.8036	0.7466
NEWMIN	0.7918	0.7863	0.7286
SMC-HNCL	**0.8511**	**0.8625**	**0.7709**

global knowledge of heterogeneous networks is conducive to obtaining higher quality drug representation. In addition, it shows that the group representation based on attention can effectively represent the drug combination, which verifies the effectiveness of SMC-HNCL method in synergistic drug pair prediction. Particularly, the NEWMIN method, which is also based on graph representation learning, uses a variety of drug information to build a multi-layer similar network, but noise may be introduced and get unsatisfactory result due to missing drug information.

Furthermore, we conduct comparative experiments of different multi-drug combinations on the CDCDB dataset, and the results are shown in Table 12. In the table, SMC-HNCL--all represents the experiment of drug pair and multi-drug combination, SMC-HNCL--multi represents the experiment of using only multi-drug combination, SMC-HNCL--2 represents the experiment of using only drug pair, SMC-HNCL--3 represents the experiment of using only three drug combination, SMC-HNCL--4 represents the experiment of using only four drug combination. Specifically, in the synergistic multi-drug combination data of CDCDB, there are 3671 synergistic three drug combinations and 1436 synergistic four drug combinations. The results show that SMC-HNCL not only performs well in the synergistic drug pair prediction task, but also can be well used in the synergistic multi-drug combination prediction, which verifies the effectiveness of SMC-HNCL in the synergistic multi-drug combination prediction. It can be seen that the prediction result of SMC-HNCL in the combination of two drugs is the best. With the increase of the number of drugs in the drug combination, the prediction result gradually decreases. The possible reason is that the combination of multiple drugs has a huge search space or the amount of relevant training data decreases, which leads to the decline of the prediction result.

Table 12 Comparative experiment results of different drug combinations on CDCDB.

Method	AUC	AUPR	F1
SMC-HNCL^{-2}	**0.9015**	**0.9021**	**0.8311**
SMC-HNCL^{-3}	0.8517	0.8364	0.7730
SMC-HNCL^{-4}	0.7941	0.7802	0.7385
SMC-HNCL^{-all}	0.8832	0.8904	0.8086
SMC-HNCL^{-Multi}	0.7974	0.7895	0.7533

4.5 Ablation experiment

To validate the effectiveness of the two different feature extraction modules and the attention-based group representation method in SMC-HNCL, ablation experiments were conducted on balanced data samples from the two datasets. Specifically, SMC-HNCL-ATC excludes the drug feature extraction module based on heterogeneous network representation learning with contrastive learning. SMC-HNCL-NETWORK removes the drug feature extraction module based on drug ATC codes. SMC-HNCL-Attention removes the drug combination representation method based on attention mechanism and only utilizes the sum of drug representations as the drug combination representation. Tables 13 and 14 present the experiment results of 5-fold cross-validation on balanced data samples from the DDInter and CDCDB datasets, respectively. Bold values indicate the best results, and underlined values indicate the second-best results.

From Tables 13 and 14, it can be observed that the performance of the models decreases to some extent when using only a single drug feature, which indicates that different drug features are effective for the prediction. The contribution of the drug heterogeneous network-based features is more significant among them, which suggests that contrastive learning-based heterogeneous network representation learning can utilize both local and global knowledge in the form of contrastive learning and capture rich information in the drug-target heterogeneous network to learn high-quality drug representations effectively. Furthermore, removing the drug combination representation method

Table 13 Ablation experiment results on DDInter balance data samples.

Method	AUC	AUPR	F1
SMC-HNCL$^{-Attention}$	0.9835	0.9786	0.9573
SMC-HNCL^{-ATC}	0.8991	0.8829	0.8392
SMC-HNCL$^{-NETWORK}$	0.9814	0.9749	0.9546
SMC-HNCL	**0.9843**	**0.9802**	**0.9587**

Table 14 Ablation experiment results on CDCDB balance data samples.

