**OXFORD**

# Network embedding in biomedical data science

## Chang Su, Jie Tong, Yongjun Zhu, Peng Cui and Fei Wang

Corresponding author: Fei Wang, Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, Cornell University, 425 East 61 Street, New York, NY 10065, USA. Tel.: +1-646-962-9405; E-mail: few2001@med.cornell.edu

## Abstract

Owning to the rapid development of computer technologies, an increasing number of relational data have been emerging in modern biomedical research. Many network-based learning methods have been proposed to perform analysis on such data, which provide people a deep understanding of topology and knowledge behind the biomedical networks and benefit a lot of applications for human healthcare. However, most network-based methods suffer from high computational and space cost. There remain challenges on handling high dimensionality and sparsity of the biomedical networks. The latest advances in network embedding technologies provide new effective paradigms to solve the network analysis problem. It converts network into a low-dimensional space while maximally preserves structural properties. In this way, downstream tasks such as link prediction and node classification can be done by traditional machine learning methods. In this survey, we conduct a comprehensive review of the literature on applying network embedding to advance the biomedical domain. We first briefly introduce the widely used network embedding models. After that, we carefully discuss how the network embedding approaches were performed on biomedical networks as well as how they accelerated the downstream tasks in biomedical science. Finally, we discuss challenges the existing network embedding applications in biomedical domains are faced with and suggest several promising future directions for a better improvement in human healthcare.

**Key words:** biomedical networks; biomedical knowledge graphs; biomedical informatics; network embedding; graph embedding; network-based learning

## Introduction

Recent advances in biomedical research as well as computer software and hardware technologies have led to an inrush of a large number of relational data interlinking drugs, genes, proteins, chemical compounds, diseases and medical concepts extracted from clinical data [1–3]. The representation of a biomedical object contains its relationship to other objects; in other words, the data is in the form of a network comprised of nodes (biomedical entities) and edges (relations between nodes). The availability of such relational data has greatly facilitated the biomedical studies, such as network biology [4–6], network medicine [7, 8], pharmacogenomics [9], disease diagnosis [10, 11], clinical phenotyping [12], etc.

Analyzing and modeling the biomedical data with network structure rely on a thorough understanding of network topology. Numerous network-based learning methods have been developed to explore reliable tools for multiple applications. Although existing methods show capacity of processing networks and demonstrate great promises [2, 13–16], they usually suffer from high computational and space cost, owning to high dimensionality and sparsity of the networks. The challenges are further complicated by various emerging heterogeneous biomedical networks, including the biomedical knowledge graphs
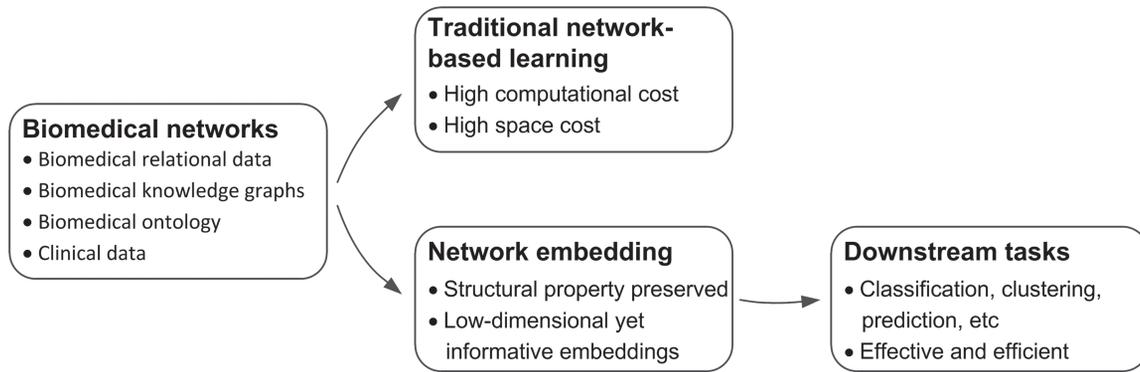
**Figure 1**. Illustration of network embedding in biomedical research. Traditional network-based learning conforms to network structure, hence suffers from high computational and space cost. In contrast, network embedding projects biomedical network into a low-dimensional space while preserving structural properties, hence traditional machine learning methods can be easily applied to the low-dimensional embedding vectors for downstream biomedical tasks.

(e.g. PharmGKB [17], DrugBank [18] and TTD [19]), biomedical ontologies [i.e. gene ontology (GO) [20], human phenotype ontology [21] and disease ontology [22]] and heterogeneous networks extracted from clinical data, which commonly consist of multiple types of nodes and edges and complex biomedical rules.

Network embedding provides another effective yet efficient way to address the network analysis problem (as shown in Figure 1). Specifically, network embedding aims at converting the network into a low-dimensional space while structural information of the network is preserved [23–26]. In this way, nodes and/or edges of the network can be represented as compacted yet informative vectors in the embedding space. Therefore, typical non-network-based machine learning methods such as linear regression, Support Vector Machine (SVM) and decision forest, which have been demonstrated to be effective and efficient as the state-of-the-art techniques, can be applied to such vectors. Network embedding methods have showed effectiveness and potential on network analysis and hence have introduced exciting opportunities for biomedical data science. Efforts of applying network embedding to improve biomedical data analysis are already planned or underway. However, network embedding has not been extensively evaluated for a broad range of biomedical issues that could benefit from its capabilities. The biomedical networks are sparse, noisy, incomplete, heterogeneous and usually consist of biomedical text and other domain knowledge. It makes embedding tasks more complicated than other application fields. To address this, it is important to understand and compare the existing network embedding models, as well as to investigate how they were implemented on biomedical data. Therefore, it can help us gain better insights on directions for future work.

In this article, we discuss existing and forthcoming applications of network embedding in biomedical informatics, while highlighting the key aspects to significantly accelerate biomedical data science. Here we do not provide a comprehensive network embedding background on technical details that has been well reviewed by previous works [23–26]. Instead of general applications of network embedding, we focus on biomedical data only, including drug-related networks and knowledge graphs, multi-omics networks, biomedical knowledge graphs and heterogeneous networks extracted from clinical data. To the best of our knowledge, there is no detailed review discussing any insights of impacts of network embedding techniques on biomedical science. To fill in this gap, we briefly introduce

the state-of-the-art network embedding models and review their applications in biomedical domain. We further discuss challenges and future research directions toward a better usage of network embedding to improve the human healthcare research.

## Network embedding methodologies

In this section, we introduce the state-of-the-art network embedding methods and propose taxonomy by grouping the methods into two categories: non-attributed network embedding and attributed network embedding (as shown in Table 1 and Figure 2).

### Non-attributed network embedding

A non-attributed network is also known as homogeneous network, of which all nodes and edges belong to a unique type, respectively. In practice, learning embeddings is to preserve local and/or global structural property measured by the 1st-order proximity and/or high-order proximity, respectively. We next introduce the non-attributed network embedding methods lying in how they define the proximity to preserve.

#### *Matrix factorization-based methods*

The 1st category of non-attributed network embedding is the matrix factorization-based methods. The pioneer efforts, such as the locally linear embedding (LLE) [27] and Laplacian eigenmaps (LE) [28], first construct the network from the non-relational data by using some constructing strategies, e.g. k-nearest neighbor approach. Then they extract the adjacency matrix that holds proximity in terms of similarity between nodes and their neighbors and factorize it to obtain the embedding vectors of nodes. The distinction is that LLE defines objective based on a linear neighborhood combination assumption, while LE transforms the embedding task into eigenvector problem of graph Laplacian matrix.

Some other works directly factorize the proximity matrices. A simple version is the graph factorization (GF) [29], which models the proximity matrix regarding the presence of each edge. GraRep [30] is a further work similar to GF, which constructs the high-order proximity matrix based on transition probability by a random walk with specific length. HOPE [31] aims at preserving high-order proximity according to the asymmetric transitivity for directed networks and

**Table 1.** Network embedding models surveyed in this study. The 1st column is the subcategory of network embedding models. The 2nd and 3rd columns present the names and release years of network embedding models, respectively. The 4th column introduces the main architectures of the network embedding models. The 5th column presents the learning methods of the network embedding models. The 6th column gives the time complexities of the methods. The last column lists URLs linked to the source codes of the network embedding models. $n$ and $m$ are numbers of nodes (entities) and edges (relations) in the network, respectively; $d$ and $k$ are dimensions of embedding spaces of node and edge, respectively; $l$ is the predefined length of random walk, $\mu$ is the average degree of node

