


KGANSynergy: knowledge graph attention network for drug synergy prediction

Ge Zhang, Zhijie Gao, Chaokun Yan, Jianlin Wang, Wenjuan Liang, Junwei Luo and Huimin Luo 

Corresponding author. Huimin Luo, School of Computer and Information Engineering, Henan University, Jinming Street, 475004 Kaifeng, China.

E-mail: luohuimin@henu.edu.cn

Abstract

Combination therapy is widely used to treat complex diseases, particularly in patients who respond poorly to monotherapy. For example, compared with the use of a single drug, drug combinations can reduce drug resistance and improve the efficacy of cancer treatment. Thus, it is vital for researchers and society to help develop effective combination therapies through clinical trials. However, high-throughput synergistic drug combination screening remains challenging and expensive in the large combinational space, where an array of compounds are used. To solve this problem, various computational approaches have been proposed to effectively identify drug combinations by utilizing drug-related biomedical information. In this study, considering the implications of various types of neighbor information of drug entities, we propose a novel end-to-end Knowledge Graph Attention Network to predict drug synergy (KGANSynergy), which utilizes neighbor information of known drugs/cell lines effectively. KGANSynergy uses knowledge graph (KG) hierarchical propagation to find multi-source neighbor nodes for drugs and cell lines. The knowledge graph attention network is designed to distinguish the importance of neighbors in a KG through a multi-attention mechanism and then aggregate the entity's neighbor node information to enrich the entity. Finally, the learned drug and cell line embeddings can be utilized to predict the synergy of drug combinations. Experiments demonstrated that our method outperformed several other competing methods, indicating that our method is effective in identifying drug combinations.

Keywords: drug synergy prediction, knowledge graph, multi-head attention network, deep learning

INTRODUCTION

In complex diseases such as cancer, HIV and cardiovascular disease [1–3], many cellular mechanisms commonly change in different tissues and organ systems [4]. Therefore, conventional 'single compound, single target' approaches to treating these diseases are typically ineffective [5]. Many studies have suggested that a combination therapy of multiple drugs can significantly improve therapeutic efficacy and reduce drug toxicity [6]. At the same time, drug combination therapy can reduce the risk of drug resistance, which can improve the success rate of drug repositioning [7]. For instance, amiloride and hydrochlorothiazide are used to treat hypertension [8, 9]. The food and drug administration (FDA) approved the combination of the BRAF inhibitor dabrafenib and the MEK inhibitor trametinib for treating patients with metastatic melanoma [10]. Similarly, drug combination therapy has become the primary treatment strategy for many complex diseases [11]. Traditional drug combination discovery is mainly based on clinical trials, which have been proved to be time-consuming, expensive and potentially harmful to patients [12, 13]. Recently, with the

emergence and application of high-throughput screening techniques (HTS) [14], ample verified drug combination data have been exploited and accumulated. However, today's rapidly growing biomedical data create difficulties in testing the whole combination space via the above techniques [5, 15, 16]. In addition, the development of drug combinations is also a costly process for pharmaceutical companies [5]. Thus, there is an urgent need for efficient and economical methods to screen drug combinations.

A variety of computational methods have been developed to predict drug combinations and thereby address the above problems, including methods in the following categories [17]: (i) systems biology methods, (ii) mathematical methods, (iii) kinetic models and (iv) machine learning methods. Systems biology methods primarily apply biological knowledge to analyze biological networks [18]. Mathematical methods require the application of mathematical models and statistical tests, as well as rely on the reliability of model assumptions. Kinetic models use kinetic equations to simulate the dynamic changes of nodes in biological networks [19]. The above three methods are suitable

Ge Zhang is an associate professor in the School of Computer and Information Engineering, Henan University, Kaifeng, China. His research interests include artificial intelligence and computational biology.

Zhijie Gao is a graduate student in the School of Computer and Information Engineering, Henan University, Kaifeng, China. Her research interests include bioinformatics and computational drug repositioning.

Chaokun Yan is a professor in the School of Computer and Information Engineering, Henan University, Kaifeng, China. His research interests include artificial intelligence and computational biology.

Jianlin Wang is an associate professor in the School of Computer and Information Engineering, Henan University, Kaifeng, China. His research interests include artificial intelligence and computational biology.

Wenjuan Liang is a lecturer in the School of Computer and Information Engineering, Henan University, Kaifeng, China. Her research interests include artificial intelligence and data mining.

Junwei Luo is an associate professor in the College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China. His research interests include genome assembly, scaffolding, gap filling and structural variant detection.

Huimin Luo is an associate professor in the School of Computer and Information Engineering, Henan University, Kaifeng, China. Her research interests include bioinformatics and computational drug repositioning.

Received: September 11, 2022. Revised: March 10, 2023. Accepted: April 3, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

for small datasets and depend on prior knowledge. Machine learning methods such as Random Forest (RF), Support Vector Machine (SVM) and Naive Bayesian methods can explore vast combinatorial spaces and expedite the identification of drug combinations [20–22].

In recent years, deep learning methods have equipped scientists with powerful machine learning tools for abstracting important data features from large-scale datasets and have also been demonstrated to be effective in biomedical fields [23, 24]. For instance, Preuer et al. [25] proposed a deep learning model called DeepSynergy to predict cancer drug combinations with synergistic effects. This model utilizes both compound and genomic information as inputs and predicts drug combinations using the Hill curve structure with tanh normalization. Kuru et al. [26] used the fully connected layer to predict whether a drug pair has synergistic effects by combining medicinal chemical structure data and cell line gene expression profiling features. Wang et al. [27] suggested a DeepDDS model to predict the synergistic effects of drug combinations for cancer. The model uses a graph convolutional network and an attention mechanism to determine drug embeddings. Drug chemical structures and gene expression patterns are embedded as features in a multilayer feedforward neural network to predict synergistic drug combinations. However, most of the aforementioned methods focus on the molecular structure of drugs and cell lines rather than the interaction information.

Some studies have reported that drug combinations interfering synergistically with protein networks can more effectively block oncogene activity [18, 28]. This topic has raised concern among bioinformaticians in the treatment of complicated diseases where the topological link between drugs and diseases in protein association networks is utilized. Xu et al. [29] integrated six features, including molecular structure, ATC Code similarity and protein–protein associations; they used SVM, NB and SGB to predict drug combinations. Liu et al. [30] applied the random walk with the restart algorithm on a protein–protein association network to extract new drug–target profiles as drug features for drug combination prediction. However, the mentioned network-based methods only consider entities related to drugs or diseases directly; they usually neglect local connections and lack interpretability. Yang et al. [31] used information propagation and attention mechanisms to explore the higher order topological information of drug proteins and cell line proteins in protein–protein interaction networks to predict drug combinations. We found that any one of protein association information, graph attention network (GAT) or neighbor node information could improve the efficacy of drug synergy prediction. However, the propagation-based approach uses a single type of neighbor information rather than multiple types of neighbor information, which may negatively impact the capture performance of a particular drug or cell line.

Knowledge graphs (KGs) are a type of heterogeneous graph in which nodes represent entities, and edges represent relations between entities. Rich semantic linkages between entities in the graph can benefit the system in discovering potential relations between entities. In recent years, KGs have been successfully applied in natural language processing, recommender systems and many other fields [32, 33]. Specifically, propagation-based recommendation methods—which are propagated over the whole KG to find the recommended auxiliary information—are the dominant approaches for KG-based recommender systems, such as CKAN and KGAT [34, 35]. Inspired by these, bioinformaticians have begun to use KGs to explicitly combine heterogeneous data

for drug repositioning downstream tasks. This approach not only enables the extraction of fine-grained multimodal knowledge elements from omics data but also captures information between entities and their neighbors more completely. Drug repositioning and adverse drug reaction prediction based on biomedical KGs increase the opportunities for accelerated drug discovery [36, 37]. For instance, a knowledge graph neural network (KGNN) [38] combines the graph neural network (GNN) and KG to determine the interactions between drug pairs. This model aggregates information from neighborhood entities to learn representations for drug entities to capture higher order structural and semantic relationships. Mohamed et al. [39] introduced the KG embedding model TriModel to learn vector representations for all drugs and proteins, inferring new drug–target interactions based on the scores computed by the model. It is evident that using KGs to learn entities such as drugs, proteins and their topological neighbors can obtain a more accurate embedding representation, but previously developed KG-based methods may perform weakly in terms of attention stability and node aggregation capability [35, 40]. Furthermore, KGs have not been widely used for drug combination prediction.

