Directed graph attention networks for predicting asymmetric drug–drug interactions

Yi-Yang Feng, Hui Yu (iD), Yue-Hua Feng and Jian-Yu Shi (iD)

Corresponding authors: Hui Yu, School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China. Tel.: +86 29 88431537;
E-mail: huiyu@nwpu.edu.cn; Jian-Yu Shi, School of Life Sciences, Northwestern Polytechnical University, Xi'an 710072, China. Tel.: +86 29 88460332;
E-mail: jianyushi@nwpu.edu.cn

Abstract

It is tough to detect unexpected drug–drug interactions (DDIs) in poly-drug treatments because of high costs and clinical limitations. Computational approaches, such as deep learning-based approaches, are promising to screen potential DDIs among numerous drug pairs. Nevertheless, existing approaches neglect the asymmetric roles of two drugs in interaction. Such an asymmetry is crucial to poly-drug treatments since it determines drug priority in co-prescription. This paper designs a directed graph attention network (DGAT-DDI) to predict asymmetric DDIs. First, its encoder learns the embeddings of the source role, the target role and the self-roles of a drug. The source role embedding represents how a drug influences other drugs in DDIs. In contrast, the target role embedding represents how it is influenced by others. The self-role embedding encodes its chemical structure in a role-specific manner. Besides, two role-specific items, aggressiveness and impressionability, capture how the number of interaction partners of a drug affects its interaction tendency. Furthermore, the predictor of DGAT-DDI discriminates direction-specific interactions by the combination between two proximities and the above two role-specific items. The proximities measure the similarity between source/target embeddings and self-role embeddings. In the designated experiments, the comparison with state-of-the-art deep learning models demonstrates the superiority of DGAT-DDI across a direction-specific predicting task and a direction-blinded predicting task. An ablation study reveals how well each component of DGAT-DDI contributes to its ability. Moreover, a case study of finding novel DDIs confirms its practical ability, where 7 out of the top 10 candidates are validated in DrugBank.

Keywords: attention, asymmetry, directed graph neural network, drug–drug interaction

Introduction

Single-drug therapy usually fails to treat complex diseases, which involves sophisticated biological processes, while poly-drug therapy is as one of the promising treatments for complex diseases [1]. The primary task in poly-drug treatment is to detect unexpected drug–drug interactions (DDIs) [2]. Because the pharmacokinetic or pharmacodynamic behaviors of a drug are changed by its interacting partners, possible adverse reactions would push patients in danger and even death [3]. Nevertheless, the identification of DDI in the wet lab is still costly and time-consuming. In recent years, computational approaches, especially deep learning-based approaches, are vigorously developed to perform preliminary DDI screening on a large scale with significantly low cost and less time. Current computational approaches (e.g. deep learning-based) can predict DDIs, to the best of our knowledge, by commonly treating two drugs in interaction as two equal roles in pharmacology. However, they ignore the pharmacological asymmetry of interacting drugs.

Many biological experiments have proved the asymmetry among DDIs. Wicha *et al.* [4] demonstrates that the majority of drug combinations between antifungal and nonantifungal drugs are of monodirectional interactions. In the case study, they validated that Terbinafine can mediate monodirectional antagonism through its effect in the ergosterol pathway and works as the perpetrator significantly increasing the INT value of Amphotericin B, while its EC50 was not significantly altered. This result is consistent with the statement in DrugBank [5], 'The risk or severity of myopathy, rhabdomyolysis, and myoglobin-uria can be increased when Amphotericin B is combined with Terbinafine.' Formally, we define the asymmetric interaction as the monodirectional interaction where the perpetrator drug influences the victim drug, such as

'Terbinafine (perpetrator)→Amphotericin B (victim)'. In addition, it is remarkable that the roles of perpetrator/victim are interaction-specific since a drug can be the perpetrator in one interaction and can be the victim in another interaction.

More importantly, such an asymmetric interaction further determines the taking sequence of drugs in a poly-drug treatment. The study of the optimum time sequence for the administration of vincristine and cyclophosphamide *in vivo* [6] showed that the additive effect did not appear when the drugs were administered at the same time. In contrast, with the extension of the time interval, the additive effect appeared. Specifically, taking Vincristine first can have better antitumor activity because the metabolism of Vincristine can be increased when combined with Cyclophosphamide (i.e. Cyclophosphamide→Vincristine). Similarly, a clinical trial compared different administration sequences of Cisplatin and 5-Fluorouracil on the efficacy and safety of chemotherapy [7]. In detail, the overall effective rate (31.3%), median survival time (239d) and Time-To-Progression (175d) in group A (i.e. Cisplatin→5-Fluorouracil) were significantly higher than those (13.9%, 174d and 140d accordingly) in group B (i.e. 5-Fluorouracil→Cisplatin). The observation is also consistent with DrugBank, which states 'the risk or severity of adverse effects can be increased when Cisplatin is combined with Fluorouracil.' More similar strategies are proved in various treatments, such as the sequential medication strategy of six MET inhibitor for non-small-cell lung cancer [8], decreasing drug resistance of antibiotics [9], eradication of helicobacter pylorie [10] and acute kidney injury [11]. In summary, it is crucial to determine the drug sequence that is an indispensable task in poly-drug treatments. Therefore, it is urgent to develop novel deep learning-based approaches to infer asymmetric DDIs.

Related works

Various approaches based on deep learning for DDI prediction have been developing in recent years. They can be approximately classified into two groups: molecule feature-based and network structure-based.

Early molecule feature-based approaches (e.g. Deep-DDI [12]) directly use deep neural networks to predict interactions by taking drug pairs as samples represented by feature engineering. After that, graph neural networks (GNNs), including graph convolutional networks (GCNs) and graph attention networks (GATs) as well as generative adversarial networks (GANs), are particularly appropriate to characterize drug molecules in an end-to-end manner. For example, Arnold *et al.* [13] designed multi-layer GATs to represent drug molecules individually and a joint con-attention layer to capture how the substructure pairs of two drugs contribute to their interaction. Deng *et al.* [14] used four types of drug features to construct deep neural network-based submodels

and learned cross-modality representations of drug pairs. Chen *et al.* [15] utilized a Siamese GCN, which can find important local atoms with the attention mechanism, to learn pairwise drug representations. Wang *et al.* [16] designed a bond-aware attentive message propagating method to capture drug molecular structure information under the framework of contrastive learning. However, molecule feature-based approaches ignore the dependence between DDIs.

To address this issue, network structure-based approaches organize DDI entries into an interaction network, where nodes are drugs and edges are interactions. Node embedding techniques are leveraged to represent drugs, while DDI prediction is regarded as link prediction. For example, Feng *et al.* [17] used a deep graph autoencoder [18, 19] to obtain drug latent representations, of which the operation of inner product is used to infer potential DDIs. Yu *et al.* [20] integrated the relation-aware network structure information in a multirelational DDI network to obtain the drug embedding. Lin *et al.* [21] utilized knowledge graph (KG) with rich bio-medical information (including enzymes, targets, genes) to learn drug representations without considering drug molecular structure information. Although all these state-of-the-art approaches exhibit the encouraging DDI prediction, they cannot handle pharmacological asymmetric interactions between drugs because they treat two drugs in an interaction equally.

It is crucial to infer asymmetric interactions because they determine the taking sequence of drugs in a poly-drug treatment. This paper aims to address the representation of asymmetry DDIs. By organizing a set of asymmetry DDIs into a directed DDI network, the prediction of asymmetry DDIs can be regarded as directed link prediction. Current works in directed link prediction can be roughly categorized into random walk-based and graph deep learning-based approaches. Random walk-based approaches usually develop random walk variants to infer directed links between nodes. For example, APP [22] captured asymmetric pairwise similarities and high-order similarities between nodes based on a directional random walk. NERD [23] applied an alternating random walk strategy, which can walk forward and backward to learn node embeddings in their corresponding source/target roles while fully exploiting the semantics of directed graphs. Ghorbanzadeh *et al.* [24] proposed a local similarity measure based on Hyperlink-Induced Topic Search. However, these approaches fail to capture the highly nonlinear characteristics in the graph.

Graph deep learning-based approaches provide a new sight of node embedding in a directed graph by generating two role-specific embeddings of a node (i.e. source role and target role). One accounts for its source role emitting links, while another is its target role absorbing links. Gravity GraphVAE [25] directly extends the graph variant autoencoder on the directed graph to learn asymmetric embeddings. DGGAN [26], a GAN-based directed graph embedding framework, leverages

adversarial mechanisms to learn each node's source and target embeddings. However, these methods cannot achieve the satisfactory performance of asymmetric DDI prediction because of no consideration of the association between drug features and node embedding.