| Category | Algorithm | Year | Architecture | Learning method | Time complexity | Source code |
|---|---|---|---|---|---|---|
| *Non-attributed network embedding* | | | | | | |
| Matrix Factorization | LLE [27] | 2000 | Eigenvector problem | Unsupervised | $O\left(d^2 m\right)$ | http://cseweb.ucsd.edu/~saul/matlab/manifolds.tar.gz |
| | LE [28] | 2002 | Laplacian eigenvector problem | Unsupervised | $O\left(d^2 m\right)$ | http://scikit-learn.org/stable/modules/manifold.html |
| | GF [29] | 2013 | Adjacency matrix factorization | Unsupervised | $O(dm)$ | - |
| | GraRep [30] | 2015 | Transition probability-based proximity matrix, SVD | Unsupervised | $O\left(n^3\right)$ | https://github.com/ShelsonCao/GraRep |
| | HOPE [31] | 2016 | Asymmetric transitivity-based proximity matrix, SVD | Unsupervised | $O\left(d^2 m\right)$ | http://git.thumedia.org/embedding/HOPE |
| Random walk | DeepWalk [32] | 2014 | Truncated random walk + SkipGram | Unsupervised | $O(dn)$ | https://github.com/phanein/deepwalk |
| | node2vec [34] | 2016 | BFS, DFS modified random walk + SkipGram | Unsupervised | $O(dn)$ | https://github.com/aditya-grover/node2vec |
| | Walklets [35] | 2016 | Random walk with skips + SkipGram | Unsupervised | $O(dn)$ | - |
| | DCA [41] | 2015 | Diffusion state by random walk with restart | Unsupervised | $O\left(dn^2\right)$ | https://github.com/hhcho/diffusion-component-analysis |
| Deep learning | Structual Deep Network Embedding (SDNE) [38] | 2016 | Deep autoencoder + Laplacian eigenmaps | Semi-supervised | $O(nm)$ | https://github.com/suanrong/SDNE |
| | Deep Neural Networks for Graph Representation (DNGR) [39] | 2016 | Deep autoencoder + random surfing | Unsupervised | $O\left(n^2\right)$ | https://github.com/ShelsonCao/DNGR |
| Others | MDS [40] | 1995 | Euclidean distance | Unsupervised | $O\left(n^2\right)$ | - |
| | Isomap [41] | 2000 | Euclidean distance | Unsupervised | $O\left(d^2 m\right)$ | http://web.mit.edu/cocosci/isomap/isomap.html |
| | LINE [42] | 2015 | Local and global context | Unsupervised | $O(dm)$ | https://github.com/tangjianpku/LINE |
| *Attributed network embedding* | | | | | | |
| Semantic Matching Models | RESCAL [46,47] | 2011 | Bilinear model | Supervised | $O\left(d^2\right)$ | https://github.com/mnick/rescal.py |
| | DistMult [49] | 2014 | Bilinear model | Supervised | $O(d)$ | - |
| | HolE [50] | 2016 | Holographic model | Supervised | $O\left(d \log(d)\right)$ | https://github.com/mnick/holographic-embeddings |

**Table 1.** (continued)

| Category | Algorithm | Year | Architecture | Learning method | Time complexity | Source code |
|---|---|---|---|---|---|---|
| | SME [51] | 2014 | Neural network | Supervised | $O\left(d^3\right)$ | https://github.com/glorotxa/SME |
| | MLP [52] | 2014 | Multi-layer perceptron | Supervised | $O\left(d^2\right)$ | - |
| | NTN [53] | 2013 | Neural tensor network | Supervised | $O\left(d^2 k\right)$ | - |
| Translational Distance Models | SE [54] | 2011 | Naive distance model | Supervised | $O\left(d^2\right)$ | https://github.com/glorotxa/SME |
| | TransE [55] | 2013 | Translation model | Supervised | $O\left(d\right)$ | https://github.com/glorotxa/SME |
| | TransH [56] | 2014 | Translation model; relation-specific hyperplane | Supervised | $O\left(d\right)$ | https://github.com/mrlyk423/relation_extraction |
| | TransR/CTransR [57] | 2015 | Translation model; relation-specific space | Supervised | $O\left(dk\right)$ | https://github.com/mrlyk423/relation_extraction |
| | TransD [58] | 2015 | Translation model; entity and relation diversity | Supervised | $O\left(\max\left(d, k\right)\right)$ | https://github.com/thunlp/TensorFlow-TransX |
| | TransF [60] | 2016 | Flexible translation model | Supervised | $O\left(d\right)$ | - |
| | TranSparse [59] | 2017 | Translation model; adaptive sparse matrices | Supervised | $O\left(dk\right)$ | - |
| Meta-path | PGHNE [62] | 2017 | Meta-path specific matrices | Supervised | $O\left(dm\right)$ | https://github.com/chentingpc/GuidedHeteEmbedding |
| | HINE [63] | 2017 | Heterogeneous proximity | Unsupervised | $O\left(nl^{\mu}\right)$ | - |
| | metapath2vec [64] | 2017 | Meta-path-based random walk + SkipGram | Unsupervised | $O\left(dn\right)$ | https://ericdongyx.github.io/metapath2vec/m2v.html |
| Others | LANE [65] | 2017 | Laplacian matrix | Unsupervised | $O\left(n^2\right)$ | - |
| | EOE [66] | 2017 | Based on LINE, harmonious matrix | Unsupervised | $O\left(d^2 n\right)$ | http://www2.comp.polyu.edu.hk/~cslcxu/#publications |

defines the proximity matrix using different global structural measurements. GraRep and HOPE optimize the objectives by introducing the singular value decomposition (SVD) technique.

*Random walk-based methods*

In graph theory, random walk is exploited to capture structural relationships between nodes. By performing truncated random walks, a network is transformed into node sequences, i.e. paths, which preserve structural proximity of the network. Inspired by SkipGram [32], a famous deep model for neuro-linguistic programming (NLP) that embeds words into a low-dimensional space by incorporating the context of words in sentences, DeepWalk [33] considers the paths as sentences and implements SkipGram to learn embedding of each node. Compared to DeepWalk, node2vec [34] introduces a more

flexible random walk strategy with a trade-off of breadth-first searching and depth-first searching. Therefore, global and local proximities are encoded in the sampled paths. Walklets [35], another extension to DeepWalk, modifies the basic random walk strategy by skipping some nodes in each walk, analogous to constructing the proximity matrix of GraRep. Hence, Walklets is confident in preserving global structural information. Besides, diffusion component analysis (DCA) [36] was proposed to deal with biological networks, which encodes inherent structural properties as diffusion state by random walk with restart (RWR) [37]. Particularly, for each node $v$ in a biological network, DCA computes its diffusion state that is defined as probability distribution that a diffusion path starting from $v$ will reach other nodes based on RWR strategy. RWR captures both global and local structural properties and enables DCA to overcome noise and sparsity of biology networks.
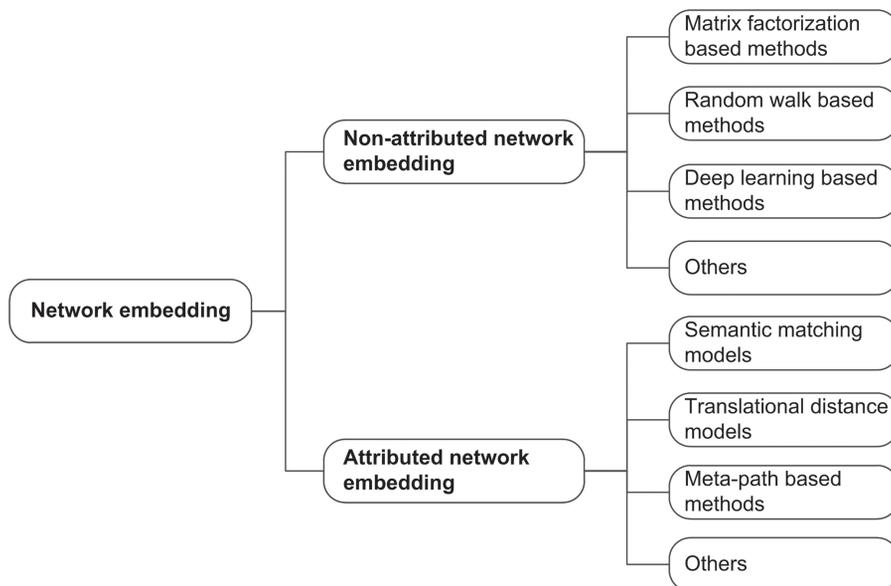
**Figure 2**. Taxonomy of network embedding models.

### Deep learning-based methods

Over the past years, deep learning methods have shown impressive improvement across diverse domains. The idea of building deep architecture was also introduced to deal with the network embedding issue. SDNE [38] and DNGR [39] were designed based on the deep autoencoder architecture. Specifically, SDNE represents nodes by their high-dimensional neighborhood vectors and feeds to the autoencoder to preserve high-order proximity; meanwhile, it also incorporates LE's proximity measure into the autoencoder to preserve 1st-order proximity. On the other hand, DNGR constructs a positive pointwise mutual information (PPMI) matrix by using random surfing, which can capture more global information than random walk. DNGR achieves embeddings by applying autoencoder to the PPMI matrix and shows better performance in preserving high-order proximity than DeepWalk.

### Other methods

Two previous works, multidimensional scaling (MDS) [40] and Isomap [41], learn node embedding by preserving the Euclidean distances of node pairs in the embedding spaces. A common drawback of them is that they need to compute the shortest lengths of node pairs. Another widely used method, LINE [42], aims at embedding by preserving both local and global structure properties. To this end, it defines 1st-order proximity and 2nd-order proximity as connection weight and node's context similarity, respectively.

## Attributed network embedding

The attributed networks, also known as heterogeneous networks, allow nodes and/or edges to belong to multiple types, including the multimedia networks, knowledge graphs, e.g. Freebase [43], DBpedia [44] and YAGO [45], as well as recently emerged biomedical knowledge graphs, e.g. PharmGKB [17], DrugBank [18] and TTD [19]. To embed an attributed network, people should explore structural consistency between different types of objects. The semantic matching models and translational distance models try to address this issue by building energy functions. Specifically, they define a fact as a triple $(h, r, t)$

such that $h$ and $t$ are head and tail entities (i.e. nodes) and $r$ is a relation (i.e. edge) connecting $h$ to $t$. Let $D^+$ denote the collection of facts observed from the network, and $D^-$ the collection of false or missing facts. Then the task of network embedding is to train a model based on an energy function $f(h, r, t)$ to preserve the ranking of facts in $D^+$ over $D^-$. In addition, some other efforts are also able to capture heterogeneity of network by using other insightful techniques, e.g. meta-path.

