Graph convolutional network approach to discovering disease-related circRNA-miRNA-mRNA axes

Graph convolutional network approach to discovering disease-related circRNA-miRNA-mRNA axes

Chengxin He[a], Lei Duan[a,*], Huiru Zheng[b], Jesse Li-Ling[c], Linlin Song[d,e] and Longhai Li[a]

[a]School of Computer Science, Sichuan University, Chengdu 610065, China
[b]School of Computing, Ulster University, Northern Ireland, United Kingdom
[c]State Laboratory of Biotherapy, Sichuan University, Chengdu 610041, China
[d]Department of Ultrasound, West China Hospital of Sichuan University, Chengdu 610041, China
[e]Frontiers Science Center for Disease-related Molecular Network, West China Hospital of Sichuan University, Chengdu 610041, China

ARTICLE INFO

Keywords:
disease-related association
circRNA-miRNA-mRNA axis
graph convolutional network

ABSTRACT

Non-coding RNAs are gaining prominence in biology and medicine, as they play major roles in cellular homeostasis among which the circRNA-miRNA-mRNA axes are involved in a series of disease-related pathways, such as apoptosis, cell invasion and metastasis. Recently, many computational methods have been developed for the prediction of the relationship between ncRNAs and diseases, which can alleviate the time-consuming and labor-intensive exploration involved with biological experiments. However, these methods handle ncRNAs separately, ignoring the impact of the interactions among ncRNAs on the diseases. In this paper we present a novel approach to discovering disease-related circRNA-miRNA-mRNA axes from the disease-RNA information network. Our method, using graph convolutional network, learns the characteristic representation of each biological entity by propagating and aggregating local neighbor information based on the global structure of the network. The approach is evaluated using the real-world datasets and the results show that it outperforms other state-of-the-art baselines on most of the metrics.

1. Introduction

At present, gene-targeted therapy for various diseases has been extensively studied, which is a great potential treatment [1, 2]. Therefore, to discover disease-related ribonucleic acids (RNAs, for short), including coding RNAs and non-coding RNAs, as well as their interactions is a crucial task for understanding complex human diseases at the molecular level, which could help researchers to focus further efforts on the most promising biomarkers. In this way, it can facilitate disease diagnosis, monitoring, therapy and prognosis.

Meanwhile, with technological breakthroughs at multiple levels of biology, multifarious convenient methods have been developed for the discovery of disease biomarkers. The past few decades have witnessed a rapid growth of researches on the relevance of RNAs and human disease pathogenesis, among which non-coding RNAs (ncRNAs) have attracted much attention for their crucial role in translational regulation and involvement in several biological processes [3–5].

The straightforward way to identify new disease-related ncRNAs is through biology or high-throughput experiments, but such experiments are expensive as well as time-consuming, and the number of discovered associations is still small. Fortunately, as researchers have strived to develop various databases to produce ever-increasing amounts of biomedical data, which has provided a better opportunity to explore the relationships between ncRNAs and human diseases by computational methods. Consequentially, comprehensive and time-efficient methods are increasingly needed as more and more biological data have become available.

*Corresponding author. Email address: leiduan@scu.edu.cn

Recently, many computational approaches have been developed for predicting the interactions between ncRNAs and diseases. The results may be validated by further biology experiments, which narrows the search space of experiment for researchers and provides them with decision support from the data level. Generally, most of the current studies on the ncRNAs-diseases association prediction can be divided into three main categories: disease-related miRNA [6–17], lncRNA [18–27], and circRNA [28–38].

However, there are still limitations to the aforementioned studies. Most methods handle ncRNAs separately, that is, only focused on exploring the relationship between diseases and a certain type of ncRNA, ignoring the regulatory relationships among different ncRNAs and the influence of this inherent interaction on disease. Existing evidence indicated that associations between divers ncRNAs play a significant role in transcriptional control, which involves a series of disease-related pathways such as apoptosis, vascularization, invasion and metastasis [39]. Some of in-depth studies revealed that some circRNAs may regulate miRNA function like “miRNA sponge”, which can decrease the cytoplasmic levels of target miRNAs by absorbing miRNAs and thus liberate mRNA transcripts that are targeted by the miRNAs [40]. For example, circTP63 can sponge miR-873-3p and upregulates CENPA and CENPB which may facilitate cell cycle progression [41]. Another research found hsa_circ_0091570 can sponge miR-1307 to regulate ISM1 expression which is associated with hepatocellular cancer (HCC) progression [42]. These all suggested that the circRNA-miRNA-mRNA axes play a crucial role in cancer-related or other disease-related pathways.

So far, there are few computational methods for identi-

fying the circRNA-miRNA-mRNA axes related to diseases, which motivates us to develop a novel approach to infer such associations.

First of all, there are three questions which need to be carefully considered for exploring the circRNA-miRNA-mRNA axes associated with diseases:

- *How to handle with the imbalance issues?* Among the existing bioinformatic data resources, the number of relationships between mRNA (i.e., gene) and disease have far exceeded the number of associations between disease and miRNA or circRNA.
- *How to reasonably capture the internal associations?* The relationships between circRNA, miRNA, mRNA and disease have derived from different data sources. It is necessary to organize these scattered relations and infer the circRNA-miRNA-mRNA axes related to the disease.
- *How to distinguish such relationships with heterogeneous strengths?* These link formation from different sources could result in various connection strengths (e.g., strong and weak connections are mixed together). Considering these relations equally could lead to the indistinguishability of the predicted results.

To address the above challenges, we proposed a novel approach, entitled *DRAMA* (short for disease-related circRNA-miRNA-mRNA axes), for detecting disease-related circRNA-miRNA-mRNA axes [43]. The characteristics of *DRAMA* include: (1) it explores the potential information among disease, mRNA, miRNA and circRNA through graph convolution; (2) it captures the inherent associations between various biological entities by training towards the path of "circRNA-miRNA-mRNA-disease" as the goal; and (3) it designs a measurement to infer the set of circRNA-miRNA-mRNA axes related to a given disease. Furthermore, we have utilized additional annotation information to classify and tag the predicted axes to further analyze the effectiveness of *DRAMA*.

The main contributions of our work include:

- We proposed a novel approach that is to explore the circRNA-miRNA-mRNA axis related to diseases instead of just considering the relationship between the disease and a single entity.
- We designed a framework, *DRAMA*, to acquire the features representation of disease, mRNA, miRNA and circRNA based on the graph convolutional network, and used a reasonable measurement to infer the relationship between them.
- We evaluated *DRAMA* on real-world datasets, and compared it with other state-of-the-art baselines to demonstrate its effectiveness.

A preliminary version of this work has been published in the proceedings of BIBM 2020 conference [43]. Compared with our previous work, we made several major improvements. Firstly, we enrich the related work of disease and ncRNA association prediction and further summarize it. Secondly, we add the work of classifying and tagging the axes predicted by *DRAMA*, which could assist researchers in follow-up biological experiment analysis. At the same time, the classification and tagged results of the prediction axes can further demonstrate the effectiveness of *DRAMA*. Finally, we conduct more extensive empirical evaluations: (1) comparing more baselines in effectiveness experiment; (2) adding detailed analysis of *DRAMA* variants; (3) providing an extra case study; and (4) testing parameter sensitivity.

The rest of the paper is organized as follows. The related work is reviewed in Section 2 and Section 3 presents the mathematical formalization of the problem. Section 4 depicts the overall framework in detail, followed by experimental settings and results discussed in Section 5. The paper is concluded by summary and future work in Section 6.

2. Related Work

2.1. Disease and ncRNA Association Prediction

The use of computational methods to predict biological entities associated with diseases has always been a research hotspot in the field of bioinformatics. In recent years, various computational models have been proposed to predict potential disease-related ncRNAs and also achieved decent performance. These methods mainly focus on targeting three kinds of ncRNA, which are miRNA, lncRNA and circRNA.

To detect novel disease-associated miRNA, from the technical perspective, current methods can be roughly divided into three categories: decomposition-based, random walk-based and machine learning-based. For example, Ding *et al*. [6] decomposed the constructed disease-miRNA-target heterogeneous network into two bipartite graphs, and employed the principle of resource-allocation to infer the relations between disease and miRNA. Chen *et al*. [7] proposed the RLSMDA model to capture the relationship between diseases and miRNAs by decomposing the similarity matrix of disease and miRNA respectively. In addition, in order to make full use of the attributes of the nodes in the constructed network, Xuan *et al*. [8] adopted non-negative matrix factorization to predict disease-related miRNA.

Most prediction approaches rely on the guilt-by-association paradigm. Therefore, the idea based on random walk is widely used in these work. RWRMDA [9] model inferred potential miRNA-disease interactions by implementing random walk on the miRNA-miRNA functional similarity network. Based on this, Xuan *et al*. [10] added the semantic and functional similarity of diseases and Shi *et al*. [11] introduced protein-protein interactions to further identify miRNA-disease associations. Some approaches have also been developed for predicting miRNA-disease associations by random walking on constructing heterogeneous biological networks [12–14].