To address the above issue, this work organizes asymmetric DDIs into a directed graph and characterizes them under two underlying assumptions. If drug u (perpetrator) influences drug v (victim), we first assume that the source role of drug u and the self-role of drug v are close in the source embedding space. Parallelly, we assume that the target role of drug v and the self-role of drug u are close in the target embedding space. Based on these assumptions, this work proposes a novel architecture of directed GATs for predicting asymmetric drug–drug interactions (DGAT-DDI). Our contributions are summarized as follows:

- DGAT-DDI generates two asymmetric embeddings of the source role and the target role for a drug pair, respectively. Its source role indicates how it influences other drugs in DDIs. Its target role represents how it is influenced by others. Its self-role is aligned into the source role space and the target role space, respectively, to reflect the proximity of the drug pair being an asymmetric interaction.
- Moreover, DGAT-DDI learns the aggressiveness of the source role and the impressionability of the target role to reflect how the number of interaction partners of a drug affects its interaction tendency.
- To the best of our knowledge, DGAT-DDI is the first approach for predicting asymmetric interactions among drugs. Its superiority is demonstrated by a direction-specific predicting task, a direction-blinded task as well as a case study of novel asymmetric DDI prediction.

Method
Problem formulation

Considering asymmetric interactions among drugs, we organize a set of DDI entries into a directed interaction graph and attempt to predict its potential edges upon node embeddings. Formally, denote a directed graph as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the node set (drugs) and $\mathcal{E}$ is the directed edge set (asymmetric interactions between drugs). For nodes $u, v \in \mathcal{V}$, $(u, v) \in \mathcal{E}$ represents a directed edge from u to v. In other words, drug u influences v. For convenience, we term u as the perpetrator drug and v as the victim drug in such an asymmetry interaction.

The task of prediction is to find a model $\mathcal{F}$ that can predict the occurrence of any directed edges by embedding nodes. Specifically, the task contains two subtasks. One is the direction-specific task, which judges how possibly an edge (u, v) with a specific direction from u to v occurs (Fig. 1A). Another, a more important but harder one, is a direction-blinded task, which determines how possible

Table 1. Notations of DGAT-DDI

p	p=1024, which the number of pre-defined substructures in Morgan fingerprint
d	d=16, which is the output dimension of embeddings from GATs and MLP
W_r	$W_r \in \mathbb{R}^{(d+1) \times p}$, which is the linear transformation matrix in the GATs
$\vec{a}_r$	$\vec{a}_r \in \mathbb{R}^{2(d+1)}$, which is the single-layer feedforward neural network in the GATs
z_u^0	$z_u^0 \in \mathbb{R}^p$, which is the initial feature of the node u
s_u / t_u	$s_u/t_u \in \mathbb{R}^d$, represents the source/target role of the node u
z_u	$z_u \in \mathbb{R}^d$, represents the self role of the node u
W_s' / W_t'	$W_s'/W_t' \in \mathbb{R}^{d \times d}$, represent role alignment matrices between the source/target role and the self-role
z_u^*	$z_u^* \in \mathbb{R}^d$, which is the self-role of node u after alignment

Table 2. Pseudo-codes of DGATDDI

Algorithm 1. DGATDDI algorithm

1: Input: DDI matrix Y, feature matrix X, hyper-parameter: α, β
2: Output: DDI network $\hat{Y}$ reconstructed by DGATDDI
3: Initialize:
4: for $u = 1, 2, \ldots, n$ do
  5: $z_u^0 \leftarrow x_u$
6: end for
7: while DGATDDI not converge do
  8: for $u = 1, 2, \ldots, n$ do
    9: $s_u, m_u^s \leftarrow sourceGAT(z_v^0) \mid u \rightarrow v \in //Eq.(4-7)$
    10: $t_u, m_u^t \leftarrow targetGAT(z_v^0) \mid u \leftarrow v \in \mathcal{E} //Eq.(4-7)$
    11: $z_u \leftarrow MLP(z_u^0)//Eq.(3)$
  12: end for
  13: Calculate predicted probability $\hat{Y} \leftarrow Eq.(8)$
  14: $\mathcal{L} \leftarrow Eq.(9)$
 15: Back-propagation to update parameters
16: end for

the edge between u and v is $(u \rightarrow v)$, $(v \rightarrow u)$, bidirectional edge or even a nonedge (Fig. 1B).

DGAT-DDI architecture

To address the above task, this section designs a DGAT-DDI for predicting asymmetric DDIs. The overall architecture of DGAT-DDI is shown in Fig. 2, its notations used in the following sections are listed in Table 1 and the pseudo-codes of its algorithm is provided in Table 2.

Its encoder module learns the source role, the target role and the self-role of a drug. The source role indicates how the drug influences other drugs in the directed interaction graph by a source GAT. In contrast, the target role reflects how it is influenced by others by a target GAT. The self-role encodes its additional properties (e.g. chemical structure) by an MLP. Moreover, DGAT-DDI represents the aggressiveness of the source role and the impressionability of the target role to capture how the number of interaction partners of a drug affects its interaction tendency w.r.t the role.
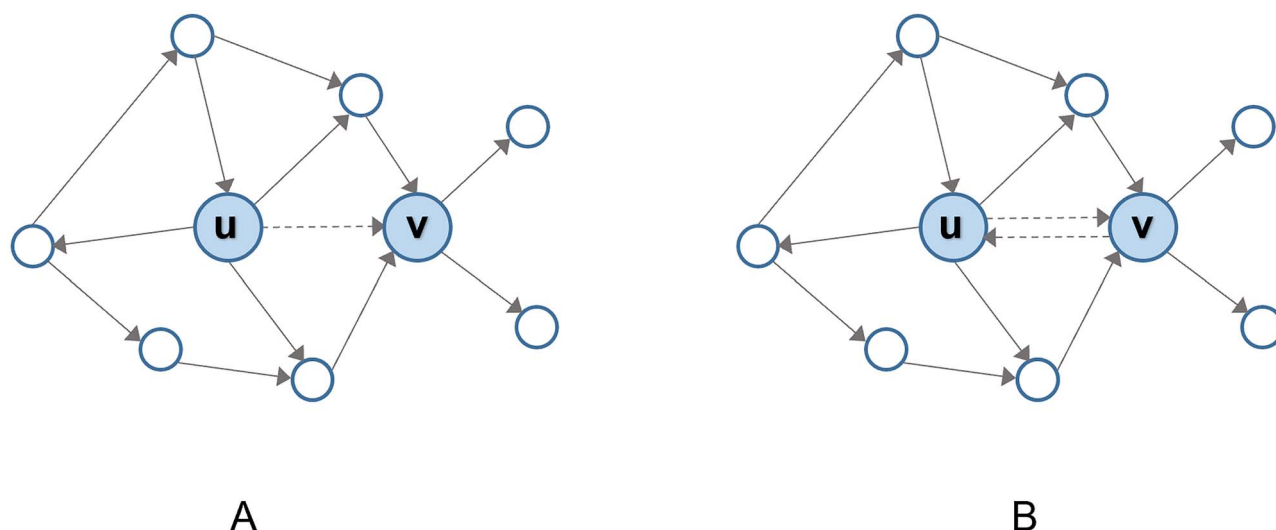
Figure 1. Two tasks of predicting asymmetric DDIs. (**A**) The direction-specific task. The scenario judges how likely an interaction ($u \rightarrow v$) from u to v occurs. (**B**) The direction-blinded task. The scenario determines how possible the interaction between u and v is ($u \rightarrow v$), ($u \leftarrow v$), or even a non-interaction. Round nodes are drugs and solid directed lines between them indicate their asymmetric interactions. Dashed directed lines indicate drug pairs of interest to be determined whether they are possible DDIs.
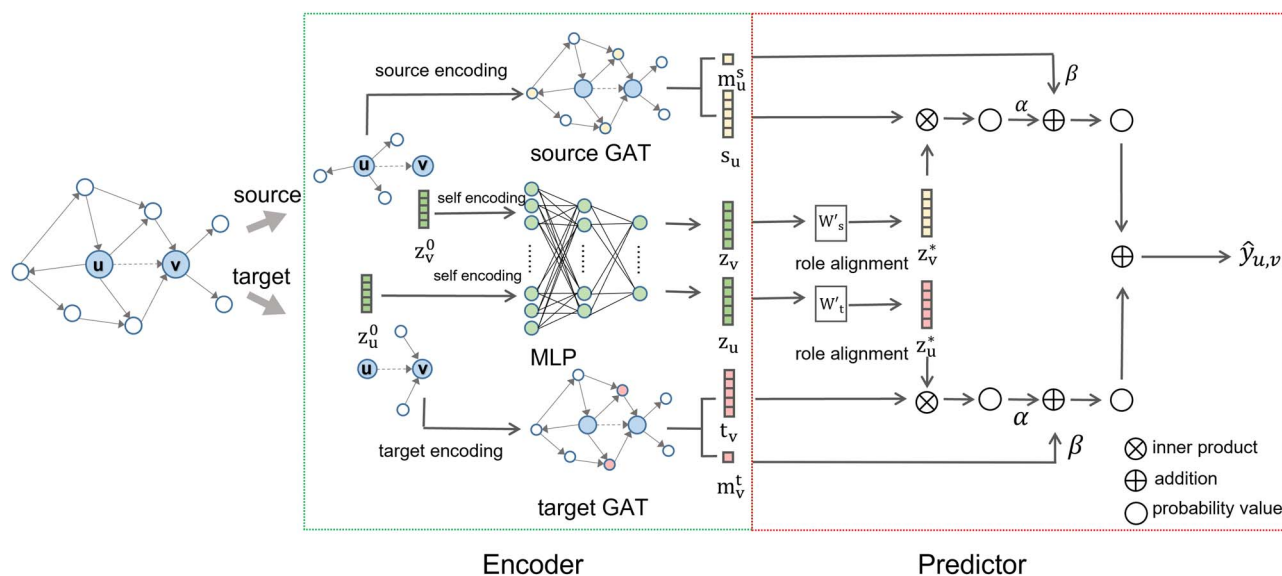

