



SeBioGraph: Semi-supervised Deep Learning for the Graph via Sustainable Knowledge Transfer

Yugang Ma¹, Qing Li^{2*}, Nan Hu³ and Lili Li^{4,5}

¹ School of Architecture and Urban Planning, Chongqing University, Chongqing, China, ² School of Computer Science, Northwestern Polytechnical University, Shaanxi, China, ³ School of Management Science and Real Estate, Chongqing University, Chongqing, China, ⁴ China Construction Science & Technology Group Co., Ltd. Shenzhen, China, ⁵ College of Civil and Environmental Engineering, Harbin Institute of Technology, Harbin, China

Semi-supervised deep learning for the biomedical graph and advanced manufacturing graph is rapidly becoming an important topic in both academia and industry. Many existing types of research focus on semi-supervised link prediction and node classification, as well as the application of these methods in sustainable development and advanced manufacturing. To date, most manufacturing graph neural networks are mainly evaluated on social and information networks, which improve the quality of network representation by integrating neighbor node descriptions. However, previous methods have not yet been comprehensively studied on biomedical networks. Traditional techniques fail to achieve satisfying results, especially when labeled nodes are deficient in number. In this paper, a new semi-supervised deep learning method for the biomedical graph via sustainable knowledge transfer called SeBioGraph is proposed. In SeBioGraph, both node embedding and graph-specific prototype embedding are utilized as transferable metric space characterized. By incorporating prior knowledge learned from auxiliary graphs, SeBioGraph further promotes the performance of the target graph. Experimental results on the two-class node classification tasks and three-class link prediction tasks demonstrate that the SeBioGraph realizes state-of-the-art results. Finally, the method is thoroughly evaluated.

Keywords: graph, semi-supervised deep learning, knowledge transfer, link prediction, node classification

OPEN ACCESS

Edited by:

Jingsha He,
Beijing University of Technology, China

Reviewed by:

Pascal Van Hentenryck,
Brown University, United States
Xiaoqiang Zhang,
Kunming University of Science and
Technology, China
Yongqin Tao,
Xi'an Jiaotong University, China

*Correspondence:

Qing Li
qingli@nwpu.edu.cn

Received: 07 February 2021

Accepted: 09 March 2021

Published: 01 April 2021

Citation:

Ma Y, Li Q, Hu N and Li L (2021)
SeBioGraph: Semi-supervised Deep
Learning for the Graph via Sustainable
Knowledge Transfer.
Front. Neurobot. 15:665055.
doi: 10.3389/fnbot.2021.665055

INTRODUCTION

Graph analysis can be used for various fields including linguistics (Akimushkin et al., 2017), social sciences (Rozemberczki et al., 2019), and biology (Theocharidis et al., 2009; Subramani et al., 2015). In biomedical graphics, the modeling of entities and their relations is indispensable for different tasks. Specifically, discovering synergistic or antagonistic effects between multiple drugs through drug-drug interaction graphs (Segura-Bedmar et al., 2011), developing new drugs for the disease through drug-disease graphs (Zhu Q. et al., 2013), and assisting doctors in clinical decision-making via disease-symptom graphs are some typical task scenarios (Li et al., 2019).

Biological graphs are notoriously complex and hard to decipher. Until now, many biomedical graph analytic methods have been proposed to analyze it (Grover and Leskovec, 2016; Fan et al., 2018; Zhang et al., 2018b). Most of these approaches transform the original data into vectorial data. In addition, the representation of the network is updated by integrating neighbor node

descriptions. Therefore, the structure information of the graph is preserved by the low-dimension representation of nodes. The various downstream tasks of the biomedical graph can be divided into three categories, as follow: clustering, link prediction, and node classification (Hamilton et al., 2017; Cai et al., 2018). Among them, the clustering analytic task aims to capture subsets of approximate nodes and then collect them together. The link prediction task is referred to predicting possible links or missing links. The node classification task is to determine the label of nodes.

However, these state-of-the-art graph analytic approaches are mainly evaluated on non-biomedical datasets. At the same time, most biomedical image analysis methods have limited receptive fields and only focus on shallow layers. These methods cannot perform medical traceability analysis. Especially, it becomes even more difficult to obtain satisfactory performance when the quantities of labeled nodes are scarce. Prediction of a link or classifying a node has been challenging, because manual annotations are often expensive, only a few nodes are involved. Most human-labeled biomedical graph features are always insufficient, while machine-labeled biomedical graph features are not sufficient to characterize entities. All these lead to the inability to build reliable and effective models. It follows that it is even more challenging to achieve semi-supervised deep learning for on biomedical graph than on independent identically distributed data (e.g., biomedical images).

More comparison details can be found in **Table 1**.

Matrix Factorization

Matrix factorization technology has been broadly utilized for graph data analysis areas, including but not limited to social networks, natural language processing, and computer vision. Through matrix factorization, different kinds of the graph can be presented as affinity. Besides, each vertex can be represented via a low-dimensional vector. Both Locally Linear Embedding (LLE) (Roweis and Saul, 2000) and Singular Value Decomposition (SVD) (De Lathauwer et al., 2000) are first focus on factorizing the 1st-order data matrix. And then, the method developed Laplacian Eigenmaps (LE) (Belkin and Niyogi, 2003) and Graph Factorization (GF) (Ahmed et al., 2013).

Due to the limitation of representation, researchers attempt to retain the graph structure by constructing various high order data proximity matrices, such as GraRep and HOPE. GraRep (Cao et al., 2015) proposes using k-step transition probability matrices to factorization. At the same time, it optimized through stochastic gradient descent, but it only applies to undirected graphs. HOPE (Ou et al., 2016) adopts network similarity measures to preserve high order network frameworks.

Random Walk

To a specified starting node and corresponding graph, the random walk approach chooses an adjacent node randomly and walk to this node. Generally, if the graph is too small or too large, this method is particularly useful to measure the graph completely.