In this study, we propose a novel deep learning model based on a KG and multi-head attention mechanism to predict drug synergy. Specifically, our model consists of three main modules. We first apply the hierarchical propagation of the KG to find drug/cell line neighbors. Then, we develop a KG attention network to obtain the corresponding neighborhood representations of drugs or cell lines. Finally, the predictive score for potential synergies is obtained via a well-defined score function that utilizes the embedded representation of drugs and the cell line. The advantage of our model lies in extending the multi-attention mechanism to prevent overfitting and stabilize the learning results. KGANSynergy also considers different types of neighbor node information.

The contributions of this work are summarized as follows:

- (i) We propose a novel KG attention network framework that can obtain high-order information from multi-source neighbors in the KG efficiently and explicitly.
- (ii) We design a neural-network-based multi-head attention method to completely utilize neighborhood information and improve the interpretability of the model. The module highlights the different levels of importance to each entity neighborhood node, which stabilizes the attention learning process.
- (iii) To enhance node embedding aggregation, we apply different types of aggregators in terms of neighboring nodes and hierarchical aggregation, highlighting different higher order connections and similarity.
- (iv) We conducted experiments on two benchmark datasets, and the results show that KGANSynergy outperformed state-of-the-art methods and baseline methods in the prediction of drug combination synergy.

MATERIALS AND METHODS

Problem formulation

We describe the problem of drug combination prediction based on KG attention.

Definition 1. (Drug–drug–cell line) Typically, drug i , drug j and cell line k are included in the drug synergy prediction. The drug combination for a particular cell line is represented as $Y \in (0, 1)^{|N_d| \times |N_d| \times |N_k|}$, where $|N_d|$ denotes the number of drugs, and $|N_k|$

denotes the number of cell lines. When $y_{i,j,k} = 1$ ($i, j \in N_d, k \in N_k$, and $i \neq j$), the combination of the drug pair has a synergistic effect on the cell line; otherwise, the combination has an antagonistic effect on the cell line.

Definition 2. (KG) In addition to drug synergy data, we also incorporate auxiliary information for the drug and cell-line-related entities (e.g. proteins) via a KG. In this work, we construct a KG to illustrate the associations between drugs and cell lines: $\mathcal{G} = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$, where \mathcal{E} and \mathcal{R} represent the set of entities and relations, respectively. The KG consists of entity-relation-entity triples (h, r, t) . Here, $h \in \mathcal{E}, r \in \mathcal{R}$ and $t \in \mathcal{E}$ denote the head, relation and tail of the knowledge triple, respectively. For example, the triplet (Miconazole, DPI, P35228) indicates that Miconazole interacts with protein P35228. In the graph, entities are represented as nodes, and relations are represented as edges from the head entity node to the tail entity node.

Dataset description

Drug synergy score dataset

In this paper, we focus on the combinations of two drugs. To investigate the performance of our model on the small samples and large samples, we conducted experiments on two drug combination datasets:

(1) **DrugCombDB.** DrugCombDB [41] is the first comprehensive database with the largest number of drug combinations to date, which integrates drug combinations from high-throughput screening (HTS) analyses, external databases and PubMed literature. The raw data contains 69,436 drug combinations, including 764 drugs and 76 cancer cell lines.

(2) **Oncology-Screen.** Oncology-Screen [42] was collected by O'Neil *et al.* through large-scale tumor screening and includes 4176 drug combinations involving 21 drugs and 29 cancer cell lines in the raw data. Furthermore, Preuer *et al.* [25] integrated the dataset by calculating the Loewe additivity value.

Each sample comprised of two drugs and one cell line, as well as the relevant synergy score. In general, the degree of synergy shown by the data was quantified based on the synergy deviation simulated by the theoretical model, such as Loewe Additivity [43], bliss independence [44], highest single agent (HSA) [45] and zero interaction potency (ZIP) [46]. The synergy scores discussed in DrugCombDB is the ZIP value, and Oncology-Screen is the Loewe Additivity Value computed by Combenefit [47]. As suggested by Liu *et al.* [41], we used the quartile as the threshold to exclude low-confidence drug combination samples, and the drug combinations with synergy scores distributed in the top quartile and bottom quartile are classified as synergistic and antagonistic effects. In this experiment, synergy was considered a positive sample, while antagonism was considered a negative sample. Table 1 displays the basic statistical results of these two datasets after processing. The DrugCombDB dataset contains 17404 synergistic drug pairs and 16 624 antagonistic drug pairs, which are used to evaluate the performance of the model on a large sample dataset; the Oncology-Screen dataset contains 1044 synergistic drug pairs and 916 antagonistic drug pairs, which are used to evaluate the performance of the model on a small sample dataset.

Knowledge graph-related data

Protein associations are a significant source of molecular information. Their associations (with one another or certain small molecules) have a role in metabolism, signaling, immunity and gene regulatory networks [48]. Aberrant associations produce aberrant cellular behaviors and diseases. Therefore, they should be major targets for molecular-based research of biological

Table 1. Details about the two drug combination datasets used after processing

Datasets	Drugs	Cell lines	Positive	Negative
DrugCombDB	475	76	17 404	16 624
Oncology-Screen	21	29	1044	916

Table 2. Statistics of the KG data

	DrugCombDB	Oncology-Screen
Entities	16 832	16 026
Relations	4	4
KG triples	250 256	226 475

disease states. In this study, we mainly used protein-related data to construct the KG, and the data were mainly from [31]. Table 2 shows the detailed data information of the KG.

Drug-protein association. Drug-protein associations of FDA-approved or clinical research drugs were collected from six commonly used databases, including DrugBank, BindingDB, ChEMBL and PharmGKB [18]. The datasets contain 15 051 drug-protein associations, including 4428 drugs and 2256 human proteins.

Cell line-protein association. Cell line-protein associations were obtained from Cancer Cell Line Encyclopedia (CCLE) gene expression and drug response data [49]. This is a large-scale analytical dataset of nearly 1000 cell lines from different tissues. The data were collected and processed by Rouillard *et al.* [50], which converted the cell line-gene expression data matrix in CCLE to a cell line-protein association matrix. Specially, the average and standard deviation are computed across all cell lines, and the genes with z-scores larger than a specific threshold suggested in [50] is considered to be associated with cell lines. The cell line-protein associations include 18 022 protein-coding genes, 1035 cancer cell lines and 749 551 associations.

Protein-protein association. Protein-protein associations were taken from the human protein-protein interactome [18], which is collected from 15 commonly used databases. It excludes evolutionary analyses, gene expression data and metabolic association-related associations. The dataset consists of 217 160 protein-protein interactions connecting 15 790 independent proteins. Based on Genecards [51], each protein is mapped to its coding gene.

Cell line-tissue association. Tissue consists of cells that divide and differentiate. As there may be connections between the cell lines of the tissues, cell line-tissue information can lead to the identification of important features between cell lines so that more accurate cell line characterization can be learned. For example, DrugCombDB contains lung, liver, large intestine, breast, prostate and eight other types of tissues. Oncology-Screen contains six tissues, including breast, lung and ovary. When drug i and drug j interact with a cell line in a particular tissue, the tissue may have a therapeutic effect.

Knowledge graph construction

To construct a KG, we integrated four categories of data, such as drug-protein association, cell line-protein association, protein-protein association and cell line-tissue data from the aforementioned datasets in Section 2.2.2 Knowledge graph-related data. The KG contains entities and relations from the above four categories. As shown in Figure 1, we assume that the drug

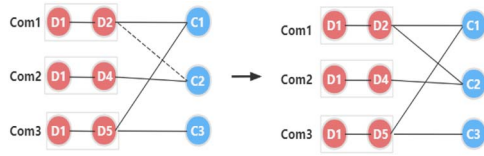


Figure 1. Example of drug combination prediction. Com1, Com2,... represents a set of drug combinations; D1, D2,... represents the drugs; and C1, C2,... represents the drug combination's cell lines.

combination Com1 and the cell line C2 are potentially predictive associations. The KG reveals information about drug–protein, cell line–protein and cell line–tissue associations as well as protein–protein associations. After a message is passed for target drugs and cell lines through the KG attention network, their embeddings include neighbor information from other relations, such as proteins or tissues. Thus, we can infer that some indirectly related entities potentially contribute to the therapeutic process. The KG-based method can integrate multiple types of entity information to make accurate predictions regarding the synergistic effects of drug combinations.

The KG of the drug i –drug j –cell line is represented as $\mathcal{G} = (\varepsilon, R)$, where ε denotes a variety of entities (drugs, cell lines, tissues and proteins), and R denotes the association between various entities (such as drug–protein association, cell line–protein association, protein–protein association and cell line–tissue). The form of the KG can be represented as multiple entity–relation–entity triples $T = (h, r, t)$, where $h, t \in \varepsilon$ represent the head entity and the tail entity, respectively, and $r \in R$ represents the relation between the two entities.