### Semantic matching models

The semantic matching models exploit similarity-based energy functions by matching latent semantics of entities and relations in embedding spaces. RESCAL [46, 47] was proposed based on the idea that entities are similar if connected to similar entities via similar relations [48]. By associating each relation $r$ with a matrix $\mathbf{M}_r$, it defines the energy function by a bilinear model $f(h, r, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t}$, where $\mathbf{h}, \mathbf{t} \in R^d$ are $d$-dimensional ($d \ll n$) embedding vectors for entities $h$ and $t$, respectively. RESCAL jointly learns embedding results for entities by $\mathbf{h}$ and $\mathbf{t}$ and for relation by $\mathbf{M}_r$. DistMult [49] simplifies RESCAL by restricting matrix $\mathbf{M}_r$ for relation $r$ as a diagonal matrix. Though DistMult is more efficient than RESCAL, it can only deal with the undirected networks. To address this, HolE [50] composes $h$ and $t$ by their circular correlation. Consequently, power of RESCAL and efficiency of DistMult are inherited by HolE.

Other works refer to the neural network architecture by considering embedding as the input layer and energy function as the output layer. For example, semantic matching energy (SME) model [51] designs the hidden layer as $\mathbf{g}_{left}(h, r) = \mathbf{M}_1 \mathbf{h} + \mathbf{M}_2 \mathbf{r} + \mathbf{b}_h$ and $\mathbf{g}_{right}(t, r) = \mathbf{M}_3 \mathbf{t} + \mathbf{M}_4 \mathbf{r} + \mathbf{b}_t$. Then its energy function is given as inner product of $\mathbf{g}_{left}(h, r)$ and $\mathbf{g}_{right}(t, r)$. Since all facts share $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3$ and $\mathbf{M}_4$, the number of parameters of SME to learn is much less than RESCAL. Multi-layer perceptron (MLP) [52] associates each relation $r$ with a vector $\mathbf{r}$ and designs a hidden layer with weight $\mathbf{w} \in R^d$. MLP defines the energy function as $f(h, r, t) = \mathbf{w}^T \tanh(\mathbf{M}_1 \mathbf{h} + \mathbf{M}_2 \mathbf{r} + \mathbf{M}_3 \mathbf{t})$ with shared $\mathbf{M}_1, \mathbf{M}_2$ and $\mathbf{M}_3$. Neural tensor network [53] constructs the hidden layer by specifying each relation $r$ a tensor $\underline{\mathbf{M}}_r$. Therefore, it is expressive but has more parameters to learn compared to RESCAL.

*Translational distance models*

The basic idea of the translational distance models is that, for each fact $(h, r, t)$, relation $r$ is considered as a translation from head entity $h$ to tail entity $t$, namely $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$ in embedding space. They exploit distance-based energy functions to model the facts. A former approach with analogous idea is the structured embedding (SE) model [54] that assumes that relation $r$ can project $h$ close to $t$ by using $\mathbf{M}_r^1$ and $\mathbf{M}_r^2$ and defines $f(h, r, t) = \|\mathbf{M}_r^1 \mathbf{h} - \mathbf{M}_r^2 \mathbf{t}\|_1$. However, massive number of $\mathbf{M}_r^1$ and $\mathbf{M}_r^2$ usually lead to inefficiency in training. TransE [55] is the pioneer of translational distance models. Given an observed fact $(h, r, t)$, TransE represents relation $r$ as translation vector $\mathbf{r}$, such that $\mathbf{h}$ and $\mathbf{t}$ is closely connected by $\mathbf{r}$. Therefore, energy function is defined as $f(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2$. Since all parameters to learn are entity and relation embedding vectors lying in a same low-dimensional space, TransE is obviously easy to train. A drawback of TransE is that it cannot do well with N-to-1, N-to-1 and N-to-N structures. To address this issue, TransH [56] extends TransE by introducing a hyperplane for each relation $r$ and projecting $h$ and $t$ into the hyperplane before constructing the translation scheme. TransH improves model capacity while preserving efficiency. Similarly, TransR [57] extends TransE by introducing relation-specific space. $h$ and $t$ are projected by a matrix $\mathbf{M}_r$ w.r.t. relation $r$. Further, for more fine-grained embedding, TransD [58] extends TransE by constructing two matrices $\mathbf{M}_r^1$ and $\mathbf{M}_r^2$ for each $r$ to project $h$ and $t$, respectively. In this way, TransD captures not only diversity of relations but also diversity of entities. TranSparse [59] is a simplified version of TransR by using adaptive sparse matrices to model different types of relations, and TransF [60] achieves a flexible embedding result by relaxing the translation restriction to $\mathbf{h} + \mathbf{r} \simeq \alpha \mathbf{t}$.

*Meta-path-based methods*

A meta-path is defined as a sequence of node types separated by edge types [61]. For example, a meta-path of length $l$ is in form of $a_1 \xrightarrow{b_1} a_2 \xrightarrow{b_2} \cdots \xrightarrow{b_{l-1}} a_l$, where $\{a_1, a_2, \cdots, a_l\}$ and $\{b_1, b_2, \cdots, b_{l-1}\}$ are sets of node type and relation type, respectively. Therefore, a meta-path is able to capture both structure and attribute information. Several attributed network embedding models have been proposed by using the meta-path conception. By defining an adjacency matrix $\mathbf{M}^p$ by node connectivity under meta-path $p$, path-augmented general heterogeneous network embedding model [62] learns node embeddings by using a neighbor prediction framework on adjacency matrices $\{\mathbf{M}^p\}$ of selected meta-paths. Following this idea, HINE [63] defines meta-path-based proximity in two ways: count of specific path between nodes or probability of meta-path-based random walks linking two nodes. HINE preserves heterogeneous structure by minimizing difference between meta-path-based proximity and expected proximity in embedding space. Moreover, similar to DeepWalk, metapath2vec [64] formalizes meta-path-based random walks and introduces a heterogeneous-version SkipGram to learn node embeddings.

*Other methods*

Like LE, LANE [65] constructs proximity matrices by incorporating node attributes, network structure and labels and learns embeddings based on Laplacian matrix. In addition, EOE [66] was designed to embed network coupled by two non-attribute networks. Particularly, EOE first embeds the non-attribute networks separately by LINE and next jointly embeds them by introducing a harmonious embedding matrix.

## Connection to machine learning

Intuitively, network embedding is proposed to bridge the gap between network topology and traditional machine learning, which is only able to process subjects in vector space. A usual use of the network embedding techniques is to translate network structural information into low-dimensional vectors and feed to machine learning models to address downstream tasks such as link prediction, node classification and clustering and network visualization, etc. In this case, embedding model and the downstream machine learning model are trained separately. For ease of use, some integrated open-source software packages of network embedding have been developed as shown in Table 2. Moreover, in many domains such as biomedicine, a network or relational data usually contains non-topological information, e.g. texts, images and domain roles. To comprehensively incorporate such heterogeneous information attached to the network, there arise increasing needs of deep combination of network embedding and machine learning. For example, as deep learning has achieved great success in representation learning of text and image [67], a deep architecture was designed to simultaneously incorporate and train network embedding, text embedding and image embedding components [68]. As network embedding has been the focus of network analysis, how to adapt network embedding to data and applications in practice has become a crucial point. In brief, a network embedding method should not only efficiently learn informative network representation but also adapt to practical application. With this in mind, we will introduce how the network embedding is applied to the biomedical data to advance the biomedical study in the next section.

## Applications in biomedical data science

The use of network embedding for biomedical data analysis is recent and not thoroughly explored. In this section, we will review some of the main literatures related to applications of network embedding techniques to pharmaceutical data analysis, multi-omics data analysis and clinical data analysis. Table 3 lists all the papers mentioned in this literature review.

### Pharmaceutical data analysis

*Drug repositioning*

Computational drug repositioning, also known as drug repurposing, is a promising and efficient tool for exploring new usage for existing drugs to save drug development cost and increase productivity [3, 69]. Drugs bind with target proteins and affect their downstream activity, consequently lead to impact on human body to treat the disease. A drug repositioning tool usually aims at predicting unknown drug–target or drug–disease interactions. The reviewed studies introduced network embedding into the drug–target and drug–disease interaction network analysis to facilitate drug repositioning.