DeepWalk (Perozzi et al., 2014) is a recently proposed method, which only suits social graphs with binary edges. In DeepWalk,

random walks are mainly adopted to enlarge the neighbor of every vertex. However, it fails to provide a clear goal that definitely expresses which graph properties are retained. At the same time, it only applies to the un-weighted graph. Similarly, Node2vec (Grover and Leskovec, 2016) reserves the higher-order proximity between various nodes. The node2vec uses a biased random walk. It can balance the depth-first and breadth-first search, so it can get more graph information than DeepWalk. Additionally, Struc2vec (Ribeiro et al., 2017) first utilizes a hierarchy weighted graph to encode the similarity between nodes. In this structure, each layer k is decided by the k-hop neighbor nodes.

Graph Neural Networks

Recently, GNNs are broadly adopted for data analysis (Kipf and Welling, 2016; Ravi and Larochelle, 2016; Finn et al., 2017; Huang et al., 2019; Liu et al., 2019; Zhang et al., 2019; Tang et al., 2020). It aims to encode the nodes with signals that lie in the receptive fields (Kipf and Welling, 2016). There are three lines of GNNs methods: non-supervised methods, semi-supervised methods, and supervised methods. All of these three approaches have gained great breakthroughs in diverse graph-based tasks, such as graph classification and node classification. However, these progressive methods are most analyzed and evaluated on non-biomedical graphs (e.g., social graphs) (Tang et al., 2015, 2016; Wang et al., 2016; Velickovic et al., 2017). Therefore, only a few studies have targeted biomedical networks (Wang et al., 2017; Gligorijevic et al., 2018; Ma et al., 2018; Zitnik et al., 2018).

In LINE (Finlayson et al., 2014), two functions are defined which include a 1st-order and a 2nd-order proximities function. And then, it minimizes the combination of the two functions. The first-order proximity function is much the same as that of the GF model (Ahmed et al., 2013). However, the LINE differs in that there are two joint probability distributions for each vertices pair, one using the embedding and the other using the adjacency matrix. GAE (Tang et al., 2016) input an adjacency matrix that relies on graph convolutional network encoder to obtain the higher-order dependencies of nodes. They have proved that the use of variational autoencoders can promote performance. Structural Deep Network Embedding (SDNE) (Wang et al., 2016) adopts auto-encoders to embedding graph nodes and acquire highly non-linear dependencies. In this model, there are two portions including supervised and unsupervised. For the first supervised portion, it imposes punishment when similar vertices are projected too far away from each other in the vector space. For the latter, it is equivalent to an auto-encoder and aims to find a representation for each node that can regenerate its neighbor.

We adopt a biomedical graphs analytic method that which has both excellent performance and enhanced interpretability. We are proposed to leverage the prior knowledge acquired from auxiliary graphs to enhance the performance of the target graphs. In addition to local topological structures, the auxiliary graphs and target graphs may share class-dependent node features. For this purpose, we proposed SeBioGraph, a new semi-supervised deep learning method for the biomedical graphs via knowledge transfer. Base on semi-supervised metric few-shot learning, the

TABLE 1 | A summary of 12 representative graph methods and existing work using them for a biomedical graph task.

Model	Node classification tasks			Link prediction tasks		
	Medical term type classification	Protein function prediction	Drug-disease association prediction	Drug-drug interaction prediction	Protein-protein interaction prediction	Chemical-protein interaction prediction
Matrix factorization						
Singular Value Decomposition (De Lathauwer et al., 2000)	N	Y (Cho et al., 2016)	Y (Dai et al., 2015)	N	Y (You et al., 2017)	N
Locally Linear Embedding (Roweis and Saul, 2000)	N	N	N	N	N	Y (Pliakos et al., 2019)
Laplacian (Belkin and Niyogi, 2003)	N	Y (Fan et al., 2018)	Y (Zhang et al., 2018b)	Y (Zhang et al., 2018a)	Y (Zhu L. et al., 2013)	N
GF (Ahmed et al., 2013)	N	N	Y (Yang et al., 2014; Zhang et al., 2018b)	Y (Zhang et al., 2018a)	N	N
GraRep (Cao et al., 2015)	N	N	N	N	N	N
HOPE (Ou et al., 2016)	N	N	N	N	N	N
Random walk						
DeepWalk (Perozzi et al., 2014)	N	Y (Cho et al., 2016; Kulmanov et al., 2018)	N	N	N	N
node2vec (Grover and Leskovec, 2016)	N	Y (Grover and Leskovec, 2016; Zitnik and Leskovec, 2017)	N	N	N	N
struc2vec (Ribeiro et al., 2017)	N	N	N	N	N	N
Neural network						
LINE (Tang et al., 2015)	N	N	N	N	N	N
GAE (Tang et al., 2016)	N	N	N	Y (Ma et al., 2018; Zitnik et al., 2018)	N	N
SDNE (Wang et al., 2016)	N	Y (Gligorijevic et al., 2018)	N	N	Y (Wang et al., 2017)	N

SeBioGraph intends to learn a transferable metric space, which predicts the label of each node through the class of the closest prototype to the node. It aims to optimize this mapping so that geometric relationships in the metric space reflect the structure of the original biomedical graphs. The metric space is to combine two parts: the embedded node and the prototype of each class.