Method	AUC	AUPR	F1
SMC-HNCL$^{-Attention}$	0.8964	0.8997	0.8273
SMC-HNCL^{-ATC}	0.8016	0.8143	0.7485
SMC-HNCL$^{-Network}$	0.8930	0.8972	0.8163
SMC-HNCL	**0.9015**	**0.9021**	**0.8311**

based on attention mechanism results in a decrease in experimental performance, indicating that the attention-based drug combination representation method can capture the importance of different drugs in the drug combination, thereby improving prediction accuracy and enabling the prediction of synergistic multi-drug combinations.

4.6 Parameter sensitivity experiment

This section primarily investigates the impact of important parameters in SMC-HNCL, which are mainly concentrated in the drug feature extraction module based on heterogeneous network representation learning with contrastive learning.

These parameters include the embedding dimension and the size of drug subgraph sampling. The specific experimental setup involves conducting experiments on balanced data samples from two datasets. The range of drug embedding dimensions for contrastive learning-based heterogeneous network representation learning is {8, 16, 32, 64, 128, 256}, and the range of drug subgraph sampling sizes is {5, 10, 15, 20, 25, 30}. Furthermore, in order to better present the trends of different parameters on SMC-HNCL, only the AUC is compared in this case. Figures 4 and 5 show the experiment results, where the AUC values on the left vertical axis correspond to the results on the balanced data samples from DDInter, and the AUC values on the right vertical axis correspond to the results on the balanced data samples from CDCDB.

In SMC-HNCL, a key hyperparameter is the size of drug subgraph sampling, denoted as K. In this study, the size of drug subgraph sampling was changed from

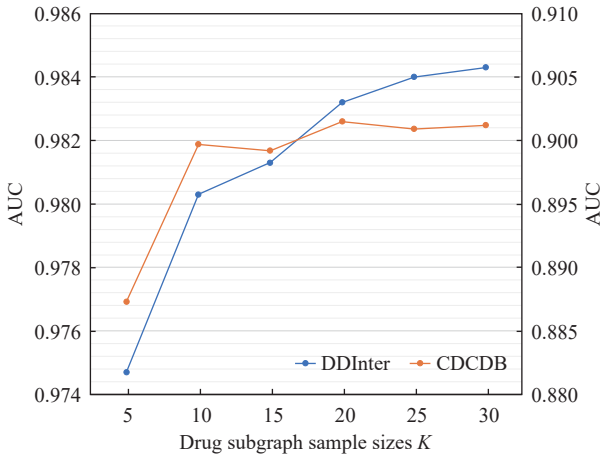

Fig. 4 Impact of different drug subgraph sizes on the performance of SMC-HNCL.

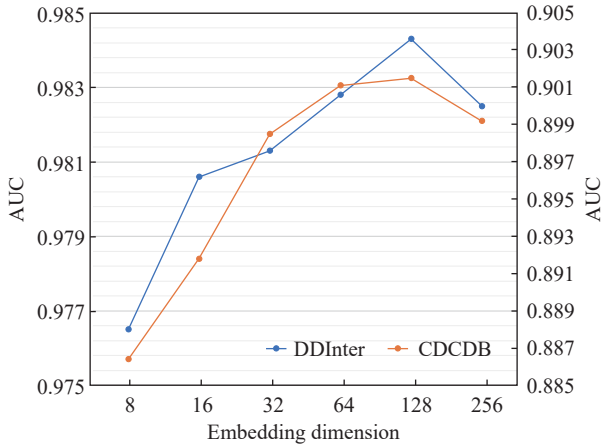

Fig. 5 Impact of different embedding dimensions on the performance of SMC-HNC.

5 to 30, and the experiment results are shown in Fig. 4. It can be observed that the model achieves the highest AUC when K is 20 on the CDCDB dataset, indicating the best performance. As K continues to increase, the AUC value does not further improve but slightly decreases and remains relatively stable. On the DDInter dataset, the AUC increases as K increases, reaching the highest value when K is 30, indicating the best model performance. These results suggest that appropriately increasing the size of drug subgraph sampling is beneficial for learning local drug knowledge and embedding drugs more effectively. However, excessively large subgraph sampling sizes may introduce redundant neighbor information, which could impact model performance and slow down the model's speed.

