Bioinformatics, 33, 2017, i190–i198
doi: 10.1093/bioinformatics/btx252
ISMB/ECCB 2017

OXFORD

Predicting multicellular function through multi-layer tissue networks

Marinka Zitnik and Jure Leskovec*

Department of Computer Science, Stanford University, Stanford, CA 94305, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Understanding functions of proteins in specific human tissues is essential for insights into disease diagnostics and therapeutics, yet prediction of tissue-specific cellular function remains a critical challenge for biomedicine.

Results: Here, we present *OhmNet*, a hierarchy-aware unsupervised node feature learning approach for multi-layer networks. We build a multi-layer network, where each layer represents molecular interactions in a different human tissue. *OhmNet* then automatically learns a mapping of proteins, represented as nodes, to a neural embedding-based low-dimensional space of features. *OhmNet* encourages sharing of similar features among proteins with similar network neighborhoods and among proteins activated in similar tissues. The algorithm generalizes prior work, which generally ignores relationships between tissues, by modeling tissue organization with a rich multiscale tissue hierarchy. We use *OhmNet* to study multicellular function in a multi-layer protein interaction network of 107 human tissues. In 48 tissues with known tissue-specific cellular functions, *OhmNet* provides more accurate predictions of cellular function than alternative approaches, and also generates more accurate hypotheses about tissue-specific protein actions. We show that taking into account the tissue hierarchy leads to improved predictive power. Remarkably, we also demonstrate that it is possible to leverage the tissue hierarchy in order to effectively transfer cellular functions to a functionally uncharacterized tissue. Overall, *OhmNet* moves from flat networks to multiscale models able to predict a range of phenotypes spanning cellular subsystems.

Availability and implementation: Source code and datasets are available at http://snap.stanford.edu/ohmnet.

Contact: jure@cs.stanford.edu

1 Introduction

A unified view of human diseases and cellular functions across a broad range of human tissues is essential, not only for understanding basic biology but also for interpreting genetic variation and developing therapeutic strategies (Greene *et al.*, 2015; GTEx *et al.*, 2015; Okabe and Medzhitov, 2014; Yeger-Lotem and Sharan, 2015). In particular, the precise functions of proteins frequently depend on the tissue, and different proteins can have different cellular functions in different tissues (Fagerberg *et al.*, 2014; Guan *et al.*, 2012; Hu *et al.*, 2016; Lois *et al.*, 2002; Magger *et al.*, 2012; Rakyan *et al.*, 2008; Yeger-Lotem and Sharan, 2015).

While our view of the human protein–protein interaction (PPI) network as a key source for studying protein function is constantly expanding, much less is known about networks that form in biologically important environments such as within distinct tissues or in specific diseases (Yeger-Lotem and Sharan, 2015). Although incredibly influential, current computational methods for extracting functional information from protein interaction networks lack tissue specificity as they assume that cellular function is constant across organs and tissues (Barutcuoglu *et al.*, 2006; Kramer *et al.*, 2014; Mostafavi *et al.*, 2008; Radivojac *et al.*, 2013; Stojanova *et al.*, 2013; Žitnik and Zupan, 2015). In other words, cellular functions in heart are assumed to be the same as functions in skin. The methods are, hence, less successful in constructing *accurate maps of both where and how proteins act*. In particular, existing network-based methods are probably not the ultimate representation of human tissues for three reasons. (1) First, current methods for cellular function prediction on networks (Mostafavi and Morris, 2009; Radivojac *et al.*, 2013; Vidulin *et al.*, 2016; Žitnik and Zupan, 2015) do not model networks with regards to patterns that span tissues, organs, and cellular systems. This means that a complex tissue involving a multiscale hierarchy of cellular subsystems is not readily captured by current models (Carvunis and Ideker, 2014; Dutkowski *et al.*, 2012). (2) Second, many genome-scale functional maps

(Costanzo *et al.*, 2016; Kitsak *et al.*, 2016; Kotlyar *et al.*, 2015; Lopes *et al.*, 2011; Rolland *et al.*, 2014; Wang *et al.*, 2016b) are descriptive maps of physical or functional protein connectivity that do not, by themselves, predict cellular function. (3) Third, only few computational approaches (Antanaviciute *et al.*, 2015; Ganegoda *et al.*, 2014; Guan *et al.*, 2012; Magger *et al.*, 2012) used tissue-specific information to identify novel genes and relationships between genes. However, their focus was to leverage tissue specificity to improve prediction of global cellular functions and global gene-disease associations. As such, these approaches account for tissue specificity, but they do not resolve the challenge of predicting gene–function relationships that might be specific to a particular tissue. To be able to predict a range of tissue-specific functions one needs to design scalable multiscale models that can relate tissues to each other, extract rich feature representations for proteins in each tissue-specific network, and then use the extracted features for tissue-specific cellular function prediction.

1.1 Present work

We present *OhmNet*, an algorithm for hierarchy-aware unsupervised feature learning in multi-layer networks. Our focus is on learning features of proteins in different tissues. We represent each tissue as a network, where nodes represent proteins. Tissue networks act as layers in a multi-layer network, where we use a hierarchy to model dependencies between the layers (i.e. tissues) (Fig. 1). We then develop a computational framework that learns features of each node (i.e. protein) by taking into consideration connections between the nodes within each layer, together with inter-layer relationships between proteins active on different layers. More precisely, our approach embeds each protein in each tissue in a d-dimensional feature space such that proteins with similar network neighborhoods in similar tissues are embedded closely together.