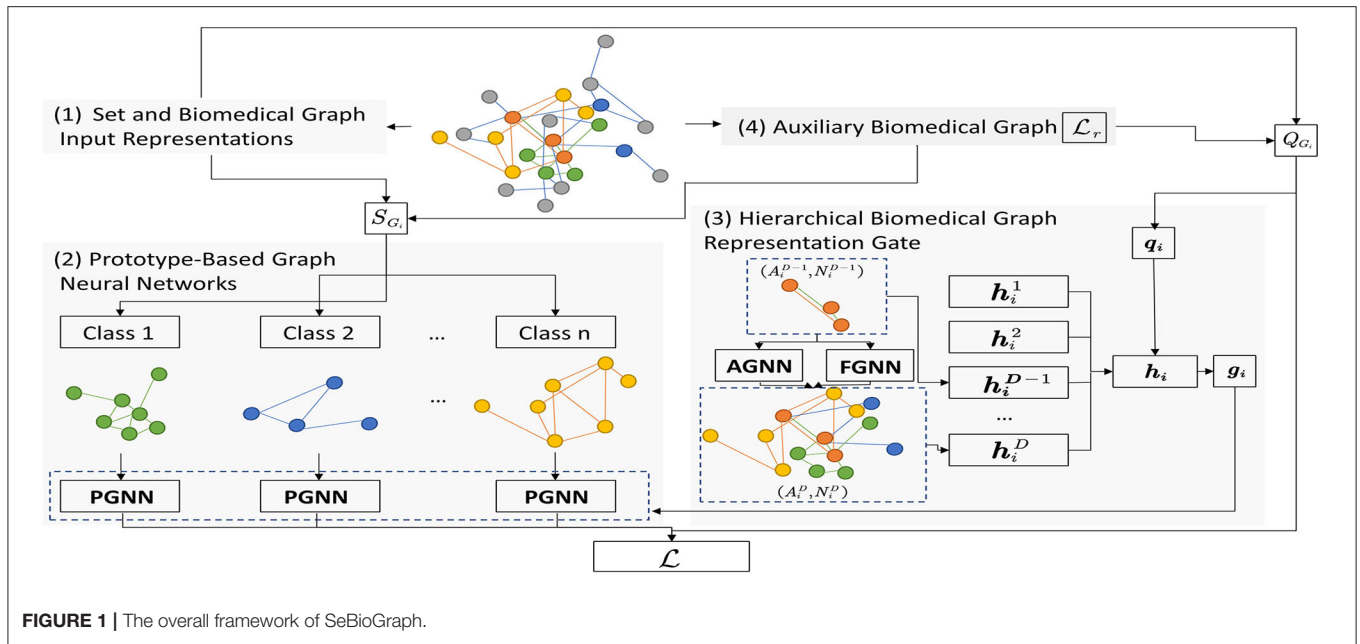
The construction of SeBioGraph consists of the following steps. At first, a graph encoder, which is mainly Graph Neural Networks (GNNs) (Kipf and Welling, 2016), is utilized to learn the information of every node. Accordingly, multiple node features (e.g., disease feature, drug chemical substructure features, and target protein feature) are mapped into a common subspace. In this subspace, it maintains the immutability of the original indication labels of nodes. Then, to obtain biomedical graphs' global information sufficiently, we construct a relational framework for all identical category samples. Through the embedding function of these two types of encrypted structured knowledge, the problem of lack of labeled nodes is compensated. After that, we design hierarchical biomedical graph representations gate to emphasize the analogous biomedical graphs having close metric spaces. Finally, in order to enhance the quality of node representation and robustness of training, we design an auxiliary graph constraint.

To sum up, our contributions can be outlined as follows:

- To the best of our awareness, it is the pioneering work to successfully perform the sustainable knowledge transfer to improve semi-supervised deep learning for the biomedical graphs;
- We propose a novel SeBioGraph to address the issue, which can simultaneously transfer all-graph-level and part-node-level structures across different graphs;
- SeBioGraph outperforms baseline models in two benchmark datasets in node classification tasks and five biomedical link prediction tasks, showing its potential to serve as an effective general-purpose representation learning algorithm for biomedical graph data.

METHODOLOGY

In this part, we introduce our proposed method SeBioGraph detailed. An illustration of the framework is shown in **Figure 1**. Here, we describe four parts of the proposed structure: set and biomedical graph input representations, prototype-based graph neural networks, hierarchical biomedical graph representation gate, and auxiliary biomedical graph.



Set and Biomedical Graph Input Representations

The input biomedical graph neural networks $G = (A, N)$ contain a collection of links and nodes, where is $A \in \{0, 1\}^{m \times m}$ the adjacent matrix, and $N = \{n_1, \dots, n_m\} \in R^{m \times h}$ is the node feature matrix. We set a batch of graphs $\{G_1, \dots, G_{N_i}\}$ sampled from a probability distribution ϵ .

Each node has two different functions in a biomedical graph: first is local interactions with different classes of neighbors; second is the same classes of neighbors. For example, (a) the structure between drug-disease nodes describing their co-association, as well as the structure between chemical-protein nodes describing their co-interaction, (b) the local interactions between protein nodes, chemical nodes, disease nodes, and drug nodes. So we will use S_i to denote a support-nodes set and Q_i to denote a query-nodes set, where:

Where m^{S_i} is a small set of labeled, and $y_{ij} \in \{1, \dots, K\}$ is the corresponding label.

In graph G_i , the effectiveness on Q_i is evaluated by the loss function L_i for every node j , as shown in Equation (3). where $|S_i^k|$ is the number of samples in S_i of class k , and Q_i^k denotes the sample set in Q_i of class k . We then predict its relevant label by jointing its embedding $f_\theta(A, n_{ij}^{q_i}) : R^h \rightarrow R^{h'}$ with representation $(f_\theta(A, n_{ij}^{s_i}), y_{ij}^{s_i})$ in support nodes set S_i through the similarity d .

Prototype-Based Graph Neural Networks

For each node in graph, the relation structure of the samples set belonging to class is extracted firstly. It is constructed based on similarity metrics (e.g, the inverse topological distance between nodes and the number of k -hop common neighbors). We denote the graph neural networks structured prototype as:

$$L_i = - \sum_k \sum_{(n_{ij}, y_{ij}^{q_i}) \in Q_i^k} \log \frac{\exp \left(-d \left(f_\theta(A, n_{ij}^{q_i}), \sum_{n_{ij}^s \in S_i^k} f_\theta(A, n_{ij}^s) / |S_i^k| \right) \right)}{\sum_{k'} \exp \left(-d \left(f_\theta(A, n_{ij}^{q_i}), \sum_{n_{ij}^s \in S_i^{k'}} (f_\theta(A, n_{ij}^{s_i}), y_{ij}^{s_i}) / |S_i^{k'}| \right) \right)} \quad (3)$$

$$c_i^k = \text{MaxPooling}_{j=1}^{m_i^k} \left(\text{PGNN}_\alpha \left(D_i^k, f_\theta(S_i^k) \right) [j] \right) \quad (4)$$

$$S_i = \left\{ \left(n_{ij}^{s_i}, y_{ij}^{s_i} \right) \right\}_{j=1}^{m^{S_i}} \quad (1)$$

$$Q_i = \left\{ \left(n_{ij}^{q_i}, y_{ij}^{q_i} \right) \right\}_{j=1}^{m^{Q_i}} \quad (2)$$

Where m^{S_i} is the number of nodes in $S_i^k, \text{PGNN}_\alpha(D_i^k, f_\theta(S_i^k))$ is the representation matrix in j -th node.