The framework design

In this paper, we design KGANSynergy as a framework for predicting drug synergy. The framework of KGANSynergy is briefly shown in Figure 2. The model consists of the following steps

- (1) KG hierarchical propagation. First, the entities and relations in KG are mapped to vectors. Then, the KG recursively explores the set of neighbor nodes directly and indirectly associated with the drug/cell line.
- (2) KG attention layer. We propose a neural-network-based knowledge attention mechanism to iteratively propagate multiple layers of information to update entity representations. This layer uses information propagated by the KG to encode the drug and cell line features among drug–protein, cell line–protein and other associations, as well as the neighborhood structure between them (i.e. entities and relations). Finally, each layer entity (h) and its neighbor node embedding set are aggregated to obtain the embedding representation set of the l -layer. The multi-head attention mechanism is used to obtain the neighbor weight information and stabilize the learning process.
- (3) Prediction layer. This layer utilizes drug and cell line representations obtained after a series of polymerization updates to calculate the prediction scores.

Knowledge graph hierarchy propagation

Adjacent entities in the KG are always strongly correlated. KGANSynergy's knowledge graph propagation layer is used to obtain multihop neighbor node sets for drugs/cell lines. Figure 3 illustrates the delivery process of Miconazole at different depths in the KG. To facilitate understanding, the protein coding gene is

mapped to the corresponding protein. The triplet (Miconazole, DPI, P35228) suggests a drug–protein association between Miconazole and protein P35228; the triplet (Q12809, CPI, K562) indicates a cell line–protein association between Q12809 and cell line K562. Take miconazole as an example. The drug Miconazole collects information of protein P35228 and protein Q12809 from the first layer, which are the first-order direct neighbors connecting Miconazole. In the second layer, Miconazole obtains information from its second-order neighbors: P00352, TC32 and K562. The first-order neighbor entity stores a representation of the prior layer's second-order neighbor entity, which captures information from second-order neighbors. Analogously, entities can obtain information from their l -th order neighbors at the l -th layer of knowledge propagation. KGANSynergy conducts hierarchical propagation of drug and cell line entity information in the KG. Therefore, we can obtain the expanded entity set of L layers, which can effectively enrich the entity latent vector representation.

First, we find the set of entities associated with a drug/cell line in the KG. Here, the drug/cell line is abbreviated as o . The set of adjacent entities of o can be expressed recursively as

$$\varepsilon_o^l = \{t \mid (h, r, t) \in \mathcal{G} \text{ and } h \in \varepsilon_o^{l-1}\}, l = 1, 2, \dots, L \quad (1)$$

where l denotes the depth of KG propagation, (h, r, t) denotes a set of triples, t denotes the tail entity and \mathcal{G} denotes the KG.

To reduce the computational burden, we choose a fixed-size set of neighbors for each entity [33]. Subsequently, the entity adjacency matrix \mathcal{A}_e and the relational adjacency matrix \mathcal{A}_r ($\mathcal{A}_e, \mathcal{A}_r \in \mathbb{R}^{u \times n}$) are constructed based on the entity set, respectively (u denotes the number of entities, and n denotes the number of fixed size neighbor entities). Particularly, each row in the entity adjacency matrix \mathcal{A}_e represents n neighbor entities of an entity. The relational adjacency matrix contains the association information between each entity and its neighbors, and the relations include drug–protein, protein–protein, cell line–protein and cell line–tissue.

The KG is used to expand the neighboring nodes at each layer, propagating from far to near, layer by layer. Knowledge-based information propagation can efficiently collect higher order association information about drugs and cell lines, thus enhancing the capability of potential vector representation of entities.

Knowledge graph attention layer

Each tail entity has different meanings and potential vector representations when it has different head entities and relations in the KG. Additionally, there are intricate associations between each adjacent tail entity. To accurately capture entity embeddings, we also take into account the relation between the head entity's various neighbors based on the attention network. The module assigns different importance levels to the entity's neighbor set through a multi-head attention mechanism. It updates and aggregates the importance levels to obtain the embedded representation of the corresponding drug or cell line. More specifically, each layer of KG attention consists of two main components: multi-head attention embedding propagation and neighborhood information aggregation.

Multi-head attention embedding propagation

Entity h may participate in multiple triples as the bridge connecting two triplets and propagating information. Therefore, entity h has multiple different neighbors, and there is an

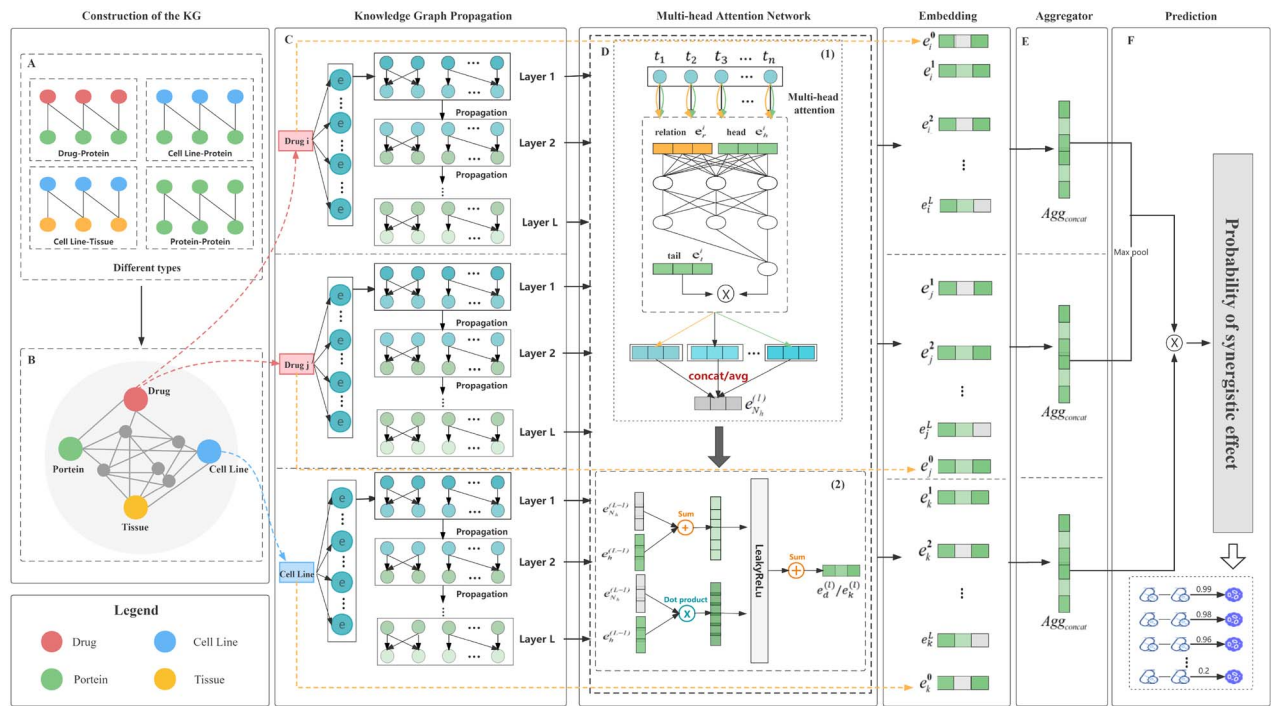


Figure 2. Architecture of KGANSynergy. A and B: Construction of the KG. **A.** Four different types of data are collected: drug-protein association, cell line-protein association, cell line-tissue association, and protein-protein association. **B.** Use the data in A to build a KG that includes four types of entities: drugs, cell lines, proteins and tissues. **C.** KG propagation layer. Hierarchical propagation is utilized to identify the set of drug/cell line entity neighbor nodes. The symbol e represents the initial entity associated with the drug/cell line entity, which is the set of tail entities directly related to the drug/cell line. **D.** KG attention layer. The figure shows a multi-head attention network that acts on each knowledge propagation process. **D (1).** The multi-head attention part of the KG. $\{t_1, t_2, \dots, t_n\}$ is the set of n neighboring nodes to entity h in this layer. Arrows with different colors represent multiple attention learning using the triple (h, r, t) , which is composed of entity h and each neighbor in the neighborhood set. The multiple learned embedding representations are concatenated/averaged to generate the final neighborhood set embedding representation $e_{N_h}^l$. **D (2).** Using the Bi-Interaction aggregator, the l -th layer entity h is aggregated with the neighbor node embedding $e_{N_h}^l$ acquired in the preceding step, and finally obtain the embedding of drug or cell line entity at layer l . **E.** Each layer of node embedding learned by the drug/cell line is aggregated to obtain the drug/cell line's final embedding representation. **F.** Prediction layer. The probability of synergy between the drug pair on a particular cell line is outputted.

