



# To Embed or Not: Network Embedding as a Paradigm in Computational Biology

Walter Nelson<sup>1,2</sup>, Marinka Zitnik<sup>3</sup>, Bo Wang<sup>3,4,5</sup>, Jure Leskovec<sup>3,6</sup>, Anna Goldenberg<sup>1,5,7\*</sup> and Roded Sharan<sup>8\*</sup>

<sup>1</sup> Genetics and Genome Biology, SickKids Research Institute, Toronto, ON, Canada, <sup>2</sup> Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada, <sup>3</sup> Department of Computer Science, Stanford University, Stanford, CA, United States, <sup>4</sup> Peter Munk Cardiac Center, University Health Network, Toronto, ON, Canada, <sup>5</sup> Vector Institute, Toronto, ON, Canada, <sup>6</sup> Chan Zuckerberg Biohub, San Francisco, CA, United States, <sup>7</sup> Department of Computer Science, University of Toronto, Toronto, ON, Canada, <sup>8</sup> School of Computer Science, Tel Aviv University, Tel Aviv, Israel

## OPEN ACCESS

### Edited by:

Marco Pellegrini,  
Italian National Research Council  
(CNR), Italy

### Reviewed by:

Noel Malod-Dognin,  
Barcelona Supercomputing Center,  
Spain

Gregorio Alanis-Lobato,  
Francis Crick Institute,  
United Kingdom

### \*Correspondence:

Anna Goldenberg  
anna.goldenberg@utoronto.ca;  
anna.goldenberg@gmail.com  
Roded Sharan  
roded@tau.ac.il

### Specialty section:

This article was submitted to  
Systems Biology,  
a section of the journal  
Frontiers in Genetics

**Received:** 05 February 2019

**Accepted:** 09 April 2019

**Published:** 01 May 2019

### Citation:

Nelson W, Zitnik M, Wang B,  
Leskovec J, Goldenberg A and  
Sharan R (2019) To Embed or Not:  
Network Embedding as a Paradigm  
in Computational Biology.  
*Front. Genet.* 10:381.  
doi: 10.3389/fgene.2019.00381

Current technology is producing high throughput biomedical data at an ever-growing rate. A common approach to interpreting such data is through network-based analyses. Since biological networks are notoriously complex and hard to decipher, a growing body of work applies graph embedding techniques to simplify, visualize, and facilitate the analysis of the resulting networks. In this review, we survey traditional and new approaches for graph embedding and compare their application to fundamental problems in network biology with using the networks directly. We consider a broad variety of applications including protein network alignment, community detection, and protein function prediction. We find that in all of these domains both types of approaches are of value and their performance depends on the evaluation measures being used and the goal of the project. In particular, network embedding methods outshine direct methods according to some of those measures and are, thus, an essential tool in bioinformatics research.

**Keywords:** network biology, network embedding, network alignment, community detection, protein function prediction

## INTRODUCTION

Network biology is a powerful paradigm for representing, interpreting and visualizing biological data (Barabási and Oltvai, 2004). One of the standard approaches to computing on networks is to transform such data into vectorial data, aka *network embedding*, to facilitate similarity search, clustering and visualization (Hamilton et al., 2017b; Cai et al., 2018).

In a network embedding problem, one is given a network and an induced similarity (or distance) function between its nodes; the goal is to find a low dimensional representation of the network nodes in some metric space so that the given similarity (or distance) function is preserved as much as possible. For example, if the input network is unweighted and the distance between nodes is defined to be the graph geodesic distance, then a possible goal could be to find an embedding into Euclidean space that minimizes the sum of squared differences between graph distances and the corresponding Euclidean distances (Tenenbaum, 2000).

The classical approach to network embedding employs matrix factorization and is based on the fact that if the desired similarity matrix is positive semi-definite then it can be decomposed into the

product of a real matrix and its transpose. Thus, if one represents each node by a row of that matrix then the given similarity is completely captured by the dot-product between the corresponding vector representations. Similarly, if one is given distances between nodes that satisfy the triangle inequality then double centering the distance matrix gives a positive semi-definite matrix whose decomposition yields vector representations that respect the given distances. This approach is precisely the multidimensional scaling procedure (Cox and Cox, 2000).

Embedding approaches have several potential advantages. Algorithms making use of embeddings are frequently faster than their counterparts which operate on the original networks. Additionally, the learned embeddings are often applicable for downstream analysis, either by direct interpretation of the embedding space or through the application of machine learning techniques which are designed for vectorial data. Beyond its computational advantages, network embedding is natural to use in biological problems that concern physical entities (such as proteins) that function in 3D space. In such scenarios, Euclidean representations may capture many of the functional properties of those entities. Finally, by working in lower dimensional space, the results are more likely to be robust to the noise inherently present in the networks. Indeed, recent network denoising approaches employed embedding for this purpose (Wang et al., 2018).

In this review, we describe several current approaches for graph embedding including spectral-based, diffusion-based and deep-learning-based methods. We provide comparisons applying representative embedding approaches to fundamental problems in network biology with using the networks directly in three distinct tasks: protein network alignment, protein module detection, and protein function prediction (**Figure 1**). We further review network embedding methods and their application to network denoising and pharmacogenomics. We conclude that network embedding methods are an essential component in the bioinformatics tool box.

## METHODOLOGY

Methods for network embedding aim to optimize the difference between the node similarities/distances in the original network space and their similarities/distances under the embedding, which is typically constrained to have a low dimension. In the following, we describe various methods for embedding a given network in Euclidean space. For a graph  $G$  with  $n$  nodes, a weighted adjacency matrix  $W$  and a diagonal degree matrix  $D$ , we define its Laplacian matrix as  $L = D - W$ .