In *OhmNet*, we define an objective function that is independent of the downstream prediction task, meaning that the feature representations are learned in a purely unsupervised way. This results in task-independent features, that, as we show, outperform task-specific approaches in predictive accuracy. Furthermore, since our features are not designed for a specific downstream prediction task, they generalize across a wide variety of tasks and tissues. For example, we use the learned features to study protein functions across different cellular systems (e.g. cell types, tissues, organs and organ systems).

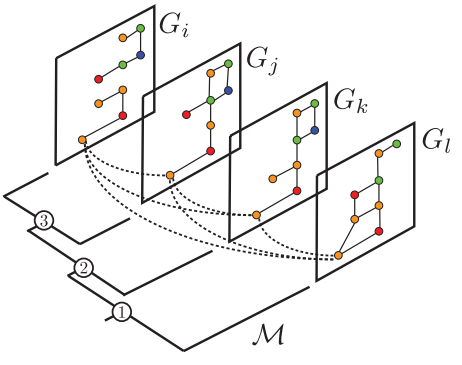

Fig. 1. A multi-layer network with four layers, where each layer represents a tissue-specific PPI network. The hierarchy M encodes biological similarities between the tissues at multiple scales. *OhmNet* embeds each node in a d-dimensional feature space, which we use for tissue-specific protein function prediction. For example, layers G_i, G_j, G_k and G_l might represent brain tissue-specific interaction networks in cerebrum, hypothalamus, tegmentum and medulla

OhmNet builds on recent success of unsupervised representation learning methods based on neural architectures (Grover and Leskovec, 2016; Mikolov *et al.*, 2013). In particular, we develop a new form of structured regularization, which makes *OhmNet* especially suitable for multi-layer interdependent networks. Our key contribution lies in modeling the tissue taxonomy constraints by encoding relationships between the tissues in a tissue hierarchy and then using the structured regularization with the tissue hierarchy (Fig. 1). This way *OhmNet* effectively learns multiscale feature representations for proteins that are consistent with the tissue hierarchy.

Our experiments focus on three tasks defined on a multi-layer tissue network: (1) a multi-label node classification task, where every protein is assigned zero, one or more tissue-specific cellular functions; (ii) a transfer learning task, where we predict cellular functions for a protein in one tissue based on classifiers trained on features from other tissues; and (iii) a network-embedding visualization task, where we create meaningful tissue-specific visualizations that lay out proteins on a 2D space. Since the multiscale protein feature vectors returned by *OhmNet* are task-independent, we use *OhmNet* one time only to learn the features for proteins in every tissue and at every scale of the tissue hierarchy. We can then solve the cellular function prediction task for any tissue using the appropriate tissue-specific protein features.

We contrast *OhmNet*'s performance with that of state-of-the-art approaches for feature learning (Cannistraci *et al.*, 2013; Grover and Leskovec, 2016; Nickel *et al.*, 2011; Tang *et al.*, 2015;), approaches for tissue-independent cellular function prediction (Mostafavi *et al.*, 2008; Zuberi *et al.*, 2013), and approaches for prioritization of disease-causing genes in tissue-specific protein interaction networks (Guan *et al.*, 2012; Magger *et al.*, 2012), which we adapted for the cellular function prediction task. We experiment with a multi-layer network having 107 genome-wide tissue-specific protein interaction layers, and we consider a tissue hierarchy describing 219 cellular systems in the human body. Experiments demonstrate that tissue-specific protein interaction layers provide the necessary protein and tissue context for predicting cellular function. *OhmNet* outperforms alternative approaches by up to 14.9% on multi-label classification and up to 20.3% on transfer learning. Another notable finding is that *OhmNet* outperforms alternative approaches, which are based on non-hierarchical versions of the same dataset, alluding to the benefits of modeling hierarchical tissue organization. We observe that neglecting the existence of tissues or aggregating tissue-specific interaction networks into a single network discards important biological information and affects performance on multi-label classification and transfer learning tasks. Finally, we exemplify the utility of *OhmNet* for exploring the multiscale structure of tissues. In a case study on nine brain tissue networks, we show that *OhmNet*'s features inherently encode a multiscale brain organization.

The rest of the article is organized as follows. In Section 2, we briefly survey related work in feature learning for networks. We present the technical details of *OhmNet* in Section 3. In Section 4, we describe the multi-layer tissue network and the tissue hierarchy. We empirically evaluate *OhmNet* in Section 5 and conclude with directions for future work in Section 6.

2 Related work

We have seen in Section 1 that despite the abundance of methods for cellular function prediction, only a few, if any, take into account

biologically important contexts given by human tissues. We now turn our focus to the problem of feature learning in networks.