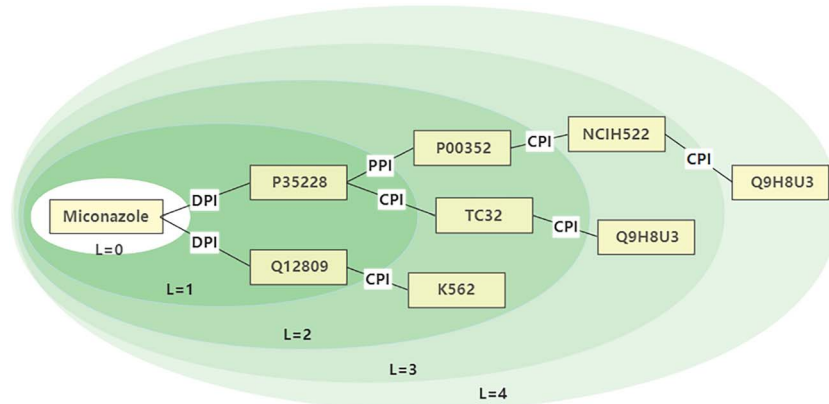


Figure 3. Illustrations of different layers of Miconazole in the KG. When $L=1$, the drug Miconazole is directly related to protein P35228 and protein Q12809, respectively. When $L=2$, the one-hop node protein P35228 is directly related to protein P00352 and cell line TC32, the one-hop node protein Q12809 is directly related to cell line K562, and these two-hop entities are indirectly adjacent to Miconazole. With the propagation of information through layer, the information contained in the node is integrated by its layer. Concentric circles depict nodes with different levels. The weaker the association between the center and surrounding entities, the lighter the hue of green.

association between these neighbors. To avoid providing the same weight to each neighbor when aggregating information, we adopted the idea of a graph attention network, which assigns different importance levels to the neighbors of each entity h and generates embedded attention weights for propagation. The process not only focuses on the network, but also considers the complex associations between the neighbor nodes of entity h .

Considering the instability of the original GAT's learning process, we add a multi-head attention mechanism to the original study's foundation. As shown in Figure 4, we draw the propagation of multi-head attention. The figure shows n neighbor nodes of h , where $\{t_1, t_2, \dots, t_n\}$ denotes the n neighbors of h in that layer. Different arrow colors indicate that several triples have collectively performed multiple iterations of attention

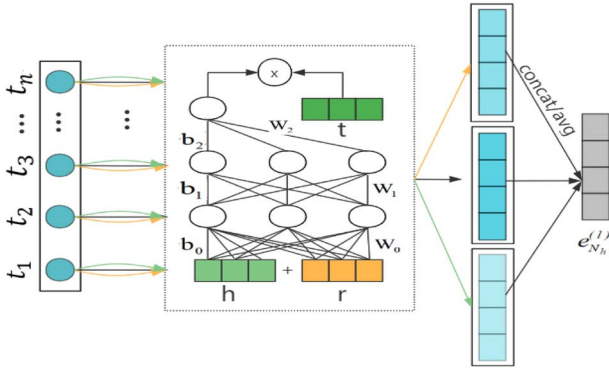


Figure 4. Diagram of the multi-head attention propagation process. Different arrow colors in multi-head attention represent independent attention computations. The aggregated features from each head are concatenated or averaged to obtain the set of neighborhood representations directly connected to entity h .

learning. Finally, multiple learned neighbor representations are aggregated to obtain a set of neighbor embedding representations. The following section describes the multiple attention propagation in detail.

First, the neighborhood representation of entity h is learned based on the neighborhood entities. When considering h , the neighbors of entity h are denoted by $N_h = \{(h, r, t) \mid (h, r, t) \in \mathcal{G}\}$, where t denotes the neighboring entities of the head entity h , and r denotes the relation. The purpose of attention embedding propagation is to encode N_h , and the output vector is represented as a set of embedding representations of neighbors. To improve the node representation capacity, the graph attention network uses e_{N_h} to represent the set of entity embedding representations directly connected to the entities h . In this study, we take the triple (h, r, t) as an example to perform a weighted sum of each neighbor node of entity h and finally obtain the neighborhood representation of entity e_{N_h} as follows:

$$e_{N_h} = \sum_{(h,r,t) \in N_h} \pi(h, r, t) e_t, \quad (2)$$

where e_t denotes the embedding representation of entity h 's neighboring entity t . $\pi(h, r, t)$ denotes the attention weight of each neighbor t . The weight controls the decay factor on each propagation on $\pi(h, r, t)$, which represents the amount of information propagated from entity t to h under relation r . The larger the attention weight of the triple, the more important the neighbor entity. Owing to the existence of the attention mechanism, the model is capable of learning different weights for different neighbors [40]. Next, we implement the function $\pi(\cdot)$ through a neural network similar to the attention mechanism, which is formulated as follows:

$$z_0 = \text{ReLU}(W_0(e_h \| e_r) + b_0), \quad (3)$$

$$\pi(h, r, t) = \sigma(W_2 \text{ReLU}(W_1 z_0 + b_1) + b_2), \quad (4)$$

where ReLU is the nonlinear activation function, and the last activation function is Sigmoid. $\|$ is the concatenation operation. W and b are the trainable weight weights and biases, respectively. In particular, W_0 and b_0 in Equation (3) represent the weights and deviations of the first layer neural network, whereas W_1, b_1 and W_2, b_2 in Equation (4) represent the weights and deviations used in the second layer neural network and the output layer.

According to the principle of the graph attention mechanism, we use the softmax function to normalize the coefficients of the triples associated with entity h to make it easier to compare the attention coefficients between different entities (the sum of attention of all neighboring nodes is 1) [40]. The final attention score can highlight which neighboring nodes should be given more attention to capture entity embeddings. The Softmax function is expressed as follows:

$$\begin{aligned} \pi(h, r, t) &= \text{softmax}(\pi(h, r, t)) \\ &= \frac{\exp(\pi(h, r, t))}{\sum_{(h,r',t')} \exp(\pi(h, r', t'))}. \end{aligned} \quad (5)$$

To stabilize the attention learning process and encapsulate more information about the neighborhood, we also apply the multi-head attention mechanism as done for GAT, converting Equation (2) [40]. Multi-head attention is computed M times by a separate attention mechanism to update the embedding representation of entity h . The multi-head attention layer also requires an aggregator to integrate all embeddings. In this study, we adapt two types of aggregators:

- **Concat aggregator.** Splicing operation. Concatenate all embeddings of multi-head graph attention and then apply a nonlinear transformation:

$$e_{N_h} = \text{LeakyReLU} \left(\big\|_{m=1}^M \sum_{(h,r,t) \in N_h} \pi(h, r, t) e_t \right) \quad (6)$$

- **Average aggregator.** Mean operation. Sum all embeddings of the multi-head graph attention and then apply the average to calculate the final embedding:

$$e_{N_h} = \text{LeakyReLU} \left(\frac{1}{M} \sum_{m=1}^M \sum_{(h,r,t) \in N_h} \pi(h, r, t) e_t \right) \quad (7)$$

In Equation (6), $\|$ denotes the connection operation. $\pi(h, r, t)$ of Equations (6) and (7) is the normalized attention coefficient calculated by the m -th attention embedding propagation, and M is the number of multi-heads.

Neighborhood information aggregation

Next, the neighbor set embedding e_{N_h} obtained in the previous step is combined with the entity embedding e_h to obtain the new entity embedding representation $e_{h'}$. We use the Bi-interaction aggregator [35] to implement the function $f(e_h, e_{N_h})$:

$$\begin{aligned} f(e_h, e_{N_h}) &= \text{LeakyReLU}(W_3(e_h + e_{N_h})) \\ &\quad + \text{LeakyReLU}(W_4(e_h \odot e_{N_h})), \end{aligned} \quad (8)$$

where LeakyReLU is a ReLU-based activation function that can assign a non-zero slope to all negative values. The parameter $W_3, W_4 \in \mathbb{R}^{d_e \times d_e'}$ are the weight matrices determined by parameter learning, and d_e' is the transformation size. \odot denotes the product of elements. The aggregator considers the interaction of two features between e_h and e_{N_h} . This function captures more messages from related entities, which enhances its ability to combine e_h and e_{N_h} .

The embedded propagation layer enables explicit use of first-order connectivity information to associate drugs, cell lines

and knowledge entity representations. However, using only the first-order neighbors of entities may result in the loss of important information. We extend multiple knowledge propagation layers to investigate the influence of higher order neighbors in greater depth. After we aggregate first-order neighbors, each entity includes information from its first-order neighbors, and then the process is repeated. As the first-order neighbor entity has saved the representation of the second-order neighbors in the preceding layer, it can collect data from second-order neighbors.