The globally shared parameter α of the PGNN is defined as a gate function g ; (more detail in section Results and Discussion) is

defined as:

$$\alpha_i = \mathbf{g}_i \circ \alpha = \sigma(\mathbf{W}_g \mathbf{h}_i + \mathbf{b}_g) \circ \alpha \quad (5)$$

Where \circ represents element-wise multiplication, W_g is a learnable weight parameter, and b_g is a learnable bias parameter.

Hierarchical Biomedical Graph Representation Gate

In order to show the different topologies specific to the graph, we following the popular method of hierarchical graph modeling (Ying et al., 2018). Compare the PGNN with globally shared parameters α , and the hierarchical biomedical graph representation gate combines two-level detail. There are biomedical graph node assignment and representation fusion.

Biomedical Graph Node Assignment

In this step, each low-level node k^d (in d -th level) is assigned to high-level node k^{d+1} community. The biomedical graph node assignment value is calculated by applying a softmax function, which is defined as follows:

$$p_i^{k^d \rightarrow k^{d+1}} = \frac{\exp\left(\text{AGNN}\left(\mathbf{A}_i^d, \mathbf{N}_i^d\right)\left[k^d, k^{d+1}\right]\right)}{\sum_{k^{d+1}=1}^{K^{d+1}} \exp\left(\text{AGNN}\left(\mathbf{A}_i^d, \mathbf{N}_i^d\right)\left[k^d, k^{d+1}\right]\right)} \quad (6)$$

where AGNN is the assigned value of the biomedical graph node, which is from the node k^d in the bottom layer d to the node k^{d+1} in the high layer $d + 1$, the $\text{AGNN}\left(\mathbf{A}_i^d, \mathbf{N}_i^d\right)\left[k^d, k^{d+1}\right] \in \mathbb{R}^1$. So we could be getting the biomedical graph node assignment matrix $P_i^{K^d \rightarrow K^{d+1}} \in \mathbb{R}^{K^d \times K^{d+1}}$. It includes each level of biomedical graph node assignment value $p_i^{k^d \rightarrow k^{d+1}}$.

Representation Fusion

For level $d + 1$, the adjacent matrix \mathbf{A}_i^{d+1} and the node feature matrix \mathbf{N}_i^{d+1} are defined as follows:

$$\mathbf{A}_i^{d+1} = \left(\mathbf{P}_i^{d \rightarrow d+1}\right)^T \mathbf{A}_i^d \mathbf{P}_i^{d \rightarrow d+1} \quad (7)$$

$$\mathbf{N}_i^{d+1} = \left(\mathbf{P}_i^{d \rightarrow d+1}\right)^T \text{FGNN}\left(\mathbf{A}_i^d, \mathbf{N}_i^d\right) \quad (8)$$

where FGNN is the fusion GNN. Then, the feature representation h_i^{d+1} can be obtained through jointing the information of all nodes, which is defined as follows:

$$\mathbf{h}_i^{d+1} = \text{MaxPooling}_{K^{d+1}=1}^{K^{d+1}} \left(\left(\mathbf{P}_i^{d \rightarrow d+1} \right)^T \text{FGNN}\left(\mathbf{A}_i^d, \mathbf{N}_i^d\right)\left[k^{d+1}\right] \right) \quad (9)$$

So we could be getting the biomedical graph structure representation set $\{h_i^1, \dots, h_i^D\}$ from varied levels. After that, the overall biomedical graph structure representation h_i is represented by the aggregator AGG of each level. We

use attention aggregators to represent different levels of contributions to the whole representation, which is defined as:

$$\mathbf{h}_i = \text{AttAGG}\left(\{h_i^1, \dots, h_i^D\}\right) = \sum_{d=1}^D \frac{\mathbf{q}_i^T \mathbf{h}_i^d}{\sum_{d'=1}^D \mathbf{q}_i^T \mathbf{h}_i^{d'}} \mathbf{h}_i^d \quad (10)$$

Where q_i is a learnable query vector.

The biomedical graph representation gate g_i maps the specific graph representation h_i to the identical space of parameter α_i as follow:

$$\mathbf{g}_i = \mathcal{T}\left(\mathbf{h}_i\right) = \sigma\left(\mathbf{W}_g \mathbf{h}_i + \mathbf{b}_g\right) \quad (11)$$

Thus, Equation (5) would be updated.

Auxiliary Biomedical Graph

Graph semi-supervised deep learning aims to learn a well-generalized embedding function from previous graphs. This function can be used to a new graph with a small support set. At the same time, we need to design a new constraint loss function to optimize the training robustness and the quality of node embedding.

$$L_d\left(\mathbf{A}_i, \mathbf{N}_i\right) = \left\| \mathbf{A}_i - \text{GNN}_{dec}\left(\mathbf{A}_i, \mathbf{H}_i\right) \text{GNN}_{dec}^T\left(\mathbf{A}_i, \mathbf{H}_i\right) \right\|_F^2 \quad (12)$$

Where $\|\cdot\|_F$ represents the Frobenius norm.

In the end, the optimization problem of SeBioGraph is defined as follows:

$$\text{Min } \Phi \leftarrow \Phi - \gamma \nabla_{\Phi} \sum_{i=1}^{N_i} L_i\left(\mathbf{A}_i, \mathbf{N}_i\right) + \beta L_d\left(\mathbf{A}_i, \mathbf{N}_i\right) \quad (13)$$

where Φ represents all learnable parameters.

EXPERIMENTS

Tasks and Dataset

In this section, we evaluate the quality of SeBioGraph for two-class biomedical graph tasks in eight datasets. The first-class tasks are node classification, i.e., protein-protein interaction with functional annotations and semantic type classification of medical term. The second-class tasks are link prediction, i.e., chemical-disease interaction prediction, drug-drug interaction prediction, chemical-protein interaction prediction.