*Drug–target interaction prediction.* Previous drug–target interaction (DTI) prediction efforts performed matrix factorization based embedding methods on proximity matrices of the bipartite DTI networks and made predictions based on distances in the learned low-dimensional embedding spaces. For example, Yamanishi *et al.* [70] constructed the graph-based proximity matrix by known DTIs and developed an eigenvalue factorization algorithm similar to LLE. Cobanoglu *et al.* [71] directly applied probabilistic matrix factorization to the DTI network to learn

**Table 2.** Open-source software packages of the network embedding techniques. The 1st column is the software package names. The 2nd column presents the network embedding algorithms included in the package. The 3rd column is the platforms that the software runs. The 4th column presents the URLs linked to the software packages

| Package name | Algorithms included | Platform | URLs |
|---|---|---|---|
| OpenNE | DeepWalk, LINE, node2vec, GraRep, TADW, GCN, HOPE, GF, SDNE and LE | Python | https://github.com/thunlp/OpenNE |
| TensorFlow-TransX | TransE, TransH, TransR and TransD | C++ | https://github.com/thunlp/TensorFlow-TransX |
| Fast-TransX | TransE, TransH, TransR, TransD and TranSparse | C++ | https://github.com/thunlp/Fast-TransX |
| knowledge-graph-embeddings | RESCAL, TransE, DistMult, HoLE, etc. | Python | https://github.com/mana-ysh/knowledge-graph-embeddings |
| scikit-kge | RESCAL, TransE, HoLE, etc. | Python | https://github.com/mnick/scikit-kge |
| Graph-Embedding | DeepWalk, LINE, node2vec, etc. | Python | https://github.com/dedekinds/Graph-Embedding |
| Graph-Embedding-Methods (GEM) | LLE, LE, GF, HOPE, SDNE, node2vec | Python | https://github.com/palash1992/GEM |

embeddings. Ezzat *et al.* [72] applied LE, SVD-based matrix factorization and another dimensionality reduction technique, Partial Least Squares [73], to the DTI network embedding. Further, many studies tried to integrate external information into the factorization. For example, in the further work by Yamanishi *et al.* [74], drug side effect and protein domain information were integrated into the proximity matrix. Zheng *et al.* [75] incorporated external chemical and genomic information as regularization terms of the factorization to improve embedding and prediction. For the purpose of incorporating new drugs and targets that do not have any DTI record, Ezzat *et al.* [76] modified proximity matrix of DTI network by using k-nearest known neighbors' interaction profiles of each new drug or target.

More recent works focused on heterogeneous frameworks that contain diverse types of drug-related interactions besides DTIs. Luo *et al.* [77] proposed DTINet by extending DCA by separately performing RWR on drug–drug, drug–disease, drug side effect and drug similarity networks for drug embedding and on protein–protein, protein–disease and protein similarity networks for target protein embedding. After that, DTINet projected drugs into the embedding space of target proteins and made prediction based on geometric proximity. Other works implemented embedding on heterogeneous networks integrated from heterogeneous interaction data. For example, Zong *et al.* [78] introduced DeepWalk to a tripartite network consisting of drug–target, drug–disease and target–disease interactions. Alshahrani *et al.* [79] integrated GO, protein–protein interactions (PPIs), DTIs, gene–disease interactions, drug side effect and disease–phenotype pairs into a heterogeneous biological knowledge graph. To capture the heterogeneity, they modified DeepWalk by incorporating the types of relations into the random walk sequences. Therefore, structural properties combining with relation-type information were preserved when projecting biological entities into the embedding space. Afterwards, a logistic regression classifier was trained for prediction. The results showed that implementation of network embedding on such heterogeneous frameworks effectively integrates chemical, genomic, pharmacological and phenotypic information, and hence accelerates accurate DTI prediction and provides new insights into drug repositioning.

*Drug–disease interaction prediction.* Other studies upon drug repositioning focused on computationally predicting drug–

disease associations, in which network embedding techniques were also involved. Dai *et al.* [80] first embedded genes by applying eigenvalue decomposition to a gene–gene interaction network and next calculated genomic representations for drugs and diseases from the gene embedding vectors via neighboring information of drug–gene and disease–gene interaction networks, respectively. Afterwards, they developed a matrix factorization method, of which the genomic representations of drugs and disease served as initial states of the final embedding vectors during training. The results revealed that introducing genomic space produced by network embedding provides rich molecular-level biological information and helps learn more informative representations for drugs and diseases. Wang *et al.* [81] proposed to detect unknown drug–disease interactions from the medical literature by using NLP and network embedding techniques. Using treatment and inducement drug–disease pairs extracted from 27 million PubMed articles, they first constructed a heterogeneous network. They next expanded the network embedding method, LINE, by modifying the 1st-order proximity to encode treatment and inducement relations as positive and negative effects to the objective function, respectively. The result showed that the embeddings lead to significant improvement in predictions of both types of drug–disease interactions.

*Adverse drug reaction analysis*

An adverse drug reaction (ADR) is defined as any undesirable effect from the medical use of drugs beyond its anticipated therapeutic effects that occurs at a usual dosage [82]. The study of ADRs is the concern of drug development especially before a drug is launched on clinical application. Detecting potential ADRs is always time consuming and expensive. To address this, computational methods based on network embedding have been introduced to ADR analysis. Stanovsky *et al.* [83] proposed to recognize ADR mentions in social media by infusing a knowledge graph, DBpedia [44]. Similar to translational distance models, such as SE, TransE and TransH, they trained a deep learning model by incorporating distance-based energy function. The embedding was infused into a recurrent neural network (RNN) transducer model [84] that was then trained for recognizing ADR mentions. The results showed that embedding of DBpedia knowledge graph is able to provide additional improvements to RNN.

**Table 3.** Biomedical applications of network embedding surveyed in this study. The 1st column presents the biomedical tasks. The 2nd column lists authors and references of the studies. The 3rd column introduces data used in the biomedical tasks. The 4th column presents the concrete applications. The 5th column introduces the network embedding methods used in the biomedical studies

| Tasks | Authors | Data | Application | Embedding method |
|---|---|---|---|---|
| *Computational drug development and discovery* | | | | |
| Drug repositioning | Yamanishi *et al.* [70,74] | DTI network, external drug and protein domain information | DTI prediction | Eigenvalue factorization algorithm |
| | Cobanoglu *et al.* [71] | DTI network | DTI prediction | Probabilistic matrix factorization |
| | Zheng *et al.* [75] | DTI network, external chemical and genomic information | DTI prediction | Matrix factorization |
| | Ezzat *et al.* [76] | Modified DTI network, external chemical and genomic information | DTI prediction | Matrix factorization |
| | Ezzat *et al.* [72] | DTI network | DTI prediction | LE, SVD, PLS |
| | Luo *et al.* [77] | Heterogeneous drug related network | DTI prediction | DCA |
| | Zong *et al.* [78] | tripartite drug-related network | DTI prediction | DeepWalk |
| | Alshahrani *et al.* [79] | biological knowledge graph | DTI prediction | Modified DeepWalk |
| | Dai *et al.* [80] | Gene–gene, gene–drug, gene–disease interactions | Drug–disease interaction prediction | Eigenvalue decomposition and matrix factorization |
| | Wang *et al.* [81] | Drug–disease pairs | Drug–disease interaction prediction | Modified LINE |
| Adverse drug reaction analysis | Stanovsky *et al.* [83] | Drug knowledge graph | Recognizing ADR mentions in social media | Distance-based model similar to SE, TransE and TransH |
| | Zitnik and Zupan [85] | DTIs and DDIs | DDI prediction | Extended RESCAL |
| | Abdelaziz *et al.* [86] | Drug knowledge graph | DDI prediction | TransH and HolE |
| | Wang *et al.* [87] | Drug knowledge graph and biomedical text information | DDI prediction | Extended TransH |
| | Zitnik *et al.* [88] | Drug knowledge graph | DDI prediction | Deep autoencoder similar to SDNE and DNGR |
| *Multi-omics data analysis* | | | | |
| Genomics data analysis | Cho *et al.* [36] | Biological network | Learning informative but low-dimensional representations for nodes in biological networks | DCA, a model based on RWR |
| | Wang *et al.* [91] | Biological network | Gene function prediction | clusDCA, an extension of DCA |
| | Wang *et al.* [92] | Heterogeneous network comprised of gene expression and drug response–gene information | Pathway identification associated with chemosensitivity data | DCA |
| | Li *et al.* [93] | Cell-ContexGene and Gene-ContexGene networks | Learning representation for single cell RNA-seq data | Extended LINE |
| | Zeng *et al.* [95] | Gene–disease network | Prediction of pathogenic human genes | Matrix factorization |
| Proteomics data analysis | Airoldi *et al.* [99] | PPI networks | Learning latent representation for proteins | Mixed membership stochastic block model |
| | Kuchaiev *et al.* [101] | PPI networks | PPI network de-noising | Extended MDS |
| | You *et al.* [102] | PPI networks | Assessing and predicting PPIs | Isomap |
| | Lei *et al.* [103] | PPI network, genomic and proteomic data | PPI network embedding | Extended Isomap |

Continued

**Table 3.** (continued)

| Tasks | Authors | Data | Application | Embedding method |
|---|---|---|---|---|
| Cannistraci *et al.* [105,106] | PPI networks | Assessing and predicting PPIs | Minimum curvilinear embedding | |
| Zhu *et al.* [107] | PPI networks | Assessing PPIs | Logistic metric embedding | |
| Josifoski and Trivodaliev [109] | PPI networks | Protein function prediction | node2vec | |
| Wang *et al.* [110] | PPI networks | Protein function prediction | Extended DCA based on meta-path | |
| Transcriptomics data analysis | Shen *et al.* [114] | miRNA–disease bipartite network, miRNA functional similarity and disease semantic similarity | Prediction of Esophageal Neoplasms-related miRNAs | Matrix factorization |
| | Li *et al.* [117] | miRNA–disease bipartite network | Prediction of associated miRNAs of 22 disease | DeepWalk |
| *Clinical data analysis* | | | | |
| Medical knowledge graph embedding | Zhao *et al.* [118] | Bipartite medical knowledge graphs | Learning medical entity embeddings | A method by extending RESCAL and TransE |
| | Wang *et al.* [119] | Medical knowledge graph | Recommending proper medicine to patients | A method by extending TransR and LINE |
| | Zhao *et al.* [120] | Symptom–disease network extracted from medical forum data | Representation learning for disease prediction, disease category prediction and disease clustering | Extended TransE |
| Electronic health/medical record embedding | Choi *et al.* [121] | EHR + medical ontology graph | Learning EHR representation with the help of medical ontologies. | GRAM |
| | Huang *et al.* [122] | EMR + biomedical knowledge graph | Visualizing EMR of patient | ProSNet, i.e. extended DCA |
| | Liu *et al.* [12] | Medical temporal graphs extracted from EHR | Learning representations of EHR record sequences, i.e., temporal phenotyping | Graph reconstruction |
| | Choi *et al.* [126] | Medical concepts | Medical concept embedding | Factorization of PPMI matrix analogous to DNGR |