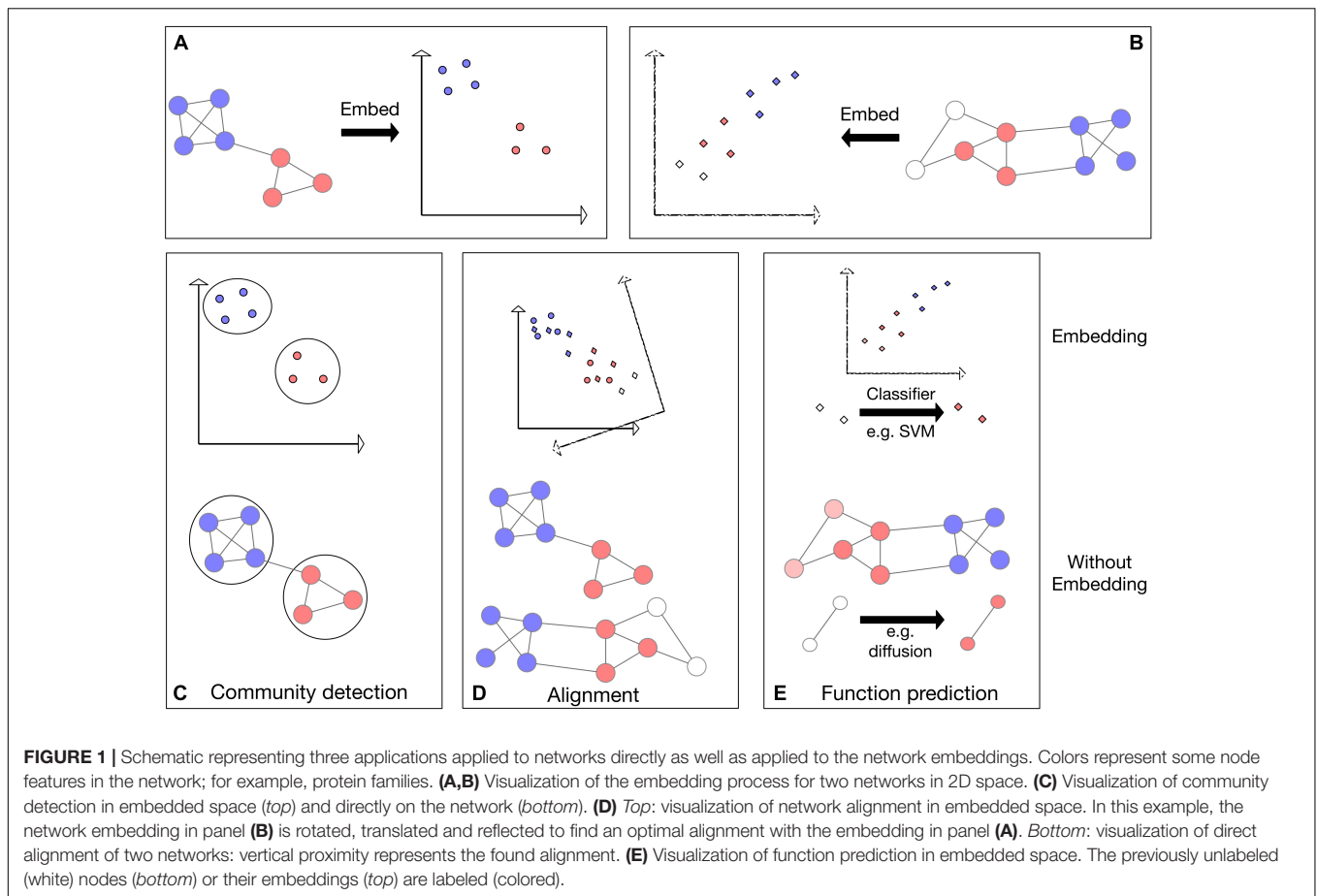
Graph drawing algorithms are perhaps the best-known embedding techniques, commonly used to visualize a graph in 2D space. Initially proposed in (Eades, 1984) as an extension of (Tutte, 1963), and further developed in (Fruchterman and Reingold, 1991), the spring-embedder model is a particularly elegant example: one can imagine that connected pairs of nodes are attached to springs which bring them closer together, while all nodes repel each other so as not to be placed too closely together. Other classes of graph drawing algorithms, including

multi-level and dimensionality reduction-based techniques, are described in detail in a recent review (Gibson et al., 2013). Spatial analysis of functional enrichment (Baryshnikova, 2018) is one recent application of force-directed graph drawing algorithm, designed for the annotation and visualization of large, complex biological networks.

One of the fundamental methods to decompose a matrix is spectral decomposition, i.e., decomposing the matrix into its eigenvectors and eigenvalues. Given a network, the principal eigenvectors  $Q$  of its Laplacian matrix capture membership of nodes in implicit network clusters, commonly used for embedding (Belkin and Niyogi, 2003). The matrix  $Q$  is obtained by optimizing  $\min_{Q \in \mathbb{R}^{n \times c}} \text{Trace}(Q^T L^+ Q)$ , s.t.  $Q^T Q = I$ , where  $L^+ = I - D^{-1/2} W D^{-1/2}$  is a normalized Laplacian and  $C$  is the number of clusters. However, this spectral embedding reflects the global structure in the network without taking into consideration more fine-grained local structures and is therefore sensitive to noise. Wang et al. (2017a) recently introduced the Vicus matrix as a local-neighborhood version of the Laplacian matrix. Each cell of the Vicus matrix represents the probability of node  $j$  having the same label as node  $i$  if we did a random walk around the local neighborhood of node  $i$ . Encoding local neighborhoods in this fashion does not only preserve the geometric properties of the original Laplacian matrix but also reduces the noise and improves the quality of the embedding. Wang et al. showed that for a variety of tasks, including network clustering of single-cell RNA-seq data, cluster stability, identification of rare cell populations, and ranking of genes associated with cancer subtypes, Vicus-based spectral methods outperformed Laplacian-based spectral methods on a wide variety of biological tasks.

Diffusion-based approaches focus on embedding nodes into low-dimensional vector spaces by first using random walks to construct a network neighborhood of every node in the network, and then optimizing an objective function with network neighborhoods as input (Perozzi et al., 2014a; Tang et al., 2015; Grover and Leskovec, 2016). The objective function is carefully designed to preserve both the local and global network structures. For example, a popular method, Mashup, complements traditional random walks, which yield only diffusion states, with a dimensionality reduction step that is aimed at reducing the noise in these diffusion computations. To this end, Mashup approximates each diffusion state  $s_i$  with a multinomial logistic model based on a latent vector representation of nodes that uses far fewer dimensions than the original,  $n$ -dimensional state. Specifically, if the latent vector representation for node  $i$  is denoted by  $x_i$ , Mashup also constructs a contextual vector  $w_i$  that has the same dimensionality as  $x_i$  and captures the topology of the subnetwork around node  $i$ . To this end, Mashup computes the probability assigned to node  $j$  in the diffusion state of node  $i$  as  $\hat{s}_{ij} = \frac{\exp(x_i^T w_j)}{\sum_k \exp(x_i^T w_k)}$ , so that these computed diffusion states align with the original diffusion states. Mashup constructs an optimization framework to minimize the KL-divergence of these two diffusion states and applies standard gradient descent methods to solve for the latent representations.

Another widely used network embedding algorithm that uses random walks is node2vec (Grover and Leskovec, 2016).



Node2vec learns node embeddings so that a node's embedding can predict nearby (neighborhood) nodes. Technically, the network neighborhood  $N(u)$  is a set of nodes that appear in an appropriately biased, short random walk from node  $u$  (Grover and Leskovec, 2016). The goal of the algorithm is to find an embedding  $f(u)$  such that the conditional probability of observing  $u$ 's network neighbors  $N(u)$  is maximized. This conditional probability is modeled using a softmax function, leading to the following log likelihood:  $\sum_u \sum_{v \in N(u)} \log \frac{\exp(f(u) \cdot f(v))}{\sum_w \exp(f(w) \cdot f(u))}$ , across all nodes  $u$  in the network. Once embeddings are learned, one can use them for any downstream prediction task, including node classification, link prediction, and clustering. A similar network embedding algorithm is DeepWalk (Perozzi et al., 2014b). DeepWalk has been originally proposed to embed nodes in a social network setting, taking ideas from the linguistics literature (Perozzi et al., 2014b). In DeepWalk, the embeddings are learned based on truncated random walks which can be intuitively thought of as putting words (nodes) into sentences (sequences of nodes visited by a random walk). In the biological context, DeepWalk has been used to associate miRNAs with diseases (Li et al., 2017), predict drug target associations (Zong et al., 2017), and predict protein function (Kulmanov et al., 2017).

With the advent of deep learning methods, several deep learning approaches were proposed to embed networks. An

important class of deep learning methods for network embedding are graph neural networks that generalize the notion of convolutions typically applied to image datasets to operations that can operate on arbitrary graphs (Defferrard et al., 2016; Kipf and Welling, 2016; Gilmer et al., 2017; Hamilton et al., 2017a). One can see graph neural networks as an embedding methodology that distills high-dimensional information about each node's neighborhood into a dense vector embedding without requiring manual feature engineering (Defferrard et al., 2016; Kipf and Welling, 2016; Gilmer et al., 2017; Hamilton et al., 2017a). A graph neural network has two main components. First, the encoder, maps a node  $u$  to a low-dimensional embedding  $f(u)$ , based on  $u$ 's local neighborhood structure, its position in the graph, and/or its attributes. Next, the decoder takes the embeddings and extracts user-specified predictions from these embeddings. In contrast to embedding approaches that use random walks (reviewed above), graph neural networks support end-to-end learning. One can jointly optimize all trainable parameters and propagate gradients of the objective function through the encoder as well as the decoder. End-to-end learning can lead to substantial improvements in performance (Defferrard et al., 2016; Zitnik et al., 2018).