Analogously, at the l -th knowledge propagation layer, the entity may collect information from l -th order neighbors. As we aggregate L layers, each entity's embedding eventually contains information from its L -layer neighbors. The representation of entity h in each layer during propagation is defined as follows:

$$e_h^{(l)} = f(e_h^{(l)}, e_{N_h}^{(l)}), \quad l = 1, 2, \dots, L, \quad (9)$$

$$e_{N_h}^{(l)} = \text{LeakyReLU} \left(\sum_{(h,r,t) \in N_h} \pi(h, r, t) e_t^{(l)} \right), \quad (10)$$

where $e_h^{(l)}$ and $e_{N_h}^{(l)}$ denote entity h embedding and the embedding set of entity h 's neighbors at layer l , respectively. The l -th layer entity h is the neighbor of entity at layer $l-1$. The embedding of entity h at layer l is obtained by aggregating the embedding of its neighbor entities at layer $l-1$. $e_t^{(l)}$ is the embedding representation of entity h 's neighbor t , whose embedding has information from its $(l-1)$ -layer neighbors.

After executing L KG attention layers, we can obtain multiple embedding representations for drug d and cell line k . Importantly, the initial entities of drugs and cell lines in the KG are the closest nodes in the potential space to the drugs and cell lines with a strong association. Therefore, it is also important to enrich the entity embedding with the original item representation for each drug and cell line. The initial embedding vector of entity is defined by random initialization, and the length of the embedding is one parameter set in our study, which can be optimized based on experiments. The resulting drug embedding representation set (T_d) and cell line embedding representation set (T_k) are as follows:

$$T_d = \{e_d^{(0)}, e_d^{(1)}, \dots, e_d^{(L)}\}, \quad (11)$$

$$T_k = \{e_k^{(0)}, e_k^{(1)}, \dots, e_k^{(L)}\}, \quad (12)$$

where $e_d^{(0)}$ or $e_k^{(0)}$ is the initial embedding of the drug or cell line. $\{e_d^{(1)}, \dots, e_d^{(L)}\}$ and $\{e_k^{(1)}, \dots, e_k^{(L)}\}$ are the embeddings learned from the first KG attention layer to the last layer. Next, we define the original representation of the drug and cell line as follows:

$$e_d^0 = \frac{\sum_{e \in \{e | (e,d) \in B\}} e}{|\{e | (e,d) \in B\}|}, \quad (13)$$

$$e_k^0 = \frac{\sum_{e \in \{e | (e,k) \in B\}} e}{|\{e | (e,k) \in B\}|}, \quad (14)$$

where B represents the entity set obtained after mapping the entity d in the drug combination dataset to the entity e in the KG, (e,d) and $(e,k) \in B$. $d \in D$ (D denotes all drugs) and $k \in K$ (K denotes all cell lines) denote a drug and a cell line, respectively.

Using the multi-head attention network in the KG can help extract the topological neighbor structure of the entities in the neighbor set, learn the high-order entity representation and construct a refined neighbor set embedding representation.

Drug synergy prediction

The output representation of different layers can be interpreted as the potential impact of different layers, which emphasizes different higher order connectivity and similarity. After L -layer embedding propagation for two drugs and cell line, the corresponding vector representation sets (T_o) are aggregated. The output is the final drug/cell line entity embedding generated by KGANSynergy. To retain the latent embedding information more completely, we aggregate the representations of all layers into a single vector by a concatenate operation and then apply a nonlinear transformation as follows:

$$agg_{\text{concat}}^d = \sigma(W \cdot (e_d^0 \| e_d^1 \| \dots \| e_d^L) + b), \quad (15)$$

$$agg_{\text{concat}}^k = \sigma(W \cdot (e_k^0 \| e_k^1 \| \dots \| e_k^L) + b), \quad (16)$$

where e_d and e_k are the embedding representations from the set T_o of drug and cell line embedding representations, respectively. W and b are the trainable weight and bias, respectively. The nonlinear activation function σ is set as Sigmoid.

We use two drug and cell line representations obtained after aggregation to calculate the prediction scores. Specifically, max pooling uses the element-wise maximum of two drug representations as the combined drug representation [31]. The synergy score is calculated as follows:

$$\hat{y}(d_i, d_j, k) = \max(e_i, e_j) \odot e_k, \quad (17)$$

Objective function

Youden's J statistic [52] is primarily used to capture the performance of dichotomous diagnostic tests. To ensure the rationality of the results, we use Youden's J statistic to search for the most appropriate cutoff for positive and negative labels. Finally, we set the labels with synergy probabilities greater than this threshold to 1 and those less than this threshold to 0. Given the drug combination data and the KG, our goal is to learn a prediction function to predict whether drug i and drug j have a synergistic effect on cell line k :

$$\text{Loss} = \sum_{i,j \in N_d, i \neq j, k \in N_k} (-y_{i,j,k} \log \hat{y}_{i,j,k} - (1 - y_{i,j,k}) \log (1 - \hat{y}_{i,j,k})) + \lambda \|\Theta\|_2^2, \quad (18)$$

where $\hat{y}_{i,j,k}$ is the predicted value of the model, and $y_{i,j,k}$ is the ground-truth value of the drug pair-cell line synergy. Θ represents a set of model parameters. $\|\Theta\|_2^2$ is a L_2 -regularizer to prevent overfitting. The hyperparameter λ is used to balance the regularizer. The model training process is optimized using the Adam optimizer.

EXPERIMENTS AND RESULTS

In this section, we compare our results with recent methods on two datasets to evaluate the performance of KGANSynergy's drug combination synergy prediction.

Comparison with previous studies

To demonstrate the effectiveness of KGANSynergy, we compared it with some state-of-the-art and baseline methods, including DeepWalk, GCN, Deepsynergy, KGNN and GraphSynergy. In addition, we adopted the stratified nested cross-validation, and all test

Table 3. Performance comparison of KGANSynergy and baselines

Model	DrugCombDB			Oncology-Screen		
	AUC-ROC	AUC-PR	ACC	AUC-ROC	AUC-PR	ACC
DeepWalk	0.7101	0.6923	0.6674	0.6987	0.6785	0.6496
GCN	0.6961	0.6797	0.6531	0.6748	0.7076	0.6534
Deepsynergy	0.7436	0.7192	0.6851	0.7360	0.7205	0.6695
KGNN	0.7610	0.7649	0.7146	0.7283	0.7478	0.6749
GraphSynergy	0.7882	0.7725	0.7216	0.7801	0.7945	0.6776
KGANSynergy	0.8951	0.8921	0.8174	0.8911	0.8983	0.8221

The bold values mean the best result for each performance metric.

drug pairs have not appeared in the training set. Below are brief descriptions of these comparison methods.

- (1) **DeepWalk** [53]: Random walk-based methods are frequently used for link prediction and knowledge representation. Here, we used DeepWalk, an algorithm for mining graph-structured data that combines the random walk and Word2vec [54] algorithms. The algorithm can discover the network's hidden information and represent the graph's nodes as vectors holding latent information.
- (2) **GCN** [55]: GCN is a multilayer graph convolutional neural network that incorporates graph structures into convolution.
- (3) **KGNN** [38]: KGNN is a KG-based GNN framework that is mostly used in drug-drug interaction prediction tasks.
- (4) **Deepsynergy** [25]: Deepsynergy is a deep-learning-based drug synergy prediction method that utilizes the drug chemical structure and genetic information as input and uses conical layers to model drug synergy.
- (5) **GraphSynergy** [31]: GraphSynergy is a state-of-the-art drug synergy prediction method that utilizes a spatial graph-based convolutional network and attention mechanism to encode higher order structural information of drug and cell line protein modules to enrich entity embedding representations.

The prediction results of the two datasets are shown in Table 3. From the results, KGANSynergy has achieved AUC value of 0.895 and AUPR value of 0.892 on the DrugCombDB dataset, and AUC value of 0.891 and AUPR value of 0.898 on the Oncology-Screen dataset, which are superior to other methods. Specifically, KGANSynergy achieves superior prediction performance compared to GraphSynergy because the combination of KG and multi-head attention stabilizes the capability to capture node embeddings. It contains multi-hop nodes and enhances entity embedding representations. Moreover, the aggregated neighbor nodes contain not only protein information but also other neighbor information in the KG, which demonstrates the effectiveness of the KG and graph attention strategies for predicting drug combination synergy. GraphSynergy and KGANSynergy may have performed better than KGNN because they explore drugs, cell line features and related entities. In addition, they acquire the embedding representation of each node by distinguishing the significance of neighboring information using an attention mechanism. KGNN surpasses GCN and DeepSynergy because KG and higher order neighbors enable better exploration of entity representations.