Multiple machine learning algorithms have been applied to the prediction of miRNA-disease associations, such as agglomerative hierarchical clustering [15], deep auto-encoder neural network [16] and manifold learning [17].

Similarly, computational approaches about inferring

disease-related lncRNA and circRNA also adopted the ideas of these aforementioned methods to implement prediction.

Chen *et al.* [18] and Liu *et al.* [19] ranked candidate disease-lncRNA pairs by integrating the information of known disease-lncRNA associations and lncRNA-related expression profiles, respectively. And Ding *et al.* [20] also proposed a framework, namely TPGLDA, to predict lncRNA-disease associations via lncRNA-disease-gene tripartite graph, which integrated gene-disease associations with lncRNA-disease associations. Moreover, most methods employed single random walk or with restart on diverse constructed networks to identify the interaction between disease and lncRNA, such as RWRlncD [21], KRWRH [22], LncDisAP [23] and the work of Zhang *et al* [24]. There are also some novel prediction methods. For example, Lu *et al*. [25] proposed a method (named SIMCLDA) for predicting potential lncRNA-disease associations based on inductive matrix completion. Cui *et al*. [26] developed a novel model based on bipartite local model with nearest profile-based association inferring for inferring this association. Yuan *et al*. [27] designed a framework to measure the strength of potential lncRNA-disease associations by calculating cluster association scores to evaluate the strength of inner relationships between disease and gene.

To predict circRNA-disease associations, most methods introduced the sequence information of circRNA and the computed Gaussian interaction profile kernel similarity to achieve, such as PWCDA [28], MRLDC [29], SIMCCDA [30] and iCDA-CGR [31]. Besides, Deng *et al.* [32] proposed a method named KATZCPDA based on the KATZ [44] method and the integrations among circRNAs, proteins, and diseases to predict circRNA-disease associations. Zhao *et al.* [33] integrated the bipartite network projection algorithm and KATZ measure for predicting potential circRNA-disease associations. Lei *et al.* [34] also designed a framework named RWRKNN, which integrated the random walk with restart and k-nearest neighbors to predict the associations between circRNAs and diseases. Similar to this method also are DWNCPCDA [35] and RWLR [36]. Many other methods have also been developed for identifying potential circRNA-disease associations by using matrix factorization [37] or decision tree [38].

2.2. Graph Neural Network

Recently, deep neural network models had a great impact on various artificial intelligence and machine learning tasks, such as speech recognition, computer vision (CV), and natural language processing (NLP). The powerful advantage of deep learning is that it can be extended to accommodate graph-level applications (e.g., graph neural network), so that it can extract potential information from complex data, especially irregular graph data.

Graph neural network (GNN), based on the deep learning models, has become increasingly popular in addressing many graph-based applications, including semi-supervised node classification [45], link prediction [46] and recommender systems [47]. Based on GNN to extend the convolution operator to non-Euclidean structured data, resulting in the so-called graph convolutional network (GCN). It can naturally integrate node information and topological structure, have shown their power in learning meaningful representations for graph data. The advantages of GCN provide great potential for extracting latent factors in complex and heterogeneous biomedical data.

3. Preliminaries

In this section, we introduce some basic concepts and formalize the task of discovering disease-related circRNA-miRNA-mRNA axes.

Definition 1 (Disease-RNA Information Network). *A disease-RNA information network is defined as an undirected graph $\mathcal{G} = (V, E, \phi, \psi)$, in which V and E are the sets of entities and links, respectively. Each entity and link are respectively associated with their type mapping functions $\phi : V \to \mathcal{A}$ and $\psi : E \to \mathcal{R}$, where $\mathcal{A}$ refers to the set of biomedical entity types, including disease and disease-related RNA, and $\mathcal{R}$ denotes the set of relations between these entities. Each entity $v \in V$ and link $e \in E$ respectively belongs to an entity type $\phi(v) \in \mathcal{A}$ and a relation $\psi(e) \in \mathcal{R}$, where $\mathcal{A} + \mathcal{R} > 2$.*

Definition 1 indicates that a disease-RNA information network is easy to expand. Considering that our aim is to detect disease-related circRNA-miRNA-mRNA axes, the disease-RNA information network we constructed includes four types of biomedical entities i.e., disease, mRNA, miRNA and circRNA. We use D, G, O, C to denote the set of diseases, mRNAs, miRNAs and circRNAs entities, respectively. As a result, $V = D \cup G \cup O \cup C$ and E is the five relations formed by these four biomedical entity types in the disease-RNA information network we constructed. The detailed information of these five relations and the specific process of building a disease-RNA information network from given data sources will be discussed in Section 4.1.

Definition 2 (Biomedical Entity Feature). *For a biomedical entity $v \in V$, the biological feature information it carries can be represented by a vector and defined as a biomedical entity feature $\mathbb{R}^N$, where N is the dimension of the biomedical entity feature.*

Take the biomedical entity type, mRNA, as an example, its sequence information can be encoded by the "one-hot" way to form the biomedical entity feature of mRNA, and the dimension of the biomedical entity feature is the sequence length. In this work, we use similarity information among the same type biomedical entities as their biomedical entity feature. The specific details will be discussed in Section 4.2.1.

For a disease-RNA information network $\mathcal{G}$, in order to facilitate the organization of the features of all its biomedical entities, we use $X \in \mathbb{R}^{|V| \times N}$ to present the feature matrix corresponding to these biomedical entities where $|V|$ is the number of entities in $\mathcal{G}$.

Definition 3 (Biomedical Entity Embedding). *Given a disease-RNA information network $\mathcal{G}$ and the feature matrix X of entities in the network, the goal of biomedical entity embedding is to project entities in the network into a latent low-dimensional representation space while preserving the network structure and semantic properties. Formally, we aim to develop a mapping function $f : (\mathcal{G}, X) \rightarrow Z$ to obtain an embedding table $Z \in \mathbb{R}^{|V|\times F}$ for all biomedical entities in $\mathcal{G}$ as follows:*

$$Z = [\underbrace{\cdots, z_d, \cdots}_{D-embeddings}, \underbrace{\cdots, z_g, \cdots}_{G-embeddings}, \underbrace{\cdots, z_o, \cdots}_{O-embeddings}, \underbrace{\cdots, z_c, \cdots}_{C-embeddings}] \tag{1}$$

where $z_{(\cdot)}$ is a low-dimensional vector and its size $F(F \ll |V|)$ is the predefined embedding dimension.

The focus of this paper is to explore disease-related circRNA-miRNA-mRNA axes via the constructed disease-RNA information network. Based on the above definitions, we formulate this task as follows:

Let a triple $\langle c, o, g\rangle$ denote a circRNA-miRNA-mRNA axis, where $c \in C$, $o \in O$ and $g \in G$. And we use $\mathcal{T}\langle c, o, g\rangle$ to represent the set of the circRNA-miRNA-mRNA axes. Given a disease $d \in D$, the embedding table Z can be obtained from the disease-RNA information network $\mathcal{G}$ and the input feature matrix X to infer the set of circRNA-miRNA-mRNA axes $\mathcal{T}\langle c, o, g\rangle$ related to d.

4. Methodologies

In this section, we discuss the proposed *DRAMA* framework in detail. Then we present how to classify and tag the axes produced by *DRAMA*.

4.1. Construct Disease-RNA Information Network

For the task of exploring disease-related circRNA-miRNA-mRNA axes, where we have both the attributes of these biomedical entities (i.e., disease, mRNA, miRNA and circRNA) and some of interactive information between them, it is challenging to explicitly capture the disease-related circRNA-miRNA-mRNA axes since not only need to consider the association between the disease and the other three types of biomedical entities but also need to analyze the impact of the interactions between mRNA, miRNA and circRNA on the disease. In other words, a disease-related circRNA-miRNA-mRNA axis is build upon the transitive relation of circRNA-to-miRNA, miRNA-to-mRNA and mRNA-to-disease. To address this challenge, the primary step of *DRAMA* is to construct the disease-RNA information network for clearly organizing the associations between diseases and RNAs.

In this work, we extract data from the following data sources to construct the disease-RNA information network by connecting five relations via shared biomedical entities and removing duplicate entities and relations. The detailed access way of these data sources are listed in Table 1.

Table 1
The detailed access ways of databases used in the disease-RNA information network we constructed.

Data Sources	URL
DisGeNET	https://www.disgenet.org/
HMDD	http://www.cuilab.cn/hmdd/
dbDEMC	https://www.picb.ac.cn/dbDEMC/
CircR2Disease	http://bioinfo.snnu.edu.cn/CircR2Disease/
circRNADisease	http://cgga.org.cn:9091/circRNADisease/
Circ2Disease	http://bioinformatics.zju.edu.cn/Circ2Disease/
miRTarBase	http://mirtarbase.cuhk.edu.cn/
CircBank	http://www.circbank.cn/

- *Disease-mRNA relation.* Disease-mRNA associations are collected from the DisGeNET [48] database. DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated with human diseases (including disorders, traits, and clinical or abnormal human phenotypes). The database integrates data from expert curated repositories, GWAS catalogues, animal models and the scientific literature. DisGeNET data are homogeneously annotated with controlled vocabularies and community-driven ontologies.