Figure 5 illustrates the impact of different embedding dimensions on the performance of SMC-HNCL on the balanced data samples of the DDInter and CDCDB datasets. The results indicate that as the drug embedding dimension increases up to 128, the AUC consistently improve, indicating enhanced model performance. When the drug embedding dimension reaches 128, the AUC reaches its maximum value. However, as the drug embedding dimension continues to increase, the AUC start to decline. This suggests that appropriately increasing the dimension of drug features helps improve prediction accuracy since higher-dimensional embedding representations can more effectively encode drug information. However, excessively large drug embedding dimensions may lead to overfitting issues and adversely affect model performance.

5 Conclusion

Combination therapy is an important treatment strategy. Despite the promising results achieved by previous methods for synergistic drug combination prediction, these methods still have limitations and shortcomings. They have not fully exploited the rich information in the drug network and can only predict pairwise synergistic combinations. In this paper, we propose the SMC-HNCL method, which introduces contrastive learning-based heterogeneous network representation learning to capture richer information in the drug-target heterogeneous network and obtain high-quality drug features. Additionally, we incorporate the unique ATC codes information specific to western medicine, and calculate the ATC code-based features using the Jaccard coefficient. Subsequently, we employ an attention mechanism to fuse the two drug features and use an attention-based group representation method to learn the importance of different drugs in drug combinations, thereby learning the representation of drug combinations and achieving innovative prediction of synergistic multi-drug combinations. We evaluate the proposed method on two synergistic drug combination datasets, demonstrating its effectiveness and superiority over other baseline methods. The inclusion of molecular structure information and the utilization of multi-feature models will be a future work for further research.

Acknowledgments

This work was supported by National Key R&D Program of China (No. 2019YFC1711000) and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

[1] L. Wu, Y. Wen, D. Leng, Q. Zhang, C. Dai, Z. Wang, Z. Liu, B. Yan, Y. Zhang, J. Wang, et al., Machine learning methods, databases and tools for drug combination prediction, *Brief Bioinform*, vol. 23, no. 1, p. bbab355, 2022.

[2] T. Roemer and C. Boone, Systems-level antimicrobial drug and drug synergy discovery, *Nat. Chem. Biol.*, vol. 9, no. 4, pp. 222–231, 2013.

[3] F. Cheng, I. A. Kovács, and A.-L. Barabási, Network-based prediction of drug combinations, *Nat. Commun.*, vol. 10, no. 1, p. 1197, 2019.

[4] G. Xiong, Z. Yang, J. Yi, N. Wang, L. Wang, H. Zhu, C. Wu, A. Lu, X. Chen, S. Liu, et al., DDInter: An online drug-drug interaction database towards improving clinical decision-making and patient safety, *Nucleic Acids Res.*, vol. 50, no. D1, pp. D1200–D1207, 2022.

[5] G. Shtar, L. Azulay, O. Nizri, L. Rokach, and B. Shapira, CDCDB: A large and continuously updated drug combination database, *Sci. Data*, vol. 9, p. 263, 2022.

[6] X. M. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. van Noort, and P. Bork, Prediction of drug combinations by integrating molecular and pharmacological data, *PLoS Comput. Biol.*, vol. 7, no. 12, p. e1002323, 2011.

[7] L. Chen, B. Q. Li, M. Y. Zheng, J. Zhang, K. Y. Feng, and Y. D. Cai, Prediction of effective drug combinations by chemical interaction, protein interaction and target enrichment of KEGG pathways, *Biomed Res. Int.*, vol. 2013, p. 723780, 2013.

[8] Y. Sun, Y. Xiong, Q. Xu, and D. Wei, A hadoop-based method to predict potential effective drug combination, *Biomed Res. Int.*, vol. 2014, p. 196858, 2014.