Parameter sensitivity analysis

Effect of dimension of embedding

We evaluated the performance of KGANSynergy by changing the dimension of entity embedding in the Oncology-Screen dataset.

As shown in Figure 5, the AUC and AUPR are at their maximum when the embedding dimension is set to 64. Then, AUC and AUPR decrease steadily as the entity embedding dimension increases. This result indicates that increasing the dimension of entity embedding within a particular range can efficiently encode more KG information, while exceeding the threshold may result in overfitting. Therefore, AUC and AUPR present a trend of first rising and then falling.

Effect of knowledge propagation layers depths

We evaluated the influence of knowledge propagation layer depth on the Oncology-Screen dataset by changing the number of layers (L) of KGANSynergy. As shown in Figure 6, the experiments are in the range of {1, 2, 3, 4}.

The findings indicate that the optimal performance is reached when L is 3. The phenomenon may be the result of a trade-off between the positive signal's dependence on distance and the negative signal's noise. Specifically, when only first-order neighborhood entities are considered, correlations and dependencies between entities are not fully considered. As L is too large, the model provides more knowledge information but may introduce more noise.

Effect of aggregators

To explore the impact of multi-head attention mechanism aggregators, we conducted experiments on KGANSynergy using different aggregators. Here, we refer to the average and concat aggregators as KGANSynergy-avg and KGANSynergy-concat, respectively. As seen in Figure 7, the best result is obtained by using the concat aggregator.

Effect of the number of attention heads

We examined the effect of the number of attentional heads on the model. As shown in Figure 8, AUC and AUPR all display a trend of rising first and then falling. The KGANSynergy performance reaches its maximum when the number of attention heads is 2. The results indicate that the number of attention heads influences the performance of the model. This is because multiple heads are equivalent to multiple single heads acting together to stabilize the learning process. A single attention head is not sufficient to capture all of the semantic information about an entity, while using too many attention heads may introduce redundant information.

Ablation study

We verify how different parts of KGANSynergy affect the performance through an ablation study and design the following several of its variants:

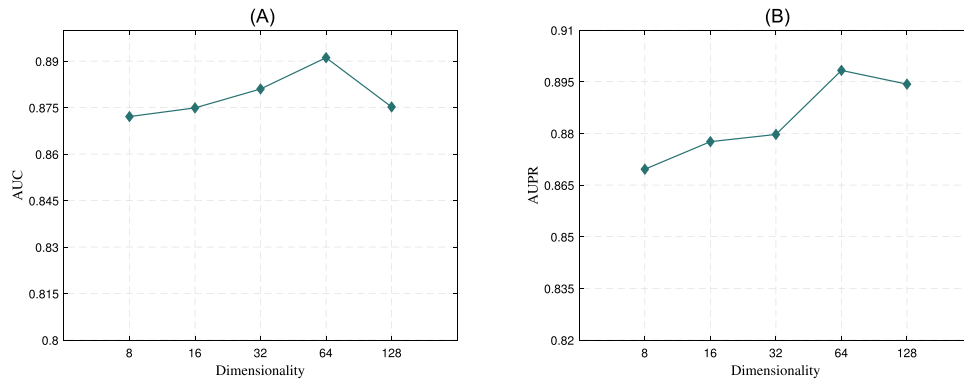


Figure 5. Embedding dimension of AUC and AUPR on Oncology-Screen.

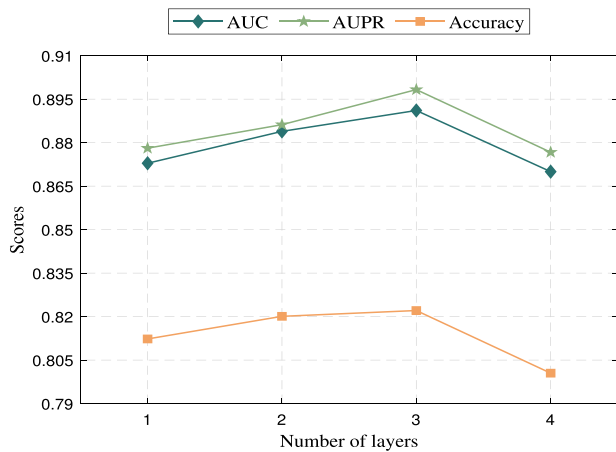


Figure 6. Impact of different layers on the model performance.

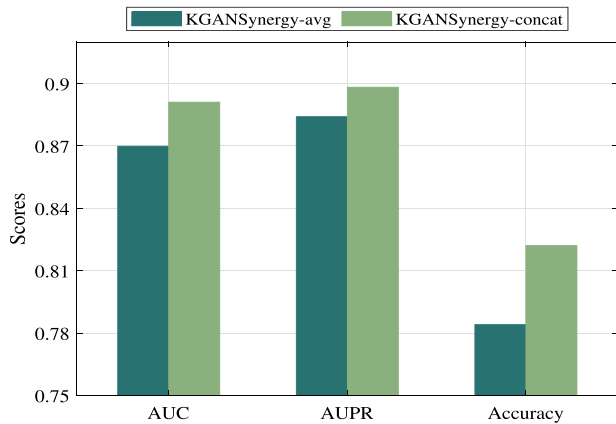


Figure 7. Effect of different attention aggregators on model performance.

- (1) KGANSynergy/-att: KGANSynergy without multi-head attention mechanism to update the entity representation and set $\pi(h, r, t)$ as $1/|N_h|$.
- (2) KGANSynergy/-m: KGANSynergy without multi-head attention mechanism to update the entity representation and replaced by single-head attention.
- (3) KGANSynergy/-d: KGANSynergy without neighbor entity embedding and replaced by the original embedding of the entity.

In addition, considering the guilt-by association principle, we removed protein-protein information and cell line-tissue

Table 4. Performance comparison between different variants

Methods	AUC-ROC	AUC-PR	ACC
KGANSynergy/-att	0.8762	0.8608	0.8155
KGANSynergy/-m	0.8882	0.8976	0.8132
KGANSynergy/-d	0.8396	0.8471	0.7662
KGANSynergy/-q	0.8581	0.8675	0.7880
KGANSynergy	0.8911	0.8983	0.8221

information, termed KGANSynergy/-q. We compared KGANSynergy with several of its variants, and the results are given in Table 4. The performance achieved by the model in different cases can be summarized as follows:

- Removing the attention mechanism or ignoring the knowledge graph neighbor entities can degrade the model performance. Moreover, KGANSynergy/-d is consistently the worst performer. This is because KGANSynergy/-d solely employs entity embedding and disregards neighbor entities and attention mechanisms. This reflects the importance and necessity of considering attention mechanisms and higher order neighbors.
- KGANSynergy/-m always performs better than KGANSynergy/-att. This may be because KGANSynergy/-att sets the weights of all triads to the same value during propagation, making it impossible to distinguish the different contributions of the triads. Thus, the addition of the attention mechanism aids in determining the weights of neighboring messages. Additionally, the result of KGANSynergy/-m indicates the superiority of the multi-head attention mechanism. Compared with the single-head attention mechanism, the multi-head attention mechanism enriches the model's capability and stabilizes the training procedure.

Case studies

To further examine the effectiveness of KGANSynergy, we forecast new drug combinations on the Oncology-Screen dataset. We used drugs and cell lines from Oncology-Screen to build drug candidate combinations and excluded existing drug pairs from the raw training data. Then, we utilized the training model to predict drug pairs with unknown synergy status and ranked all prospective drug pair-cell line combinations based on their prediction scores. The model predicted specific drug combinations in the cell lines and finally selected drug pairs with high prediction synergy scores in each tissue. We select three important cancers – lung, ovary and

Table 5. Top synergistic drug pairs predicted by KGANSynergy

Cell line	Tissue	Drug 1	Drug 2	PMID
NCIH2122	Lung	ERLOTINIB	ABT-888	22005537
NCIH2122	Lung	METHOTREXATE	ERLOTINIB	NA
NCIH2122	Lung	METHOTREXATE	LAPATINIB	NA
SKMES1	Lung	LAPATINIB	ZOLINZA	25896603
SKOV3	Ovary	DEXAMETHASONE	ETOPOSIDE	22932097
SKOV3	Ovary	DEXAMETHASONE	5-FU	34575034
SKOV3	Ovary	DEXAMETHASONE	VINORELBINE	25462205
CAOV3	Ovary	SUNITINIB	ERLOTINIB	24041628
KPL1	Breast	ABT-888	SN-38	26842236
KPL1	Breast	ZOLINZA	SN-38	26571493
KPL1	Breast	ETOPOSIDE	SN-38	31383812
KPL1	Breast	ETOPOSIDE	ABT-888	30327308

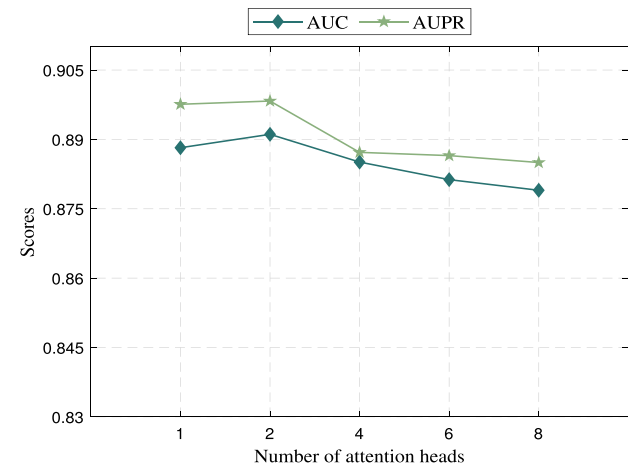


Figure 8. Effect of different number of attention heads on model performance.

breast for case study. Lung cancer is the cancer with the highest mortality rate. Breast cancer and ovarian cancer are two leading malignant tumors of women. We examined the top four drug pairs based on predicted scores for those three cancers, respectively (Table 5).