Most approaches for automatic (i.e. non-hand-engineered) feature learning in networks can be categorized into matrix factorization and neural network embedding-based approaches. In matrix factorization, a network is expressed as a data matrix where the entries represent relationships. The data matrix is projected to a low-dimensional space using linear techniques based on SVD (Tang *et al.*, 2012), or nonlinear techniques based on multi-dimensional scaling (Belkin and Niyogi, 2001; Hou *et al.*, 2014; Tenenbaum *et al.*, 2000). These methods have two important drawbacks. First, they do not account for important structures typically exhibited in networks such as high sparsity and skewed degree distribution. Second, matrix factorization methods perform a global factorization of the data matrix while a local-centric method might often yield more useful feature representations (Kramer *et al.*, 2014).

Limitations of matrix factorization are overcome by neural network embeddings. Recent studies focused on embedding nodes into low-dimensional vector spaces by first using random walks to construct the network neighborhood of every node in the graph, and then optimizing an objective function with network neighborhoods as input (Grover and Leskovec, 2016; Perozzi *et al.*, 2014; Tang *et al.*, 2015). The objective function is carefully designed to preserve both the local and global network structures. A state-of-the-art neural network embedding algorithm is the Node2vec algorithm (Grover and Leskovec, 2016), which learns feature representations as follows: it scans over the nodes in a network, and for every node it aims to embed it such that the node's features can predict nearby nodes, that is, node's feature predict which other nodes are part of its network neighborhood. Node2vec can explore different network neighborhoods to embed nodes based on the principles of homophily (i.e. network communities) as well as structural equivalence (i.e. structural roles of nodes).

However, a challenging problem for neural network embedding-based methods is to learn features in multi-layer networks. Existing methods can learn features in multi-layer networks either by treating each layer independently of other layers, or by aggregating the layers into a single (weighted) network. However, neglecting the existence of multiple layers or aggregating the layers into a single network, alters topological properties of the system as well as the importance of individual nodes with respect to the entire network structure (De Domenico *et al.*, 2016). This is a major shortcoming of prior work that can lead to a wrong identification of the most versatile nodes (De Domenico *et al.*, 2015) and overestimation of the importance of more marginal nodes (De Domenico *et al.*, 2014). As we shall show, this shortcoming also affects predictive accuracy of the learned features. Our approach *OhmNet* overcomes this limitation since it learns features in a multi-layer network in the context of the entire system structure, bridging together different layers and generalizing methods developed for learning features in single-layer networks.

In biological domains, measures based on similarities of nodes' extended network neighborhoods are well established for predicting protein functions. Several approaches use graphlets (Pržulj, 2007) to systematically describe network structure around each node. This is done by counting how many instances of small subgraph patterns occur in the network neighborhood of a given node. Graphlet-based methods, such as graphlet degree vectors (Hayes *et al.*, 2013), can thus be seen as an alternative approach for extracting feature representations for nodes. In contrast to neural embedding-based methods, such as *OhmNet*, which learn continuous feature representations, graphlet-based methods return discrete counts of motif occurrences. Further, graphlet-based methods in their current form cannot be applied to multi-layer networks without collapsing the network layers into one network.

Finally, there exists recent work for task-dependent feature learning based on graph-specific deep network architectures (Li *et al.*, 2015; Xiaoyi *et al.*, 2014; Wang *et al.*, 2016a; Zhai and Zhang, 2015). Our approach differs from those approaches in two important ways. First, those architectures are task-dependent, meaning they directly optimize the objective function for a downstream prediction task, such as cellular function prediction in a particular tissue, using several layers of nonlinear transformations. Second, those architectures do not model rich graph structures, such as multi-layer networks with hierarchies.

3 Feature learning in multi-layer networks

We formulate feature learning in multi-layer networks as a maximum likelihood optimization problem. Let V be a given set of N nodes (e.g. proteins) $\{u_1, u_2, \ldots, u_N\}$, and let there be K types of edges (e.g. protein interactions in different tissues) between pairs of nodes $u_1, u_2, \ldots, u_N$. A multi-layer network is a general system in which each biological context is represented by a distinct *layer* i (where $i = 1, 2, \ldots, K$) of a system (Fig. 1). We use the term *single-layer network (layer)* for the network $G_i = (V_i, E_i)$ that indicates the edges E_i between nodes $V_i \subseteq V$ within the same layer i. Our analysis is general and applies to any (un)directed, (un)weighted multi-layer network.

We take into account the possibility that a node u_k from layer i can be related to any other node u_h in any other layer j. We encode information about the dependencies between layers in a hierarchical manner that we use in the learning process. Let the hierarchy be a directed tree $\mathcal{M}$ defined over a set M of elements by the parent-child relationships given by $\pi : M \rightarrow M$, where $\pi(i)$ is the parent of element i in the hierarchy (Fig. 1). Let $T \subset M$ be the set of all leaves in the hierarchy. Let T_i be the set of all leaves in the sub-hierarchy rooted at i. We assume that each layer G_i is attached to one leaf in the hierarchy. As a result, the hierarchy $\mathcal{M}$ has exactly K leaves. For convenience, let C_i denote the set of all children of element i in the hierarchy.

The problem of feature learning in a multi-layer network is to learn functions $f_1, f_2, \ldots, f_K$, such that each function $f_i : V_i \rightarrow \mathbb{R}^d$ maps nodes in V_i to feature representations in $\mathbb{R}^d$. Here, d is a parameter specifying the number of dimensions in the feature representation of one node. Equivalently, f_i is a matrix of $|V_i| \times d$ parameters.