- *Disease-microRNA relation.* The disease-microRNA data we used are from the merging two well-known databases: HMDD [49] and dbDEMC [50]. These databases that curated experiment-supported evidence for human microRNA (miRNA) and disease associations. HMDD (Human MicroRNA Disease Database) represents the first miRNA disease database in the world according to literature record and presents detailed and comprehensive annotations to the human miRNA-disease association data, including miRNA-disease association data from the evidence of genetics, epigenetics, circulating miRNAs, and miRNA-target interactions. dbDEMC (A Database of Differentially Expressed MiRNAs in Human Cancers) is an integrated database that designed to store and display differentially expressed microRNAs (miRNAs) in human cancers detected by high-throughput methods.

- *Disease-circRNA relation.* Disease-circRNA associations are downloaded from the CircR2Disease [51], circRNADisease [52] and Circ2Disease [53]. Also, these are manually curated databases of experimentally supported circRNA and disease associations. CircR2Disease is a platform for investigating mechanism of the disease-related circRNAs. circRNADisease records detailed information on a circRNA-disease association, including circRNA and disease name, the circRNA expression pattern, a brief functional description about circRNA and so on. Circ2Disease provides a comprehensive associations between of circRNA and human diseases and also integrates experimentally verified miRNAs and miRNA targets from several known databases.

- *microRNA-mRNA relation.* We select microRNA-mRNA associations from the miRTarBase [54] database, which

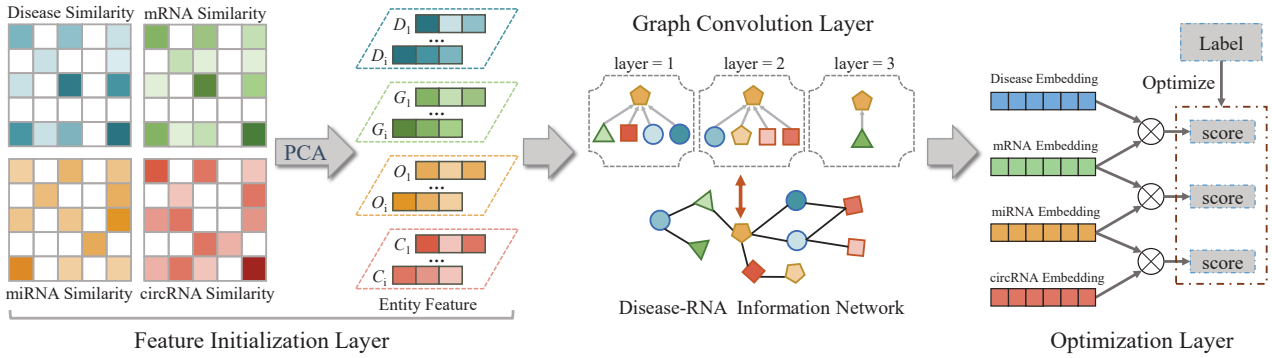

Figure 1: The main steps of the graph embedding model. Firstly, the input features of diseases, mRNAs, miRNAs, and circRNAs are extracted through the feature initialization layer. Then the embeddings of such biological entities are derived through the graph convolution layer to propagate and aggregate messages based on the disease-RNA information network constructed. Next, the optimization layer is used to refine embeddings by minimizing the objective function.

serves as more comprehensively annotated, experimentally validated miRNA-target interactions databases in the field of miRNA related research. It has accumulated more than three hundred and sixty thousand miRNA-target interactions (MTIs), which are collected by manually surveying pertinent literature after NLP of the text systematically to filter research articles related to functional studies of miRNAs.

- *circRNA-microRNA relation.* CircRNA-microRNA associations are from the CircBank [55]. This database is a novel nomenclature system based on the host gene name, start position and end position was applied for the naming of circRNA. In addition to the basic information of circRNA such as chromosome location and host gene, it adds multiple new feature of circRNA including binding miRNA, circRNA conservation, circRNA mutation and circRNA methylation.

4.2. Design Graph Embedding Model

Based on the constructed disease-RNA information network, the key step of *DRAMA* is to obtain biomedical entity embeddings in the network by the designed graph embedding model. To complete this task, it mainly consists of three components: (1) the feature initialization layer; (2) the graph convolution layer; and (3) the optimization layer. Figure 1 illustrates the main steps of the graph embedding model.

4.2.1. Feature initialization layer

Considering that the relationships between biomedical entities in the disease-RNA information network are scattered, we adopt GCN to capture the intrinsic associations among these entities. The core idea of GCN is to use graph structure to propagate the characteristics of entities to neighbors and reflect the implicit relationship between entities. However, specifying the complex relationships between biomedical entities requires sufficient domain knowledge. Thus, in order to avoid the strong requirements for the graph, we explore potential relationships by providing additional features for the biomedical entity and employing GCN to gather information about the entity's neighbors.

For the four different types of biomedical entities in the network, in order to characterize the differences between entities of the same type, we carry out principal component analysis (PCA) for similarity between the same type entities to initialize the biomedical entity features. Specifically, these messages are collected from: *MimMiner* [56], which organizes the disease similarity information from the phenotype network; *HumanNet* [57] is a functional gene network dataset, which collects 21 sources to measure the functional associations between genes; *MISIM* [58] offers the human microRNA functional similarity; and obtains circRNA similarity matrix from the work of *Peng et al.* [59]. And these generated PCA vectors are unified into the same dimension N, which are used as the initial features X for input.

4.2.2. Graph convolution layer

In this layer, there are two operations, propagate entity's messages and aggregate neighbor's information, to capture the explicit and implicit signals of entities in the disease-RNA information network. Specifically, let $Z^{(l)}$ be the output embeddings of l-th convolutional layer and given the graph adjacency matrix A corresponding to the disease-RNA information network $\mathcal{G}$, the updating rule of the graph convolution layer as:

$$Z^{(l+1)} = ReLU(\widetilde{\mathbb{D}}^{-\frac{1}{2}}\widetilde{A}\widetilde{\mathbb{D}}^{-\frac{1}{2}}Z^{(l)}W^{(l)}) \tag{2}$$

where $W^{(l)} \in \mathbb{R}^{N\times F}$ is the learnable weight matrix of l-th layer in the graph convolution layer, $\widetilde{\mathbb{D}}$ is the degree matrix of $\widetilde{A}$ and $\widetilde{A} = A + I$. In particular, $Z^{(0)} = X$.

According to the structural characteristics of the network that we build, for any type of node to gather information on the other types of nodes, we set the layer of the graph convolution l to 2. Finally, $f : (\mathcal{G}, X) \to Z = Z^{(2)}$. By this way, it could produce the embeddings of diseases z_d, mRNAs z_g, miRNAs z_o and circRNAs z_c.

4.2.3. Optimization layer

To determine the model parameters, we adopt the pairwise bayesian personalized ranking (BPR) [60] loss to optimize the objective function. BPR assumes that, compared with the unobserved ones, the observed interaction has higher scores. As our task is to infer the circRNA-miRNA-mRNA axis related to the disease, we assume that the interactions formed on the axis are more likely to be observed in the disease-RNA information network. At the same time, we also consider that, in actual biological processes, circRNA is mostly associated with miRNA and then influences the interaction of miRNA with mRNA, which may cause disease. In this process, circRNA does not directly interact with mRNA, so we design the objective function by considering circRNA-miRNA-mRNA as a whole with a fixed order. Furthermore, the linkage between this whole and disease is constituted through the entity "mRNA" as a bridge, and circRNA and miRNA have no direct interaction with disease. Thus, the design of the objective function is oriented to the axis-related interaction, which consists of three parts. The objective function is defined as follows:

$$Loss = \arg\min_{\Theta} \sum_{(u,i,j)\in\mathcal{O}} \mathcal{L}_{DG} + \mathcal{L}_{GO} + \mathcal{L}_{OC} + \lambda \left\|\Theta\right\|_2^2 \quad (3)$$

$$\mathcal{L}_{DG} = -\ln \mu(z_{d_u}^{\top} z_{g_i} - z_{d_u}^{\top} z_{g_j}) \quad (4)$$

$$\mathcal{L}_{GO} = -\ln \mu(z_{g_u}^{\top} z_{o_i} - z_{g_u}^{\top} z_{o_j}) \quad (5)$$

$$\mathcal{L}_{OC} = -\ln \mu(z_{o_u}^{\top} z_{c_i} - z_{o_u}^{\top} z_{c_j}) \quad (6)$$

where $\mathcal{O} = \left\{(u,i,j)|(u,i) \in E^+, (u,j) \in E^-\right\}$ denotes the pairwise training data, E^+ indicates the observed interactions in the network, and E^- is the unobserved interactions; $\mu(\cdot)$ is the sigmoid function; Θ is the set of model parameters, and λ controls the L2 regularization strength to prevent overfitting. And then we use the Adam optimization method [61] to optimize the model and update the model parameters.