[9] Q. Xu, Y. Xiong, H. Dai, K. M. Kumari, Q. Xu, H. Y. Ou, and D. Q. Wei, PDC-SGB: Prediction of effective drug combinations using a stochastic gradient boosting algorithm, *J. Theor. Biol.*, vol. 417, pp. 1–7, 2017.

[10] Y. Y. Wang, K. J. Xu, J. Song, and X. M. Zhao, Exploring drug combinations in genetic interaction network, *BMC Bioinform.*, vol. 13, no. 7, p. S7, 2012.

[11] J. Zou, P. Ji, Y. L. Zhao, L. L. Li, Y. Q. Wei, Y. Z. Chen, and S. Y. Yang, Neighbor communities in drug combination networks characterize synergistic effect, *Mol. Biosyst.*, vol. 8, no. 12, pp. 3185–3196, 2012.

[12] N. Yin, W. Ma, J. Pei, Q. Ouyang, C. Tang, and L. Lai, Synergistic and antagonistic drug combinations depend on network topology, *PLoS One*, vol. 9, no. 4, p. e93960, 2014.

[13] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, NLLSS: Predicting synergistic drug combinations based on semi-supervised learning, *PLoS Comput. Biol.*, vol. 12, no. 7, p. e1004975, 2016.

[14] F. Cheng, I. A. Kovács, and A. L. Barabási, Network-based prediction of drug combinations, *Nat. Commun.*, vol. 10, no. 1, p. 1197, 2019.

[15] P. Ding, R. Yin, J. Luo, and C. K. Kwoh, Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological, and network knowledge, *IEEE J. Biomed. Health Inform.*, vol. 23, no. 3, pp. 1336–1345, 2019.

[16] P. Ding, C. Shen, Z. Lai, C. Liang, G. Li, and J. Luo, Incorporating multisource knowledge to predict drug synergy based on graph co-regularization, *J. Chem. Inf. Model.*, vol. 2020, pp. 37–64, 2020.

[17] P. Ding, C. Liang, W. Ouyang, G. Li, Q. Xiao, and J. Luo, Inferring synergistic drug combinations based on symmetric meta-path in a novel heterogeneous network, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 4, pp. 1562–1571, 2021.

[18] X. Chen, L. Luo, C. Shen, P. Ding, and J. Luo, An in silico method for predicting drug synergy based on multitask learning, *Interdiscip. Sci.*, vol. 13, no. 2, pp. 299–311, 2021.

[19] L. Yu, M. Xia, and Q. An, A network embedding framework based on integrating multiplex network for

drug combination prediction, *Brief. Bioinform.*, vol. 23, no. 1, p. bbab364, 2022.

[20] P. Zhang and S. Tu, MGAE-DC: Predicting the synergistic effects of drug combinations through multi-channel graph autoencoders, *PLoS Comput. Biol.*, vol. 19, no. 3, p. e1010951, 2023.

[21] T. C. Chou, Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies, *Pharmacol. Rev.*, vol. 58, no. 3, pp. 621–681, 2006.

[22] J. Chen, L. Wu, K. Liu, Y. Xu, S. He, and X. Bo, EDST: A decision stump based ensemble algorithm for synergistic drug combination prediction, *BMC Bioinform.*, vol. 24, no. 1, p. 325, 2023.

[23] X. Wang, D. Bo, C. Shi, S. Fan, Y. Ye, and P. S. Yu, A survey on heterogeneous graph embedding: Methods, techniques, applications and sources, *IEEE Trans. Big Data*, vol. 9, no. 2, pp. 415–436, 2023.

[24] A. Grover and L. Jure, Node2vec: Scalable feature learning for networks, in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (*KDD '16*), New York, NY, USA, 2016, pp. 855–864.

[25] Y. Dong, N. V. Chawla, and S. Ananthram, Metapath2vec: Scalable representation learning for heterogeneous networks, in *Proc. 23rd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, New York, NY, USA, 2017, pp. 135–144.