We proceed by describing *OhmNet*, our approach for feature learning in multi-layer networks. *OhmNet* has two components:

- *single-layer network objectives*, in which nodes with similar network neighborhoods in each layer are embedded close together,
- *hierarchical dependency objectives*, in which nodes in nearby layers in the hierarchy are encouraged to share similar features.

We start by describing the model that considers the layers independently of each other. We then extend the model to encourage nodes which are nearby in the hierarchy to have similar features.

3.1 Single-layer network objectives

We start by formalizing the intuition that nodes with similar network neighborhoods in each layer should share similar features. For that, we specify one objective for each layer in a given multi-layer network. We shall later discuss how *OhmNet* incorporates the dependencies between different layers.

Our goal is to take layer G_i and learn f_i which embeds nodes from similar network regions, or nodes with similar structural roles, closely together. In *OhmNet*, we aim to achieve this goal by specifying the following objective function for each layer G_i. Given a node $u \in V_i$, the objective function ω_i seeks to predict, which nodes are members of u's network neighborhood $N_i(u)$ based on the learned node features f_i:

$$\omega_i(u) = \log Pr(N_i(u)|f_i(u)), \quad (1)$$

where the conditional likelihood of every node-neighborhood node pair is modeled independently as $Pr(N_i(u)|f_i(u)) = \prod_{v \in N_i(u)} Pr(v|f_i(u))$. The conditional likelihood is a softmax unit parameterized by a dot product of nodes' features, which corresponds to a single-layer feed-forward neural network: $Pr(v|f_i(u)) = \exp(f_i(v)f_i(u)) / \sum_{z \in V_i} \exp(f_i(z)f_i(u))$. Given a node u, maximization of $\omega_i(u)$ tries to maximize classification of nodes in u's network neighborhood based on u's learned representation.

The objective Ω_i is defined for each layer i:

$$\Omega_i = \sum_{u \in V_i} \omega_i(u), \quad \text{for} \quad i = 1, 2, \ldots, K. \quad (2)$$

The objective is inspired by the intuition that nodes with similar network neighborhoods tend to have similar meanings, or roles, in a network. It formalizes this intuition by encouraging nodes in similar network neighborhoods to share similar features.

We found that a flexible notion of a network neighborhood N_i is crucial to achieve excellent predictive accuracy on a downstream cellular function prediction task (Grover and Leskovec, 2016). For that reason, we use a randomized procedure to sample many different neighborhoods of a given node u. Technically, the network neighborhood $N_i(u)$ is a set of nodes that appear in an appropriately biased random walk defined on layer G_i and started at node u (Grover and Leskovec, 2016). The neighborhoods $N_i(u)$ are not restricted to just immediate neighbors but can have vastly different structures depending on the sampling strategy.

Next, we expand *OhmNet*'s single-layer network objectives to leverage information provided by the tissue taxonomy and this way inform embeddings across different layers.

3.2 Hierarchical dependency objectives

So far, we specified K layer-by-layer objectives each of which estimates node features in its layer independently of node features in other layers. This means that nodes in different layers representing the same entity have features that are learned independently of each other.

To harness the dependencies between the layers, we expand *OhmNet* with terms that encourage sharing of protein features between the layers. Our approach is based on the assumption that nearby layers in the hierarchy are semantically close to each other and hence proteins/nodes in them should share similar features. For example, in the tissue multi-layer network, we model the fact that the "medulla" layer is part of the "brainstem" layer, which is, in turn, part of the "brain" layer. We use the dependencies among the layers to define a joint objective for regularization of the learned features of proteins.

We propose to use the hierarchy in the learning process by incorporating a recursive structure into the regularization term for every element in the hierarchy $\mathcal{M}$. Specifically, we propose the following form of regularization for node u that resides in element i of the hierarchy $\mathcal{M}$:

$$c_i(u) = \frac{1}{2} ||f_i(u) - f_{\pi(i)}(u)||_2^2. \quad (3)$$

This recursive form of regularization enforces the features of node u in the hierarchy i to be similar to the features of node u in i's parent $\pi(i)$ under the Euclidean norm. When regularizing features of all nodes across all elements of the hierarchy, we obtain:

$$C_i = \sum_{u \in L_i} c_i(u), \quad \text{where} \quad L_i = \cup_{j \in T_i} V_j \quad (4)$$

In words, we specify the features for both leaf as well as internal, i.e. non-leaf, elements in the hierarchy, and we regularize the features of sibling (i.e. sharing the same parent) hierarchy elements toward features in the common parent element in the hierarchy.

3.2.1 Node features at multiple scales

It is important to notice that *OhmNet*'s structured regularization allows us to learn feature representations at multiple scales. For example, consider a multi-layer network in Figure 2, consisting of four layers that are interrelated by a two-level hierarchy. *OhmNet* learns the mappings f_i, f_j, f_k and f_l that map nodes in each layer into a d-dimensional feature space. In addition, *OhmNet* also learns the mapping f_2 representing features for nodes appearing in the hierarchy leaves T_2, i.e. $V_i \cup V_j$, at an intermediate scale, and the mapping f_1 representing features for nodes appearing in the hierarchy leaves T_1, i.e. $V_i \cup V_j \cup V_k \cup V_l$, at the highest scale.