4.3. Query over Embeddings

After deriving the optimal mapping function f through the graph embedding model, we can obtain the embeddings Z of all entities in the disease-RNA information network. Since each entity is now represented by a vector, all entities are in the same latent space, the correlation between mRNA g, miRNA o and circRNA c as well as the association of the three types of entities and their combinations with the given disease d can be easily calculated through a favorable measurement.

Here, we design a measurement to calculate the correlation score of candidate mRNA-miRNA-circRNA axis $\langle c, o, g\rangle$ related to a given disease $d \in D$, as follows:

$$\underset{Z}{Score}(d, \langle c, o, g\rangle) = \frac{(z_c \cdot z_o) + (z_o \cdot z_g)}{\left|(z_c \cdot z_o) * (z_o \cdot z_g)\right|} + \sqrt[3]{(z_c \cdot z_d) * (z_o \cdot z_d) * (z_g \cdot z_d)} \quad (7)$$

where the left part of the plus sign reflects the possibility of forming a circRNA-miRNA-mRNA axis and the right side selects the most relevant c, o and g to d to form the axis. The candidate axes can be sorted based on this score to generate a set of circRNA-miRNA-mRNA axes $\mathcal{T}\langle c, o, g\rangle$ in relation with a given disease d, the size of which can be predefined according to the requirements. It is assumed that the axis with a higher ranking is more likely to be related to the disease query.

4.4. Classification and Tagging Axes

Given a query disease, *DRAMA* can get user-defined top-k axes associated with this disease. In order for researchers to conduct further biological experimental analysis and exploration of these predicted axes, we use the annotation information of the biomedical entities in the circRNA-miRNA-mRNA axes to classify and tag these axes.

First, for the predicted top-k axes, we use the three databases GOA [62], HMDD [49] and CircBank [55] to add annotation information to the mRNA, miRNA, and circRNA entities on each axis, respectively. Then the annotation information of each axis of the top-k axes is spliced together, and we use C^k to represent the annotation corpus obtained after splicing. With each annotation in C^k as the minimum word segmentation unit, the number of keywords that need to be extracted is set to K through the TF-IDF algorithm, so as to obtain K classifications for these axes. Finally, the word2vec method is employed to calculate the most relevant one of the K classifications for an axis, and this classification is used to tag the axis.

5. Experiments

In this section, we present the details of experiments including experimental settings and results. First, we introduce the dataset, baselines, metrics and experimental setup. The performance comparisons are then demonstrated in detail. Finally, we discuss the case study and the influence of several critical parameters.

5.1. Experimental Settings

5.1.1. Dataset

We construct the disease-RNA information network based on the four types of biomedical entities and the five relationships from the eight databases as discussed in the Section 4.1. In order to maintain the connectivity of the constructed network, we preprocess these extracted relations and entities. Table 2 lists the statistics of the entities and links in the disease-RNA information network. In summary, there are 175,081 entities and 1,186,908 links.

Table 2
Statistics of the constructed disease-RNA information network.

Relation	Type	# Entities	# Links
Disease-mRNA relation	disease	21775	781622
	mRNA	20158	
Disease-miRNA relation	disease	715	16526
	microRNA	1248	
Disease-circRNA relation	disease	101	1240
	circRNA	552	
miRNA-mRNA relation	microRNA	2604	380650
	mRNA	150150	
circRNA-miRNA relation	circRNA	353	6870
	microRNA	351	

5.1.2. Baselines

To demonstrate the effectiveness, we compare *DRAMA* with ten popular graph representation learning methods including GF, SVD, GAE, LINE, HOPE, GraRep, DeepWalk, Laplacian, node2vec and struc2vec. The detailed characteristics of these graph embedding methods can be referred to [63]. In addition, we also compare with the two graph neural network methods to further verify the effectiveness of *DRAMA*. One is graph attention network (GAT) [64], which integrates attention mechanism into graph neural networks. It employs the self-attention strategy, which can self-adaptively learn the importance of different neighbor nodes in the process of aggregating neighbor information, to address the shortcomings of graph convolutions based on spectral approaches. The other is GraphSAGE [65], which uses local aggregation functions for embedding computation.

5.1.3. Evaluation metrics

Since the path "circRNA-miRNA-mRNA-disease" we explored is a novel problem, there are no reliable benchmark databases containing large amount of relevant data. Thus, all these baselines and our method are designed to predict the interaction of circRNA-miRNA, miRNA-mRNA and mRNA-disease. These interactions are likely to be parts of the path we aim to detect. To evaluate the effectiveness of *DRAMA*, we employ several performance evaluation metrics that are widely used, including the area under the receiver operating characteristic (ROC) curve (**AUC-ROC**), the area under the precision-recall (PR) curve (**AUC-PR**), accuracy (**ACC**), precision (**PRE**), sensitivity, i.e. recall (**REC**) and F1-score (**F1**).

5.1.4. Experimental setup

DRAMA was implemented in PyTorch. The dimension of biomedical entity feature N is set to 500. The reason for this setting is that the number of circRNA in the dataset used in this work is the smallest relative to other biomedical entities, so an integer smaller than its amount (552) is selected as the dimension of all biomedical entity features. For a fair comparison, the embedding size (i.e. the size of biomedical entity embedding F) is fixed to 128 for all models. Moreover, for DeepWalk, node2vec and struc2vec based on random walk, we set the walk length is 80, the window size is 10 and the number of walks per node is 10. For *DRAMA*, GF, GAE, GAT, LINE and GraphSAGE based on neural network, we set the epoch as 800, the learning rate as 0.01 and the layer of the graph convolution as 2. Thereafter, the embeddings of biomedical entities obtained by these methods are used for predicting the corresponding links.

All experiments were conducted on a PC with four Intel Xeon E5-2698 GHz CPUs, four GeForce RTX 2080 Ti GPUs and 512 GB memory, running Ubuntu 20.04. The algorithms were implemented in Python and compiled by Python 3.7.

5.2. Effectiveness

To the best of our knowledge, currently there is no other solution to the problem that we address. To evaluate the performance of *DRAMA*, we design the following experiments to prove it.

In the experiment compared with graph representation learning methods, we extract the circRNA-miRNA, miRNA-mRNA and mRNA-disease interactions from the disease-RNA information network we constructed as the positive samples of training set and testing set with a ratio of 8:2. The negative training set and testing set were generated by randomly selecting the same number of unobserved links as the positive samples size. Such unobserved links are unlikely to have interactions, because they are pair-wise combinations (in the same order) between the entities involved in the aforementioned three types of pairs from the entire dataset and the known interactions are excluded. To reduce the occasionality and ensure the fairness of the experiments, we iterate 100 times to obtain the average results for the comparison.

The results of *DRAMA* and other graph embedding-based baselines are shown in Table 3. Based on such results, we provide the following analyses.

In general, different graph embedding methods have different performances, but all of them can achieve AUC-PR scores > 0.60. Therefore, graph representation learning methods can extract potentially useful information from the disease-RNA information network, and have different performances in the "circRNA-miRNA", "miRNA-mRNA" and "mRNA-disease" interactions prediction. We also find that *DRAMA* has superior performance in this task with the AUC-PR score of 0.9411, AUC-ROC score of 0.9052, ACC score of 0.8434 and F1 score of 0.8836. It also achieves decent scores in the REC and PRE evaluation metrics with the scores of 0.9093 and 0.8591. This demonstrates the effectiveness of *DRAMA* for capturing potential information about "circRNA-miRNA", "miRNA-mRNA" and "mRNA-disease" interactions, which also further indicates the effectiveness of *DRAMA* in mining disease-related circRNA-miRNA-mRNA axes.

5.3. Ablation Study

In our model, for each type of biomedical entity, we use the similarity property to initialize their features and then predict the associations between them. In order to

Table 3
Overall performances of comparison with baselines.