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

[27] C. Zhang, J. Yin, and X. Zhu, MetaGraph2Vec: Complex semantic path augmented heterogeneous network embedding, in *Advances in Knowledge Discovery and Data Mining*, D. Phung, V. Tseng, G. Webb, B. Ho, M. Ganji, and L. Rashidi, eds. Cham, Switzerland: Springer, vol. 10938, pp. 196–208, 2018.

[28] H. X. Chen, H. Z. Yin, W. Q. Wang, H. Wang, Q. V. H. Nguyen, and X. Li, PME: Projected metric embedding on heterogeneous networks for link prediction, in *Proc. 24th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, New York, NY, USA, 2018, pp. 1177–1186.

[29] H. Gui, J. Liu, F. Tao, M. Jiang, B. Norick, and J. Han, Large-scale embedding learning in heterogeneous event data, in *Proc. IEEE 16th Int. Conf. Data Mining* (*ICDM*), Barcelona, Spain, 2016, pp. 907–912.

[30] C. X. Zhang, D. J. Song, C. Huang, A. Swami, N. V. Chawla, Heterogeneous graph neural network, in *Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, New York, NY, USA, pp. 793–803, 2019.

[31] X. Wang, H. Y. Ji, C. Shi, B. Wang, Y. F. Ye, P. Cui, and P. S. Yu, Heterogeneous graph attention network, in *Proc. World Wide Web Conf. 2019*, New York, NY, USA, pp. 2022–2032, 2019.

[32] Z. N. Hu, Y. X. Dong, K. S. Wang, and Y. Z. Sun, Heterogeneous graph transformer, in *Proc. Web Conference 2020*, New York, NY, USA, pp. 2704–2710, 2020.

[33] S. Yun, M. Jeong, S. Yoo, S. Lee, S. S. Yi, R. Kim, J. Kang, and H. J. Kim, Graph transformer networks: Learning meta-path graphs to improve GNNs, *Neural Netw.*, vol. 153, pp. 104–119, 2022.

[34] J. Y. Guo, L. Du, W. D. Bi, Q. Fu, X. J. Ma, X. Chen, S. Han, D. M. Zhang, Y. Zhang, Homophily-oriented heterogeneous graph rewiring, in *Proc. ACM Web Conference 2023* (*WWW '23*), New York, NY, USA, 2023, pp. 511–522.

[35] Q. Lv, M. Ding, Q. Liu, Y. Chen, W. Feng, S. He, C. Zhou, J. Jiang, Y. Dong, and J. Tang, Are we really making much progress? Revisiting, benchmarking and refining heterogeneous graph neural networks, in *Proc. 27th ACM SIGKDD Conf. Knowledge Discovery & Data Mining*, Virtual Event, Singapore, 2021, pp. 1150–1160.

[36] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio', and Y. Bengio, Graph attention networks, arXiv:1710.10903, 2017.

[37] Q. H. Mao, Z. M. Liu, C. H. Liu, and J. L. Sun, HINormer: Representation learning on heterogeneous information networks with graph transformer, in *Proc. ACM Web Conf. 2023* (*WWW '23*), New York, NY, USA, 2023, pp. 599–610.

[38] L. Yu, L. Sun, B. Du, C. Liu, W. Lv, and H. Xiong, Heterogeneous graph representation learning with relation awareness, *IEEE Trans. Knowl. Data Eng.*, p. 1, 2022.

[39] S. Y. Chang, W. Han, J. L. Tang, G. J. Qi, C. R. C. Aggarwal, and T. S. Huang, Heterogeneous network embedding via deep architectures, in *Proc. 21th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (*KDD '15*), New York, NY, USA, 2018, pp. 119–128.

[40] C. X. Zhang, C. Huang, L. Yu, X. L. Zhang, and N. V. Chawla, Camel: Content-aware and meta-path augmented metric learning for author identification, in *Proc. 2018 World Wide Web Conf.* (*WWW'18*), Lyon, France, 2018, pp. 709–718.