Other works on ADR analysis aimed at predicting drug–drug interactions (DDIs) because the majority of preventable ADRs occur between pairs of drugs. Zitnik and Zupan [85] proposed a collective relational learning method, Copacar, based on the intuition of RESCAL to identify the most meaningful relations from multi-relational data. To predict novel DDIs, Copacar was applied to medical relational data composed of known DTIs and DDIs. Most recent works implemented network embedding on knowledge graphs that contain drug-related entities and relations. For example, Abdelaziz *et al.* [86] proposed Tiresias that utilizes TransH and HolE to embed a drug knowledge graph. The embedding results then served as global features for DDI prediction. The predictive results showed that combination of network embeddings, text embeddings and similarity-based local features helps reach a significant prediction. Wang *et al.* [87] developed a new framework, PRD, which aims at encoding drug knowledge graph and biomedical text information into a common embedding space for DDI prediction. In particular, TransH was extended by replacing each fact $(h, r, t)$ as $(h, I, t)$, where $I$ keeps text information over relation $r$. Accordingly, a deep autoencoder model was developed. The results showed that joint learning of embedding results in PRD outperforming Tiresias, TransE and TransR in DDI prediction. In a recent work by Zitnik *et al.* [88], they proposed a deep autoencoder method, Decagon, following the intuition of SDNE and DNGR to predict labeled DDIs. Decagon consists of two components: an encoder by convolutional network for producing embeddings [89, 90] and a decoder by tensor factorization model for prediction by using the embeddings. The results showed that Decagon outperformed baselines, including RESCAL and DeepWalk, up to 69%.

## Multi-omics data analysis

Omics aims at quantitatively and qualitatively studying structures, functions and dynamics of molecules of the organisms. Network embedding is a valuable tool for implementations of relational data analysis in omics. The reviewed studies tried to introduce network embedding methods to accelerate computational tasks in multi-omics from the following subclasses: genomics, proteomics and transcriptomics.

### Genomics data analysis

Several works applied network embedding to predictive tasks in genomics data analysis. A widely used biological network embedding method, DCA, was introduced to interaction prediction studies in genomics analysis. For example, Wang *et al.* [91] proposed clusDCA to predict gene function, by applying DCA to gene–gene interaction and GO to learn low-dimensional

representations for genes and GO labels, respectively. Based on such embedding results, they trained a projection model from gene space to GO space such that genes geometrically closed to their known GO labels. It bridges latent gene features and GO labels and results in desirable prediction of sparsely annotated gene functions. In a recent work by Wang's group [92], they developed PACER that introduced DCA to embed a heterogeneous network comprised of gene expression and drug response-gene information. As genes and pathways are embedded in to a unified space, PACER is able to rank pathways by similarities to response-correlated genes for specific compound. PACER was applied to pathway identification associated with chemosensitivity data. Other embedding techniques were also involved in genomics data analysis. Li *et al.* [93] proposed SCRL to address representation learning for single cell RNA-seq data by network embedding. The basic idea is to extend the network embedding method, LINE, to two bipartite networks, Cell-ContexGene and Gene-ContexGene networks. The low-dimensional representations of cells and genes were jointly learned while context genes appeared in both networks bridging information from gene expression data and pathway priors. The experimental results showed that SCRL outperforms traditional dimensional reduction methods, e.g. PCA [94]. Besides, Zeng *et al.* [95] extended matrix factorization to embed gene–disease network for prediction of pathogenic human genes.

### Proteomics data analysis

PPIs produced by high-throughput experimental technologies [96–98] play crucial roles in most cell functions. Network embedding has also been introduced to PPI networks for proteomics data analysis, such as assessing and predicting PPIs and predicting protein functions, etc. Airoldi *et al.* [99] applied their mixed membership stochastic block (MMSB) model to learn embeddings for PPI networks [100]. The MMSB was originally proposed to detect community structure in complex networks. For each protein, MMSB generated a latent representation vector of which each element denotes the probability that the protein belongs to a specific cluster/community. Such latent representation vectors construct the embedding space for the proteins. To address the high false positive and false negative rate of PPIs by the high-throughput experimental techniques, Kuchaiev *et al.* [101] proposed an embedding algorithm based on MDS for PPI network de-noising. By using the embeddings of proteins, they predicted new PPIs and assessed the confidence of existing PPIs. You *et al.* [102] proposed to use Isomap to embed PPI networks by preserving geodesic distances between protein nodes. The task of assessing and predicting PPIs was transformed into measuring similarity between proteins in the embedding space. Lei *et al.* [103] constructed PPI network by incorporating both genomic and proteomic data and extended Isomap for PPI network embedding. Czekanowski–Dice distance index [104] was applied to the protein embeddings for PPI assessment and prediction. Cannistraci *et al.* [105, 106] proposed minimum curvilinear embedding (MCE) that encoded structural properties by extracting the minimum spanning tree (MST; MST is a subset of edges of a connected (un)directed graph that connects all nodes in the graph without cycles). The results showed that MCE can result in a better performance compared to MDS and Isomap. Zhu *et al.* [107] developed a logistic metric embedding (LME) model based on Euclidean distance analogous to SE. LME can also outperform MDS and Isomap in assessing PPIs. Besides, network embedding was also used to predict protein functions. For example, Kulmanov *et al.* [108] used modified DeepWalk to

learn protein embeddings, which were further input into a deep model to predict protein functions. Josifoski and Trivodaliev [109] proposed to adapt node2vec to PPI networks to embed proteins by preserving both local and global topologies. The embeddings were then used to train a binary classifier for protein function prediction. Wang *et al.* [110] proposed ProSNet for protein function prediction by introducing DCA to a heterogeneous molecular network. A meta-path was introduced to modify DCA to preserve heterogeneous structural information. The prediction performance was greatly improved due to embeddings of the heterogeneous network.

### Transcriptomics data analysis

Transcriptomics focuses on the study of an organism's transcriptome. MicroRNAs (or miRNAs), a class of short non-coding RNA molecules, normally regulate gene expression and have been found to highly associate with complex human diseases [111–113]. Identifying miRNA-disease associations has become a crucial component of the study of pathogenicity. Network embedding has been also involved in transcriptomics for prediction of miRNA-disease associations. Shen *et al.* [114] developed CMFMDA that introduced matrix factorization to bipartite miRNA-disease network for embedding to predict new associations. In CMFMDA, miRNA functional similarity and disease semantic similarity were involved in factorization in terms of regularizations to improve embedding. The evaluation was performed to discover esophageal neoplasms-related miRNAs that were previously confirmed by miR2Disease [115] and dbDEMC [116]. The results showed that CMFMDA can outperform other computational methods. Besides, Li *et al.* [117] proposed a method by using DeepWalk to embed the bipartite miRNA-disease network. After that, the topological similarities of disease pairs were calculated by using the low-dimensional embedding vectors of diseases. The method was applied to prediction of associated miRNAs of 22 diseases. The results showed that, by preserving both local and global topology of miRNA-disease network, DeepWalk can result in significant improvements in association prediction, especially AUC ranging from 0.805 to 0.973.

## Clinical data analysis

Recent network embedding-based computational methods were applied to the clinical data, such as medical knowledge graph, electronic health records (EHRs) and electronic medical records (EMRs), to provide useful assistance for clinicians.

### Medical knowledge graph embedding

Embedding of medical knowledge graph is similar to other knowledge graphs. For example, Zhao *et al.* [118] derived a new method to learn embeddings of medical entities in medical knowledge graph. By modifying the energy functions of RESCAL and TransE, two arbitrarily derivable energy functions were proposed and resulted in better performances than RESCAL and TransE. Wang *et al.* [119] recently proposed to learn embeddings from a heterogeneous medical knowledge graph to recommend proper medicine to patients. They constructed objective by using both TransR's energy and Line's 2nd-order proximity measurement. Upon a bipartite symptom–disease network, Zhao *et al.* [120] proposed ContexCareto to learn representation of medical forum data. Specifically, they defined energy function by considering the relation between the symptoms of a patient and a specific disease as a translation vector

analogous to TransE. To alleviate sparseness, they incorporated symptom co-occurrence and disease evolution networks into the construction of objective function. Evaluations on real medical forum data revealed significance of ContexCare in disease prediction, disease category prediction and disease clustering.