Figure 2. The overall architecture of DGAT-DDI. The encoder module is framed by green dotted lines, while the predictor is framed by red dotted lines. Yellow elements indicate the source embedding, red elements indicate the target embedding, green elements indicate the self-role embedding. Though the architecture of the source GAT is same as that of the target GAT, they handle different neighborhoods in terms of source/target nodes to encode the asymmetry of interactions. Specifically, the source GAT characterize drug U by aggregating its outgoing neighbors (3 nodes highlighted by yellow), while the target GAT characterizes drug V by aggregating its incoming neighbors (2 nodes highlighted by red in Fig. 2). In contrast, if the two drugs are characterized in an undirected graph, drug U needs to aggregate all its neighbors (5 nodes except for V), while drug V needs to aggregate all its neighbors (4 nodes except for U).

Its predictor contains three steps to determine potential asymmetric interaction between two drugs of interest. For example, we attempt to determine whether there is a potential interaction from u (a possible perpetrator drug) to v (a possible victim drug), denoted as $u \rightarrow v$. First, it determines how likely u influences v by the proximity between the source role of the perpetrator u and the self-role of the victim v in addition to the aggressiveness of the perpetrator. It then determines how likely v is influenced by u by the proximity between the target role of v and the self-role of u in addition to the impressionability of the victim. Last, bi-directional proximities are averaged as the final measure of how

likely the interaction from u to v occurs. In other words, it determines how likely u is a perpetrator drug, and meanwhile, v is a victim drug.

Source role encoder and target role encoder

Because of our emphasis on the difference between a perpetrator drug and a victim drug, we characterize the source role and the target role separately. In the directed interaction graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, we first define two kinds of neighborhoods of a node u according to its source role and target role, respectively. One is the neighborhood, $\mathcal{N}_s = \{v \in \mathcal{V} \mid u \rightarrow v \in \mathcal{E}\}$, which represents the first-order outgoing neighbors of u w.r.t. its source role. Another is
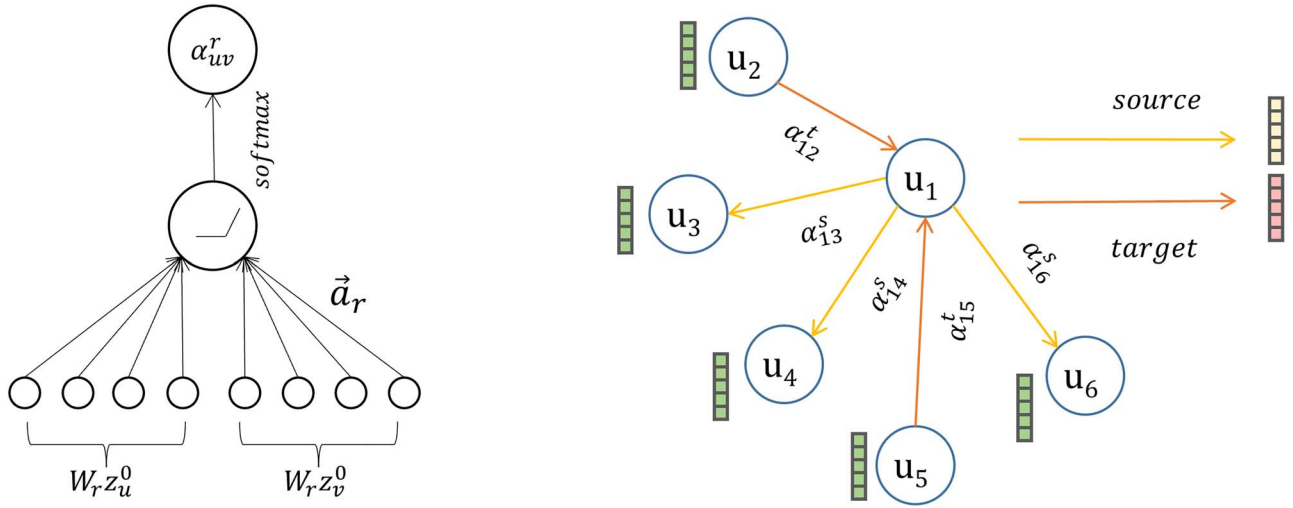
Figure 3. Illustration of role embeddings. Left Panel: The attention mechanism employed by our model. Right Panel: The center node u_1 aggregates the information from its neighbors by directed specific graph attention convolutions to generate its source embedding and target embedding respectively. Note that the center node u_1 has no self-loop. Yellow arrows indicate neighboring nodes accounting for source embeddings, while red arrows indicate neighboring nodes accounting for target embeddings.

$\mathcal{N}_t = \{v \in \mathcal{V} \mid u \leftarrow v \in \mathcal{E}\}$, which represents the first-order incoming neighbors of u w.r.t its target role.

Then, we leverage GATs [27] to aggregate the information from neighboring nodes of u via the attention mechanism. The attention score α_{uv}^r of v to u with respect to the role $r \in \{s, t\}$ is defined as follows:

$$\alpha_{uv}^r = \frac{\exp\left(\sigma\left(\vec{a}_r^\top \left[W_r z_u^0 \| W_r z_v^0\right]\right)\right)}{\sum_{k \in \mathcal{N}_r(u)} \exp\left(\sigma\left(\vec{a}_r^\top \left[W_r z_u^0 \| W_r z_k^0\right]\right)\right)}, \quad (1)$$

where $\|$ is the concatenation operation, $\mathcal{N}_r(u)$ is the neighbors of node u w.r.t $r \in \{s, t\}$, s is the source role and t is the target role. $W_r \in \mathbb{R}^{d \times p}$ is the weight matrix, p is the dimension of drug properties, $\vec{a}_r \in \mathbb{R}^{2d}$ is a single-layer feedforward neural network and σ is a nonlinear activation function (i.e. LeakyReLU). The attention score α_{uv}^r aggregates different neighbors v under the same role r of u. Differently from the original GAT, we discard self-loops of the center node when using the attention aggregation because its source role and its target role are supposed to be distinguishable. In other words, the center node cannot be its own outgoing neighbor and its own incoming neighbor simultaneously in the directed graph (Fig. 3).

Last, the embedding representations of the source role and the target role are defined in a similar manner (i.e. an operation of graph attention convolution) as follows:

$$s_u = \sum_{v \in \mathcal{N}_s(u)} \alpha_{uv}^s W_s z_v^0$$

$$t_u = \sum_{v \in \mathcal{N}_t(u)} \alpha_{uv}^t W_t z_v^0 \quad (2)$$

In such a manner, the source role represents the outgoing node neighborhood, while the target role represents the incoming node neighborhood.

Self-role encoder

For a given asymmetric interaction $u \rightarrow v$, we believe that v, as one of the outgoing neighbors of u, is similar to other outgoing neighbors of u. Meanwhile, u, as one of the incoming neighbors of v, is similar to other incoming neighbors of v. Thus, we introduce the self-role $z_v(u)$ of v attached to the source role of u as well as the self-role $z_u(v)$ of u attached to the target role of v.

For short, we simplify $z_v(u)$ as z_v and $z_u(v)$ as z_u. Considering that the source role s_u of u aggregates its outgoing neighbors, we assume that z_v and s_u are close in the embedding space. Likewise, considering that the target role t_v of v gathers its incoming neighbors, we assume that the target role t_v and the self-role z_u are close in the embedding space. The proximity between s_u and z_v or that between t_v and z_u can be measured by similarity metrics, such as inner product.

Suppose that $\mathcal{H}$ is the self-role encoder, defined as $z_u = \mathcal{H}(z_u^0)$, where z_u^0 is the raw representation vector of a node u, and z_u is its self-role embedding. $\mathcal{H}$ is implemented by a multilayer perceptron (MLP) with a nonlinear activation layer as follows:

$$\mathcal{H}\left(z_u^0\right) = \sigma\left(W_2\left(\sigma\left(W_1 z_u^0 + b_1\right)\right) + b_2\right), \quad (3)$$

where σ is the exponential linear unit [28], W_1, W_2, b_1 and b_2, are weights and bias items. The MLP is shared by all the training nodes.

It is worth noting that self-role embeddings cannot be directly compared with source role embeddings or target role embeddings. The underlying reason is that self-role embeddings are obtained by the MLP, while source/target role embeddings are obtained by the GAT. In other words,

they are from different embedding spaces. Therefore, an additional alignment of embedding space is needed. See also Section Predictor and loss function for details.