The modeling of relationships between layers in a multi-layer network has several implications:

- First, the model encourages nodes which are in nearby layers in the hierarchy to share similar features.
- Second, the model shares statistical strength across the hierarchy as nodes in different layers representing the same protein share features through ancestors in the hierarchy.
- Third, this model is more efficient than the fully pairwise model. In the fully pairwise model, the dependencies between layers are modeled by pairwise comparisons of nodes across all pairs of layers, which takes $O(K^2N)$ time, where K is the number of layers and N is the number of nodes. In contrast, *OhmNet* models inter-layer dependencies according to the parent-child relationships specified by the hierarchy, which takes only $O(|M|N)$ time. Since *OhmNet*'s hierarchy is a tree, it holds that $|M| \ll K^2$, meaning that the proposed model scales more easily to large multi-layer networks than the fully pairwise model.
- Finally, the hierarchy is a natural way to represent and model biological systems spanning many different biological scales (Carvunis and Ideker, 2014; Greene *et al.*, 2015; Yu *et al.*, 2016).

3.3 Full *OhmNet* model

Given a multi-layer network consisting of layers $G_1, G_2, \ldots, G_K$, and a hierarchy encoding relationships between the layers, the

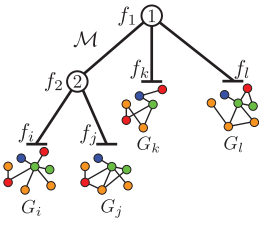

Fig. 2. A multi-layer network with four layers. Relationships between the layers are encoded by a two-level hierarchy $\mathcal{M}$. Leaves of the hierarchy correspond to the network layers. Given networks G_i and hierarchy $\mathcal{M}$, *OhmNet* learns node embeddings captured by functions f_i

OhmNet's goal is to learn the functions $f_1, f_2, \ldots, f_K$ that map from nodes in each layer to feature representations. *OhmNet* achieves this goal by fitting its feature learning model to a given multi-layer network and a given hierarchy, i.e. by finding the mapping functions $f_1, f_2, \ldots, f_K$ that maximize the data likelihood.

Given the data, *OhmNet* aims to solve the following maximum likelihood optimization problem:

$$\max_{f_1, f_2, \ldots, f_{|M|}} \sum_{i \in T} \Omega_i - \lambda \sum_{j \in M} C_j, \tag{5}$$

which includes the single-layer network objectives for all network layers, and the hierarchical dependency objectives for all hierarchy elements. In Equation (5), parameter λ is a user-specified parameter representing the regularization strength. While the optimization problem in Equation (5) is non-convex due to the non-convexity of the single-layer objective (Grover and Leskovec, 2016), stochastic gradient with negative sampling can be used to efficiently solve the problem.

One appealing property of *OhmNet* is that by solving the problem in Equation (5) we obtain estimates for functions $f_1, f_2, \ldots, f_K$ located in the leaf elements of the hierarchy (i.e. layers of a given multi-layer network), as well as estimates for functions $f_{K+1}, f_{K+2}, \ldots, f_{|M|}$ located in the internal elements of the hierarchy.

3.4 The *OhmNet* algorithm

The pseudocode for *OhmNet* is given in Algorithm 1.

In the first phase, *OhmNet* applies the Node2vec's algorithm (Grover and Leskovec, 2016) to construct network neighborhoods for each node in every layer. Given a layer G_i and a node $u \in V_i$, the algorithm simulates a user-defined number of fixed length random walks started at node u (step 4 in Algorithm 1).

In the second phase, *OhmNet* uses an iterative approach in which features associated with each object in the hierarchy are iteratively updated by fixing the rest of the features. The iterative approach has the advantage that it can easily incorporate the closed-form updates developed for the internal objects of the hierarchy (step 11 in Algorithm 1), thereby accelerating the convergence of *OhmNet* algorithm. For each leaf object i, *OhmNet* isolates the terms in the optimization problem in Equation (5) that depend on the model parameters defining function f_i. *OhmNet* then optimizes Equation (5) by performing one epoch of stochastic gradient descent (SGD1) over f_i's model parameters (step 15 in Algorithm 1).

The two phases of *OhmNet* are executed sequentially. The *OhmNet* algorithm scales to large multi-layer networks because each phase is parallelizable and executed asynchronously. The choice to model the dependencies between network layers using the hierarchical model requires $O(|M|N)$ time instead of the fully pairwise model, which requires $O(K^2N)$ time.