### Electronic health/medical record embedding

EHR and EMR commonly include medical and clinical information of patients. Using network embedding techniques to learn representations for EHR and EMR can help both medical research and clinical decision. Yet, EHR and EMR data are heterogeneous because they contain multiple types of information and usually have no obvious relationships. To address this issue, some works incorporated external knowledge. Choi *et al.* [121] developed GRAM to learn EHR representation with the help of hierarchical information inherent to medical ontologies. Specifically, embedding vector of a node (i.e. a medical code) was generated by its ancestors in the medical ontology graph by using an attention mechanism. Using representations by GRAM to predict heart failure resulted in 10% higher accuracy and 3% higher AUC than RNN. For visualizing EMR of patient, Huang *et al.* [122] introduced ProSNet to an integrated biomedical knowledge graph to learn the embeddings of medical entities. Afterwards such embeddings were used to calculate a similarity matrix of medical features to enrich a profile matrix. The proposed method was applied to the visualization of Parkinson's disease data set. In addition, some other efforts aimed at constructing network from EHR and EMR directly. As EHR of a patient typically records sequence of his/her medical events, it can be represented as a temporal graph [123–125]. By assuming that each medical temporal graph can be reconstructed by multiple latent graph bases, Liu *et al.* [12] proposed to extract latent graph bases and learn embedding vectors for the temporal graphs. The results showed effectiveness on personalized medicine, disease diagnosis and patient segmentation in heart failure. In another work by Choi's group [126], they tried to embed medical concepts, including diseases, medications, procedures and laboratory tests, into a unified space with dimensionality around 100. To this aim, they proposed to introduce two strategies: one is to sample connected concepts as word pairs to put into word2vec [127] and another is to factorize the shifted PPMI matrix analogous to DNGR, which has been demonstrated to be equivalent to word2vec. The learned embeddings were then applied to the study of medical relatedness property. Figure 3 provides an overview of the different applications of network embedding in biomedicine.

## Challenges and opportunities

Despite the promising results obtained using network embedding techniques, there remain several unsolved challenges the biomedical application is faced with. In particular, we highlight the following key issues:

- **Data quality.** Unlike other domain where the data are clean and well structured, networks constructed from the biomedical data are usually noisy and incomplete. For example, the PPI data produced by high-throughput techniques, such as Y2H and TAP-MS, suffer from high false negative rates up to 70% and high false positive rates up to 64% [128]. Meanwhile, relational data extracted from EHRs are usually highly incomplete. Though efforts have focused on the issues including network sparsity, redundancy and incompleteness, training an effective model to thoroughly overcome undesirable data

quality and accurately embed the biomedical networks is still challenging.

- **Local and global.** Performances of the network embedding model and its downstream tasks rely on the type of structural property to preserve. Preserving local property will gather connected nodes in the embedding space, while preserving global property will project topologically similar (even far separated) nodes together. Designing embedding method by properly considering local and global structure properties according to application scenarios is an important aspect that will require the development of novel solutions.
- **Network evolution.** Networks are always not static, especially in the biomedical domain. For example, increasing number of omics data are being produced thanks to the well-developed high-throughput experimental techniques and database systems. Existing network embedding models mainly focused on the static networks, and the settings of network evolution were overlooked. To learn embeddings for a dynamic network, existing methods should be trained repeatedly for each timestamp, which is definitely time consuming and may not capture the temporal properties. Therefore, most of the existing network embedding methods cannot be directly applied to evolving biomedical networks.
- **Domain complexity.** Different from network embedding application in other domains, the issue on biomedicine and health care is much more complicated. For example, in a biomedical network, each interaction between entities usually represents a complex genetic, pathological or pharmacological event or process, and there is usually no complete knowledge on how it progresses. When applying an embedding model, the biomedical domain knowledge is also needed to better understand the network structure.

All above challenges introduce several opportunities and future research possibilities to improve biomedical informatics. Therefore, with all of them in mind, we point out the following directions, which we believe would be promising for the future application of network embedding in biomedical field.

- **Local and global trade-off embedding.** Preserving local and global structure properties will result in distinct embedding results. It is hard to assert which conception works better due to the complexity of application scenarios. In fact, some network embedding methods such as LINE and node2vec aim at preserving both local and global structure properties. Yet, how to better balance local and global information to benefit biomedical informatics is rarely discussed. Therefore, designing embedding model that is able to flexibly reach a trade-off local and global structure properties according to application scenarios, especially biomedicine, would be a promising direction of our future work.
- **Dynamic embedding.** Considering that networks in biomedicine and health care are growing rapidly, embedding results should also evolve following the changes of network topology. Therefore, training a time-sensitive model for network embedding is crucial for a better understanding of temporal properties and for settings of downstream applications. For example, learning embeddings from a drug knowledge graph in real time helps involve newly released *in vitro* experimental results to improve analysis *in silico*, such as drug repositioning and drug side effect prediction. Unlike static network embedding, the models for dynamic networks need to be scalable and flexible to deal with the changes of networks effectively and efficiently and remains a promising issue.
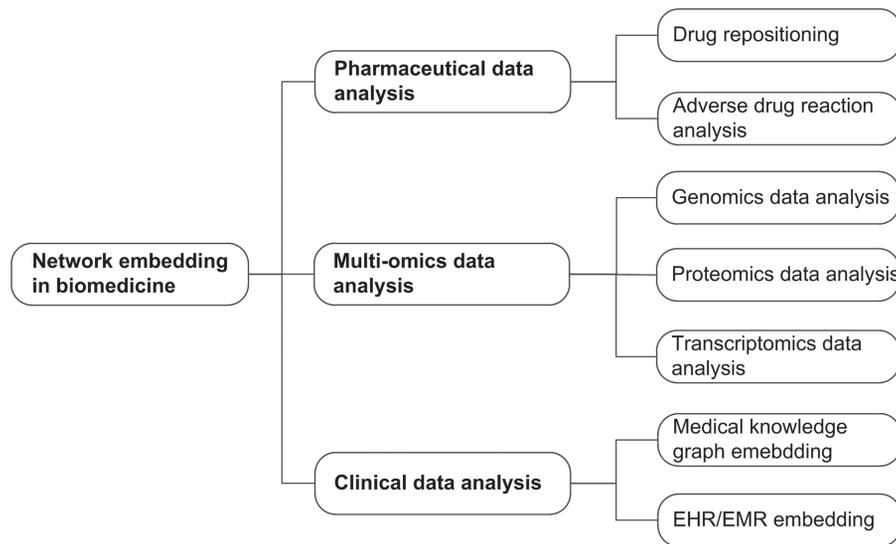
**Figure 3**. Illustration of applications of network embedding in biomedical data science.

- **Text-associated embedding.** Networks in biomedical domain, especially the well-organized biomedical knowledge graphs, always contain rich text information such as descriptions of entities and relations, which would have high potential to address network incompleteness and improve understanding of topological properties. However, to the best of our knowledge, text information is rarely used to assist the network embedding applications in biomedical domain. Taking full advantage of biomedical text information for network embedding needs to properly concatenate network embedding with NLP techniques, which is still challenging for researchers and needs more efforts in the future.

- **Domain-knowledge-associated embedding.** The existing expert knowledge is invaluable for computational analysis in biomedical informatics. Incorporating domain knowledge into the network embedding process to guide it toward the right direction is an important research topic in addressing undesirable data quality and domain complexity. For example, some previous works have incorporated external information, such as drug similarity and protein similarity, into matrix factorization-based embedding methods of DTI prediction [70, 74–76]. In fact, well-developed ICD-9, ICD-10, GO, online medical encyclopedia, PubMed abstracts, etc., also provide abundant biomedical domain knowledge but are rarely involved in network embedding applications. There remain large room and desirable potential for incorporating such external domain knowledge into network embedding in biomedical informatics, and we expect more consummate domain-knowledge-associated embedding models will be launched soon.

## Conclusions

Network is an important data format for data-driven issues in biomedical science. Network embedding approaches, lying in the overlapping of network analytics and representation learning, are powerful tools to learn compact yet informative representations for networks and raise the possibility of using efficient traditional machine learning to solve network-based problems. These methods have been used in numerous biomedical applications. All the results available in the literatures reviewed in this work illustrate the capabilities of network embedding for biomedical network analysis. In fact, processing biomedical networks with network embedding increased the predictive power for several specific applications in different biomedical domains. By carefully reviewing and comparing applications of network embedding in biomedicine, we summarize the challenges the current network embedding applications are faced with and consequently point promising future directions in this domain.

> **Key Points**
>
> - Advances in biomedical research have generated a large volume of biomedical networks, which are high dimensional, sparse, noisy and heterogeneous.
> - Early applications of network-based learning to biomedical networks helped understand topology and knowledge from the complex networks and benefited human healthcare research but suffered from high computational and space cost.
> - Network embedding can open a new way toward effective yet efficient network analysis, which projects network into the low-dimensional yet informative space that is friendly to state-of-the-art machine learning methods.
> - Network embedding has been widely applied to biomedical data science, including pharmaceutical data analysis, multi-omics data analysis and clinical data analysis and showed robust performances in biomedical tasks.
> - Balancing local and global structural properties, handling dynamics of evolving networks as well as incorporating rich text and domain knowledge would be promising directions of network embedding for better improving human healthcare in future study.