Method category	Method name	AUC-PR	AUC-ROC	ACC	F1	REC	PRE
Matrix factorization-based	GraRep	0.6323	0.4982	0.4946	0.5612	0.4910	0.6547
	HOPE	0.7233	0.4977	0.4987	0.5713	0.5031	0.6566
	Laplacian	0.8874	0.7955	**0.7260**	**0.8032**	0.8423	**0.7672**
	SVD	**0.9125**	**0.8141**	0.7092	0.7998	**0.8753**	0.7311
Random walk-based	struc2vec	0.7604	0.6412	0.6425	0.7820	0.9688	0.6557
	node2vec	0.8186	0.7251	0.6695	0.7986	**0.9955**	0.6653
	DeepWalk	**0.8201**	**0.7312**	**0.6854**	**0.8043**	0.9882	**0.6769**
Neural network-based	LINE	0.6074	0.4166	0.6376	0.7768	0.9508	0.6557
	GAE	0.9122	0.8351	**0.6929**	0.8007	0.9430	**0.6957**
	GF	**0.9206**	**0.8829**	0.6828	**0.8046**	**0.9973***	0.6755
Graph neural network-based	GAT	0.8998	0.8583	0.7434	0.7823	0.6957	0.8936
	GraphSAGE	0.9379	0.8942	0.7705	0.8048	0.7136	**0.9226***
	DRAMA	**0.9411***	**0.9052***	**0.8434***	**0.8836***	**0.9093**	0.8591

Table 4
Performance evaluation of variants.

Variants	AUC-PR	AUC-ROC	F1
DRAMA-ALL	0.6628	0.5000	0.7972
DRAMA-DGO	0.7615	0.6378	0.7951
DRAMA-DOC	0.9015	0.8274	0.8289
DRAMA-GOC	0.8602	0.7577	0.8132
DRAMA-DG	0.8915	0.7981	0.7818
DRAMA-DO	0.9091	0.8383	0.8191
DRAMA-DC	0.9343	0.8832	0.8198
DRAMA-GO	0.8093	0.6894	0.8138
DRAMA-GC	0.9045	0.8293	0.8003
DRAMA-OC	0.9059	0.8398	0.8415
DRAMA-D	0.9347	0.8838	0.8446
DRAMA-G	0.9171	0.8515	0.8190
DRAMA-O	0.9133	0.8524	0.8402
DRAMA-C	0.9388	0.8985	0.8487

investigate whether conditional and random feature initialization is helpful to the effectiveness of *DRAMA*, we have conducted an ablation study by randomly initializing the partial biomedical entity feature in *DRAMA*. Specifically, we tested the performances of variants by randomly initializing only one type of biomedical entity feature (named *DRAMA-D*, *DRAMA-G*, *DRAMA-O* and *DRAMA-C*), two types of biomedical entity features (named *DRAMA-DG*, *DRAMA-DO*, *DRAMA-DC*, *DRAMA-GO*, *DRAMA-GC* and *DRAMA-OC*), three types of biomedical entity features (named *DRAMA-DGO*, *DRAMA-DOC* and *DRAMA-GOC*) and all biomedical entity features (named *DRAMA-ALL*).

Table 4 shows the performance of *DRAMA* could be weakened when the biomedical entity types of random initialization increase. Interestingly, we have observed that when randomly initializing the same number of biomedical entity types, as long as the randomly initialized type contains 'mRNA', the impact on performance is most distinct.

5.4. Case Study

We select two diseases *Cervical Cancer*, and *Bladder Cancer* as the query objects and analyze for further evaluation of the effectiveness of *DRAMA*. Through the literature [66–68], the circRNA-miRNA-mRNA axes related to *Cervical Cancer* and *Bladder Cancer* can be obtained and used as the benchmark data for the query objects.

The axes composed of each mRNA, miRNA and circRNA entity in the disease-RNA information network all may be candidates for the query disease, we select 1000 circRNAs, miRNAs and mRNAs that are most relevant to the query disease in a ratio of 1:3:6, forming $1.8 * 10^7 (100 * 300 * 600)$ axes as candidates. Obviously, this number is very large, but *DRAMA* can narrow the range of candidates and detect the most relevant ones. We compare *DRAMA* with other baselines by conducting the experiment which is to detect how many of the literature-mentioned circRNA-miRNA-mRNA axes in the top-k results. As shown in Tables 5 and 6, *DRAMA* can clearly detect more effective circRNA-miRNA-mRNA axes than other methods.

From the results in Table 5, we see that for the *Cervical Cancer*, *DRAMA* can detect the effective circRNA-miRNA-mRNA axes related to it when selected top 0.056% , while none of the other methods can detect it. As the selected top range increases, *DRAMA* obviously detects more circRNA-miRNA-mRNA axes mentioned in the literature. When it reaches the top of 11.1%, more than 90% of the axes have been detected, while the best performance of the other methods has detected 71.64% of the axes. Also, it is observed from Table 6 that when *Bladder Cancer* is used as the query object, *DRAMA* has a similar performance comparing to other methods.

5.5. Axes Classification and Tagging Analysis

Based on the analysis results of the case study in the previous subsection, we select the cervical cancer-related top-400 circRNA-miRNA-mRNA axes and the bladder cancer-related top-800 circRNA-miRNA-mRNA axes predicted by *DRAMA* for axes classification and tagging. Specifically, we

Table 5
Case study results of cervical cancer.

Top-k/Top-ratio	Method name	Num	Ratio (%)
10000 / 0.056%	**DRAMA**	**6**	**0.05**
	others	0	0.00
100000 / 0.56%	GAT	18	0.16
	DRAMA	**12**	**0.11**
	LINE	11	0.10
	struc2vec	4	0.04
	HOPE	1	0.01
	others	0	0.00
200000 / 1.11%	**DRAMA**	**37**	**0.34**
	GAT	29	0.27
	LINE	23	0.21
	struc2vec	8	0.07
	HOPE	1	0.01
	others	0	0.00
1500000 / 8.33%	**DRAMA**	**9698**	**88.65**
	GAT	7007	64.05
	node2vec	376	3.44
	LINE	77	0.70
	GraRep	64	0.59
	GAE	22	0.20
	struc2vec	14	0.13
	GraphSAGE	10	0.09
	HOPE	4	0.04
	others	0	0.00
2000000 / 11.1%	**DRAMA**	**9862**	**90.15**
	GAT	7837	71.64
	node2vec	733	6.70
	GraRep	119	1.09
	LINE	99	0.90
	GraphSAGE	70	0.64
	GAE	42	0.38
	struc2vec	14	0.13
	HOPE	5	0.05
	others	0	0.00
5000000 / 27.8%	**DRAMA**	**9914**	**90.62**
	GAT	8078	73.84
	node2vec	6758	61.77
	GraphSAGE	5132	46.91
	DeepWalk	334	3.05
	GAE	150	1.37
	GraRep	137	1.25
	LINE	126	1.15
	HOPE	107	0.98
	struc2vec	65	0.59
	GF	28	0.26
	others	0	0.00

Table 6
Case study results of bladder cancer.

Top-k/Top-ratio	Method name	Num	Ratio (%)
100000 / 0.56%	**DRAMA**	**19**	**0.11**
	LINE	8	0.04
	others	0	0.00
200000 / 1.11%	**DRAMA**	**55**	**0.31**
	LINE	53	0.30
	struc2vec	3	0.02
	others	0	0.00
1500000 / 8.33%	**DRAMA**	**3333**	**18.73**
	LINE	933	5.24
	GraRep	290	1.63
	HOPE	149	0.84
	struc2vec	22	0.12
	GraphSAGE	3	0.02
	node2vec	1	0.01
	others	0	0.00
2000000 / 11.1%	**DRAMA**	**8367**	**47.01**
	LINE	1258	7.07
	GraRep	290	1.63
	HOPE	280	1.57
	struc2vec	50	0.28
	GraphSAGE	15	0.08
	node2vec	7	0.04
	others	0	0.00
5000000 / 27.8%	**DRAMA**	**11418**	**64.16**
	LINE	2423	13.61
	HOPE	2297	12.91
	node2vec	495	2.78
	GraphSAGE	394	2.21
	GraRep	294	1.65
	struc2vec	173	0.97
	GAE	22	0.12
	DeepWalk	1	0.01
	others	0	0.00

divide the annotation information of GOA into three categories, i.e., cellular component, molecular function, and biological process to classify and tag these axes respectively. We set the number of classification K for each category to 5, and then make statistics on these tagged axes. The analysis result of cervical cancer is shown in Figure 2 and bladder cancer is illustrated in Figure 3.

In order to demonstrate the effectiveness of the classification and tagging results of these axes, we further make use of the Database for Annotation, Visualization and Integrated Discovery (DAVID) [69] to analyze Kyoto Encyclopedia of Gene and Genome (KEGG) [70] pathway on the predicted cervical cancer and bladder cancer-related circRNA-miRNA-mRNA axes. Figure 4 depicts the analysis result.