Algorithm 1. The *OhmNet* algorithm

Input: Multi-layer network, $(G_1, G_2, \ldots, G_K)$ with $G_i = (V_i, E_i)$, Hierarchy, $\mathcal{M}$, Feature representation size, d, Network neighborhood strategy, S, Regularization strength, λ

1: **for** $i \in T$ **do**
2: **for** $u \in V_i$ **do**
3: $N_i(u) = \text{Node2vecWalk}(G_i, u, S)$ (Grover and Leskovec, 2016)
4: **end for**
5: **end for**
6: **while** $f_1, f_2, \ldots, f_{|M|}$ not converged **do**
7: **for** $i \in M$ **do**
8: **if** $i \in T$ **then**
9: **for** $u \in V_i$ **do**
10: $f_i(u) = \text{SGD1}(N_i(u), d, \lambda)$ by Equation (5)
11: **end for**
12: **else**
13: **for** $u \in \cup_{j \in T_i} V_j$ **do**
14: $f_i(u) = \frac{1}{|C_i|+1}\left(f_{\pi(i)}(u) + \sum_{c \in C_i} f_c(u)\right)$
15: **end for**
16: **end if**
17: **end for**
18: **end while**
19: **return** $f_1, f_2, \ldots, f_{|M|}$

4 Tissue-specific interactome data

To construct the human PPI network, tissue-specific network layers, tissue hierarchy and tissue-specific gene–function relationships, we downloaded and used standard protein, tissue and function information from various reputable data sources.

4.1 Tissue hierarchy

We retrieved the mapping of tissues in the Human Protein Reference Database (HPRD) (Prasad *et al.*, 2009) to tissues in the BRENDA Tissue Ontology (Chang *et al.*, 2014) from Greene *et al.* (2015). The data is provided as a supplementary dataset in Greene *et al.* (2015). The hierarchical relationships between tissues were then determined by the directed acyclic graph structure of the BRENDA Tissue Ontology. Examples of tissues included: muscle, adrenal cortex, bone marrow and spleen (Fig. 3).

4.2 Tissue-specific interaction networks

We took the gene-to-tissue mapping compiled by Greene *et al.* (2015). Greene *et al.* mapped genes to HPRD tissues based on low-throughput tissue-specific gene expression data. The gene-to-tissue mapping was then combined with the human PPI network. The resulting multi-layer tissue network had 107 layers, each layer corresponded to a PPI network specific to a particular tissue. Details are provided next.

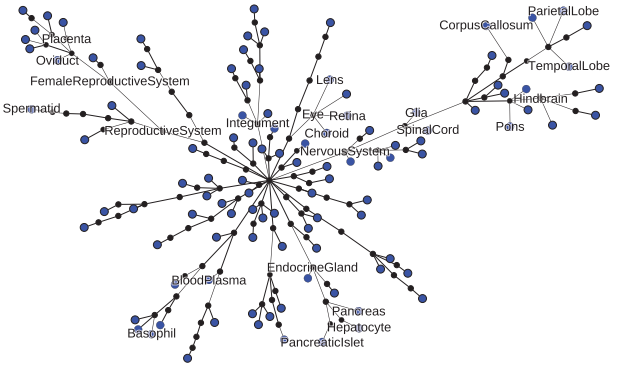

Fig. 3. The tissue hierarchy considered in this study. The tissue hierarchy is a directed tree defined over $|M| = 219$ tissue terms from the BRENDA Tissue Ontology. Edges in the tree point from children to parents based on ontological relationships: "develops_from", "is_a", "part_of" and "related_to". The $K = 107$ tissues with tissue-specific protein interaction networks are the blue leaves in the tree

The human PPI network was collected from Orchard *et al.* (2013), Rolland *et al.* (2014), Chatr-Aryamontri *et al.* (2015), Prasad *et al.* (2009), Ruepp *et al.* (2010) and Menche *et al.* (2015). Considered were physical PPIs with supported by experimental evidence. It should be noted that interactions based on gene expression and evolutionary data were not considered. The global (unweighted) human PPI network has 21 557 proteins interconnected by 342 353 interactions. The reader is referred to Menche *et al.* (2015) for a detailed description of the data.

For each of 107 tissues, a tissue-specific human PPI network was constructed based on the global PPI network. For a given tissue, every edge in the global PPI network was labeled as specifically co-expressed in that tissue using the criterion developed by Greene *et al.* (2015). Greene *et al.* labeled each edge as specifically co-expressed if either both proteins are specific to that tissue or one protein is tissue-specific and the other is ubiquitous. Lists of specifically co-expressed proteins were retrieved from Greene *et al.* (2015). Finally, the PPI network specific to a particular tissue is a subnetwork of the global PPI network, induced by the set of specifically co-expressed edges in that tissue.

4.3 Tissue-specific cellular functions and gene annotations

Associations between tissues and cellular functions were retrieved from Greene *et al.* (2015). Greene *et al.* manually curated biological processes in the Gene Ontology (GO) (Ashburner *et al.*, 2000) and mapped them to tissues in the BRENDA Tissue Ontology (Chang *et al.*, 2014) based on whether a given biological process is specifically active in a given tissue. The data is provided as a supplementar dataset in Greene *et al.* (2015). An example of a cellular function-tissue pair is "low-density lipoprotein particle remodeling" in the blood plasma tissue.

All gene annotations were propagated along the ontology hierarchy. Considered are functions with at least 15 annotated proteins (Guan *et al.*, 2012). In total, there are 584 tissue-specific cellular functions covering 48 distinct tissues. Each tissue-specific function is assigned to one or more leaves in the tissue hierarchy (Section 4.1).

5 Results

The *OhmNet*'s objective in Equation (5) is independent of any downstream task. This flexibility offered by *OhmNet* makes the learned feature representations suitable for a variety of analytics tasks discussed below.