There has been significant recent interest in graph embeddings in non-Euclidean spaces. In particular, hyperbolic spaces have attracted much attention due to successful natural language

processing models which use them for embedding words (Chamberlain et al., 2017). Muscoloni et al. (2017) describe a general algorithm termed “coalescent embedding” for embedding vertices in hyperbolic spaces. The algorithm proceeds by pre-weighting the network and applying a non-linear dimension reduction technique, followed by computing and adjusting the angular positions of the Euclidean embeddings and radial positioning according to node degree. More generally, networks and their respective embeddings can be interpreted geometrically, as described in recent reviews (Barthélemy, 2011; Papadopoulos et al., 2015; Moyano, 2017). These geometric models have been used successfully in applications to biological networks, particularly protein–protein interaction (PPI) networks (Serrano et al., 2012; Alanis-Lobato et al., 2016, 2018).

## APPLICATIONS

### Network Alignment

A basic operation in biological research is to transfer knowledge across species. Indeed, sequence alignment has been the power horse of computational biology for almost five decades now. With the availability of physical interaction data, it was suggested to generalize alignment concepts to the network level (Kelley et al., 2003; Sharan and Ideker, 2006). There are several types of network alignment problems, here we focus on global network alignment where given the networks of two species (typically, PPI networks) one wishes to identify a 1–1 correspondence between the proteins of the two species under which the networks are most similar (**Figure 1D**).

A leading approach to this problem is the IsoRank algorithm (Singh et al., 2008) which is based on Google’s PageRank method, essentially measuring the correspondence, or similarity, between two proteins from different species based on the similarities of their neighboring nodes in the two corresponding networks. Thus, if we denote by  $R_{ij}$  the similarity between proteins  $i$  and  $j$  (from two different species), and we let  $N(i)$  denote the (open) neighborhood of protein  $i$  in its network, then:

$$R_{ij} = \frac{1}{|N(i)||N(j)|} \sum_{u \in N(i), v \in N(j)} R_{uv}$$

These recursive equations give rise to an eigenvalue problem and their solution is used as an input to a maximum matching algorithm to compute the eventual correspondence.

Another, more recent approach is MAGNA (Saraph and Milenkovi a, 2014) and its successor MAGNA++ (Vijayan et al., 2015). MAGNA uses a genetic algorithm to find the optimal alignment, where individuals are viewed as permutations of the nodes. Crossover relies on the notion of adjacency, where a pair of permutations is adjacent if they differ only by a single swap of two nodes; the crossover of two permutations is then the midpoint of the shortest path between the two permutations in the graph constructed from these adjacencies. Selection can be based on any metric, such as EC. MAGNA++ augments this approach by including cross-species node similarity information.

An extensive review of methods for biological network alignment can be found in (Guzzi and Milenkovic, 2018) that mentions over thirty different approaches. Comparative network analysis methods are further reviewed in (Emmert-Streib et al., 2016).

A recent work by Fan et al. (2017) uses an embedding-based approach, MuNK, to compare networks across species by assessing similarity via embedded network topologies. The idea is to project the nodes of the two networks into the same Euclidean space in a way that preserves their intra-species network similarity and inter-species sequence similarity. For each species separately, a kernel similarity function is defined, and the corresponding embedding is computed by matrix decomposition. To tie the projections together, Fan et al. (2017) assume a given set of known matches, regarded as landmarks, between the two networks. A similar embedding approach that does not require a known subset of correspondences was suggested in (Heimann et al., 2018).

As a test case for network embedding, we evaluated the two algorithms, IsoRank and MuNK, using metrics of alignment quality. A common and simple metric is the edge correctness (EC), defined as the percentage of edges conserved under the mapping  $f$  (Kuchaiev et al., 2009; Clark and Kalita, 2014):

$$EC = \frac{|f(E_A) \cap E_B|}{|E_A|} \times 100\%$$

Note that the EC metric is asymmetric, and the order of the networks is traditionally chosen to maximize EC, i.e.,  $A$  is chosen to be the smaller of the two networks. Beyond topological similarity, one can use different biological annotations, such as the Gene Ontology (GO) functional annotation, to compute biologically relevant measures of alignment quality such as GO functional consistency (Aladag and Erten, 2013), defined as the proportion of aligned pairs with more than  $k$  GO terms in common.

Similar to the use of landmarks in MuNK, IsoRank can incorporate additional similarity information in its computation of the score matrix, so the landmark pairs are provided as a binary information matrix to the IsoRank algorithm. In our experiments, we produce two outputs for method comparison: cross-species pairwise similarity scores and the node-to-node mappings. Thus, in addition to the two measures described above that use the node-to-node mappings, we also evaluated IsoRank and MuNK using AUPR as a measure of enrichment of GO functional consistency with respect to the cross-species pairwise similarity scores. When comparing MuNK to the more recent MAGNA++, MAGNA++ performs very well according to EC (as it optimizes EC directly), but it does not output node scores so we could not directly compare MuNK to MAGNA++ according to AUPR and other metrics. Per the author recommendation, the regularization parameter for the Laplacian in MuNK was fixed at  $\lambda = 0.05$ . Damping can be used in the PageRank step of the IsoRank algorithm, and therefore we performed a grid search with step size 0.05 over possible convexity parameters  $\alpha \in (0, 1)$ , optimizing for EC score. As input data, we use the PPI networks for two species of yeast, *S. cerevisiae* and

*S. pombe*, extracted from the BioGRID interaction database (Oughtred et al., 2018).

IsoRank performs better on the measures directly related to the node mapping (Table 1). This may be due to the fact that the cross-species similarity coefficients in IsoRank directly incorporate local neighborhood (i.e., topological) information, a fact that the IsoRank greedy algorithm is designed to take advantage of. The MuNK scores predict functional correctness better than the scores produced by IsoRank, suggesting that MuNK's learned embedded space is biologically meaningful potentially even beyond alignment. In comparing network alignment methods (Guzzi and Milenkovic, 2018) also found that methods that do very well according to the topological quality measures are not very good as far as functional quality is concerned. The interpretability of the embedding space is one of the primary benefits of embedding techniques over standard approaches in the case of network alignment. For example, the embedding space learned by MuNK captures biological information beyond pairwise node alignment, specifically, cross-species synthetic lethal interactions (Fan et al., 2017).

## Community Detection

One of the natural uses of a network is the identification of clusters, or modules of similar nodes, a task known as community detection (Fortunato, 2010). Community detection methods (Figure 1C) have great uses in biology from protein

module identification to disease subnetwork discovery (Ghiassian et al., 2015; Menche et al., 2015). Among the most popular community detection methods on networks are random walk-based approaches including Louvain (Blondel et al., 2008), Infomap (Rosvall and Bergstrom, 2011), Label propagation (Raghavan et al., 2007), and Walktrap (Pons and Latapy, 2005), that came up as best performers in a review comparing these and other approaches (Yang et al., 2016). Originally developed for community detection in social networks, these methods are frequently used in biology (Barabási et al., 2011), for example to identify cancer drivers (Cantini et al., 2015).

Network embedding for the purpose of community detection was covered in a recent review (Hamilton et al., 2017b). The authors hypothesized that due to vector-like embedding representation of a network, there is a wider range of clustering and community detection methods that can be applied to embedded networks as compared to graphs directly. The authors further introduced an encoder-decoder framework that unifies many of the recently popularized approaches, including DeepWalk (Perozzi et al., 2014a) and node2vec (Grover and Leskovec, 2016). A geometric approach, not covered in the review, suggests a scalable embedding of networks in a hyperbolic circle and show that the popular random walk-based community detection methods (Louvain, Infomap, Label propagation, and Walktrap) can be significantly boosted when applied to hyperbolic distances (Muscoloni et al., 2017).