Axes classified by our approach are highly coincident with pathway enrichment analysis from several aspects. Human papillomavirus (HPV) is the most well-known risk factor for the development of cervical cancer. High-risk HPV types cause almost all cervix malignancy, at the present time, HPV of types 16, 18, and 45 have been identified as the most common causes of cancers of the cervix. Oncoproteins of HPV, especially E5, E6, and E7, largely changing the cellular function through a cascade of amplifying signals among cross-talking pathways. PI3K/Akt, Wnt/β-catenin, ERK/MAPK, NF-κB, and JAK/STAT signaling pathways have a significant function in the progression of cervical cancer in HPV infected individuals, and further deregulate several cellular processes, mostly apoptosis, cell cycle control, migration, immune evasion, and even induction of genetic instability [71]. Furthermore, viruses have been shown to

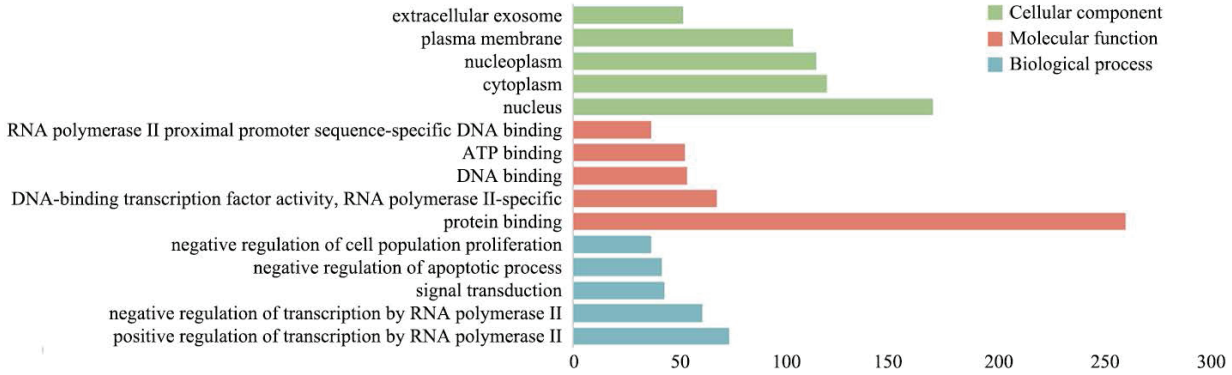

Figure 2: The classification and tagging results of the predicted top-400 circRNA-miRNA-mRNA axes related to cervical cancer.

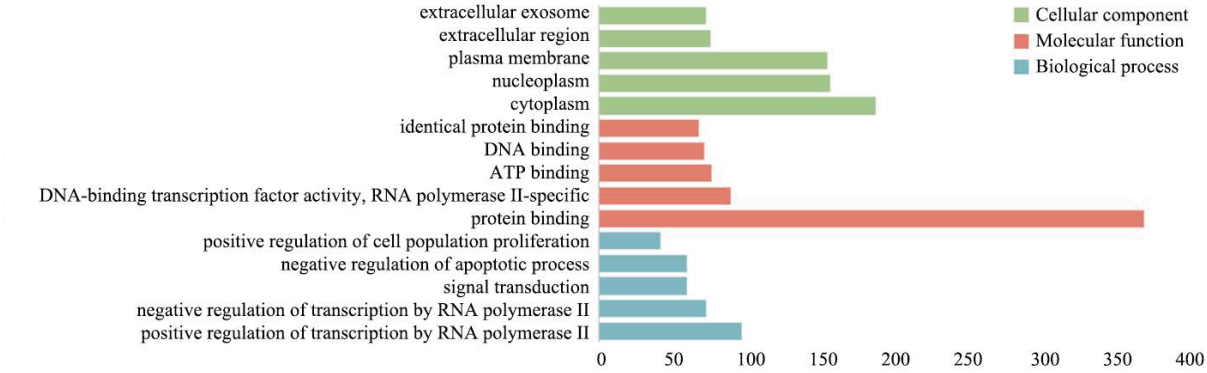

Figure 3: The classification and tagging results of the predicted top-800 circRNA-miRNA-mRNA axes related to bladder cancer.

affect miRNA processing through their viral proteins. miRNAs are transcriptional products of RNA polymerase II, and the miRNA processing proteins including Drosha, DGCR8, Dicer, and AGO2 are deregulated in HPV-induced carcinoma [72].

Variant molecular sub-types, mutation sub-types, and epigenetic modification markers predict different prognosis of bladder cancer. The predicted bladder cancer-related axes classification demonstrates the sub-types by function enrichment, which consists of the KEGG pathway analysis. Compared with other solid tumors, bladder cancer develops relatively high frequency of somatic mutations, which changes several signaling pathways, including the PI3K-AKT pathway, TERT promoter, cell cycle, RAS-RAF-MAPK pathway [73]. Also, miRNAs aberrantly expressed in bladder cancer.

From a biomedical aspect, *DRAMA* is an effective algorithm, which provides the circRNA-miRNA-mRNA landscape view of specific diseases and predicts potential circR-

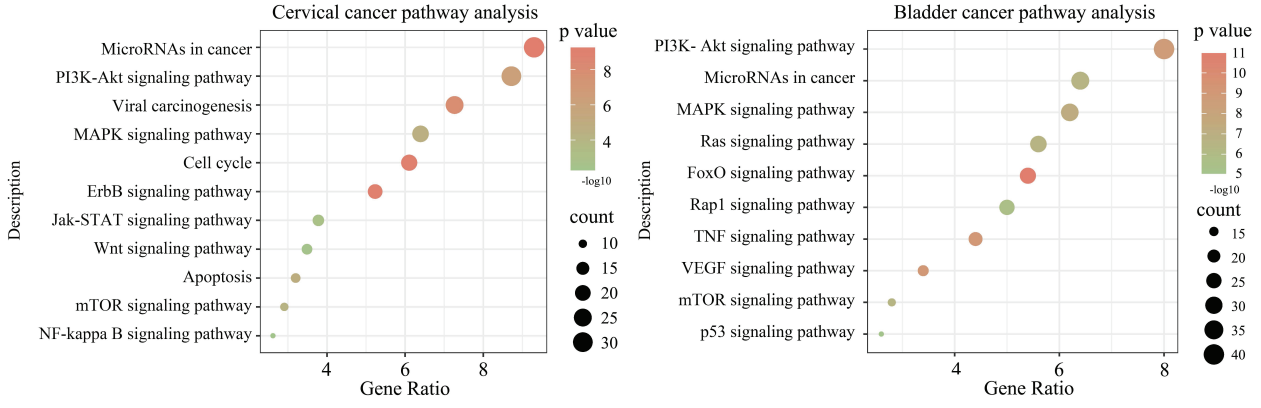

Figure 4: Pathway enrichment analysis results of cervical cancer and bladder cancer.

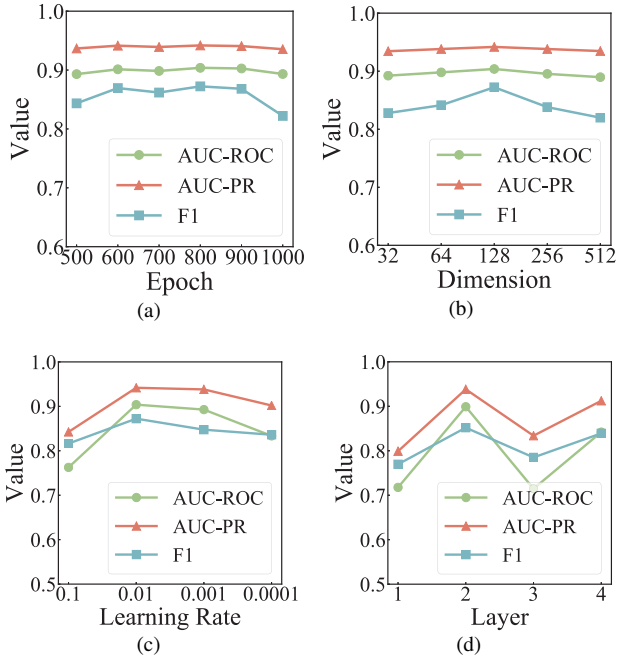

Figure 5: Performance analysis of parameters (a) epoch, (b) dimension, (c) learning rate and (d) layer from three evaluation metrics: AUC-ROC, AUC-PR and F1.

NAs and related axes.

5.6. Parameter Sensitivity

Our proposed model contains several hyper-parameters, which have been tested to evaluate their impacts on *DRAMA* measured by AUC-PR, AUC-ROC and F1 scores, including the dimension of output embeddings, the epoch and learning rate in the optimization process, and the layer of the graph convolution layer. In this part, all experiments are performed on the whole dataset i.e., the disease-RNA information network we built. When comparing a parameter, we keep the other parameters unchanged. Their performances are presented in Figure 5.

5.6.1. Effect of epoch

To verify the effect of epoch on *DRAMA*, we change the epoch in {500, 600, 700, 800, 900, 1000}. From Figure 5 (a), we see that with the change of epoch, all results under the evaluation metrics of AUC-PR and AUC-ROC have very trivial differences, and F1 score also becomes stable in the range from 600 to 900. In general, the three evaluation scores reach the best performance when the epoch was 800.

5.6.2. Effect of dimension

Similarly, in order to analyze the influence of the dimensions on the *DRAMA*, we vary the dimension from {32, 64, 128, 256, 512}. It is easily observed from Figure 5 (b) that the overall performance has become stable as the change of dimension under the evaluation metrics of AUC-PR and AUC-ROC, and the three evaluation scores are the best when the dimension is 128.