Aggressiveness and impressionability

The topology and dynamics of a network indicate that a node having more connections tends to connect more nodes [29]. Inspired by this observation, we believe that the drug having more interactions tends to interact with more drugs according to the DDI network. Thus, two drugs even having similar features (e.g. chemical structure) but having significantly different numbers of interactions would have remarkably varied tendencies to interact with other drugs.

In this context, we name the aggressiveness for source roles and the impressionability for target roles. These two items capture how the number of interaction partners of a drug affects its interaction tendency. For example, the greater the in-degree of node u, the greater its impressionability to other nodes. The greater the out-degree of node u, the greater its aggressiveness.

Considering that the aggregation of GAT reflects the in-degree/out-degree of a node u, we adopt two extra scalar items, m_u^s and m_u^t (Fig. 2). The former captures its aggressiveness when calculating its source-role embeddings, while the latter captures its impressionability when calculating its target-role embeddings. To obtain the two items, we extend the output dimension into $(d+1)$ during aggregating the information of outgoing/incoming neighbors for source/target roles. The first d dimensions represent the source/target roles and the last dimension represents the aggressiveness/impressionability. Thus, $W_r \in \mathbb{R}^{(d+1)\times p}$ and $\overrightarrow{a}_r \in \mathbb{R}^{2(d+1)}$ in Eqs (1) and (2). Especially, to avoid the confusion of variable names, we rename Eq. (2) as follows:

$$s_u^* = \sum_{v \in \mathcal{N}_s(u)} \alpha_{uv}^s W_s z_v^0, s_u = s_u^*[:-1], m_u^s = s_u^*[-1], \quad (4)$$

$$t_u^* = \sum_{v \in \mathcal{N}_t(u)} \alpha_{uv}^t W_t z_v^0, t_u = t_u^*[:-1], m_u^t = t_u^*[-1] \quad (5)$$

where $s_u^*, t_u^* \in \mathbb{R}^{(d+1)}$, $[:-1]$ denotes the operation using the elements in the vector except for the last one, and $[-1]$ denotes the operation using the last element.

Predictor and loss function

For a given drug pair (u,v), we discriminate whether u influences v ($u \to v$) by four items, including the source role of u aligned with the self-role of v, the target role of v aligned with the self-role of u, the aggressiveness of u and the impressionability of v. As shown in Fig. 2, the asymmetry of the interaction $u \to v$ is characterized by the source (perpetrator) view of u and the target (victim) view of v jointly. Each view accounts for an assumption.

First, we hold the source-view-specific assumption that the source role embedding s_u and the self-role embedding z_v are close in the embedding space if u influences v. Naturally, by the inner product between

s_u and z_v, we can measure how likely such an interaction $u \to v$ is. As illustrated in Fig. 2, s_u is the embedding of u generated by the source GAT, which characterizes u by aggregating its outgoing neighbors (3 nodes highlighted by light yellow, except for v) but not incoming neighbors (2 white nodes). In contrast, z_v is obtained through the MLP. Because s_u and z_v are not in the same embedding space, their distance cannot be measured directly. To address this issue, we make a role alignment W_s' to map z_v into the space of s_u. For simplicity, we leverage a single-layer neural network as the alignment, denoted as $z_v^* = z_v W_s'^T$, such that the inner product $s_u^T z_v^*$ can be performed. It measures the proximity between u (perpetrator, represented by the source GAT) and v (victim, represented by the MLP) from the source view, where the central node is u. Finally, the inner product between s_u and z_v^* is linearly combined with the aggressiveness m_u^s (indicating the outgoing interaction tendency of u) as the possibility of (u,v) being $u \to v$ from the view of source role (i.e. $\alpha * s_u^T z_v^* + \beta * m_u^s$). Here, two hyper-parameters, α and β, are designed for tuning the tradeoff between the proximity and the interaction tendency.

Meanwhile, we consider the target-view-specific assumption that the target role embedding t_v and the self-role embedding z_u are close in the embedding space if u influences v. Likewise, $t_v^T z_u^*$ measures the proximity between u (perpetrator, represented by MLP) and v (victim, represented by the target GAT) from the target view, where the central node is v. As illustrated in Fig. 2, the target GAT characterizes drug v by aggregating its incoming neighbors (2 nodes highlighted by light red, except for u) but not outgoing neighbors (2 white nodes as well). From the view of the target role, we can calculate the possibility of (u,v) being $u \to v$ by the target role embedding t_v obtained by the target GAT, the aligned self-role embedding $z_u^* = z_u W_t'^T$ and the impressionability m_v^t (i.e. $\alpha * t_v^T z_u^* + \beta * m_v^t$).

So far, we characterize the interaction (e.g. $u \to v$) in an asymmetric manner from each view. To integrate the two views, a sum operation is naturally used. Formally, the two possibilities from two views are averaged as the final possibility of (u,v) being an asymmetric interaction $u \to v$ (denoted as $\hat{y}_{u,v}$). The formal definition of $\hat{y}_{u,v}$ is as follows:

$$\hat{y}_{u,v} = \sigma \left(\alpha * s_u^T z_v^* + \beta * m_u^s + \alpha * t_v^T z_u^* + \beta * m_v^t \right), \quad (6)$$

where $z_v^* = z_v W_s'^T$, $z_u^* = z_u W_t'^T$, s_u is the source role embedding of u, t_v is the target role embedding of v, z^* is the self-role of a node, m_u^s is the aggressiveness of u and m_v^t is the impressionability of v, W_s' and W_t' are the role alignment matrices between the source/target role and the self-role, σ is the Sigmoid function and α and β are the coefficients of the linear combination of four items. Especially, α and β are the hyper-parameters

which control the weights of node proximity and aggressiveness/impressionability. Moreover, we enforce $\alpha + \beta = 0.5$ to meet the numerical constraint of probability.

We determine the drug pair (u, v) as an asymmetric interaction $u \rightarrow v$ if $\hat{y}_{u,v} \geq 0.5$, otherwise no such an asymmetric interaction. This formula is directly used in Task 1, while both $\hat{y}_{u,v}$ and $\hat{y}_{v,u}$ are considered simultaneously in Task 2. In details, for a given unlabeled pair of u and v, the trained model performed the prediction twice in Task 2. One prediction determines how possible the interaction is from u to v, while another determines how possible the interaction is from v to u. Two predicting scores $\hat{y}_{u,v}$ and $\hat{y}_{v,u}$ (being a direction-specific interaction) were jointly to give the final decision. According to the classification threshold 0.5, four combinations of the two scores are ($\hat{y}_{u,v} \geq 0.5$ and $\hat{y}_{v,u} \geq 0.5$) denoting $u \rightarrow v$ and $v \rightarrow u$, ($\hat{y}_{u,v} \geq 0.5$ and $\hat{y}_{v,u} < 0.5$) denoting $u \rightarrow v$, ($\hat{y}_{u,v} < 0.5$ and $\hat{y}_{v,u} \geq 0.5$) denoting $v \rightarrow u$ and ($\hat{y}_{u,v} < 0.5$ and $\hat{y}_{v,u} < 0.5$) denoting a noninteraction. Since there is no bi-directional interaction among asymmetric DDI entries, we cannot find the first combination of scores. The remaining three were used to indicate the final prediction for a given drug pair. In such a manner, we determine the drug pair (u, v) as an asymmetric interaction $u \rightarrow v$ in the direction-blind predicting task.

Finally, the whole DGAT-DDI model is trained in an end-to-end manner with the binary cross-entropy loss as follows:

$$\mathcal{L} = \frac{1}{N} \sum_{e_{u,v} \in \mathcal{E}} -y_{u,v} \log(\hat{y}_{u,v}) - (1 - y_{u,v}) \log(1 - \hat{y}_{u,v}), \quad (7)$$

where $\mathcal{E}$ is the edge set, $y_{u,v}$ is the ground truth label, $y_{u,v} = 1$ indicates existing asymmetric interactions or $y_{u,v} = 0$ indicates nonexisting interactions w.r.t. the specific direction.

Experiments
Dataset and setup

We collected asymmetric DDI entries from version 5.17 of DrugBank released on 2 July 2020. The original dataset contains 603 816 asymmetric interactions among 1974 approved small-molecular drugs. After a double-check, we removed some drugs which have incorrect SMILES strings or cannot be represented by Morgan fingerprints. As a result, we obtained 1752 approved drugs and 504 468 asymmetric interactions among them.

Then, we organize these DDI entries into a directed interaction network, where nodes are drugs and directed edges are asymmetric interactions. In terms of network, the maximum in-degree of a node is up to 1289, while the maximum out-degree of a node is 1062. Both the minimum out-degree and the minimum in-degree are 0 since some nodes have only one link. In addition, the average clustering coefficient of the network is 0.346.

For each encoder in our DGAT-DDI, we set the dimensions of input, embedding, output vectors in the following. Since GNN requires the vector representation of each node in the network, we initially represented each drug by Morgan fingerprints. Known as one of the most popular circular fingerprints, it represents a drug as a 1024-bit input vector, of which each bit indicates a specific local structure presenting in the molecule. In the source/target role embedding, we empirically set the dimension of the output embedding to 16. The aggressiveness/impressionability is just a scalar. Moreover, in the self-role embedding, we empirically considered two hidden layers, whose nodes are 64 and 16 in addition to its raw self-role $z_u^0 \in \mathbb{R}^{1024}$ (Morgan fingerprints) in the input layer.