## Funding

## References

1. Goh K-I, Cusick ME, Valle D, *et al*. The human disease network. *Proc Natl Acad Sci U S A* 2007;**104**:8685–90.
2. Chen H, Ding L, Wu Z, *et al*. Semantic web for integrated network analysis in biomedicine. *Brief Bioinform* 2009;**10**:177–92.
3. Zhu Y, Elemento O, Pathak J, Wang F. Drug knowledge bases and their applications in biomedical informatics research. *Brief Bioinform* 2018;bbx169. doi:10.1093/bib/bbx169
4. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;**5**:101–13.
5. Albert R. Scale-free networks in cell biology. *J Cell Sci* 2005;**118**:4947–57.
6. Buchanan M, Caldarelli G, De Los Rios FR, *et al. Networks in Cell Biology*. Cambridge, UK: Cambridge University Press, 2012.
7. Barabási A-L. Network medicine—from obesity to the "diseasome". *N Engl J Med* 2007;**357**:404–7.
8. Janjic V, Przulj N. Biological function through network topology: a survey of the human diseasome. *Brief Funct Genomics* 2012;**11**:522–32.
9. Weinshilboum R. Inheritance and drug response. *N Engl J Med* 2003;**348**:529–37.
10. Choi E, Schuetz A, Stewart WF, *et al*. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc* 2017;**24**:361–70.
11. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: a deep learning approach. In: *Proceedings of the 2016 SIAM International Conference on Data Mining, 2016*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
12. Liu C, Wang F, Hu J, Xiong H. Temporal phenotyping from longitudinal electronic health records: a graph based framework. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015*, pp. 705–14. ACM, Sydney, NSW, Australia.
13. Tsuyuzaki K, Nikaido I. Biological systems as heterogeneous information networks: a mini-review and perspectives, 2017. In: *1st International Workshop on Heterogeneous Networks Analysis and Mining (HeteroNAM), 2018*. Los Angeles, California, USA.
14. Wu Z, Wang Y, Chen L. Network-based drug repositioning. *Mol Biosyst* 2013;**9**:1268.
15. Zhang W, Chien J, Yong J, Kuang R. Network-based machine learning and graph theory algorithms for precision oncology. *NPJ Precis Oncol* 2017;**1**(1):25.
16. Kito K, Ito T. Methods for protein–protein interaction analysis. In: *Introduction to Systems Biology*. New York, NY, USA: Humana Press, 160–82.
17. Klein TE, Chang JT, Cho MK, *et al*. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics research network and knowledge base. *Pharmacogenomics J* 2001;**1**:167–70.
18. Wishart DS. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006;**34**:D668–72.
19. Chen X. TTD: therapeutic target database. *Nucleic Acids Res* 2002;**30**:412–5.
20. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–9.
21. Köhler S, Doelken SC, Mungall CJ, *et al*. The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res* 2013;**42**:D966–74.
22. Kibbe WA, Arze C, Felix V, *et al*. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res* 2015;**43**:D1071–8.
23. Cai H, Zheng VW, Chang K. A comprehensive survey of graph embedding: problems, techniques and applications. *IEEE Trans Knowl Data Eng* 2018;**30**:1616–37.
24. Wang Q, Mao Z, Wang B, *et al*. Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans Knowl Data Eng* 2017;**29**:2724–43.
25. Cui P, Wang X, Pei J, *et al*. A survey on network embedding. *IEEE Trans Knowl Data Eng* 2018.
26. Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: a survey. *Knowl Based Syst* 2018;**151**:78–94.
27. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science* 2000;**290**:2323–6.
28. Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: *Advances in Neural Information Processing Systems*. British Columbia, Canada: Vancouver, pp. 585–91.
29. Ahmed A, Shervashidze N, Narayanamurthy S, *et al*. Distributed large-scale natural graph factorization. In: *Proceedings of the 22nd International Conference on World Wide Web—WWW '13, 2013*, pp. 37–48. Rio de Janeiro, Brazil.
30. Cao S, Lu W, Xu Q. GraRep: learning graph representations with global structural information. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015*, pp. 891–900. ACM, Melbourne, Australia.
31. Ou M, Cui P, Pei J, *et al*. Asymmetric transitivity preserving graph embedding. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016*, pp. 1105–14. ACM, San Francisco, CA, USA.
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:13013781.
33. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014*, pp. 701–10. ACM, New York, NY, USA.
34. Grover A, Leskovec J. node2vec: scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016*, pp. 855–64. ACM, San Francisco, CA, USA.
35. Perozzi B, Kulkarni V, Skiena S. Walklets: multiscale graph embeddings for interpretable network classification, 2016, arXiv preprint arXiv:160502115.
36. Cho H, Berger B, Peng J. Diffusion component analysis: unraveling functional topology in biological networks. *Res Comput Mol Biol* 2015;**9029**:62–4.
37. Köhler S, Bauer S, Horn D, *et al*. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet* 2008;**82**:949–58.
38. Wang D, Cui P, Zhu W. Structural deep network embedding. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '16, 2016*, pp. 1225–1234. San Francisco, California, USA.

39. Cao S, Lu W, Xu Q. Deep neural networks for learning graph representations. In: *AAAI Conference on Artificial Intelligence*, pp. 1145–52.

40. Hofmann T, Buhmann J. Multidimensional scaling and data clustering. In: *Advances in Neural Information Processing Systems*. Denver, CO, USA, 1995, pp. 459–66.

41. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000;**290**:2319–23.

42. Tang J, Qu M, Wang M, *et al*. LINE: large-scale information network embedding. In: *24th International Conference on World Wide Web*, 2015, pp. 1067–77. Florence, Italy.

43. Bollacker K, Evans C, Paritosh P, *et al*. Freebase: a collaboratively created graph database for structuring human knowledge. In: *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 2008, pp. 1247–50. ACM, Vancouver, Canada.

44. Lehmann J, Isele R, Jakob M, *et al*. DBpedia: a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 2015;**6**:167–95.

45. Suchanek FM, Kasneci G, Weikum G. Yago: a core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*, 2007, 697–706. ACM, Banff, Alberta, Canada.

46. Nickel M, Tresp V, Kriegel H-P. A three-way model for collective learning on multi-relational data. In: *Proceedings of the 28th International Conference on Machine Learning*, pp. 809–16.

47. Jenatton R, Roux NL, Bordes A, Obozinski GR. A latent factor model for highly multi-relational data. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 3167–75.

48. Nickel M, Tresp V. Tensor factorization for multi-relational learning. *Lecture Notes in Computer Science*. New York, NY, USA: Springer, 2013, 617–21.

49. Yang B, Yih W-T, He X, *et al*. Embedding entities and relations for learning and inference in knowledge bases. In: *International Conference on Learning Representations (ICLR)*, 2015. San Diego, California, USA.

50. Nickel M, Rosasco L, Poggio TA. Holographic embeddings of knowledge graphs. In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 1955–1961. Phoenix, Arizona.

51. Bordes A, Glorot X, Weston J, *et al*. A semantic matching energy function for learning with multi-relational data. *Mach Learn* 2013;**94**:233–59.

52. Dong X, Gabrilovich E, Heitz G, *et al*. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 601–610. ACM, New York, New York, USA.

53. Socher R, Chen D, Manning CD, Ng A. Reasoning with neural tensor networks for knowledge base completion. In: *Advances in Neural Information Processing Systems, 2013*, pp. 926–34. Lake Tahoe, Nevada.

54. Bordes A, Weston J, Collobert R, Bengio Y. Learning structured embeddings of knowledge bases. In: *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011*, pp. 301–306. San Francisco, California.

55. Bordes A, Usunier N, Garcia-Duran A, *et al*. Translating embeddings for modeling multi-relational data. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013*, pp. 2787–95. Lake Tahoe, Nevada.

56. Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: AAAI Conference on Artificial Intelligence, pp. 1112–9.

57. Lin Y, Liu Z, Sun M, *et al*. Learning entity and relation embeddings for knowledge graph completion. In: AAAI Conference on Artificial Intelligence, pp. 2181–7.

58. Ji G, He S, Xu L, *et al*. Knowledge graph embedding via dynamic mapping matrix. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers), pp. 687–96. Beijing, China.

59. Ji G, Liu K, He S, Zhao J. Knowledge graph completion with adaptive sparse transfer matrix. In: AAAI Conference on Artificial Intelligence, pp. 985–91.

60. Feng J, Huang M, Wang M, *et al*. Knowledge graph embedding by flexible translation. In: *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning, 2016*, pp. 557–60. Cape Town, South Africa.

61. Sun Y, Han J, Yan X, *et al*. Pathsim: meta path-based top-k similarity search in heterogeneous information networks. *Proc VLDB Endowment* 2011;**4**(11):992–1003.

62. Chen T, Sun Y. Task-guided and path-augmented heterogeneous network embedding for author identification. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017*, pp. 295–304. ACM, Cambridge, United Kingdom.

63. Huang Z, Mamoulis N. Heterogeneous information network embedding for meta path based proximity. 2017, arXiv preprint arXiv:1701.05291.

64. Dong Y, Chawla NV, Swami A. metapath2vec: scalable representation learning for heterogeneous networks. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017*, pp. 135–44. ACM, Halifax, NS, Canada.

65. Huang X, Li J, Hu X. Label informed attributed network embedding. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017*, pp. 731–9. ACM, Cambridge, United Kingdom.

66. Xu L, Wei X, Cao J, Yu PS. Embedding of embedding (eoe): joint embedding for coupled heterogeneous networks. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 2017*, pp. 741–9. ACM, Cambridge, United Kingdom.

67. Zhong G, Wang LN, Ling X, *et al*. An overview on data representation learning: from traditional feature learning to recent deep learning. *J Finance Data Sci* 2016;**2**(4): 265–78.

68. Chang S, Han W, Tang J, *et al*. Heterogeneous network embedding via deep architectures. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015*, pp. 119–28. ACM, Sydney, NSW, Australia.