**TABLE 1** | Comparative analysis of direct vs. embedding methods across a range of problems in network biology.

	IsoRank ( $\alpha = 0.5$ )	MuNK ( $\lambda = 0.05$ )
<b>A. Network alignment</b>		
EC	<b>39.0%</b>	21.9%
GOC		
$K = 20$	<b>63.4%</b>	57.6%
$K = 50$	<b>20.7%</b>	17.9%
$K = 100$	<b>1.2%</b>	1.0%
GOC (AUPR)	0.721	<b>0.746</b>
Runtime	26 min 40 s (incl. grid search)	1 min 52 s (incl. alignment)
	<b>densityCut</b> ( $K = 4, \alpha = 0.9$ )	<b>Vicus</b> ( $K = 10, \sigma = 0.5$ )
<b>B. Community detection</b>		
Buettner ( $C = 2, C_t = 11$ )	0.256	<b>0.316</b>
Kolodziejczyk ( $C = 5, C_t = 4$ )	0.325	<b>0.552</b>
Pollen ( $C = 13, C_t = 11$ )	<b>0.931</b>	0.928
Usoskin ( $C = 9, C_t = 4$ )	0.373	<b>0.591</b>
Avg. Runtime	1 min 15 s (incl. parameter grid search)	<5 s
STRING v9.1 <i>Homo sapiens</i> . Included with the Mashup distribution.		
	<b>GeneMANIA</b>	<b>Mashup</b>
<b>C. Function prediction</b>		
AUPR		
MF	0.327	<b>0.372</b>
BP	0.213	<b>0.222</b>
CC	<b>0.514</b>	0.487
Avg. Runtime	3 min 57 s	14 min 56 s (incl. recommended SVM tuning procedure)

Running times were obtained on a 64-bit machine with Intel Core i5-8400 CPU @ 2.80 GHz  $\times$  6 with 16 GB RAM running Ubuntu 18.04. Bold refers to the most successful result, according to the referenced metric.

We compared two community detection methods, an embedding-based and a graph-based, on the problem of single-cell RNA-seq (scRNA-seq) analysis. scRNA-seq data has recently emerged as a powerful tool to decipher the heterogeneity of cell populations. This is an important and growing area of network applications where community detection methods are used to perform clustering on the constructed cell-to-cell networks (Wang et al., 2018). Given a gene expression matrix, Gaussian kernel is usually adopted to construct a pairwise similarity network in which nodes represent cells and edge weights depict the similarity between cells.

The first method is Vicus, a generalization of spectral clustering, which we combined with  $k$ -means clustering in the embedded space. For the network-based approach, we used densityCut, a random walk-based community detection method, which approximates clusters using the density of local neighborhoods. The densityCut method approximates the true network using a  $k$ -nearest neighbor graph, and selects the number of clusters using an automated procedure. Therefore, this number of clusters was used as input to the  $k$ -means step of the Vicus evaluation. We used four scRNA-seq datasets, all from *Mus musculus* (Pollen et al., 2014; Buettner et al., 2015; Kolodziejczyk et al., 2015; Usoskin et al., 2015) but which vary according to tissue of origin (neural, blood and stem cells) and have known ground truth labels. We evaluated performance using normalized mutual information (NMI). Vicus outperformed densityCut on three of the four datasets (Table 1).

## Function Prediction

Another fundamental problem in network biology is the inference of protein function from the known functions of its network neighbors (Sharan et al., 2007). The earliest approach to this problem, neighborhood counting (Schwikowski et al., 2000), predicted a protein to be involved in a certain function if a sufficient number of its direct (or up to some specified distance) neighbors had this property. Current state of the art methods are based on similar guilt-by-association principles (Figure 1E). For example, Cao et al. (2013) define a distance metric between proteins that is based on network diffusion, thus capturing similarities that are based on multiple paths in the network.

These single-network methods were generalized in several ways (Cho et al., 2016) integrate information across multiple networks and use a low rank approximation of the network diffusion based similarities to reduce potential noise. The integration challenge is also tackled by (Gligorijevic et al., 2018) who learn a compact node representation using deep autoencoders. In Fan et al. (2017), the cross-species embedding is utilized to infer protein function. Zitnik and Leskovec (2017) suggest a network embedding approach for predicting tissue-specific protein function, which encourages proteins to share features not only with their network neighbors but also with proteins that are active in similar tissues.

Two recent methods were compared on the task of protein function prediction using multiple interaction networks.

GeneMANIA performs label diffusion, while Mashup finds an embedding for each of the proteins, allowing one to use traditional classification techniques such as support vector machines (SVMs). The area under the precision-recall curve (AUPR) was used as an evaluation metric. Overall, Mashup performed better with respect to molecular function and biological process annotations, while GeneMANIA performed better on the cellular compartment annotation (Table 1).

## Network Denoising

The application of network biology techniques to experimental data depends on the accuracy and completeness of the network of interest. The challenge of noisy interaction measurements plagues many different types of biological networks, such as Hi-C interaction networks (Rao et al., 2014), cell-cell interaction networks (Wang et al., 2017b), and PPI networks (Saito et al., 2002; Przulj et al., 2004; Chua and Wong, 2008; Higham et al., 2008; Kuchaiev et al., 2009; You et al., 2010; Marras et al., 2011; Alanis-Lobato et al., 2013; Cannistraci et al., 2013; Newman, 2018a,b). Such noise adversely impacts the performance of downstream analysis, calling for methods for network denoising.

The most common approach to denoise any given network is to perform diffusions on the network to exploit high-order structures that can potentially improve the qualities of the direct links between nodes. Diffusion maps (Coifman et al., 2005) employ high-order random walks and then use spectral decomposition to construct an affinity measure. A tensor-based dynamical model (Wang et al., 2012) aims to search high-order paths between pairs of objects through their common nearest neighbors. A low-rank constraint has been employed to help denoise the network manifold (Wang and Tu, 2013). Diffusion-state distance (DSD) (Cao et al., 2013) was utilized to denoise PPI networks and improve the signal-to-noise ratio for better prediction of protein functions. To tackle the problem of transitive edges in networks in a computationally efficient way (Feizi et al., 2013) proposed a simple closed-form solution, called Network Deconvolution (ND), to infer direct links.

An alternative direction of network denoising takes embedding-based approaches. For instance, Mashup (Cho et al., 2016) aims to learn compact low-dimensional vector representation of proteins that best explains their wiring patterns for the input protein-protein association networks by applying a matrix factorization method on the diffused network. The embeddings of the nodes (proteins) reflect the relational structures of the original network, therefore facilitating the downstream applications by feeding the embeddings to a support vector machine.

A recent study (Wang et al., 2018) performed an in-depth comparison between these network denoising methods in three different experimental settings: PPI function predictions, HiC network module detection, and species identification. The study highlighted the advantages of embedding-based methods such as Mashup (Cho et al., 2016) when the network contains distinct cluster structures and the noise level is small. However, it also showed that when the cluster structures are corrupted by high

noise, existing methods usually fail to uncover the underlying network structure.

## Pharmacogenomics

Modern pharmaceutical research faces challenges with decreasing productivity in drug development and a persistent gap between therapeutic needs and available treatments (Hodos et al., 2016; Moffat et al., 2017). Network approaches have emerged as a promising direction to address these challenges and improve our understanding of the therapeutic and side effects of drugs (Hopkins, 2008; Berger and Iyengar, 2009). We review three practically important problems within the realm of pharmacogenomics that have been tackled with network embedding methods: drug-target prediction, drug-drug interaction prediction and prediction problems involving small molecules.