State-of-the-art methods in comparison

To demonstrate the superiority of our DGAT-DDI model, we selected four graph representation learning methods as the baselines, including the standard Graph Autoencoder [19], Source/Target Graph AE [25], Gravity Graph VAE [25] and DGGAN [26]. They are summarized as follows.

- Standard Graph Autoencoder [19] is a kind of unsupervised model extending autoencoder to graph structures. Their goal is to learn a node embedding, i.e. a low dimensional vector space representation of the nodes. Although the standard GAE is designed for the undirected graph, we used it to illustrate how its symmetric representation degrades the prediction performance of asymmetric interactions.
- Source/Target Graph AE [25] builds a GCN based on out-degree normalization to encode drugs, where the odd and even bits of an embedding vector account for the source role and the target role, respectively.
- Gravity Graph VAE [25], an extension of the graph variant autoencoder on the directed graph, learns asymmetric embeddings.
- DGGAN [26], a GAN-based directed graph embedding framework, leveraging the adversarial mechanism to learn each node's source and target embeddings together.

As GNN-based methods, the first three models commonly include a two-layer GCN encoder, which contains a 64-d hidden layer and a 32-d output layer. We adopted the suggested values of model parameters in the original papers [19, 25], such as Adam as the optimizer, the learning rate of 1e−2. For the GAN-based DGGAN, we set the numbers of generator and discriminator training iterations per epoch to 5 and 15, respectively. We also set the learning rate to 1e−4 and the batch size to 128. After tuning, we set the dimension of node embeddings to 128 for the best performance.

When training our DGAT-DDI, we set the dropout rate to 0.6 in the embeddings, the learning rate to 1e−2, the number of epochs per run to 200 and selected the Adam algorithm as the optimizer [30]. Besides, we set

Table 3. Comparison in direction-specific task (percentage)

Methods	AUROC	AUPRC	ACC	F1	PRECISION	RECALL
Standard GAE	72.3 ± 0.4	66.1 ± 0.6	65.5 ± 0.2	73.0 ± 0.1	60.0 ± 0.2	93.3 ± 0.4
Source/Target GAE	80.5 ± 0.3	77.8 ± 0.6	73.0 ± 0.2	75.0 ± 0.6	70.0 ± 0.9	81.1 ± 2.5
Gravity GVAE	80.2 ± 1.4	75.2 ± 1.8	72.6 ± 1.1	76.8 ± 0.5	66.8 ± 1.4	90.4 ± 1.2
DGGAN	83.1 ± 0.2	80.6 ± 0.2	60.9 ± 2.7	71.8 ± 1.3	56.2 ± 1.8	99.2 ± 0.6
DGAT-DDI(ours)	95.1 ± 0.0	94.3 ± 0.0	88.6 ± 0.2	88.4 ± 0.1	86.9 ± 0.5	90.5 ± 0.9

the hyperparameter β by tuning its value from the list {0.00, 0.05, ..., 0.45} with the interval of 0.05. See also Section Comparison results for the details.

Five-fold cross-validation was adopted to evaluate the above models in the prediction of asymmetric DDIs. Specifically, the whole dataset was randomly split into a training set (containing 60% DDIs), a validation set (containing 20% DDIs) and a testing set (containing 20% DDIs). We used the training set to train the models, the validation set to tune them and the testing set to evaluate the generalization ability of the well-trained models. Such a random split was repeated 20 times. All the methods run on the same splits of the dataset. The performance of these models was reported by an average of 20 predictions.

The prediction results were measured by 6 popular metrics, including the 'Area Under the Receiver Operating Characteristic' (AUROC), the 'Area Under the Precision Recall Curve' (AUPRC), the 'Accuracy' (ACC), F1-score, Precision and Recall.

Comparison results

To generate negative samples, we adopted the sampling strategy suggested in [22]. Considering the asymmetry of DDI, we named a new term, nonexisting interactions between drugs, as negative samples. Here, a nonexisting asymmetric interaction accounts for the pair of two non-interacting drugs or the directional inverse of a direction-specific DDI (i.e. a fake interaction). For example, there exists an asymmetric interaction (u, v) between node u and node v (i.e. $u \rightarrow v$), and there is no interaction between u and another node w. Both the noninteracting pair (u, w) and the fake interaction (v, u) are considered as negative samples, while (u, v) is the positive sample. After directly taking all true asymmetric interactions as positive samples, we randomly selected the same number of negative samples among non-existent directed interactions.

Based on the sampling strategy, we evaluated the performance of the above models in two tasks, including a direction-specific task (Table 3) and a direction-blinded task (Table 4).

(1) **Direction-specific task**. This task discriminates how possibly or whether a direction-specific interaction between two nodes occurs, as illustrated in (Fig. 1A).

Overall, the comparison in Task 1 demonstrates the significant superiority of our DGAT-DDI in terms of six metrics. Specifically, except for Recall, DGAT-DDI achieves 12 ~ 30% improvements over AUROC, AUPRC, ACC, F1 and Precision. Although Standard GAE and DGGAN exhibit better recall values (93.3 and 99.2%, respectively) than that of DGAT-DDI (90.5%), their precision values are only ~60%, which are much smaller than that of DGAT-DDI (86.9%). The high recall shows that a method can find almost all the positive samples, while the low precision indicates that it predicts many false positive cases. In other words, Standard GAE and DGGAN tend to discriminate negative samples (noninteractive pairs and fake interactions) as positive samples (direction-specific interactions) in Task 1. Thus, we pay more attention to F1, which is a more appropriate metric by balancing Precision and Recall. Compared with these two methods, DGAT-DDI exhibits 15.4 and 16.6% improvements w.r.t. F1.

Moreover, we noted an interesting phenomenon, where Standard GAE works much better than a random guess. In our original thought, Standard GAE would just work like a random guess because it ignores the direction information. This finding pushed us to dig out the underlying reason. After checking, we found that the task of direction-specific prediction degrades as a traditional binary prediction task if the number of non-interactive pairs is much more than that of fake interactions. In such a circumstance, direction information in Task 1 is ignored because it is not easy to sample fake interactions among many noninteracting pairs. This is why Standard GAE works beyond a random guess. Therefore, to address this issue, we evaluated these models on Task 2, which emphasizes the importance of direction asymmetry among DDIs.

(2) **Direction-blind predicting task**. This task determines how possible the interaction between u and v is $u \rightarrow v$, $v \rightarrow u$, bidirectional interaction or a noninteraction (Fig. 1B). Compared with the first task, Task 2 is more important but more challenging since it requires the reconstruction of interaction asymmetry.

To reflect the importance of direction asymmetry, we adopted the same training as that in Task 1 but a different testing manner as suggested in [22]. For a given unlabeled pair of u and v, the trained model performed the prediction twice. One prediction determines how possible the interaction is from u to v, while another determines how possible the interaction is from v to u. Two predicting scores $\hat{y}_{u,v}$ and $\hat{y}_{v,u}$ (being a direction-specific interaction) were jointly to give the final decision. According to the classification threshold 0.5, four combinations of the two

Table 4. Comparison in direction-blind predicting task (percentage)

Methods	AUROC	AUPRC	ACC	F1	PRECISION	RECALL
Standard GAE	50.0 ± 0.0	50.0 ± 0.0	50.0 ± 0.0	65.1 ± 0.1	50.0 ± 0.0	93.2 ± 0.3
Source/Target GAE	70.5 ± 0.5	72.3 ± 0.6	60.6 ± 0.9	67.8 ± 0.2	57.3 ± 1.0	83.2 ± 1.9
Gravity GVAE	55.4 ± 0.6	53.2 ± 0.6	52.9 ± 0.3	65.7 ± 0.4	51.7 ± 0.2	90.1 ± 1.4
DGGAN	75.3 ± 0.6	73.2 ± 0.5	63.2 ± 0.7	69.8 ± 0.5	60.1 ± 0.9	87.4 ± 1.3
DGAT-DDI (ours)	86.7 ± 0.1	85.4 ± 0.1	79.5 ± 0.1	77.1 ± 0.2	71.9 ± 0.5	89.1 ± 0.7

scores are ($\hat{y}_{u,v} \geq 0.5$ and $\hat{y}_{v,u} \geq 0.5$) denoting $u \rightarrow v$ and $v \rightarrow u$, ($\hat{y}_{u,v} \geq 0.5$ and $\hat{y}_{v,u} < 0.5$) denoting $u \rightarrow v$, ($\hat{y}_{u,v} < 0.5$ and $\hat{y}_{v,u} \geq 0.5$) denoting $v \rightarrow u$ and ($\hat{y}_{u,v} < 0.5$ and $\hat{y}_{v,u} < 0.5$) denoting a noninteraction. Since there is no bi-directional interaction among asymmetric DDI entries, we cannot find the first combination of scores. The remaining three were used to indicate the final prediction for a given drug pair. In such a manner, we accomplished the direction-blind predicting task.