Node Classification Tasks

The task of node classification is a very important first step of graph analysis. For a partly labeled graph, this task is to predict the class of unlabeled nodes. In 2018, Gligorijevic proposed to obtain the representation of proteins via developing deepNF models (Gligorijevic et al., 2018). In the same year, Lim adopts a method based on regularized Laplacian kernel, which can learn the low-dimensional graph feature of proteins (Fan et al., 2018). To evaluate the impact of semi-supervised deep learning biomedical graphs, we use classification tasks based on a single unlabeled node. Here, SeBioGraph focused on the following two kinds of node classification tasks benchmark experimental datasets.

Medical Term Semantic Type Classification

We utilize a set of medical terms that can be obtained publicly and their co-occurrence statistics datasets (Clin Term COOC) (Finlayson et al., 2014). For two terms, we compute its co-occurrence frequencies based on 1-day. Besides, we only save those edges whose PPMI is greater than two. The Clin Term COOC datasets contain 48,651 nodes.

Protein-Protein Interaction (PPI) With Functional Annotations

There are two PPI graphs datasets containing functional annotations, which are node2vec and MashUp. The first one is Node2vec (Grover and Leskovec, 2016), and it contains the 3,890 proteins node. The second one is MashUp (Cho et al., 2016), which contains six individual PPI graphs. It contains 16,143 proteins node and 300,181 protein-protein interactions.

Link Prediction Tasks

In the biomedical field, the discovery of new links (a.k.a. association, interactions) is an important task. For a series of biomedical entities and links, the purpose of this task is to predict some other hidden interactions of entities. Most previous methods focus on establishing biological feature engineering, such as graph topological similarities (Hamilton et al., 2017) and chemical substructures (Liang et al., 2017). After that, the semi-supervised graph inference model or supervised deep learning methods are utilized to predict potential interactions. In order to compare the performance of our model with the previous model more comprehensively. To compare performance with previous models, SeBioGraph focused on the following five kinds of link prediction tasks benchmark experimental datasets.

Chemical-Disease Association (CDA) Prediction

The Comparative Toxicogenomics Database (CTD) (Davis et al., 2019) is a public biomedical graph based on literature, which manually labeled associations between gene products, chemicals, diseases, and so on. We filtered the association biomedical graph between 12,765 chemical-disease nodes in the CTD graph.

Drug-Disease Association (DDA) Prediction

The DDA prediction database is NDF-RT (National Drug File Reference Terminology) (Bodenreider, 2004) produced by the U.S. Department of Veterans Affairs. The drug characteristics are including related diseases, physiologic effects, and ingredients. We filtered the association biomedical graph between 13,545 drug-disease nodes in the NDF-RT graph.

Drug-Drug Interaction (DDI) Prediction

The DDI prediction database is DrugBank (Wishart et al., 2018), which contains detailed data about drugs including mechanisms, interactions and drug targets.

Protein-Protein Interaction (PPI) Prediction

The PPI prediction database is STRING (Szklarczyk et al., 2015), which includes indirect (functional) and direct (physical) associations. We filtered the association biomedical graph between 15,131 protein-protein nodes in the STRING graph.

Chemical-Protein Interaction (CPI) Prediction

The CPI prediction database is STITCH (Kuhn et al., 2007), which includes the interaction information of more than 68,000 different chemicals and 2,200 drugs. It links them to 1.5 million genes across 373 genomes. We filtered the association biomedical graph between 4,138,421 chemical-protein nodes in the STITCH graph.

Experiments on the Parameter Settings

In these experiments, we use an open Python package of OpenNE to train the node representation in the SeBioGraph. For the link prediction tasks, our model is split the 80% for the training set and 20% for the testing set. In this work, we follow the traditional semi-supervised deep learning settings (Finn et al., 2017; Snell et al., 2017). The is a two-layer graph convolutional structure. In each layer, there are 32 neurons. For PGNN, AGNN, and FGNN, we adopt a one-layer graph convolutional structure as the substitute for GNN. Other weights are randomly initialized from a zero-mean Gaussian distribution. We tuned all the hyperparameters for our model 5-fold cross-validation for the optimization of the hyperparameters and report as final results.

Results and Discussion

Node Classification Tasks

Table 2 illustrates the result of various biomedical graph analytic methods on protein function prediction and medical term semantic type classification task. We use two F1 weighted criteria including Micro-F1 and Macro-F1 to evaluate the performance of different approaches. For the Macro-F1, it computes metrics for every label type, and then acquires their un-weighted mean. For the Micro-F1, it computes metrics globally by counting all samples.

We divided the traditional methods into four groups: matrix factorization, random walk, graph neural networks, and our model. First, the matrix factorization methods used many features to the classifier, such as SVD, LLE, LE, GF, GraRep, and HOPE. According to the result, they achieved a Micro-F1 score of $42.4 \pm 0.6\%$ (GraRep) and a Macro-F1 score of $18.6 \pm 0.7\%$. This shows that modeling the first-order proximity directly could be sufficient for basic classification nodes. The random walk model can catch more different functions for nodes in different subgraphs. The Node2vec performs better since it mostly pays attention to modeling the structural identity of each node. But the biomedical graph may not exist a clear structural role. Its accuracy is limited. The other model of graph neural network methods are an effective way for the node classification task. There are GNN-based models such as LINE (Tang et al., 2015), GAE (Tang et al., 2016), and SDNE (Wang et al., 2016). However, the graph neural network methods may have several flaws. On the one hand, it may be inaccurate. On the other hand, the parsing time will be exponentially increased by data. The last model is our model for SeBioGraph, which shows the advantage of prior knowledge obtained from the learned graphs. Experimental results show that our SeBioGraph reach an improvement of 1.2% on the Macro-F1 score and 3.8% on the Micro-F1 score. Obviously, it exceeds the second-best Node2vec.

TABLE 2 | Comparison between SeBioGraph and other node classification methods on three biomedical graph datasets.