Drugs influence biological systems by binding to target proteins and affecting their downstream activity (Imming et al., 2006). Network approaches formulate drug-target interaction prediction as a link prediction task on a graph

of drugs/chemicals and the proteins which they interact with (Yildirim et al., 2007; Yamanishi et al., 2010; Perlman et al., 2011; Chen et al., 2012; Cheng et al., 2012; Gönen, 2012; Isik et al., 2015; Zitnik and Zupan, 2016; Luo et al., 2017; Wen et al., 2017; Lee and Nam, 2018). Given such a graph (Crichton et al., 2018) use various node embedding methods, including node2vec (Grover and Leskovec, 2016), DeepWalk (Perozzi et al., 2014b), and LINE (Tang et al., 2015), to embed nodes into a compact vector space in a manner that preserves local network structure. As a result, drugs with many shared target proteins obtain similar embeddings, and vice-versa, proteins targeted by similar drugs obtain similar embeddings. These embeddings are thus well-suited for predicting drug-target interactions by calculating the similarity between embeddings representing the drug and the protein, or by using embeddings as inputs to a machine learning method (Crichton et al., 2018). Alternatively, predictions can be made in an end-to-end fashion, where a neural network learns node embeddings and predicts interactions directly from the graph (Wang and Zeng, 2013; Gao et al., 2018; Wan et al., 2018).

**TABLE 2** | A summary of network embedding tools and their applications.

Name of the tool	Availability	What was it applied to
<b>Denoising</b>		
Network enhancement	Matlab code <a href="http://snap.stanford.edu/ne/">http://snap.stanford.edu/ne/</a>	Hi-C interaction networks combining gene interaction networks across tissues
Single-cell representation learning	Binary <a href="https://github.com/SuntreeLi/SCRL">https://github.com/SuntreeLi/SCRL</a>	Single-cell RNA-seq data
Geometric denoising	<a href="http://kuchaev.com/Denoising/">http://kuchaev.com/Denoising/</a>	PPI networks
<b>Network alignment</b>		
MuNK	Python code and all Anaconda-reproducible experiments <a href="https://github.com/lrgr/munk">https://github.com/lrgr/munk</a>	Cross-species functional PPIs (yeast, mouse, human)
<b>Community detection</b>		
Minimum curvilinearity embedding II	<a href="https://sites.google.com/site/carlovittoriocannistraci/5-datasets-and-matlab-code/minimum-curvilinearity-ii-april-2012">https://sites.google.com/site/carlovittoriocannistraci/5-datasets-and-matlab-code/minimum-curvilinearity-ii-april-2012</a>	(i) Cerebrospinal fluid proteomics – neuropathic pain
Vicus		(ii) Transcription factor expressions – tissue prediction Single-cell RNA-seq: (i) Pollen – neural and stem cells (ii) Usoskin – mouse neurons, sensory subtypes (iii) Buettner – embryonic stem cells (iv) Kolodziejczyk – pluripotent cells
Coalescent embedding	<a href="https://github.com/biomedical-cybernetics/coalescent_embedding">https://github.com/biomedical-cybernetics/coalescent_embedding</a>	Non-biological
<b>Function prediction</b>		
Mashup	<a href="http://cb.csail.mit.edu/cb/mashup/">http://cb.csail.mit.edu/cb/mashup/</a>	Protein function prediction, gene ontology reconstruction, and genetic interaction prediction
OhmNet	<a href="http://snap.stanford.edu/ohmnet/">http://snap.stanford.edu/ohmnet/</a>	Tissue-specific gene function prediction
Disease gene discovery	<a href="http://snap.stanford.edu/pathways/">http://snap.stanford.edu/pathways/</a>	Disease pathway detection
<b>Pharmacogenomics</b>		
Molecular fingerprints	<a href="https://github.com/HIPS/neural-fingerprint">https://github.com/HIPS/neural-fingerprint</a>	Prediction of molecular properties, including drug efficacy, solubility, and photovoltaic efficiency
Decagon	<a href="http://snap.stanford.edu/decagon/">http://snap.stanford.edu/decagon/</a>	(i) Polypharmacy side-effect prediction (ii) Drug-drug interaction prediction
Graph convolutional policy network	<a href="https://github.com/bowenliu16/rl_graph_generation">https://github.com/bowenliu16/rl_graph_generation</a>	Molecular graph generation
Residual LSTM Embeddings	<a href="https://github.com/deepchem/deepchem">https://github.com/deepchem/deepchem</a>	(i) Drug side-effect prediction (ii) Drug toxicity prediction

Detecting drug–drug interactions, in which the activity of one drug changes, favorably or unfavorably, if taken with another drug, is an important challenge with significant implications for patient mortality and morbidity (Chan and Giaccia, 2011; Guthrie et al., 2015; Han et al., 2017). Ma et al. (2018) model each drug as a node in a multi-view drug association graph, where edges between drugs in different views encode different types of similarity between drugs. The approach uses graph convolutional networks (Kipf and Welling, 2016) to embed the multi-view graph and attentive mechanisms (Veličković et al., 2018) to fuse information from multiple views and to make learning more interpretable. By such embedding, the approach learns a similarity score between any two drugs and uses the scores to predict drug–drug interactions. While such an approach can be useful to describe drug interactions at the cellular level (Sridhar et al., 2016; Ryu et al., 2018), it cannot predict the safety or side effects of drug combinations. To identify the side effects of drug combinations and provide guidance on the development of new drug therapies (Zitnik et al., 2018) developed an embedding approach that constructs a multi-modal graph of PPIs, drug–protein interactions, and drug–drug interactions, where each drug–drug interaction is labeled by a different edge type signifying the type of the side effect. The approach takes the multi-modal graph and uses graph neural networks as an embedding methodology to distill information about each node’s network neighborhood into an embedding vector without any hand-engineering. The final approach is an end-to-end method for predicting side effects of drug combinations that considers all types of side effects at once. The approach learns embeddings of side effects that are indicative of polypharmacy in patients.

Chemical prediction problems represent another class of practically important graph problems (Ralaivola et al., 2005; Altae-Tran et al., 2017; Gilmer et al., 2017; Gómez-Bombarelli et al., 2018). One key distinction between these problems and standard network prediction tasks discussed above is that chemical prediction problems are graph-level classification problems where individual data examples are graphs (rather than nodes) representing small molecules. Typical prediction tasks aim to predict various molecular properties such as drug efficacy or solubility (Coley et al., 2017; Jin et al., 2017), predict which drugs bind to which target proteins (Morris et al., 2018), and identify sites at which a particular candidate drug binds to a target protein (Feinberg et al., 2018). The input to a predictor is a small molecule, which is commonly represented as a graph in which nodes and edges represent atoms and bonds between atoms, respectively. One difficulty with such inputs is that molecular graphs can be of arbitrary size and shape (Niepert et al., 2016; Xu et al., 2017). However, currently, most machine learning pipelines can only handle inputs of a fixed size. For this reason, state-of-the-art systems use embedding techniques to embed molecular graphs into fixed-dimensional embeddings and then use the learned representations as inputs to a fully connected deep neural network or other standard machine learning methods (Duvenaud et al., 2015; Kearnes et al., 2016). The proposed graph convolution models do not yet consistently outperform

traditional structural-based fingerprints, however, their flexibility and potential for further optimization and development have led to models that provide significant boosts in the predictive power over older fingerprints.