Again, the comparison with other methods in Task 2 demonstrates the significant superiority of our DGAT-DDI overall. Specifically, except for Recall, DGAT-DDI achieves $7 \sim 26\%$ improvements over AUROC, AUPRC, ACC, F1 and Precision. Although Standard GAE and Gravity GVAE exhibit better recall values (93.2 and 90.1%, respectively) than that of DGAT-DDI (89.1%), their precision values are only $\sim50\%$ (just like a random guess on interactions), which are much smaller than that of DGAT-DDI (71.9%).

In summary, the comparisons in the two tasks demonstrate the excellent ability of our DGAT-DDI to represent asymmetric interactions.

(3) **Direction-free predicting task**. This task determines how possible the interaction between u and v is a direction-free interaction or a noninteraction. Compared with the above two tasks, the task is just to fulfill a comprehensive comparison with advanced deep learning models (i.e. DeepDDI [12], DDIMDL [14] and KGNN [21]) for DDI prediction. Since these models were originally designed for symmetric interactions but not for asymmetric interactions, we modified our model to accommodate symmetric interaction prediction by treating symmetric interactions as bidirectional edges. The results show that our DGATDDI outperforms both DeepDDI and DDIMDL (Table 5). In addition, although KGNN (achieving <1% improvement) is slightly better than DGATDDI, it needs extra drug-related entries (e.g. enzyme, target, gene, etc., except for chemical structures) and abundant associations between entries to build a KG. In many cases, these rich entries may not be obtained easily. In contrast, our DGATDDI only needs basic chemical structures. Thus, our DGATDDI can be a competitive model for symmetric DDI prediction.

(4) **Cold-start predicting task**. The task evaluates how well DGATDDI is under a more challenging experimental setting (i.e. cold-start scenario), where the testing drugs have no overlap with the training drugs. In this cold-start scenario, we tried to predict the interaction between

Table 5. Comparison in direction-free predicting task (percentage)

Methods	AUROC	AUPRC	ACC
DeepDDI	93.2	92.5	85.6
DDIMDL	94.9	95.1	87.5
KGNN	99.1	98.9	94.6
DGAT-DDI(ours)	98.3	98.1	93.6

newly coming drugs with the known drugs in the network. Because newly coming drugs have no known interaction (no neighbor), we use only the single-side role embedding (i.e. $s_u^T z_v^*$ or $t_v^T z_u^*$ instead of $s_u^T z_v^* + t_v^T z_u^*$ in the case of $u \rightarrow v$). In other words, we can obtain all three roles of a known drug but obtain only the self-role of a new drug because it has no neighbor in the network. To keep the consistency with the ordinary scenario, DGATDDI was run in Task 1 and Task 2 accordingly. The comparison between different experiment settings illustrates $6 \sim 15\%$ declines across the measuring metrics (Fig. 7). These results demonstrate that the cold-start scenario is more difficult than the ordinary scenario. An elaborate model to handle the cold-start issue is expected.

Parameter tuning and ablation study
In this section, we first investigated how the hyper-parameter in DGAT-DDI fluence its performance based on the validation set. There is only one hyper-parameter β that reflects the trade-off between the source-role embedding and the aggressiveness as well as the trade-off between the target-role embedding and the impressionability. We tuned the value of β from the list $\{0.00, 0.05, \ldots, 0.45\}$ with the interval of 0.05. The results on the two tasks measured by AUROC and AUPRC show the increment of prediction performance when increasing β and the decrement when increasing it further (Fig. 4). Specifically, the peak is located on $\beta=0.2$ in Task 1 (Fig. 4A), while it is located on $\beta=0.15$ in Task 2 (Fig. 4B). The values were adopted by DGAT-DDI to run all experiments.

Furthermore, we investigated how well each of the major components in DGAT-DDI contributes to the prediction by an ablation study. To this purpose, we made four DGAT-DDI variants, denoted as w/o AI, w/o SR, w/o TS and w/o RA, of which each removes one component, respectively. Compared with the full model of DGAT-DDI, the variant w/o AI removes both the aggressiveness
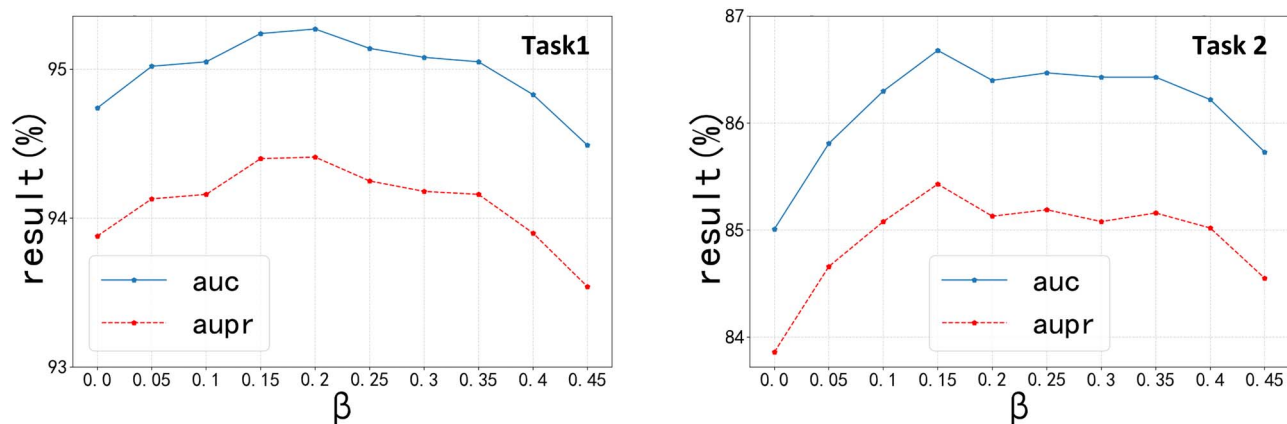
Figure 4. Hyper-parameter tuning in Task 1 and Task 2.

Table 6. Comparison between DGAT-DDI and variants in ablation study

Model	Task1			Task2		
	AUROC	AUPRC	ACC	AUROC	AUPRC	ACC
w/o AI	94.7	93.9	87.6	85.0	83.9	78.2
w/o SR	93.2	92.1	85.8	81.2	80.0	71.4
w/o TS	92.8	91.6	85.2	80.6	79.2	71.0
w/o RA	93.9	92.9	86.3	84.1	83.0	74.1
Full	95.1	94.3	88.6	86.7	85.4	79.5

and the impressionability, w/o SR removes the self-role embedding, w/o TS has no two sides of role embedding but considers only single-side role embedding (i.e. $s_u^T z_v^*$ instead of $s_u^T z_v^* + t_v^T z_u^*$ in the case of $u \rightarrow v$) and w/o RA removes the role alignment from DGAT-DDI.

Overall, the comparison shows that DGAT-DDI outperforms all its variants across two tasks in terms of AUROC, AUPRC and ACC (Table 6). The results demonstrate that each of the major components in DGAT-DDI contributes to the prediction. Especially, their contributions in Task 2 are greater than those in Task 1 because Task 2 requires more information about interaction asymmetry (direction).

Among them, w/o TS accounts for the bigger decrement of prediction performance (e.g. 6.1~8.5% decline in Task 2). This result reveals that both source role embeddings and target role embeddings exhibit the most important contribution. The underlying reason is that they capture how a drug influences other drugs and how it is influenced by others among asymmetric DDIs. Similarly, the performance decline (5.1~8.5%) caused by w/o SR reflects that self-role embeddings are crucial as well because it contains drugs' own properties.

Moreover, 2.4~5.4% decline caused by w/o RA reveals that the role alignment component is essential to represent drugs in asymmetric interactions, because source-/target role embeddings and self-role embeddings are from different spaces.

In addition, ~1.5% decline made by w/o AI reflects that both the aggressiveness and the impressionability are helpful for drug representation. The underlying

reason is that they capture how the number of interaction partners of a drug affects its interaction tendency.

In general, all these components play indispensable roles in representing and predicting asymmetric DDIs.

Case study 1: assumption visualization

To make a clear understanding of our assumptions, we performed a visualization of embedding space. Taking Idarubicin as an example, we leverage t-SNE [31] to illustrate our assumption that the source role embedding s_u and the self-role embedding z_v are close in the embedding space if u influences v. Idarubicin has 382 interactions, of which 257 are outgoing interactions and 125 are incoming interactions. As shown in Fig. 5, the source role of Idarubicin (marked by a green triangle), the source-aligned self-role of its outgoing neighbors (victims, marked green dots) and the source-aligned self-role of 300 randomly sampled drugs not interacting with Idarubicin (nonvictims, marked by blue dots) are shown. In terms of source role, we found two separate clusters, where Idarubicin is in the cluster of outgoing neighbors, but is far away from the cluster of noninteracting drugs. Similar results show that Idarubicin (marked by a red square) is near to its incoming neighbors (perpetrators, marked red dots) but far away from its noninteracting drugs (nonperpetrators, marked by light red dots) in terms of target role. In brief, the case visualization demonstrates that DGAT-DDI models the assumptions.