[41] C. X. Zhang, A. Swami, and N. V. Chawla, SHNE: Representation learning for semantic-associated heterogeneous networks, https://dl.acm.org/doi/pdf/10.1145/3289600.3291001, 2019.

[42] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, GraphGAN: Graph representation learning with generative adversarial nets, arXiv:1711.08267, 2018.

[43] T. Lan, C. Wu, C. Yu, and X. Wang, Adversarial network embedding on heterogeneous information networks, *J. Phys.*: *Conf. Ser.*, vol. 1693, no. 1, p. 012018, 2020.

[44] K. Zhao, T. Bai, B. Wu, B. Wang, Y. J. Zhang, Y. Y. Yang, and J. Y. Nie, Deep adversarial completion for sparse heterogeneous information network embedding, in *Proc. Web Conf. 2020* (*WWW '20*), New York, NY, USA, 2020, pp. 508–518.

[45] C. Wang, S. Zhou, K. Yu, D. Chen, B. Li, Y. Feng, and C. Chen, Collaborative knowledge distillation for heterogeneous information network embedding, in *Proc. ACM Web Conf. 2022*, Virtual Event, Lyon, France, 2022, pp. 1631–1639.

[46] J. Gasteiger, S. Weißenberger, and S. Günnemann, Diffusion improves graph learning, *NeurIPS*, vol. 32, pp. 13366–13378, 2019.
[47] D. Fogaras, B. Rácz, K. Csalogány, and T. Sarlós, Towards scaling fully personalized PageRank: Algorithms, lower bounds, and experiments, *Internet Math.*, vol. 2, no. 3, pp. 333–358, 2005.
[48] R. Devon Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, A. Trischler, and Y. Bengio, Learning deep representations by mutual information estimation and maximization, arXiv preprint arXiv:1808.06670, 2018.
[49] Gerhard Nahler, Anatomical therapeutic chemical classification system (ATC), https://www.who.int/tools/atc-ddd-toolkit/atc-classification, 2024.
[50] Z. Sheng, Y. Sun, Z. Yin, K. Tang, and Z. Cao, Advances in computational approaches in identifying synergistic drug combinations, *Brief. Bioinform.*, vol. 2018,19, no. 6, p. 1172, 1182.
[51] S. Zhang, et at, GBERT: Pre-training User representations for Ephemeral Group Recommendation, in Proc. *the 31st ACM International Conference on Information & Knowledge Management* (*CIKM '22*), New York, USA, 2022, pp. 2631–2639.
[52] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is All you Need, arXiv preprint arXiv:1706.03762, 2017.
[53] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., DrugBank 5.0: A major update to the DrugBank database for 2018, *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1074–D1082, 2018.
[54] T. U. Consortium UniProt: A worldwide hub of protein knowledge, *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506–D515, 2019.
[55] D. Szklarczyk, A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, et al., The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets, *Nucleic Acids Res.*, vol. 49, no. D1, pp. D605–D612, 2021.

Xin Xi received the bachelor degree in software engineering from Anhui University, China in 2020, and the master degree in computer technology from Southeast University, China in 2023. His research interests include deep learning, network representation learning, and drug synergy prediction.

Shan Lu received the MS degree in technology for computer applications and PhD degree in management science & engineering from Southeast University, China in 1999 and 2006, respectively. He is currently a chief engineer at Nanjing FiberHome Tiandi Co., Ltd. His research interests include DBMS, neural network, and complex system management.

Jinhui Yuan received the bachelor degree in software engineering from Fuzhou University, China in 2021, and he is currently a master student at Southeast University, China. His research interests include deep learning, network representation learning, and drug synergy prediction.

Jieyue He received the BSc and MSc degrees from Nanjing University, China, and the PhD degree from Southeast University, China. She is currently a professor of School of Computer Science and Engineering, Southeast University, China. Her current research interests include big data analysis technology, machine learning, data mining, and bioinformatics.