5.1 Prediction of tissue-specific cellular functions

5.1.1 Experimental setup

We view the problem of predicting cellular functions as solving a multi-label node classification task. Here, every node (i.e. protein) is assigned one or more labels (i.e. cellular functions from the GO) from a finite set of labels (i.e. all cellular functions in the GO, see Section 4.3).

We apply *OhmNet*, which for every node in every layer learns a separate feature vector in an unsupervised way. Thus, for every layer and every function we then train a separate one-versus-all linear classifier using the modified Huber loss with elastic net regularization. Using cross validation, we observe 90% of proteins and all their cellular functions across the layers during the training phase. The task is then to predict the tissue-specific functions for the remaining 10% of proteins.

We evaluate the performance of *OhmNet* against the following feature-learning approaches:

- RESCAL tensor decomposition (Nickel *et al.*, 2011): This is a tensor factorization approach that takes the multi-layer network structure into account. Given X_i, a normalized Laplacian matrix of layer G_i, matrix X_i is factorized as: $X_i = AR_iA^T$, for $i = 1, 2, \ldots, K$. Here, matrix A contains d-dimensional feature representation for nodes.
- Minimum curvilinear embedding (Cannistraci *et al.*, 2013): This is a non-linear unsupervised framework that embeds nodes in a low-dimensional space. The approach was originally developed for protein interaction prediction, aiming to embed protein pairs representing good candidate interactions closer to each other. It utilizes a network denoising method as well as structural information provided by the PPI network topology.
- LINE (Tang *et al.*, 2015): This approach first learns $d/2$ dimensions based on immediate network neighbors of nodes, and then the next $d/2$ dimensions based on network neighbors at a 2-hop distance.
- Node2vec (Grover and Leskovec, 2016): This approach learns d-dimensional features for nodes based on a biased random walk procedure that flexibly explores network neighborhoods of nodes.

In addition, we evaluate the performance of *OhmNet* against the following tissue-specific/agnostic function prediction approaches:

- GeneMania (Zuberi *et al.*, 2013): This is a supervised approach that takes a multi-layer network as input and directly predicts cellular functions in two separate phases. In the first phase, it aggregates the layers into one weighted network by weighting the layers according to their utility for predicting a given function. It then uses a label propagation algorithm on the weighted network to predict the function.
- Tissue-specific network propagation (Magger *et al.*, 2012): This approach assigns a prior score to proteins associated with known functions that are phenotypically similar to the query function. This score is then propagated through a network in an iterative process. The approach was developed for tissue-specific disease gene prioritization.
- Network-based tissue-specific support vector machine (SVM) (Guan *et al.*, 2012): This approach adopts the network-based candidate gene prediction scheme. Essentially, the connection weights in a network to all positive examples (i.e. genes already known to be related to a phenotype) are utilized as features for linear SVM classification. The approach was developed for tissue-specific phenotype and disease gene prioritization.

The parameter settings for every approach are determined using internal cross-validation procedure with a grid search over candidate parameter values. Specifically, $d = 128$ is used in all experiments.

Last, we aim to evaluate the benefit of our proposed multi-layer representation of the tissue networks. To this end we also consider two additional network representations:

- *Independent layers*: This approach learns features for nodes in each layer by running LINE or Node2vec algorithm on one layer at a time and independently of other layers in the network.
- *Collapsed layers*: This approach first aggregates the layers into a single network by connecting nodes representing the same entity in different layers to each other. It then learns feature for nodes in the aggregated network.

5.1.2 Experimental results

Table 1 and Figure 4 give the area under the curve (AUC) scores of tissue-specific protein function prediction.

Table 1. AUROC and area under precision-recall curve (AUPRC) scores for tissue-specific cellular function prediction

Approach	AUROC	AUPRC
Tensor decomposition	0.674 (±0.124)	0.235 (±0.052)
Minimum curvilinear embedding	0.674 (±0.064)	0.248 (±0.071)
Independent LINE	0.642 (±0.053)	0.261 (±0.068)
Collapsed LINE	0.663 (±0.047)	0.271 (±0.053)
Independent Node2vec	0.649 (±0.063)	0.283 (±0.052)
Collapsed Node2vec	0.697 (±0.085)	0.298 (±0.061)
GeneMania	0.683 (±0.077)	0.274 (±0.094)
Network-based tissue-specific SVM	0.701 (±0.091)	0.281 (±0.059)
Tissue-specific network propagation	0.675 (±0.051)	0.265 (±0.083)
OhmNet (Section 3)	0.756 (±0.067)	0.336 (±0.045)

Values in the brackets are halves of the interquartile distance. *OhmNet*'s results are statistically significant with a P-value of < 0.05