From the analysis of the above two parameters, it could prove the stability of *DRAMA*, that is, our results would not fluctuate too much if the dimension and the epoch are selected within an appropriate range.

5.6.3. Effect of learning rate

We also assess the sensitivity of the learning rate when the learning rate changes in {0.0001, 0.001, 0.01, 0.1}. As shown in Figure 5 (c), when the learning rate is 0.01, *DRAMA* obtains the best performance.

5.6.4. Effect of layer

We further analyze the changes in the performance of *DRAMA* as the layers increases and the layer varies from {1, 2, 3, 4}. Figure 5 (d) shows that when the layer of *DRAMA* is 2, the model has the best scores.

Also, it can be seen that *DRAMA* is sensitive to these two parameters. Reasonable explanations may be that controlling the learning rate is related to the optimal solution of the algorithm and the choice of the layer is related to the structure of the constructed network.

6. Conclusion

Discovering disease-ncRNA association is an important issue in the field of biomedicine. It can assist biologists and physicians to understand the etiology and pathogenic mechanism of disease symptoms. Different from previous work that predicting the relationship between disease and ncRNA, in this paper, we considered the impact of the interaction among ncRNAs on the disease, and thus proposed and evaluated *DRAMA*, a general framework to discover disease-related circRNA-miRNA-mRNA axes with graph convolutional network. By constructing a disease-RNA information network, *DRAMA* captured the explicit and implicit connections among disease, mRNA, miRNA and circRNA utilizing local neighbors and global structure. Experimental results have demonstrated that the effectiveness of *DRAMA*, which has outperformed baselines in real-world datasets.

For future work, improvements can be made in considering more abundant and richer domain knowledge to initialize the features of biological entities, introducing the regulation mechanism of ncRNA to construct the disease-RNA information network for distinguishing edge heterogeneity, as well as developing a system for updating the results in real time when new data sources are added.

7. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (61972268), and the Sichuan Science and Technology Program (2020YFG0034).

References

[1] E. I. Marks, N. S. Yee, Molecular genetics and targeted therapeutics in biliary tract carcinoma, World Journal of Gastroenterology 22 (2016) 1335–1347.

[2] Y. Qian, Y. Gong, Z. Fan, G. Luo, Q. Huang, S. Deng, H. Cheng, K. Jin, Q. Ni, X. Yu, C. Liu, Molecular alterations and targeted therapy in pancreatic ductal adenocarcinoma, Journal of Hematology & Oncology 13 (2020) 130.

[3] C. L. Chew, S. A. Conos, B. Unal, V. Tergaonkar, Noncoding RNAs: Master regulators of inflammatory signaling, Trends in Molecular Medicine 24 (2018) 66–84.

[4] F. J. Slack, A. M. Chinnaiyan, The role of non-coding RNAs in oncology, Cell 179 (2019) 1033–1055.

[5] J. Wang, S. Zhu, N. Meng, Y. He, R. Lu, G. Yan, ncRNA-Encoded peptides or proteins and cancer, Molecular Therapy 27 (2019) 1718–1725.

[6] L. Ding, M. Wang, D. Sun, A. Li, A novel method for identifying potential disease-related miRNAs via a disease-miRNA-target heterogeneous network, Molecular BioSystems 13 (2017) 2328–2337.

[7] X. Chen, G. Yan, Semi-supervised learning for potential human microRNA-disease associations inference, Scientific Reports 4 (2014) 5501.

[8] P. Xuan, T. Shen, X. Wang, T. Zhang, W. Zhang, Inferring disease-associated microRNAs in heterogeneous networks with node attributes, IEEE/ACM Transactions on Computational Biology and Bioinformatics 17 (2020) 1019–1031.

[9] X. Chen, M. Liu, G. Yan, RWRMDA: Predicting novel human microRNAdisease associations, Molecular BioSystems 8 (2012) 2792–2798.

[10] P. Xuan, K. Han, Y. Guo, J. Li, X. Li, Y. Zhong, Z. Zhang, J. Ding, Prediction of potential disease-associated microRNAs based on random walk, Bioinformatics 31 (2015) 1805–1815.

[11] H. Shi, J. Xu, G. Zhang, L. Xu, C. Li, L. Wang, Z. Zhao, W. Jiang, Z. Guo, X. Li, Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes, BMC Systems Biology 7 (2013).

[12] Y. Liu, X. Zeng, Z. He, Q. Zou, Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources, IEEE/ACM Transactions on Computational Biology and Bioinformatics 14 (2017) 905–915.

[13] J. Luo, Q. Xiao, A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network, Journal of Biomedical Informatics 66 (2017) 194–203.

[14] Z. Chen, X. Wang, P. Gao, H. Liu, B. Song, Predicting disease related microRNA based on similarity and topology, Cells 8 (2019) 1405.

[15] X. Chen, D. Xie, L. Wang, Q. Zhao, Z. You, H. Liu, BNPMDA: Bipartite network projection for miRNAdisease association prediction, Bioinformatics 34 (2018) 3178–3186.

[16] K. Zheng, Z. You, L. Wang, Y. Zhou, L. Li, Z. Li, MLMDA: A machine learning approach to predict and validate MicroRNA-disease associations by integrating of heterogenous information sources, Journal of Translational Medicine 17 (2019) 260.

[17] F. Yan, Y. Zheng, W. Jia, S. Hou, R. Xiao, MAMDA: Inferring microRNA-disease associations with manifold alignment, Computers in Biology and Medicine 110 (2019) 156–163.

[18] X. Chen, G. Yan, Novel human lncRNAdisease association inference based on lncRNA expression profiles, Bioinformatics 29 (2013) 2617–2624.

[19] M. Liu, X. Chen, G. Chen, Q. Cui, G. Yan, A computational framework to infer human disease-associated long noncoding RNAs, PLoS One 9 (2014) e84408.

[20] L. Ding, M. Wang, D. Sun, A. Li, TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph, Scientific Reports 8 (2018) 1065.

[21] J. Sun, H. Shi, Z. Wang, C. Zhang, L. Liu, L. Wang, W. He, D. Hao, S. Liu, M. Zhou, Inferring novel lncRNAdisease associations based on a random walk model of a lncRNA functional similarity network, Molecular BioSystems 10 (2014) 2074–2081.

[22] G. U. Ganegoda, M. Li, W. Wang, Q. Feng, Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations, IEEE Transactions on NanoBioscience 14 (2015) 175–183.

[23] Y. Wang, L. Juan, J. Peng, T. Zang, Y. Wang, LncDisAP: A computation model for LncRNA-disease association prediction based on multiple biological datasets, BMC Bioinformatics 20 (2019) 582.

[24] H. Zhang, Y. Liang, C. Peng, S. Han, W. Du, Y. Li, Predicting lncRNA-disease associations using network topological similarity based on deep mining heterogeneous networks, Mathematical Biosciences 315 (2019) 108229.

[25] C. Lu, M. Yang, F. Luo, F. Wu, M. Li, Y. Pan, Y. Li, J. Wang, Prediction of lncRNA-disease associations based on inductive matrix completion, Bioinformatics 34 (2018) 3357–3364.

[26] Z. Cui, J. Liu, Y. Gao, R. Zhu, S. Yuan, LncRNA-Disease associations prediction using bipartite local model with nearest profile-based association inferring, IEEE Journal of Biomedical and Health Informatics 24 (2020) 1519–1527.

[27] Q. Yuan, X. Guo, Y. Ren, X. Wen, L. Gao, Cluster correlation based method for lncRNA-disease association prediction, BMC Bioinformatics 21 (2020) 180.

[28] X. Lei, Z. Fang, L. Chen, F. Wu, PWCDA: Path weighted method for predicting circRNA-disease associations, International Journal of Molecular Sciences 19 (2018) 3410.

[29] Q. Xiao, J. Luo, J. Dai, Computational prediction of human disease-associated circRNAs based on manifold regularization learning framework, IEEE Journal of Biomedical and Health Informatics 23 (2019) 2661–2669.

[30] M. Li, M. Liu, Y. Bin, J. Xia, Prediction of circRNA-disease associations based on inductive matrix completion, BMC Medical Genomics 13 (2020) 42.

[31] K. Zheng, Z. You, J. Li, L. Wang, Z. Guo, Y. Huang, iCDA-CGR: Identification of circRNA-disease associations based on Chaos Game Representation, PLoS Computational Biology 16 (2020) e1007872.

[32] L. Deng, W. Zhang, Y. Shi, Y. Tang, Fusion of multiple heterogeneous networks for predicting circRNA-disease associations, Scientific Reports 9 (2019) 9605.

[33] Q. Zhao, Y. Yang, G. Ren, E. Ge, C. Fan, Integrating bipartite network projection and KATZ measure to identify novel circRNA-disease associations, IEEE Transactions on NanoBioscience 18 (2019) 578–584.