Glucocorticoids (GCs) like dexamethasone may be able to minimize acute toxicity or protect normal tissues and have been widely employed as combination agents in the treatment of solid cancers [56–59]. Erlotinib is a quinazoline derivative that selectively and reversibly inhibits the TK activity of EGFR. It is active in advanced non-small cell lung cancer, head and neck tumors, glioblastoma and other types of tumors [60]. SN-38 (an active metabolite of Irinotecan) is a camptothecin derivative targeting topoisomerase 1 and is primarily used in combination regimens for the treatment of metastatic or advanced solid tumors [61]. To further evaluate the accuracy of these predicted drug combinations, we undertook a comprehensive literature survey. For example, dexamethasone treatment in SKOV3 significantly increased SGK1 mRNA expression [62]. Studies have shown that dexamethasone sensitizes cancer stem cells (CSCs) to 5-FU by reducing NRF2 and increasing reactive oxygen species production [63]. The same mechanism works in ovarian and colonic CSCs treated with the combination of dexamethasone and chemotherapeutic agents. Vinorelbine induces apoptosis and reduces telomerase activity in the human ovarian epithelial carcinoma cell line SKOV3 [64]. A phase II study demonstrated that gemcitabine, vincristine

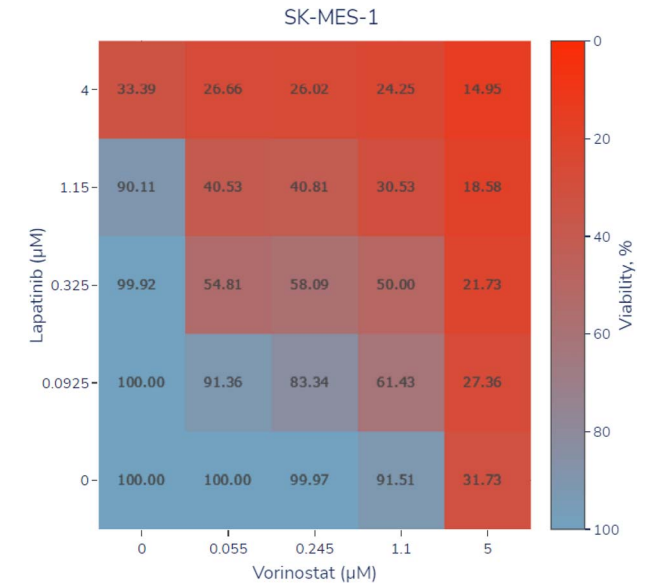


Figure 9. Heatmap of synergistic inhibition analysis of Lapatinib and Zolinza on SKMES1.

and dexamethasone are effective regimens for the treatment of relapsed/refractory Hodgkin's lymphoma (RRHL) with acceptable toxicity [65]. Both methotrexate and erlotinib have proven their value in treating lung cancer alone, but their combination has not attracted much attention [66, 67]. Subsequently, we used SynergxDB (<https://www.SYNERGxDB.ca/>) [68] to further evaluate the inhibitory effect of our predicted drug combination on cancer cells at different concentrations. For example, the combination of lapatinib and zolinza was observed to better inhibit lung cancer cell activity through the heatmap (Figure 9). Based on these results, we believe that KGANSynergy's prediction results are consistent with many previous studies, and have a strong ability to predict the ability to give candidate drugs.

CONCLUSION

In recent years, drug combination therapy has been successfully applied to the treatment of many complex diseases. Synergistic drugs can increase the efficacy and reduce the dose of a single drug. In this study, we proposed a novel drug combination synergy prediction model, KGANSynergy. Based on known drug and cell line information, KGANSynergy enhances the stability

of the traditional graph attention mechanism on the KG using a multi-head attention mechanism to update entity embedding representation. First, the method used a KG to better explore entity representations and enhance predictive capabilities. Second, entity embedding used KG attention to enrich itself through entity neighbor information and stabilized the attention learning process. Experiments on two datasets showed that KGANSynergy outperformed existing methods in predicting drug combinations.

The enrichment of biological data may help improve model performance. Therefore, our future work will focus on how to combine KGs with other biological information to extract high-quality drug and cell line embedding representations. For example, we can consider other types of multi-omics data (e.g. methylation, copy number and pathway activity) to enhance the model interpretability.

Key Points

- This paper proposes a novel KG attention network framework called KGANSynergy that can obtain high-order information from multi-source neighbors in the KG efficiently and explicitly.
- KGANSynergy uses a neural-network-based multi-head attention method to completely utilize neighborhood information and improve the interpretability of the model.
- KGANSynergy applies different types of aggregators in terms of neighboring nodes and hierarchical aggregation, highlighting different higher order connections and similarity.

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant nos 61802113 and 61802114); the Education Department of Henan Province (grant no. 222102210238) and the Science and Technology Development Plan Project of Henan Province (grant no. 212102210091).

DATA AVAILABILITY

The implementation of KGANSynergy and the datasets are available at <https://github.com/juanerzz7/KGANSynergy>.

AUTHORS' CONTRIBUTIONS STATEMENT

G.Z. and Z.G. conceived the main idea and the framework of the manuscript. Z.G. drafted the manuscript. Z.G. and H.L. collected the data and performed the experiments. C.Y., W.L. and J.W. helped to improve the idea and the manuscript. J.L. reviewed drafts of the paper. All authors read and commented on the manuscript.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**(6): 394–424.
2. Tan X, Long H, Luquette LJ, et al. Systematic identification of synergistic drug pairs targeting hiv. *Nat Biotechnol* 2012; **30**(11): 1125–30.
3. Giles TD, Weber MA, Basile J, et al. Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study. *Lancet* 2014;**383**(9932): 1889–98.
4. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011; **12**(1): 56–68.
5. Humphrey RW, Brockway-Lunardi LM, Bonk DT, et al. Opportunities and challenges in the development of experimental drug combinations for cancer. *J Natl Cancer Inst* 2011; **103**(16): 1222–6.
6. Jia J, Zhu F, Ma X, et al. Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009; **8**(2): 111–28.
7. Sun W, Sanderson PE, Zheng W. Drug combination therapy increases successful drug repositioning. *Drug Discov Today* 2016; **21**(7): 1189–95.
8. Wilson DR, Honrath U, Sonnenberg H. Interaction of amiloride and hydrochlorothiazide with atrial natriuretic factor in the medullary collecting duct. *Can J Physiol Pharmacol* 1988; **66**(5): 648–54.
9. Skolnik NS, Beck JD, Clark M. Combination antihypertensive drugs: recommendations for use. *Am Fam Physician* 2000; **61**(10): 3049.
10. Menzies AM, Long GV. Dabrafenib and trametinib, alone and in combination for braf-mutant metastatic melanoma. *Clin Cancer Res* 2014; **20**(8): 2035–43.
11. Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination therapy in the quest for clinical efficacy. *Nat Chem Biol* 2006; **2**(9): 458–66.
12. Pang K, Wan Y-W, Choi WT, et al. Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics* 2014; **30**(10): 1456–63.
13. Day D, Siu LL. Approaches to modernize the combination drug development paradigm. *Genome Med* 2016; **8**(1): 1–14.
14. He L, Kuleskiy E, Saarela J, et al. Methods for high-throughput drug combination screening and synergy scoring. In: von Stechow, L. (eds) *Cancer Systems Biology*. Humana Press, New York, NY: Springer, 2018, 351–98.
15. Bulusu KC, Guha R, Mason DJ, et al. Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discov Today* 2016; **21**(2): 225–38.
16. Morris MK, Clarke DC, Osimiri LC, Lauffenburger DA. Systematic analysis of quantitative logic model ensembles predicts drug combination effects on cell signaling networks. *CPT Pharmacometrics Syst Pharmacol* 2016; **5**(10): 544–53.
17. Lianlian W, Wen Y, Leng D, et al. Machine learning methods, databases and tools for drug combination prediction. *Brief Bioinform* 2022; **23**(1): bbab355.
18. Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. *Nat Commun* 2019; **10**(1): 1–11.
19. Sun X, Bao J, You Z, et al. Modeling of signaling crosstalk-mediated drug resistance and its implications on drug combination. *Oncotarget* 2016; **7**(39): 63995.
20. Wildenhain J, Spitzer M, Dolma S, et al. Prediction of synergism from chemical-genetic interactions by machine learning. *Cell Systems* 2015; **1**(6): 383–95.