69. Li J, Zheng S, Chen B, *et al*. A survey of current trends in computational drug repositioning. *Brief Bioinform* 2016; **17**:2–12.

70. Yamanishi Y, Araki M, Gutteridge A, *et al*. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**: i232–40.

71. Cobanoglu MC, Liu C, Hu F, *et al*. Predicting drug–target interactions using probabilistic matrix factorization. *J Chem Inf Model* 2013;**53**:3399–409.

72. Ezzat A, Wu M, Li X-L, *et al*. Drug–target interaction prediction using ensemble learning and dimensionality reduction. *Methods* 2017;**129**:81–8.

73. de Jong S. SIMPLS: an alternative approach to partial least squares regression. *Chemometr Intell Lab Syst* 1993;**18**:251–63.

74. Yamanishi Y, Kotera M, Moriya Y, *et al*. DINIES: drug–target interaction network inference engine based on supervised analysis. *Nucleic Acids Res* 2014;**42**:W39–45.

75. Zheng X, Ding H, Mamitsuka H, Zhu S. Collaborative matrix factorization with multiple similarities for predicting drug–target interactions. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '13, 2013*, pp. 1025–1033. Chicago, Illinois, USA.

76. Ezzat A, Zhao P, Wu M, *et al*. Drug–target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans Comput Biol Bioinform* 2017;**14**:646–56.

77. Luo Y, Zhao X, Zhou J, *et al*. A network integration approach for drug–target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun* 2017;**8**:573.

78. Zong N, Kim H, Ngo V, *et al*. Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations. *Bioinformatics* 2017;**33**:2337–44.

79. Alshahrani M, Khan MA, Maddouri O, *et al*. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics* 2017;**33**:2723–30.

80. Dai W, Liu X, Gao Y, *et al*. Matrix factorization-based prediction of novel drug indications by integrating genomic space. *Comput Math Methods Med* 2015;**2015**:275045.

81. Wang P, Hao T, Yan J, *et al*. Large-scale extraction of drug–disease pairs from the medical literature. *J Assoc Inf Sci Technol* 2017;**68**:2649–61.

82. James I, Henry L. Adverse drug reactions. *Gen Pract* 2007;**56**:193.

83. Stanovsky G, Gruhl D, Mendes P. Recognizing mentions of adverse drug reaction in social media using knowledge-infused recurrent models. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017*, pp. 142–51. Valencia, Spain.

84. Graves A. Sequence transduction with recurrent neural networks. In: *29th International Conference on Machine Learning (ICML), 2012*. Edinburgh, Scotland.

85. Zitnik M, Zupan B. Collective pairwise classification for multi-way analysis of disease and drug data. *Pac Symp Biocomput* 2016;**21**:81–92.

86. Abdelaziz I, Fokoue A, Hassanzadeh O, *et al*. Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions. *Web Semant* 2017;**44**:104–17.

87. Wang M, Chen Y, Qian B, *et al*. Predicting rich drug–drug interactions via biomedical knowledge graphs and text jointly embedding, 2017, arXiv preprint arXiv:171208875.

88. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018. doi:10.1101/258814

89. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 3844–52.

90. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations (ICLR), 2017*. Palais des Congrès Neptune, Toulon, France.

91. Wang S, Cho H, Zhai C, *et al*. Exploiting ontology graph for predicting sparsely annotated gene function. *Bioinformatics* 2015;**31**:i357–64.

92. Wang S, Huang E, Cairns J, *et al*. Identification of pathways associated with chemosensitivity through network embedding. 2017, bioRxiv preprint bioRxiv: 168450 doi:10.1101/168450

93. Li X, Chen W, Chen Y, *et al*. Network embedding-based representation learning for single cell RNA-seq data. *Nucleic Acids Res* 2017;**45**:e166.

94. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometr Intell Lab Syst* 1987;**2**:37–52.

95. Zeng X, Ding N, Rodríguez-Patón A, Zou Q. Probability-based collaborative filtering model for predicting gene–disease associations. *BMC Med Genomics* 2017;**10**(5):76.

96. Gavin A-C, Bösche M, Krause R, *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;**415**:141–7.

97. Krogan NJ, Cagney G, Yu H, *et al*. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* 2006;**440**:637–43.

98. Ito T, Chiba T, Ozawa R, *et al*. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 2001;**98**:4569–74.

99. Airoldi EM, Blei DM, Fienberg SE, *et al*. Mixed membership stochastic block models for relational data with application to protein–protein interactions. In: *Proceedings of the International Biometrics Society Annual Meeting, 2006*. McGill University, Montréal, Québec, Canada.

100. Airoldi EM, Blei DM, Fienberg SE, *et al*. Mixed membership stochastic block models. *J Mach Learn Res* 2008;**9**: 1981–2014.

101. Kuchaiev O, Rasajski M, Higham DJ, *et al*. Geometric de-noising of protein–protein interaction networks. *PLoS Comput Biol* 2009;**5**:e1000454.

102. You Z-H, Lei Y-K, Gui J, *et al*. Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data. *Bioinformatics* 2010;**26**:2744–51.

103. Lei Y-K, You Z-H, Ji Z, *et al*. Assessing and predicting protein interactions by combining manifold embedding with multiple information integration. *BMC Bioinformatics* 2012;**13**(Suppl 7):S3.

104. Angelelli J-B, Baudot A, Brun C, *et al*. Two local dissimilarity measures for weighted graphs with application to protein interaction networks. *Adv Data Anal Classif* 2008;**2**: 3–16.

105. Cannistraci CV, Ravasi T, Montevecchi FM, *et al*. Nonlinear dimension reduction and clustering by minimum curvilinearity unfold neuropathic pain and tissue embryological classes. *Bioinformatics* 2010;**26**:i531–9.

106. Cannistraci CV, Alanis-Lobato G, Ravasi T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* 2013;**29**: i199–209.

107. Zhu L, Deng S-P, Huang D-S. A two-stage geometric method for pruning unreliable links in protein–protein networks. *IEEE Trans Nanobioscience* 2015;**14**:528–34.

108. Kulmanov M, Khan MA, Hoehndorf R, *et al*. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* 2018;**34**:660–8.

109. Josifoski M, Trivodaliev K. Instance based learning in protein interaction networks. In: *Proceedings of the 14th International Conference for Informatics and Information Technology, 2017*, pp. 200–205. Skopje, Macedonia.

110. Wang S, Qu M, Peng J. Prosnet: integrating homology with molecular networks for protein function prediction. *Pac Symp Biocomput* 2017;**22**:27–38.

111. Alvarez-Garcia I, Miska EA. MicroRNA functions in animal development and human disease. *Development* 2005;**132**:4653–62.

112. Lynam-Lennon N, Maher SG, Reynolds JV. The roles of microRNA in cancer and apoptosis. *Biol Rev Camb Philos Soc* 2009;**84**:55–71.

113. Meola N, Gennarino V, Banfi S. microRNAs and genetic diseases. *Pathogenetics* 2009;**2**:7.

114. Shen Z, Zhang Y-H, Han K, *et al*. miRNA-disease association prediction with collaborative matrix factorization. *Complexity* 2017;**2017**:1–9.

115. Jiang Q, Wang Y, Hao Y, *et al*. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res* 2009;**37**:D98–104.

116. Yang Z, Ren F, Liu C, *et al*. dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics* 2010;**11**:S5.

117. Li G, Luo J, Xiao Q, *et al*. Predicting microRNA-disease associations using network topological similarity based on DeepWalk. *IEEE Access* 2017;**5**:24032–9.

118. Zhao C, Jiang J, Guan Y, *et al*. EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning. *Artif Intell Med* 2018;**87**:49–59.

119. Wang M, Liu M, Liu J, *et al*. Safe medicine recommendation via medical knowledge graph embedding, 2017, arXiv preprint arXiv:171005980.

120. Zhao S, Jiang M, Yuan Q, *et al*. ContextCare: incorporating contextual information networks to representation learning on medical forum data. In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017*, pp. 3497–503. AAAI Press, Melbourne, Australia.

121. Choi E, Bahadori MT, Song L, *et al*. GRAM: graph-based attention model for healthcare representation learning. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017*, pp. 787–95. ACM, Halifax, Nova Scotia, Canada.

122. Huang EW, Wang S, Zhai C. VisAGE: integrating external knowledge into electronic medical record visualization. *Pac Symp Biocomput* 2018;**23**:578–89.

123. Mörchen F, Ultsch A. Efficient mining of understandable patterns from multivariate interval time series. *Data Min Knowl Discov* 2007;**15**:181–215.

124. Moerchen F, Fradkin D. Robust mining of time intervals with semi-interval partial order patterns. In: *Proceedings of the 2010 SIAM International Conference on Data Mining, 2017*, pp. 315–26. Columbus, Ohio, USA.

125. Wang F, Lee N, Hu J, *et al*. A framework for mining signatures from event sequences and its applications in healthcare data. *IEEE Trans Pattern Anal Mach Intell* 2013;**35**:272–85.

126. Choi Y, Chiu CY-I, Sontag D. Learning low-dimensional representations of medical concepts. *AMIA Jt Summits Transl Sci Proc* 2016;**2016**:41–50.

127. Levy O, Goldberg Y. Neural word embedding as implicit matrix factorization. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, pp. 2177–85.

128. Edwards AM, Kus B, Jansen R, *et al*. Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Drug Discov Today* 2004;**9**:S32–40.