Moreover, we attempted to illustrate the assumption that the number of interaction partners of a drug affects its interaction tendency. We attempted to illustrate the
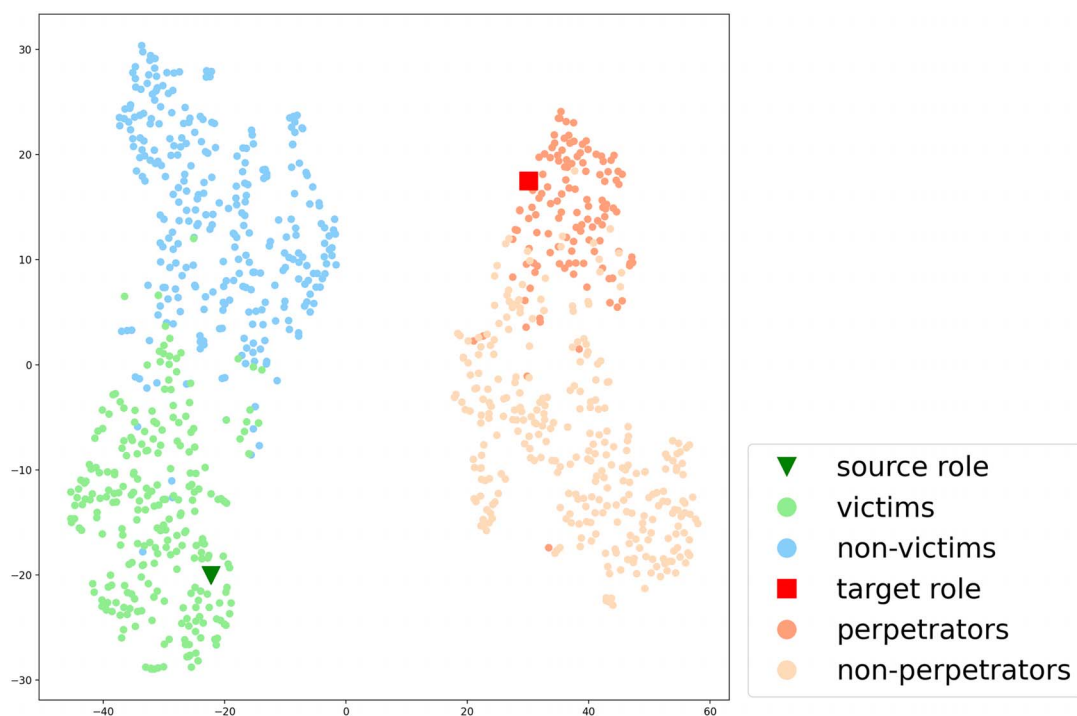
Figure 5. Visualization of Interaction Prediction. In our model, the source role of a drug and its outgoing neighbor, as well as the target role of a drug and its incoming neighbor are close in the latent space. Self-role of non-edge drug will stay away from the source role and the target role.
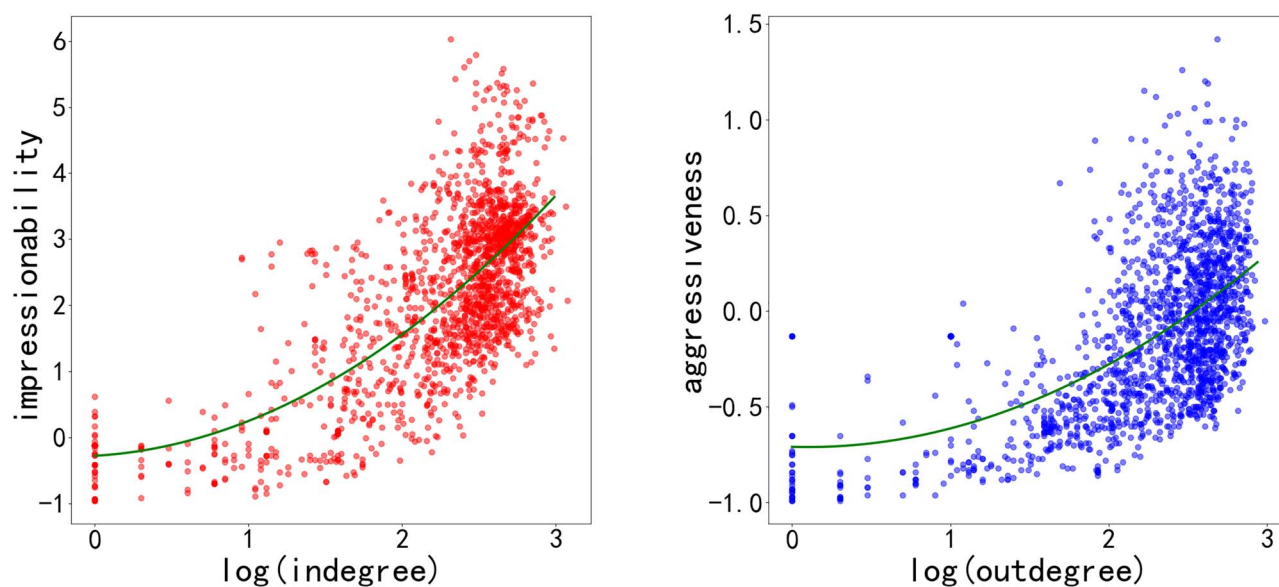

Figure 6. Interaction tendency. (**A**) Impressiveness versus in-degree, (**B**) aggressiveness versus out-degree. The number of interaction partners of a drug affects its interaction tendency. Based on the observed moderate Spearman correlation, the larger the node degree, the larger the impressiveness/aggressiveness.

potential association between the number of interaction partners and DDI tendency indicated by out-degree and in-degree (shown in the left panel and the right panel in Fig. 6, respectively). The results show a moderate correlation in both cases. In detail, the Spearman correlation coefficient between the aggressiveness and the out-degree is 0.55 (with the P-value $=1e^{-139}$), while that between the impressiveness and the in-degree is 0.67 (with the P-value $= 7e^{-232}$). The illustration supports that the aggressiveness/impressionability in DGAT-DDI

captures the information that how the number of outgoing/incoming interactions of a drug affects its interaction tendency.

Case study 2: novel prediction

In this section, we investigated the ability of DGAT-DDI in finding unobserved DDIs. Considering the update of DrugBank annotates more unobserved interactions, we performed a version-independent validation to achieve our investigation as follows.
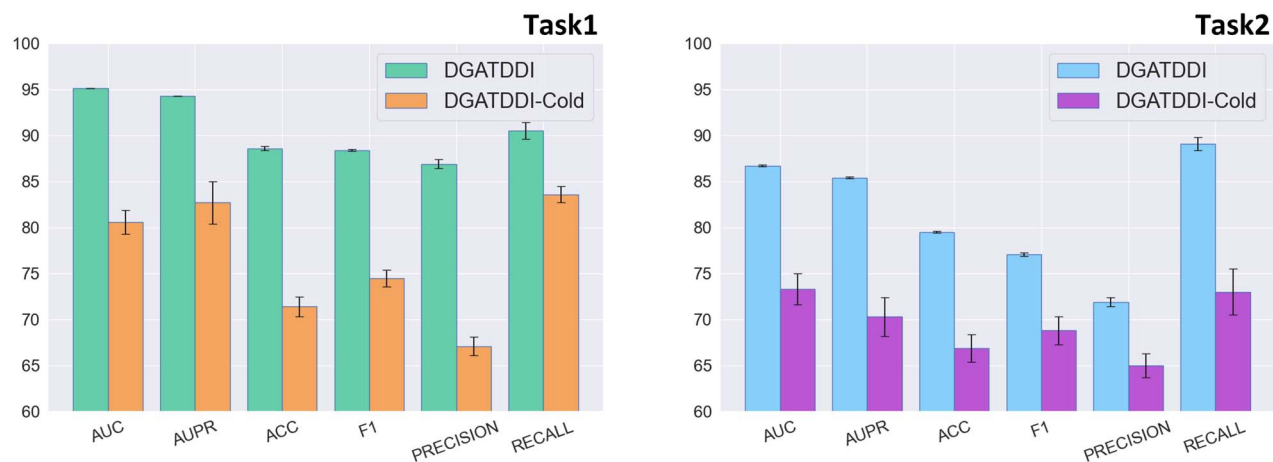
Figure 7. Comparison with cold-start scenario.