21. Doucet J-P, Barbault F, Xia H, et al. Nonlinear SVM approaches to QSQR/QSAR studies and drug design. *Curr Comput Aided Drug Des* 2007; **3**(4): 263–89.
22. Li P, Huang C, Yingxue F, et al. Large-scale exploration and analysis of drug combinations. *Bioinformatics* 2015; **31**(12): 2007–16.
23. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**(7553): 436–44.
24. Luo H, Li M, Yang M, et al. Biomedical data and computational models for drug repositioning: a comprehensive review. *Brief Bioinform* 2021; **22**(2): 1604–19.
25. Preuer K, Lewis RPI, Hochreiter S, et al. DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018; **34**(9): 1538–46.
26. Kuru HI, Tastan O, Cicek E. Matchmaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2021; **19**:2334–44.
27. Wang J, Liu X, Shen S, et al. Deepdds: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform* 2022; **23**(1): bbab390.
28. Ma J, Wang J, Ghorai LS, et al. A comparative study of cluster detection algorithms in protein–protein interaction for drug target discovery and drug repurposing. *Front Pharmacol* 2019; **10**: 109.
29. Qian X, Xiong Y, Dai H, et al. Pdc-sgb: prediction of effective drug combinations using a stochastic gradient boosting algorithm. *J Theor Biol* 2017; **417**:1–7.
30. Liu Q, Xie L. Transynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput Biol* 2021; **17**(2): e1008653.
31. Yang J, Zhongzhi X, William Ka Kei W, et al. Graphsynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J Am Med Inform Assoc* 2021; **28**(11): 2336–45.
32. Dai Y, Wang S, Xiong NN, Guo W. A survey on knowledge graph embedding: approaches, applications and benchmarks. *Electronics* 2020; **9**(5): 750.
33. Wang H, Zhao M, Xie X, et al. Knowledge graph convolutional networks for recommender systems. In: *The World Wide Web Conference*. New York, NY, USA: ACM, 2019;3307–13.
34. Wang Z, Lin G, Tan H, Chen Q, and Liu X. Ckan: collaborative knowledge-aware attentive network for recommender systems. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM, pp. 219–28, 2020.
35. Wang X, He X, Cao Y, Liu M, and Chua T-S. Kgat: knowledge graph attention network for recommendation. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 950–8, 2019. ACM, New York, NY, USA.
36. Zeng X, Xinqi T, Liu Y, et al. Toward better drug discovery with knowledge graph. *Curr Opin Struct Biol* 2022; **72**:114–26.
37. Bonner S, Barrett IP, Ye C, et al. Understanding the performance of knowledge graph embeddings in drug discovery. *Artificial Intelligence in the Life Sciences* 2022; **2**:100036.
38. Lin X, Quan Z, Wang Z-J, et al. Kgnn: knowledge graph neural network for drug-drug interaction prediction. In: *IJCAI* 2020; **380**: 2739–45.
39. Mohamed SK, Nováček V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. *Bioinformatics* 2020; **36**(2): 603–10.
40. Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. *stat* 2017; **1050**(20):10–48550.
41. Liu H, Zhang W, Zou B, et al. Drugcombdb: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020; **48**(D1): D871–81.
42. O'Neil J, Benita Y, Feldman I, et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol Cancer Ther* 2016; **15**(6): 1155–62.
43. Loewe S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* 1953; **3**:285–90.
44. Bliss CI. The toxicity of poisons applied jointly 1. *Ann Applied Biology* 1939; **26**(3): 585–615.
45. Borisy AA, Elliott PJ, Hurst NW, et al. Systematic discovery of multicomponent therapeutics. *Proc Natl Acad Sci*, **100**(13): 7977–82, 2003.
46. Yadav B, Wennerberg K, Aittokallio T, Tang J. Searching for drug synergy in complex dose–response landscapes using an interaction potency model. *Comput Struct Biotechnol J* 2015; **13**: 504–13.
47. Di Veroli GY, Fornari C, Wang D, et al. Combeneft: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics* 2016; **32**(18): 2866–8.
48. Segal E, Wang H, Koller D. Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics* 2003; **19**(suppl_1): i264–72.
49. Barretina J, Caponigro G, Stransky N, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; **483**(7391): 603–7.
50. Rouillard AD, Gundersen GW, Fernandez NF, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016; **2016**:1–16.
51. Safran M, Dalah I, Alexander J, et al. Genecards version 3: the human gene integrator. *Database* 2010; **2010**:baq020.
52. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical J: J Math Methods Biosci* 2008; **50**(3): 419–30.
53. Perozzi B, Al-Rfou R, and Skiena S. Deepwalk: online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*. New York, USA: ACM, pp. 701–10, 2014.
54. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In *Proc. of ICLR Workshops*. CoRR abs/1301.3781. 2013.
55. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *Proceedings of the International Conference on Learning Representationss (ICLR)*, Toulon. arXiv preprint arXiv:1609.02907. 2016.
56. Rutz HP, Herr I. Interference of glucocorticoids with apoptosis signaling and host-tumor interactions. *Cancer Biol Ther* 2004; **3**(8): 715–8.
57. Rutz HP. Effects of corticosteroid use on treatment of solid tumours. *Lancet* 2002; **360**(9349): 1969–70.
58. Xing K, Bingxin G, Zhang P, Xianghua W. Dexamethasone enhances programmed cell death 1 (pd-1) expression during t cell activation: an insight into the optimum application of glucocorticoids in anti-cancer therapy. *BMC Immunol* 2015; **16**(1): 1–9.
59. Herr I, Pfizenmaier J. Glucocorticoid use in prostate cancer and other solid tumours: implications for effectiveness of cytotoxic treatment and metastases. *Lancet Oncol* 2006; **7**(5): 425–30.

60. Bareschino MA, Schettino C, Troiani T, et al. Erlotinib in cancer treatment. *Ann Oncol* 2007; **18**:vi35–41.
61. Bailly C. Irinotecan: 25 years of cancer treatment. *Pharmacol Res* 2019; **148**:104398.
62. Amal Melhem S, Yamada D, Fleming GF, et al. Administration of glucocorticoids to ovarian cancer patients is associated with expression of the anti-apoptotic genes sgk1 and mkp1/dusp1 in ovarian tissues. *Clin Cancer Res* 2009; **15**(9): 3196–204.
63. Suzuki S, Yamamoto M, Sanomachi T, et al. Dexamethasone sensitizes cancer stem cells to gemcitabine and 5-fluorouracil by increasing reactive oxygen species production through nrf2 reduction. *Life* 2021; **11**(9): 885.
64. Li-Yuan SHEN, Dong-Mei XU, Xiao-Han LIU, et al. Vinorelbine induces apoptosis and decreases telomerase activity in human epithelial ovarian cancer cells line skov3. *Basic Clin Med* 2018; **38**(1): 87.
65. Ganesan P, Mehra N, Joel A, et al. Gemcitabine, vinorelbine and dexamethasone: a safe and effective regimen for treatment of relapsed/refractory hodgkin's lymphoma. *Leuk Res* 2019; **84**:106188.
66. Abdelrady H, Hathout RM, Osman R, et al. Exploiting gelatin nanocarriers in the pulmonary delivery of methotrexate for lung cancer therapy. *Eur J Pharm Sci* 2019; **133**:115–26.
67. Saito H, Fukuhara T, Furuya N, et al. Erlotinib plus bevacizumab versus erlotinib alone in patients with egfr-positive advanced non-squamous non-small-cell lung cancer (nej026): interim analysis of an open-label, randomised, multicentre, phase 3 trial. *Lancet Oncol* 2019; **20**(5): 625–35.
68. Seo H, Tkachuk D, Ho C, et al. Synergxdb: an integrative pharmacogenomic portal to identify synergistic drug combinations for precision oncology. *Nucleic Acids Res* 2020; **48**(W1): W494–501.