## CONCLUSION

We have reviewed several classes of approaches for network embedding, including spectral-based methods, random-walk based approaches and deep neural network techniques. We have demonstrated the utility of these approaches in a broad set of applications, ranging from network alignment to community detection, protein function prediction, and network denoising. We have also discussed recent embedding approaches in pharmacogenomics. We were interested in seeing whether the field of network embedding indeed enhances the types of questions that can be answered using graph-based approaches and our conclusion is that there is value in both graph-based and graph-embedding-based methods in a variety of applications.

In our experiments we found that depending on the task at hand and metric used, sometimes graph-based methods outperformed network embedding tools. This was the case with, for example, IsoRank beating MuNK with respect to edge conservation in network alignment, whereas MuNK outperformed IsoRank according to the area under the precision recall curve with respect to node mapping. In community detection experiments, our results were reversed, where the embedding method outperformed the graph-based method 3 out of 4 times. In fact, there is no single metric according to which one type of method is consistently better than the other. Even in compute time, where embedding methods outperform graph-based methods most of the time, on the function prediction task graph-based GeneMANIA outperforms the embedding method Mashup. This implies that the choice of graph-based versus embedding-based method will depend on many factors, not just the task at hand, but also the aspect or evaluation measure of highest importance to the user.

The network embedding principles create new opportunities to model large network datasets and move beyond standard prediction tasks of node classification, link prediction, and node clustering. For example, given a partially observed network of interactions between drugs, diseases, and proteins, one might be interested in posing a logical query: “What proteins are likely to be associated with diseases that have both symptoms X and Y?” Such a query requires reasoning about all possible proteins that might be associated with at least two diseases, which, in turn, clinically manifest through symptoms X and Y. Valid answers to such queries correspond to subgraphs. Since edges in the network might be missing because of biotechnological limits and natural variation, naively answering the queries requires enumeration over all possible combinations of diseases (Hamilton et al., 2018) developed a network embedding approach that answers such complex logical queries and achieves a time complexity linear in the size of a query, compared to the exponential complexity required by a naive enumeration-based approach. The approach embeds nodes into a low-dimensional space and



represents logical operators as learned geometric operations in this embedding space. They demonstrated the utility of the approach in a study involving a biomedical network of drugs, diseases, proteins, side effects, and protein functions with millions of edges.

We summarize network embedding tools that are used in the biomedical field in **Table 2**. We expect the importance of these tools to grow with the magnitude and complexity of biomedical data that are being generated.

## REFERENCES

- Aladag, A. E., and Erten, C. (2013). SPINAL: scalable protein interaction network alignment. *Bioinformatics* 29, 917–924. doi: 10.1093/bioinformatics/btt071
- Alanis-Lobato, G., Cannistraci, C. V., and Ravasi, T. (2013). Exploitation of genetic interaction network topology for the prediction of epistatic behavior. *Genomics* 102, 202–208. doi: 10.1016/j.ygeno.2013.07.010
- Alanis-Lobato, G., Mier, P., and Andrade-Navarro, M. (2018). The latent geometry of the human protein interaction network. *Bioinformatics* 34, 2826–2834. doi: 10.1093/bioinformatics/bty206
- Alanis-Lobato, G., Mier, P., and Andrade-Navarro, M. A. (2016). Manifold learning and maximum likelihood estimation for hyperbolic network embedding. *Appl. Netw. Sci.* 1:10. doi: 10.1007/s41109-016-0013-0
- Altae-Tran, H., Ramsundar, B., Pappu, A. S., and Pande, V. (2017). Low data drug discovery with one-shot learning. *ACS Cent. Sci.* 3, 283–293. doi: 10.1021/acscentsci.6b00367
- Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918
- Barabási, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272
- Barthélemy, M. (2011). Spatial networks. *Phys. Rep.* 499, 1–101.
- Baryshnikova, A. (2018). Spatial analysis of functional enrichment (SAFE) in large biological networks. *Methods Mol. Biol.* 1819, 249–268. doi: 10.1007/978-1-4939-8618-7\_12
- Belkin, M., and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 15, 1373–1396. doi: 10.1162/089976603321780317
- Berger, S. I., and Iyengar, R. (2009). Network analyses in systems pharmacology. *Bioinformatics* 25, 2466–2472. doi: 10.1093/bioinformatics/btp465
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008:10008.
- Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., et al. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* 33, 155–160. doi: 10.1038/nbt.3102
- Cai, H., Zheng, V. W., and Chang, K. C.-C. (2018). A comprehensive survey of graph embedding: problems, techniques, and applications. *IEEE Trans. Knowl. Data Eng.* 30, 1616–1637. doi: 10.1109/tkde.2018.2807452
- Cannistraci, C. V., Alanis-Lobato, G., and Ravasi, T. (2013). Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* 29, i199–i209. doi: 10.1093/bioinformatics/btt208
- Cantini, L., Medico, E., Fortunato, S., and Caselle, M. (2015). Detection of gene communities in multi-networks reveals cancer drivers. *Sci. Rep.* 5: 17386.
- Cao, M., Zhang, H., Park, J., Daniels, N. M., Crovella, M. E., Cowen, L. J., et al. (2013). Going the distance for protein function prediction: a new distance metric for protein interaction networks. *PLoS One* 8:e76339. doi: 10.1371/journal.pone.0076339
- Chamberlain, B. P., Clough, J., and Deisenroth, M. P. (2017). Neural embeddings of graphs in hyperbolic space. arXiv:1705.10359 [Preprint]. doi: 10.1371/journal.pone.0076339
- Chan, D. A., and Giaccia, A. J. (2011). Harnessing synthetic lethal interactions in anticancer drug discovery. *Nat. Rev. Drug Discov.* 10, 351–364. doi: 10.1038/nrd3374

## AUTHOR CONTRIBUTIONS

WN did the performance comparisons. All authors participated in writing the manuscript.

## FUNDING

AG and RS were supported by a TAU-UOT cooperation grant.

- Chen, X., Liu, M.-X., and Yan, G.-Y. (2012). Drug-target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst.* 8, 1970–1978.
- Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., et al. (2012). Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* 8:e1002503. doi: 10.1371/journal.pcbi.1002503
- Cho, H., Berger, B., and Peng, J. (2016). Compact integration of multi-network topology for functional analysis of genes. *Cell Syst.* 3, 540.e5–548.e5.
- Chua, H. N., and Wong, L. (2008). Increasing the reliability of protein interactomes. *Drug Discov. Today* 13, 652–658. doi: 10.1016/j.drudis.2008.05.004
- Clark, C., and Kalita, J. (2014). A comparison of algorithms for the pairwise alignment of biological networks. *Bioinformatics* 30, 2351–2359. doi: 10.1093/bioinformatics/btu307
- Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., et al. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7426–7431. doi: 10.1073/pnas.0500334102
- Coley, C. W., Barzilay, R., Green, W. H., Jaakkola, T. S., and Jensen, K. F. (2017). Convolutional embedding of attributed molecular graphs for physical property prediction. *J. Chem. Inf. Model.* 57, 1757–1772. doi: 10.1021/acs.jcim.6b00601
- Cox, T. F., and Cox, M. A. A. (2000). *Multidimensional Scaling*, 2nd Edn. Boca Raton, FL: CRC Press.
- Crichton, G., Guo, Y., Pyysalo, S., and Korhonen, A. (2018). Neural networks for link prediction in realistic biomedical graphs: a multi-dimensional evaluation of graph embedding-based approaches. *BMC Bioinformatics* 19:176. doi: 10.1186/s12859-018-2163-9
- Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016)* (Barcelona: NIPS), 3844–3852.
- Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., et al. (2015). “Convolutional networks on graphs for learning molecular fingerprints,” in *Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015)* (Barcelona: NIPS), 2224–2232.
- Eades, P. (1984). A heuristic for graph drawing. *Congr. Numer.* 42, 149–160.
- Emmert-Streib, F., Dehmer, M., and Shi, Y. (2016). Fifty years of graph matching, network alignment and network comparison. *Inf. Sci.* 34, 180–197. doi: 10.1016/j.ins.2016.01.074
- Fan, J., Cannistra, A., Fried, I., Lim, T., Schaffner, T., Crovella, M., et al. (2017). A multi-species functional embedding integrating sequence and network structure. *bioRxiv* [Preprint].
- Feinberg, E. N., Sur, D., Husic, B. E., Mai, D., Li, Y., Yang, J., et al. (2018). Spatial graph convolutions for drug discovery. arXiv:1803.04465 [Preprint].
- Feizi, S., Marbach, D., Médard, M., and Kellis, M. (2013). Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat. Biotechnol.* 31, 726–733. doi: 10.1038/nbt.2635
- Fortunato, S. (2010). Community detection in graphs. *Phys. Rep.* 486, 75–174. doi: 10.1016/j.physrep.2009.11.002
- Fruchterman, T. M. J., and Reingold, E. M. (1991). *Graph Drawing by Force-Directed Placement*. Hoboken, NJ: Wiley.
- Gao, K. Y., Fokoue, A., Luo, H., Iyengar, A., Dey, S., and Zhang, P. (2018). “Interpretable drug target prediction using deep neural representation,” in *Proceedings of the Conference: Twenty-Seventh International Joint Conference on Artificial Intelligence* (New York, NY: IJCAI), 3371–3377.