Table 7. Top 10 asymmetric DDI candidates by DGAT-DDI. Interaction indicates whether the interaction between the two drugs is predicted correctly, and the direction indicates whether the direction of asymmetric interaction between the two drugs is predicted correctly

Rank	Perpetrator drug	Victim drug	Description	Interaction	Direction
1	Flunarizine	Darifenacin	The metabolism of Flunarizine can be decreased when combined with Darifenacin.	✓	-
2	Dapagliflozin	Ephedrine	The risk or severity of Cardiac Arrhythmia can be increased when Ephedrine is combined with Dapagliflozin.	✓	✓
3	Clomipramine	Sodium aurothiomalate	Clomipramine may decrease the excretion rate of Sodium aurothiomalate which could result in a higher serum level.	✓	✓
4	Capsaicin	Hexafluronium	The therapeutic efficacy of Hexafluronium can be decreased when used in combination with Capsaicin.	✓	✓
5	Paroxetine	Cephalexin	NA	-	-
6	Hydrocortisone	Glasdegib	The metabolism of Glasdegib can be increased when combined with Hydrocortisone.	✓	✓
7	Alogliptin	Nylidrin	The risk or severity of hypoglycemia can be increased when Nylidrin is combined with Alogliptin.	✓	✓
8	Nafcillin	Olaparib	The metabolism of Olaparib can be increased when combined with Nafcillin.	✓	✓
9	Lomitapide	Quinine	The metabolism of Lomitapide can be increased when combined with Quinine.	✓	-
10	Methscopolamine	Fenoterol	The risk or severity of adverse effects can be increased when Fenoterol is combined with Methscopolamine.	✓	✓

First, we run a transductive prediction on the dataset, collected from version 5.17 of DrugBank (released on 2 July 2020). The transductive prediction inferred potential DDIs among unlabeled drug pairs. Then, after sorting these drug pairs according to their predicting scores, we picked up the top-10 asymmetric DDI candidates. For unlabeled drug pairs, the higher predicting scores, the higher probabilities to be interactions. Last, we validated the candidates according to their labels provided by version 5.18 of DrugBank (released on 3 January 2021) since we suppose the advanced version is more accurate.

Overall, the validation shows that a significant fraction of novel predicted asymmetric DDIs (7 out of 10) is confirmed (see Table 7). In detail, seven asymmetric interactions are predicted correctly, two interactions are predicted correctly but with opposite directions (the first and the ninth) and one interaction is predicted wrongly (the fifth). In summary, this investigation demonstrates the inspiring ability of DGAT-DDI on predicting asymmetric DDIs in practice.

Conclusions

This paper has proposed a novel architecture of DGAT-DDI for predicting asymmetric interactions between drugs. DGAT-DDI characterizes asymmetric DDIs by source-role embeddings, target-role embeddings and self-role embeddings. Moreover, it considers how the number of interaction partners of a drug affects its interaction tendency by the aggressiveness of the source role and the impressionability of the target role. Furthermore, it discriminates potential asymmetric interactions based on two assumptions. The first is that the source role of a drug is close to the self-role of its outgoing neighbors. The second is that the target role of a drug is close to the self-role of its incoming neighbors in the embedding space.

In the experiments, the significant superiority of DGAT-DDI is demonstrated in the comparison with four state-of-the-art approaches under both a direction-specific predicting task and a direction-blinded task. Moreover, how well each component of DGAT-DDI contributes to its ability is revealed by the ablation study. Finally, its practical ability to predict novel asymmetric interactions is demonstrated by the case study, where 7 candidates are validated by the lasted release of DrugBank.

In the coming future, DGAT-DDI will be extended to accommodate multitype asymmetric interactions and also be improved to predict asymmetric interactions for newly coming drugs (i.e. the cold-start scenario).

Code and data availability

Source codes are freely available at https://github.com/F-windyy/DGATDDI.

Key Points

- DGAT-DDI generates two asymmetric embeddings of the source role and the target role for a drug, respectively. Its source role indicates how it influences other drugs in DDIs. Its target role represents how it is influenced by others. Its self-role is aligned into the source role space and the target role space, respectively, to reflect the proximity of the drug pair being an asymmetric interaction.
- Moreover, DGAT-DDI learns the aggressiveness of the source role and the impressionability of the target role to reflect how the number of interaction partners of a drug affects its interaction tendency.
- To the best of our knowledge, DGAT-DDI is the first approach for predicting asymmetric interactions among drugs. Its superiority is demonstrated by a direction-specific predicting task, a direction-blinded task as well as a case study of novel asymmetric DDI prediction.

Acknowledgements

Funding

References

1. Han K, Jeng EE, Hess GT, *et al.* Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol* 2017;**35**(5):463–74.
2. Prueksaritanont T, Chu X, Gibson C, *et al.* Drug–drug interaction studies: regulatory guidance and an industry perspective. *AAPS J* 2013;**15**(3):629–45.
3. Tatonetti NP, Ye PP, Daneshjou R, *et al.* Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012;**4**(125):125ra31.
4. Wicha SG, Chen C, Clewe O, *et al.* A general pharmacodynamic interaction model identifies perpetrators and victims in drug interactions. *Nat Commun* 2017;**8**(1):2129.
5. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2017;**46**(D1):D1074–82.
6. Razek A, Vietti T, Valeriote F. Optimum time sequence for the administration of vincristine and cyclophosphamide in vivo. *Cancer Res* 1974;**34**(8):1857–61.
7. Koizumi W, Kurihara M, Hasegawa K, *et al.* Sequence-dependence of cisplatin and 5-fluorouracil in advanced and recurrent gastric cancer. *Oncol Rep* 2004;**12**:557–61.
8. Bahcall M, Kuang Y, Paweletz CP, *et al.* Abstract 4100: Mechanisms of resistance to type I and type II MET inhibitors in non-small cell lung cancer. *Cancer Res* 2017;**77**(13 Supplement):4100–0.
9. Batra A, Roemhild R, Rousseau E, *et al.* High potency of sequential therapy with only β-lactam antibiotics. *Elife* 2021;**10**:e68876.
10. Liou J-M, Chen CC, Fang YJ, *et al.* 14 day sequential therapy versus 10 day bismuth quadruple therapy containing high-dose esomeprazole in the first-line and second-line treatment of Helicobacter pylori: a multicentre, non-inferiority, randomized trial. *J Antimicrob Chemother* 2018;**73**(9):2510–8.
11. Chen Q, Ding F, Zhang S, *et al.* Sequential therapy of acute kidney injury with a DNA nanodevice. *Nano Lett* 2021;**21**:4394–402.
12. Ryu Jae Y, Hyun UK, Sang YL. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc Natl Acad Sci* 2018;**115**(18):E4304–11.
13. Nyamabo A, Yu H, Shi J-Y. SSI–DDI: substructure-substructure interactions for drug-drug interaction prediction. *Brief Bioinform* 2021;**22**(6):bbab133.
14. Deng Y, Xu X, Qiu Y, *et al.* A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* 2020;**36**(15):4316–22.
15. Chen X, Liu X, Wu J. *Drug-drug interaction prediction with graph representation learning*. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Piscataway, NJ: IEEE, 2019. p. 354–61.
16. Wang Y, Min Y, Chen X, *et al.* Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: *Proceedings of the Web Conference 2021*. Ljubljana, Slovenia: Association for Computing Machinery, 2021, 2921–33.
17. Feng YH, Zhang SW, Shi JY. DPDDI: a deep predictor for drug-drug interactions. *BMC Bioinformatics* 2020;**21**(1):419.
18. Kipf TN, Welling M. *Semi-supervised classification with graph convolutional networks*. In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*, OpenReview.net.
19. Kipf TN, Welling M. Variational graph auto-encoders. *NIPS Workshop on Bayesian Deep Learning* 2016.
20. Yu H, Dong W, Shi J. RANEDDI: Relation-aware network embedding for drug-drug interaction prediction. *Inform Sci* 2022;**582**:167–80.
21. Lin X, Quan Z, Wang Z-J, *et al.* KGNN: knowledge graph neural network for drug-drug interaction prediction. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*. San Francisco, CA: ijcai.org. 2020. p. 2739–45.
22. Zhou C, Liu Y, Liu X, *et al.* Scalable graph embedding for asymmetric proximity. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park, CA: AAAI, 2017.

23. Khosla M, Leonhardt J, Nejdl W, *et al. Node representation learning for directed graphs*. In: *Machine Learning and Knowledge Discovery in Databases*. Cham: Springer International Publishing, 2020.

24. Ghorbanzadeh H, Sheikhahmadi A, Jalili M, *et al.* A hybrid method of link prediction in directed graphs. *Expert Syst Appl* 2021;**165**:113896.

25. Salha G, Limnios S, Hennequin R, *et al. Gravity-inspired graph autoencoders for directed link prediction*. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing, China: Association for Computing Machinery, 2019, 589–98.

26. Zhu S, Li J, Peng H, *et al.* Adversarial directed graph embedding. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park, CA: AAAI, 2021, Vol. **35**(5). pp. 4741–8.

27. Veličković P, Cucurull G, Casanova A, *et al. Graph attention networks*. In: *International Conference on Learning Representations*. Vancouver, BC: OpenReview.net. 2018.

28. Clevert D-A, Unterthiner T, Hochreiter S. *Fast and accurate deep network learning by exponential linear units (ELUs)*. In: *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings*.

29. Barabási A-L, Albert RJS. Emergence of scaling in random networks. *Science* 1999;**286**(5439):509–12.

30. Kingma DP, Ba J. *Adam: a method for stochastic optimization*. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*.

31. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**(86):2579–605.