[34] X. Lei, C. Bian, Integrating random walk with restart and k-Nearest Neighbor to identify novel circRNA-disease association, Scientific Reports 10 (2020).

[35] G. Li, J. Luo, D. Wang, C. Liang, Q. Xiao, P. Ding, H. Chen, Potential circRNA-disease association prediction using DeepWalk and network consistency projection, Journal of Biomedical Informatics 112 (2020) 103624.

[36] Y. Ding, B. Chen, X. Lei, B. Liao, F. Wu, Predicting novel CircRNA-disease associations based on random walk and logistic regression model, Computational Biology and Chemistry 87 (2020) 107287.

[37] H. Wei, B. Liu, iCircDA-MF: Identification of circRNA-disease associations based on matrix factorization, Briefings in Bioinformatics 21 (2019) 1356–1367.

[38] X. Lei, Z. Fang, GBDTCDA: Predicting circRNA-disease associations based on gradient boosting decision tree with multiple biological data fusion, International Journal of Biological Sciences 15 (2019) 2911–2924.

[39] D. Rong, H. Sun, Z. Li, S. Liu, C. Dong, K. Fu, W. Tang, H. Cao, An emerging function of circRNA-miRNAs-mRNA axis in human diseases, Oncotarget 8 (2017) 73271–73281.

[40] L. Verduci, S. Strano, Y. Yarden, G. Blandino, The circRNAmicroRNA code: Emerging implications for cancer diagnosis and treatment, Molecular Oncology 13 (2019) 669–680.

[41] Z. Cheng, C. Yu, S. Cui, H. Wang, H. Jin, C. Wang, B. Li, M. Qin, C. Yang, J. He, Q. Zuo, S. Wang, J. Liu, W. Ye, Y. Lv, F. Zhao, M. Yao, L. Jiang, W. Qin, circTP63 functions as a ceRNA to promote lung squamous cell carcinoma progression by upregulating FOXM1, Nature Communications 10 (2019) 3200.

[42] Y. Wang, T. Wang, M. Ding, S. Xiang, M. Shi, B. Zhai, hsa_circ_0091570 acts as a ceRNA to suppress hepatocellular cancer progression by sponging hsa-miR-1307, Cancer Letters 460 (2019) 128–138.

[43] C. He, L. Duan, H. Zheng, J. Li-Ling, L. Li, DRAMA: Discovering disease-related circRNA-miRNA-mRNA axes from disease-RNA information network, in: IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020, Virtual Event, South Korea, December 16-19, 2020, pp. 269–274.

[44] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1953) 39–43.

[45] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, Conference Track Proceedings (Poster), 2017.

[46] M. Zhang, Y. Chen, Link prediction based on graph neural networks, in: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, Canada, December 3-8, 2018, pp. 5171–5181.

[47] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, J. Leskovec, Graph convolutional neural networks for web-scale recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018, pp. 974–983.

[48] J. Piñero, J. M. Ramírez Anguita, J. Saüch Pitarch, F. Ronzano, E. Centeno, F. Sanz, L. I. Furlong, The DisGeNET knowledge platform for disease genomics: 2019 update, Nucleic Acids Research 48 (2019) D845–D855.

[49] Z. Huang, J. Shi, Y. Gao, C. Cui, S. Zhang, J. Li, Y. Zhou, Q. Cui, HMDD v3.0: A database for experimentally supported human microRNAdisease associations, Nucleic Acids Research 47 (2018) D1013–D1017.

[50] Z. Yang, L. Wu, A. Wang, W. Tang, Y. Zhao, H. Zhao, A. E. Teschendorff, dbDEMC 2.0: Updated database of differentially expressed miRNAs in human cancers, Nucleic Acids Research 45 (2016) D812–D818.

[51] C. Fan, X. Lei, Z. Fang, Q. Jiang, F. Wu, CircR2Disease: A manually curated database for experimentally supported circular RNAs associated with various diseases, Database: the Journal of Biological Databases and Curation 2018 (2018).

[52] Z. Zhao, K. Wang, F. Wu, W. Wang, K. Zhang, H. Hu, Y. Liu, T. Jiang, circRNA disease: A manually curated database of experimentally supported circRNA-disease associations, Cell Death & Disease 9 (2018).

[53] D. Yao, L. Zhang, M. Zheng, X. Sun, Y. Lu, P. Liu, Circ2Disease: A manually curated database of experimentally validated circRNAs in human disease, Scientific Reports 8 (2018) 2045–2322.

[54] C. Chou, S. Shrestha, C. Yang, N. Chang, Y. Lin, K. Liao, W. Huang, T. H. Sun, S. J. Tu, W. H. Lee, M. Y. Chiew, C. Tai, T. Wei, T. R. Tsai, H. T. Huang, C. Wang, H. Y. Wu, S. Ho, P. Chen, C. H. Chuang, P. J. Hsieh, Y. S. Wu, W. Chen, M. Li, Y. C. Wu, X. Huang, F. L. Ng, W. Buddhakosai, P. C. Huang, K. C. Lan, C. Y. Huang, S. Weng, Y. N. Cheng, C. Liang, W. L. Hsu, H. D. Huang, miRTarBase update 2018: A resource for experimentally validated microRNA-target interactions, Nucleic Acids Research 46 (2017) D296–D302.

[55] M. Liu, Q. Wang, J. Shen, B. B. Yang, X. Ding, Circbank: A comprehensive database for circRNA with standard nomenclature, RNA Biology 16 (2019) 899–905.

[56] M. A. van Driel, J. Bruggeman, G. Vriend, H. G. Brunner, J. A. M. Leunissen, A text-mining analysis of the human phenome, European Journal of Human Genetics : EJHG 14 (2006) 535–542.

[57] S. Hwang, C. Y. Kim, S. Yang, E. Kim, T. Hart, E. M. Marcotte, I. Lee, HumanNet v2: Human gene networks for disease research, Nucleic Acids Research 47 (2018) D573–D580.

[58] J. Li, S. Zhang, Y. Wan, Y. Zhao, J. Shi, Y. Zhou, Q. Cui, MISIM v2.0: A web server for inferring microRNA functional similarity based on microRNA-disease associations, Nucleic Acids Research 47 (2019) W536–W541.

[59] N. Peng, L. Shi, Q. Zhang, Y. Hu, N. Wang, H. Ye, Microarray profiling of circular rnas in human papillary thyroid carcinoma, PLoS One 12 (2017) e0170287.

[60] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, in: UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009, pp. 452–461.

[61] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, Conference Track Proceedings (Poster), 2015.

[62] R. P. Huntley, T. Sawford, P. Mutowo-Meullenet, A. Shypitsyna, C. Bonilla, M. J. Martin, C. O'Donovan, The GOA database: Gene ontology annotation updates for 2015, Nucleic Acids Research 43 (2015) D1057–D31063.

[63] X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S. M. Lin, W. Zhang, P. Zhang, H. Sun, Graph embedding on biomedical networks: Methods, applications and evaluations, Bioinformatics 36 (2020) 1241–1251.

[64] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, Conference Track Proceedings (Poster), 2018.

[65] W. L. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, December 4-9, 2017, pp. 1024–1034.

[66] J. Gong, H. Jiang, C. Shu, M. Hu, Y. Huang, Q. Liu, R. Li, Y. Wei, Integrated analysis of circular RNA-associated ceRNA network in cervical cancer: Observational study, Medicine 98 (2019) e16922.

[67] Y. Yi, Y. Liu, W. Wu, K. Wu, W. Zhang, Reconstruction and analysis of circRNAmiRNAmRNA network in the pathology of cervical cancer, Oncology Reports 41 (2019) 2209–2225.

[68] W. Jiang, P. Yuan, Molecular network-based identification of competing endogenous RNAs in bladder cancer, PLoS One 14 (2019) e0220118.

[69] D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nature protocols 4 (2009) 44–57.

[70] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Research 28 (2000) 27–30.

[71] F. R. Bonab, A. Baghbanzadeh, M. Ghaseminia, N. Bolandi, A. Mokhtarzadeh, M. Amini, K. Dadashzadeh, K. Hajiasgharzadeh, B. Baradaran, H. B. Baghi, Molecular pathways in the development of HPV-induced cervical cancer, EXCLI Journal 20 (2021) 320–337.

[72] X. Wang, H. K. Wang, Y. Li, M. Hafner, N. S. Banerjee, S. Tang, D. Briskin, C. Meyers, L. T. Chow, X. Xie, T. Tuschl, Z. Zheng, microRNAs are biomarkers of oncogenic human papillomavirus infections, Proceedings of the National Academy of Sciences of the United States of America 111 (2014) 4262–4267.

[73] F. Audenet, K. Attalla, J. P. Sfakianos, The evolution of bladder cancer genomics: What have we learned and how can we use it?, Urologic Oncology 36 (2018) 313–320.