Method	Clin Term COOC		Node2vec PPI		MashUp PPI	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
Matrix factorization						
SVD (De Lathauwer et al., 2000)	42.0 ± 0.5%	18.6 ± 0.7%	22.8 ± 1.1%	17.9 ± 1.1%	34.7 ± 1.4%	29.7 ± 1.4%
LLE (Roweis and Saul, 2000)	32.5 ± 0.7%	13.9 ± 0.4%	18.1 ± 0.9%	13.8 ± 1.2%	19.4 ± 1.3%	37.5 ± 1.2%
LE (Belkin and Niyogi, 2003)	31.3 ± 0.5%	7.3 ± 0.2%	10.1 ± 0.8%	7.0 ± 0.7%	13.2 ± 0.9%	10.7 ± 0.8%
GF (Ahmed et al., 2013)	35.2 ± 0.7%	14.3 ± 0.9%	16.8 ± 1.1%	12.1 ± 1.1%	29.0 ± 1.5%	23.7 ± 1.6%
GraRep (Cao et al., 2015)	42.4 ± 0.6%	17.7 ± 0.5%	23.8 ± 1.0%	19.3 ± 1.3%	33.4 ± 1.1%	28.3 ± 1.1%
HOPE (Cao et al., 2015)	39.5 ± 0.5%	16.3 ± 0.6%	20.8 ± 1.1%	15.2 ± 1.1%	32.2 ± 1.3%	26.6 ± 1.3%
Random walk						
DeepWalk (Perozzi et al., 2014)	47.2 ± 0.5%	22.7 ± 0.7%	24.3 ± 0.1%	19.4 ± 1.1%	35.7 ± 1.1%	31.1 ± 1.2%
Node2vec (Grover and Leskovec, 2016)	47.9 ± 0.5%	23.1 ± 1.0%	24.3 ± 0.9%	19.0 ± 1.1%	36.7 ± 1.2%	31.3 ± 1.3%
Struc2vec (Ribeiro et al., 2017)	25.3 ± 0.6%	3.8 ± 0.1%	9.4 ± 0.6%	6.1 ± 0.4%	12.0 ± 1.0%	8.7 ± 0.8%
Graph Neural networks						
LINE (Tang et al., 2015)	45.3 ± 0.6%	20.5 ± 0.8%	23.6 ± 1.1%	17.6 ± 1.2%	35.2 ± 1.7%	29.6 ± 1.7%
GAE (Tang et al., 2016)	29.5 ± 1.2%	7.1 ± 0.7%	23.7 ± 1.4%	18.6 ± 1.4%	35.8 ± 1.3%	30.7 ± 1.4%
SDNE (Wang et al., 2016)	27.1 ± 1.6%	4.2 ± 0.7%	9.8 ± 1.0%	4.7 ± 0.7%	17.8 ± 1.3%	10.9 ± 1.2%
Our model						
SeBioGraph	51.7 ± 0.9%	24.3 ± 1.0%	23.6 ± 1.1%	21.7 ± 1.0%	42.4 ± 1.4%	35.4 ± 0.9%
- Auxiliary	46.5 ± 1.1%	19.2 ± 0.7%	21.9 ± 0.9%	19.9 ± 1.0%	36.8 ± 1.2%	31.5 ± 0.9%

The meaning of bold is the best F1 precision.

TABLE 3 | Comparison of accuracy value between SeBioGraph and other link prediction methods on five biomedical graph datasets.

Method	CTD CDA	NDF-RT DDA	DrugBank DDI	STRING PPI	STITCH CPI
Matrix factorization					
SVD (De Lathauwer et al., 2000)	93.6 ± 0.2%	77.9 ± 0.3%	91.9 ± 0.1%	86.7 ± 0.1%	31.7 ± 0.4%
LLE (Roweis and Saul, 2000)	86.5 ± 0.3%	89.7 ± 0.4%	89.1 ± 0.2%	79.8 ± 1.0%	29.4 ± 0.3%
LE (Belkin and Niyogi, 2003)	85.6 ± 0.4%	93.0 ± 0.3%	79.6 ± 0.2%	63.9 ± 2.1%	23.2 ± 0.5%
GF (Ahmed et al., 2013)	88.4 ± 0.4%	72.0 ± 0.6%	88.2 ± 0.3%	81.7 ± 0.5%	32.1 ± 0.3%
GraRep (Cao et al., 2015)	96.0 ± 0.1%	96.3 ± 0.1%	92.5 ± 0.1%	89.4 ± 0.1%	41.4 ± 0.4%
HOPE (Ou et al., 2016)	95.1 ± 0.1%	94.9 ± 0.1%	92.3 ± 0.1%	83.9 ± 0.1%	42.7 ± 0.2%
Random walk					
DeepWalk (Perozzi et al., 2014)	92.9 ± 0.2%	78.3 ± 0.4%	92.1 ± 0.1%	88.4 ± 0.1%	26.4 ± 0.3%
Node2vec (Grover and Leskovec, 2016)	91.1 ± 0.2%	81.9 ± 0.5%	90.2 ± 0.1%	82.8 ± 0.3%	37.7 ± 0.6%
Struc2vec (Ribeiro et al., 2017)	96.5 ± 0.1%	95.8 ± 0.1%	90.4 ± 0.1%	90.9 ± 0.1%	44.0 ± 0.1%
Graph Neural networks					
LINE (Tang et al., 2015)	96.5 ± 0.1%	96.2 ± 0.2%	90.5 ± 0.2%	85.9 ± 0.3%	36.2 ± 0.4%
GAE (Tang et al., 2016)	93.7 ± 0.1%	81.3 ± 0.7%	91.7 ± 0.1%	90.0 ± 0.1%	35.8 ± 0.1%
SDNE (Wang et al., 2016)	93.5 ± 1.0%	94.4 ± 0.4%	91.1 ± 0.6%	88.4 ± 0.8%	37.8 ± 0.8%
Our model					
SeBioGraph	97.2 ± 0.5%	96.4 ± 0.6%	93.1 ± 0.3%	89.9 ± 0.6%	48.8 ± 0.7%
- Auxiliary	93.8 ± 0.5%	87.1 ± 0.5%	88.9 ± 0.3%	85.6 ± 0.4%	39.1 ± 0.5%

To demonstrate the effect of each portion in SeBioGraph, the ablation experiments are implemented. By observing the results, we find that the auxiliary biomedical graph mechanism in SeBioGraph significantly outperforms Node2vec. Evidently, the auxiliary biomedical graph module plays an indispensable role

in the experiment. Experimental results show that our model achieved a Micro-F1 score of $51.7 \pm 0.9\%$, which performs better than other approaches. The auxiliary biomedical graph module enhances the performance by 5.2% than the model not applied it.