- Ghiassian, S. D., Menche, J., and Barabási, A.-L. (2015). A DIseAse MOdule detection (DIAMOND) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput. Biol.* 11:e1004120. doi: 10.1371/journal.pcbi.1004120
- Gibson, H., Faith, J., and Vickers, P. (2013). A survey of two-dimensional graph layout techniques for information visualisation. *Inf. Visual.* 12, 324–357. doi: 10.1177/1473871612455749
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). Neural message passing for quantum chemistry. arXiv:1704.01212v2 [Preprint].
- 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
- Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* 4, 268–276. doi: 10.1021/acscentsci.7b00572
- Gönen, M. (2012). Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. *Bioinformatics* 28, 2304–2310. doi: 10.1093/bioinformatics/bts360
- Grover, A., and Leskovec, J. (2016). “node2vec: scalable feature learning for networks,” in *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, Vol. 22 (New York, NY: ACM), 855–864.
- Guthrie, B., Makubate, B., Hernandez-Santiago, V., and Dreischulte, T. (2015). The rising tide of polypharmacy and drug-drug interactions: population database analysis 1995–2010. *BMC Med.* 13:74. doi: 10.1186/s12916-015-0322-7
- Guzzi, P. H., and Milenkovic, T. (2018). Survey of local and global biological network alignment: the need to reconcile the two sides of the same coin. *Brief. Bioinform.* 19, 472–481.
- Hamilton, W. L., Bajaj, P., Zitnik, M., Jurafsky, D., and Leskovec, J. (2018). Embedding logical queries on knowledge graphs. arXiv:1806.01445v3 [Preprint].
- Hamilton, W. L., Ying, R., and Leskovec, J. (2017a). Inductive representation learning on large graphs. *arxiv*.
- Hamilton, W. L., Ying, R., and Leskovec, J. (2017b). Representation learning on graphs: methods and applications. *arxiv*.
- Han, K., Jeng, E. E., Hess, G. T., Morgens, D. W., Li, A., and Bassik, M. C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* 35, 463–474. doi: 10.1038/nbt.3834
- Heimann, M., Shen, H., Safavi, T., and Koutra, D. (2018). “REGAL: representation learning-based graph alignment,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management – CIKM '18* (New York, NY: ACM), doi: 10.1145/3269206.3271788
- Higham, D. J., Rašajski, M., and Pržulj, N. (2008). Fitting a geometric graph to a protein-protein interaction network. *Bioinformatics* 24, 1093–1099. doi: 10.1093/bioinformatics/btn079
- Hodos, R. A., Kidd, B. A., Shameer, K., Readhead, B. P., and Dudley, J. T. (2016). In silico methods for drug repurposing and pharmacology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 8, 186–210. doi: 10.1002/wsbm.1337
- Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* 4, 682–690. doi: 10.1038/nchembio.118
- Imming, P., Sinning, C., and Meyer, A. (2006). Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discov.* 5, 821–834. doi: 10.1038/nrd2132
- Isik, Z., Baldow, C., Cannistraci, C. V., and Schroeder, M. (2015). Drug target prioritization by perturbed gene expression and network information. *Sci. Rep.* 5:17417.
- Jin, W., Coley, C., Barzilay, R., and Jaakkola, T. (2017). “Predicting organic reaction outcomes with weisfeiler-lehman network,” in *Advances in Neural Information Processing Systems 30*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (New York, NY: Curran Associates, Inc.), 2607–2616.
- Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* 30, 595–608. doi: 10.1007/s10822-016-9938-8
- Kelley, B. P., Sharan, R., Karp, R. M., Sittler, T., Root, D. E., Stockwell, B. R., et al. (2003). Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11394–11399. doi: 10.1073/pnas.1534710100
- Kipf, T. N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv:1609.02907v4 [Preprint].
- Kolodziejczyk, A. A., Kim, J. K., Tsang, J. C. H., Ilicic, T., Henriksson, J., Natarajan, K. N., et al. (2015). Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* 17, 471–485. doi: 10.1016/j.stem.2015.09.011
- Kuchaiev, O., Rasajski, M., Higham, D. J., and Pržulj, N. (2009). Geometric de-noising of protein-protein interaction networks. *PLoS Comput. Biol.* 5:e1000454. doi: 10.1371/journal.pcbi.1000454
- Kulmanov, M., Khan, M. A., Hoehndorf, R., and Wren, J. (2017). DeepGO: predicting protein functions from sequence using a deep ontology-aware classifier. *Bioinformatics* 34, 660–668. doi: 10.1093/bioinformatics/btx624
- Lee, I., and Nam, H. (2018). Identification of drug-target interaction by a random walk with restart method on an interactome network. *BMC Bioinformatics* 19:208. doi: 10.1186/s12859-018-2199-x
- Li, G., Luo, J., Xiao, Q., Liang, C., Ding, P., and Cao, B. (2017). Predicting MicroRNA-disease associations using network topological similarity based on deepwalk. *IEEE Access* 5, 24032–24039. doi: 10.1109/access.2017.2766758
- Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., et al. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* 8:573.
- Ma, T., Xiao, C., Zhou, J., and Wang, F. (2018). Drug similarity integration through attentive multi-view graph auto-encoders. arXiv:1804.10850v1 [Preprint].
- Marras, E., Travaglione, A., and Capobianco, E. (2011). Manifold learning in protein interactomes. *J. Comput. Biol.* 18, 81–96. doi: 10.1089/cmb.2009.0258
- Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., et al. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347:1257601. doi: 10.1126/science.1257601
- Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., and Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* 16, 531–543. doi: 10.1038/nrd.2017.111
- Morris, P., DaSilva, Y., Clark, E., Hahn, W. E., and Barenholtz, E. (2018). “Convolutional neural networks for predicting molecular binding affinity to HIV-1 proteins,” in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics BCB '18* (New York, NY: ACM), 220–225.
- Moyano, L. G. (2017). Learning network representations. *Eur. Phys. J. Spec. Top.* 226, 499–518.
- Muscoloni, A., Thomas, J. M., Ciucci, S., Bianconi, G., and Cannistraci, C. V. (2017). Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nat. Commun.* 8:1615.
- Newman, M. E. J. (2018a). Estimating network structure from unreliable measurements. *Phys. Rev. E* 98:062321.
- Newman, M. E. J. (2018b). Network structure from rich but noisy data. *Nat. Phys.* 14, 542–545. doi: 10.1038/s41567-018-0076-1
- Niepert, M., Ahmed, M., and Kutzkov, K. (2016). “Learning convolutional neural networks for graphs,” in *Proceedings of the International Conference on Machine Learning (jmlr.org)*, Beijing, 2014–2023.
- Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., et al. (2018). The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47, D529–D541.
- Papadopoulos, F., Aldecoa, R., and Krioukov, D. (2015). Network geometry inference using common neighbors. *Phys. Rev. E* 92:022807.
- Perlman, L., Gottlieb, A., Atias, N., Ruppim, E., and Sharan, R. (2011). Combining drug and gene similarity measures for drug-target elucidation. *J. Comput. Biol.* 18, 133–145. doi: 10.1089/cmb.2010.0213
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014a). “Deepwalk: online learning of social representations,” in *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)* (New York, NY: ACM), 701–710.
- Perozzi, B., Al-Rfou, R., and Skiena, S. (2014b). “DeepWalk: online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '14* (New York, NY: ACM), 701–710.
- Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular

- heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* 32, 1053–1058. doi: 10.1038/nbt.2967
- Pons, P., and Latapy, M. (2005). “Computing communities in large networks using random walks,” in *Computer and Information Sciences – ISICIS 2005 Lecture Notes in Computer Science*, eds P. Yolum, T. Güngör, F. Gürgeç, and C. Özturan (Berlin: Springer), 284–293. doi: 10.1007/11569596\_31
- Przulj, N., Corneil, D. G., and Jurisica, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics* 20, 3508–3515. doi: 10.1093/bioinformatics/bth436
- Raghavan, U. N., Albert, R., and Kumara, S. (2007). Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* 76:036106.
- Ralaivola, L., Swamidass, S. J., Saigo, H., and Baldi, P. (2005). Graph kernels for chemical informatics. *Neural Netw.* 18, 1093–1110. doi: 10.1016/j.neunet.2005.07.009
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021
- Rosvall, M., and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS One* 6:e18209. doi: 10.1371/journal.pone.0018209
- Ryu, J. Y., Kim, H. U., and Lee, S. Y. (2018). Deep learning improves prediction of drug–drug and drug–food interactions. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4304–E4311.
- Saito, R., Suzuki, H., and Hayashizaki, Y. (2002). Interaction generality, a measurement to assess the reliability of a protein–protein interaction. *Nucleic Acids Res.* 30, 1163–1168. doi: 10.1093/nar/30.5.1163
- Saraph, V., and Milenkovic, T. (2014). MAGNA: maximizing accuracy in global network alignment. *Bioinformatics* 30, 2931–2940. doi: 10.1093/bioinformatics/btu409
- Schwikowski, B., Uetz, P., and Fields, S. (2000). A network of protein–protein interactions in yeast. *Nat. Biotechnol.* 18, 1257–1261. doi: 10.1038/82360
- Serrano, M. Á., Boguñá, M., and Sagués, F. (2012). Uncovering the hidden geometry behind metabolic networks. *Mol. Biosyst.* 8, 843–850.
- Sharan, R., and Ideker, T. (2006). Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* 24, 427–433. doi: 10.1038/nbt1196
- Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. *Mol. Syst. Biol.* 3:88.
- Singh, R., Xu, J., and Berger, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12763–12768. doi: 10.1073/pnas.0806627105
- Sridhar, D., Fakhraei, S., and Getoor, L. (2016). A probabilistic approach for collective similarity-based drug–drug interaction prediction. *Bioinformatics* 32, 3175–3182. doi: 10.1093/bioinformatics/btw342
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2015). “LINE: large-scale information network embedding,” in *Proceedings of the 24th International Conference on World Wide Web WWW ’15* (Geneva: International World Wide Web Conferences Steering Committee), 1067–1077.
- Tenenbaum, J. B. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi: 10.1126/science.290.5500.2319
- Tutte, W. T. (1963). How to draw a graph. *Proc. Lond. Math. Soc.* 3, 743–767.
- Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnnerberg, P., Lou, D., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* 18, 145–153. doi: 10.1038/nn.3881
- Velicković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. arXiv:1710.10903v3 [Preprint].
- Vijayan, V., Saraph, V., and Milenkovic, T. (2015). MAGNA++: maximizing accuracy in global network alignment via both node and edge conservation. *Bioinformatics* 31, 2409–2411. doi: 10.1093/bioinformatics/btv161
- Wan, F., Hong, L., Xiao, A., Jiang, T., and Zeng, J. (2018). NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *bioRxiv* [Preprint].
- Wang, B., Huang, L., Zhu, Y., Kundaje, A., Batzoglou, S., and Goldenberg, A. (2017a). Vicus: exploiting local structures to improve network-based analysis of biological data. *PLoS Comput. Biol.* 13:e1005621. doi: 10.1371/journal.pcbi.1005621
- Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., and Batzoglou, S. (2017b). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* 14, 414–416. doi: 10.1038/nmeth.4207
- Wang, B., Jiang, J., Wang, W., Zhou, Z., and Tu, Z. (2012). “Unsupervised metric fusion by cross diffusion,” in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE Computer Society), 2997–3004.
- Wang, B., Pourshafeie, A., Zitnik, M., Zhu, J., Bustamante, C. D., Batzoglou, S., et al. (2018). Network enhancement as a general method to denoise weighted biological networks. *Nat. Commun.* 9:3108.
- Wang, B., and Tu, Z. (2013). “Sparse subspace denoising for image manifolds,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 468–475.
- Wang, Y., and Zeng, J. (2013). Predicting drug–target interactions using restricted Boltzmann machines. *Bioinformatics* 29, i126–i134. doi: 10.1093/bioinformatics/btt234
- Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., et al. (2017). Deep-learning-based drug–target interaction prediction. *J. Proteome Res.* 16, 1401–1409. doi: 10.1021/acs.jproteome.6b00618
- Xu, Z., Wang, S., Zhu, F., and Huang, J. (2017). “Seq2Seq fingerprint: an unsupervised deep molecular embedding for drug discovery,” in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics ACM-BCB ’17* (New York, NY: ACM), 285–294.
- Yamanishi, Y., Kotera, M., Kanehisa, M., and Goto, S. (2010). Drug–target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 26, i246–i254. doi: 10.1093/bioinformatics/btq176
- Yang, Z., Algesheimer, R., and Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Sci. Rep.* 6:30750.
- Yildirim, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L., and Vidal, M. (2007). Drug–target network. *Nat. Biotechnol.* 25:1119.
- You, Z.-H., Lei, Y.-K., Gui, J., Huang, D.-S., and Zhou, X. (2010). Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data. *Bioinformatics* 26, 2744–2751. doi: 10.1093/bioinformatics/btq510
- Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34, 457–466.
- Zitnik, M., and Leskovec, J. (2017). Predicting multicellular function through multi-layer tissue networks. *Bioinformatics* 33, i190–i198. doi: 10.1093/bioinformatics/btx252
- Zitnik, M., and Zupan, B. (2016). “Collective pairwise classification for multi-way analysis of disease and drug data,” in *Proceedings of the Pacific Symposium on Biocomputing* (Bethesda MD: NIH Public Access), 81.
- Zong, N., Kim, H., Ngo, V., and Harismendy, O. (2017). Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations. *Bioinformatics* 33, 2337–2344. doi: 10.1093/bioinformatics/btx160

**Conflict of Interest Statement:** The 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 © 2019 Nelson, Zitnik, Wang, Leskovec, Goldenberg and Sharan. 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.