Link Prediction Tasks

For link prediction tasks, we comparison accuracy values on the five biomedical graph datasets: CTD CDA, NDF-RTDDA, DrugBank DDI, STRING PPI, and STITCH CPI. We report the averaged accuracy with 95% confidence intervals on the 10-shot classification in **Table 3**. It manifests the accuracy value generated for early prediction using graph neural networks, random walk and matrix factorization methods. The results attest that our SeBioGraph achieves a high accuracy value of $97.2 \pm 0.5\%$, which excels all competing for state-of-the-art approaches.

Generally, compared to traditional methods [e.g., LLE (Roweis and Saul, 2000), LE (Belkin and Niyogi, 2003), and GF (Ahmed et al., 2013)], the existing proposed approaches have greatly enhanced the performance of link prediction. Especially in the STITCH CPI dataset with large-scale aggregation and edges, our methods are more effective. These results demonstrate that our methods can improve prediction performance in various biological link prediction tasks. Based on these results, we made the following observations: First, we can see that SeBioGraph significantly enhances the final result, which shows that transferring knowledge from learned graphs is effective. Second, our SeBioGraph achieves the best on all five datasets, indicating the robustness of prototype-based graph neural networks, auxiliary biomedical graph and hierarchical biomedical graph representation gate. In addition, as a metric distance-based semi-supervised method, SeBioGraph outperforms other existing methods and on the other hand, it achieves better performance than non-supervised methods and supervised methods.

REFERENCES

- Ahmed, A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., and Smola, A. J. (2013). "Distributed large-scale natural graph factorization," in *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro), 37–48.
- Akimushkin, C., Amancio, D. R., and Oliveira, O. N. Jr. (2017). Text authorship identified using the dynamics of word co-occurrence networks. *PLoS ONE* 12:e0170527. doi: 10.1371/journal.pone.0170527
- Belkin, M., and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 15, 1373–1396. doi: 10.1162/08997660321780317
- Bodenreider, O. (2004). The unified medical language system (umls): integrating biomedical terminology. *Nucleic Acids Res.* 32(Suppl_1), D267–D270. doi: 10.1093/nar/gkh061
- Cai, H., Zheng, V. W., and Chang, K. C.-C. (2018). A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transact. Know. Data Eng.* 30, 1616–1637. doi: 10.1109/TKDE.2018.2807452
- Cao, S., Lu, W., and Xu, Q. (2015). "Grarep: learning graph representations with global structural information," in *Proceedings of the 24th ACM International Conference on Information and Knowledge Management* (Melbourne, VIC), 891–900.
- Cho, H., Berger, B., and Peng, J. (2016). Compact integration of multi-network topology for functional analysis of genes. *Cell Syst.* 3, 540–548. doi: 10.1016/j.cels.2016.10.017
- Dai, W., Liu, X., Gao, Y., Chen, L., Song, J., Chen, D., et al. (2015). Matrix factorization-based prediction of novel drug indications by integrating genomic space. *Comput. Math. Methods Med.* 2015:275045. doi: 10.1155/2015/275045
- Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., McMorran, R., Wieggers, J., et al. (2019). The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.* 47, D948–D954. doi: 10.1093/nar/gky868

CONCLUSION

In this paper, we propose a novel framework called SeBioGraph. Our method strengthens the effectiveness of semi-supervised node classification and link prediction on a new target biomedical graph through conducting knowledge transfer which is learned from auxiliary graphs. Built upon the semi-supervised deep learning, SeBioGraph joints graph-level and local node-level global knowledge to learn a transferable metric space characterized. The experimental results show our proposed model is effective for two-class biomedical graph tasks in eight datasets.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YM is responsible for designing the framework. QL is responsible for the writing of the paper and specific experiments. NH and LL are responsible for designing the framework and idea. All authors contributed to the article and approved the submitted version.

- De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* 21, 1253–1278. doi: 10.1137/S0895479896305696
- Fan, J., Cannistra, A., Fried, I., Lim, T., Schaffner, T., Crovella, M., et al. (2018). A multi-species functional embedding integrating sequence and network structure. *bioRxiv [Preprint]*. doi: 10.1101/229211
- Finlayson, S. G., LePendu, P., and Shah, N. H. (2014). Building the graph of medicine from millions of clinical narratives. *Sci. Data* 1:140032. doi: 10.1038/sdata.2014.32
- Finn, C., Abbeel, P., and Levine, S. (2017). "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70 (JMLR.org), 1126–1135.
- Gligorijevic, V., Barot, M., and Bonneau, R. (2018). deepnf: deep network fusion for protein function prediction. *Bioinformatics* 34, 3873–3881. doi: 10.1093/bioinformatics/bty440
- Grover, A., and Leskovec, J. (2016). "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA), 855–864.
- Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Representation learning on graphs: methods and applications. *arXiv*.
- Huang, C., Wu, X., Zhang, X., Zhang, C., Zhao, J., Yin, D., et al. (2019). "Online purchase prediction via multi-scale modeling of behavior dynamics," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK), 2613–2622.
- Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv*.
- Kuhn, M., von Mering, C., Campillos, M., Jensen, L. J., and Bork, P. (2007). Stitch: interaction networks of chemicals and proteins. *Nucleic Acids Res.* 36(Suppl_1), D684–D688. doi: 10.1093/nar/gkm795

- Kulmanov, M., Khan, M. A., and Hoehndorf, R. (2018). Deepgo: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* 34, 660–668. doi: 10.1093/bioinformatics/btx624
- Li, Q., Huang, L. F., Zhong, J., Li, L., Li, Q., and Hu, J. (2019). “Data-driven discovery of a sepsis patients severity prediction in the icu via pre-training bilstm networks,” in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (San Diego, CA: IEEE), 668–673.
- Liang, X., Zhang, P., Yan, L., Fu, Y., Peng, F., Qu, L., et al. (2017). Lrssl: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics* 33, 1187–1196. doi: 10.1093/bioinformatics/btw770
- Liu, Z., Chen, C., Li, L., Zhou, J., Li, X., Song, L., et al. (2019). “Geniepath: graph neural networks with adaptive receptive paths,” in *Proceedings of the AAAI Conference on Artificial Intelligence* (New York, NY), 4424–4431.
- Ma, T., Xiao, C., Zhou, J., and Wang, F. (2018). Drug similarity integration through attentive multi-view graph auto-encoders. *arXiv*.
- Ou, M., Cui, P., Pei, J., Zhang, Z., and Zhu, W. (2016). “Asymmetric transitivity preserving graph embedding,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA), 1105–1114.
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). “Deepwalk: online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–710.
- Pliakos, K., Vens, C., and Tsoumakas, G. (2019). Predicting drug-target interactions with multi-label classification and label partitioning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2019.2951378
- Ravi, S., and Larochelle, H. (2016). *Optimization as a Model for Few-Shot Learning*. Available online at: <https://openreview.net/forum?id=rjY0-Kcll>
- Ribeiro, L. F., Saverese, P. H., and Figueiredo, D. R. (2017). “struc2vec: learning node representations from structural identity,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Halifax, NS), 385–394.
- Roweis, S. T., and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326 doi: 10.1126/science.290.5500.2323
- Rozemberczki, B., Davies, R., Sarkar, R., and Sutton, C. (2019). “Gemsec: graph embedding with self clustering,” in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (Vancouver, BC), 65–72.
- Segura-Bedmar, I., Martínez, P., and de Pablo-Sánchez, C. (2011). Using a shallow linguistic kernel for drug–drug interaction extraction. *J. Biomed. Inform.* 44, 789–804. doi: 10.1016/j.jbi.2011.04.005
- Snell, J., Swersky, K., and Zemel, R. (2017). “Prototypical networks for few-shot learning,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, CA), 4077–4087. Available online at: https://www.researchgate.net/publication/315096921_Prototypical_Networks_for_Few-shot_Learning
- Subramani, S., Kalpana, R., Monickaraj, P. M., and Natarajan, J. (2015). Hpiminer: a text mining system for building and visualizing human protein interaction networks and pathways. *J. Biomed. Inform.* 54, 121–131. doi: 10.1016/j.jbi.2015.01.006
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2015). “Line: Largescale information network embedding,” in *Proceedings of the 24th International Conference on World Wide Web* (Florence), 1067–1077.
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2016). Variational graph auto-encoders. *arXiv*.
- Tang, X., Li, Y., Sun, Y., Yao, H., Mitra, P., and Wang, S. (2020). “Transferring robust graph neural network against poisoning attacks,” in *WSDM '20: Proceedings of the 13th International Conference on Web Search and Data Mining*. 600–608. doi: 10.1145/3336191.3371851
- Theocharidis, A., Van Dongen, S., Enright, A. J., and Freeman, T. C. (2009). Network visualization and analysis of gene expression data using biolayout express 3d. *Nat. Protoc.* 4:1535. doi: 10.1038/nprot.2009.177
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. *arXiv*.
- Wang, D., Cui, P., and Zhu, W. (2016). “Structural deep network embedding,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, CA), 1225–1234.
- Wang, Y.-B., You, Z.-H., Li, X., Jiang, T.-H., Chen, X., Zhou, X., et al. (2017). Predicting protein–protein interactions from protein sequences by a stacked sparse autoencoder deep neural network. *Mol. BioSyst.* 13, 1336–1344. doi: 10.1039/C7MB00188F
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037
- Yang, J., Li, Z., Fan, X., and Cheng, Y. (2014). Drug–disease association and drugrepositioning predictions in complex diseases using causal inference–probabilistic matrix factorization. *J. Chem. Inf. Model.* 54, 2562–2569. doi: 10.1021/ci500340n
- Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., and Leskovec, J. (2018). “Hierarchical graph representation learning with differentiable pooling,” in *Proceedings of the 32th Advances in Neural Information Processing Systems* (Montréal), 4800–4810. Available online at: https://www.researchgate.net/publication/325986373_Hierarchical_Graph_Representation_Learning_withDifferentiable_Pooling
- You, Z.-H., Li, X., and Chan, K. C. (2017). An improved sequence-based prediction protocol for protein–protein interactions using amino acids substitution matrix and rotation forest ensemble classifiers. *Neurocomputing* 228, 277–282. doi: 10.1016/j.neucom.2016.10.042
- Zhang, C., Song, D., Huang, C., Swami, A., and Chawla, N. V. (2019). “Heterogeneous graph neural network,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage, AK), 793–803.
- Zhang, W., Chen, Y., Li, D., and Yue, X. (2018a). Manifold regularized matrix factorization for drug–drug interaction prediction. *J. Biomed. Inform.* 88, 90–97. doi: 10.1016/j.jbi.2018.11.005
- Zhang, W., Yue, X., Lin, W., Wu, W., Liu, R., Huang, F., et al. (2018b). Predicting drug–disease associations by using similarity constrained matrix factorization. *BMC Bioinform.* 19, 1–12. doi: 10.1186/s12859-018-2220-4
- Zhu, L., You, Z.-H., and Huang, D.-S. (2013). Increasing the reliability of protein–protein interaction networks via non-convex semantic embedding. *Neuro Comput.* 121, 99–107. doi: 10.1016/j.neucom.2013.04.027
- Zhu, Q., Freimuth, R. R., Pathak, J., Durski, M. J., and Chute, C. G. (2013). Disambiguation of pharngkb drug–disease relations with ndf-rt and spl. *J. Biomed. Inform.* 46, 690–696. doi: 10.1016/j.jbi.2013.05.005
- Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34, i457–i466. doi: 10.1093/bioinformatics/bty294
- Zitnik, M., and Leskovec, J. (2017). Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* 33, i190–i198. doi: 10.1093/bioinformatics/btx252

Conflict of Interest: LL was employed by company China Construction Science & Technology Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ma, Li, Hu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.