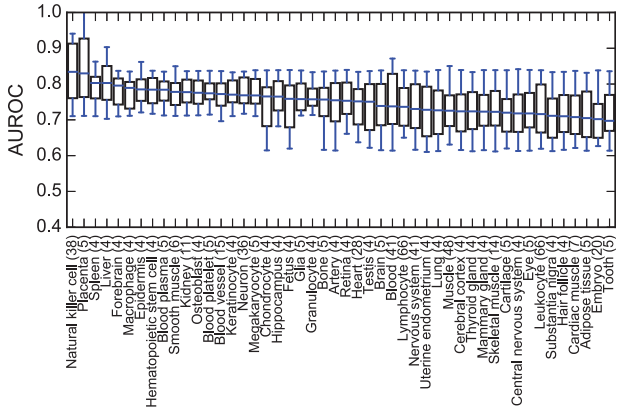

Fig. 4. Area under ROC curve (AUROC) scores for tissue-specific cellular function prediction by *OhmNet*. Numbers in the brackets are counts of tissue-specific cellular functions per tissue

From the results, we see how modeling the tissues and their hierarchy spanning multiple biological scales allows *OhmNet* to outperform other benchmark approaches. *OhmNet* outperforms GeneMania (Mostafavi *et al.*, 2008; Zuberi *et al.*, 2013) by 10.7%, which can be explained by GeneMania's inability to weight layers in the tissue network according to a multiscale tissue organization that is consistent with the tissue taxonomy constraints. We also compared *OhmNet* with two other methods (Guan *et al.*, 2012; Magger *et al.*, 2012) that were so far demonstrated as useful for mining tissue-specific protein relationships. *OhmNet* has produced more accurate predictions, surpassing other methods by up to 12.0% (AUROC) and up to 26.8% (AUPRC).

Independent modeling of the layers showed worse performance than collapsing the layers into one network. We observed that Collapsed LINE achieved a gain of 3.3% over Independent LINE, and Collapsed Node2vec achieved a gain of 7.4% over Independent Node2vec. However, approaches that neglect the existence of tissues or collapse tissue-specific protein interaction networks into a single network discard important information about the rich hierarchy of biological systems, giving *OhmNet* a 14.0% gain over Collapsed LINE, and a 8.5% gain over Collapsed Node2vec in AUC scores. This result is a good illustration of how tissue specificity is related to specialization of protein function (Greene *et al.*, 2015), and approaches able to directly profile proteins' distinct interaction neighborhoods in different tissues can leverage this specificity to generate more accurate hypotheses about tissue-specific protein actions.

Table 2. AUROC scores for transfer learning

Target tissue	AUROC (non-transfer)	AUROC (transfer)
Natural killer cell	0.834 (±0.076)	0.776 (±0.063)
Placenta	0.830 (±0.082)	0.758 (±0.068)
Spleen	0.803 (±0.030)	0.779 (±0.043)
Liver	0.803 (±0.047)	0.741 (±0.025)
Forebrain	0.796 (±0.036)	0.755 (±0.037)
Macrophage	0.789 (±0.037)	0.724 (±0.024)
Epidermis	0.785 (±0.030)	0.749 (±0.032)
Hematopoietic stem cell	0.784 (±0.035)	0.744 (±0.036)
Blood plasma	0.784 (±0.027)	0.703 (±0.039)
Smooth muscle	0.778 (±0.031)	0.729 (±0.041)
Average	0.799	0.746

Shown are the scores for ten tissues with best performance on cellular function prediction task. "Non-transfer": a classifier is trained on a target tissue and then used to predict cellular functions in the same tissue (Section 5.1). "Transfer": classifiers are trained on all non-target tissues and then used to predict cellular functions in the target tissue (Section 5.2).

5.2 Transfer of cellular functions to a new tissue

5.2.1 Experimental setup

In the transfer learning setting, we attempt to transfer knowledge learned in one or more *source layers* and use it for prediction in a *target layer*.

As before, we apply *OhmNet* to obtain a separate feature vector for every node and every layer in an unsupervised way. We then consider, in turn, every tissue as a target layer and *all other tissues* as source layers. For every function and every source layer, we train a separate classifier using the same classification model as in Section 5.1. We then predict functions for the target layer using only classifiers trained on the source layers. That is, we aim to predict cellular functions taking place in the target tissue without having access to any cellular function gene annotation in that tissue, i.e. we pretend the target tissue has no annotations. Prediction for one node in the target layer is the weighted average of predictions of the classifiers trained on source layers. Weights reflect hierarchy-based distances of source tissues from the target tissue. They are determined by the closed-form expressions mathematically equivalent to *OhmNet*'s regularization (details omitted due to space constraints).

5.2.2 Experimental results

Table 2 shows the classification accuracy results for transfer learning based on *OhmNet*. Since transfer tasks are more difficult than non-transfer tasks (Section 5.1), it is expected that the AUC scores will decrease on transfer tasks. Results in Table 2 confirm these expectations; however, we observe a very graceful degradation in performance leading to an only 7% average decrease in the AUC scores. We get the smallest performance differences for target tissues with many biologically similar source tissues (i.e. source layers) in the tissue network. For example, performance difference for the forebrain is only 5.2%, which is due to the fact that there are nine other layers in the tissue network closely related to the forebrain, such as the cerebellum and the midbrain. Considering all 48 tissues with tissue-specific cellular functions, *OhmNet* outperforms all comparison methods on most transfer tasks, achieving a gain of up to 20.3% over the closest benchmark in AUC scores (scores not shown). Notice that we exclude GeneMania in the comparison because it is not amenable to transfer learning. This result suggests

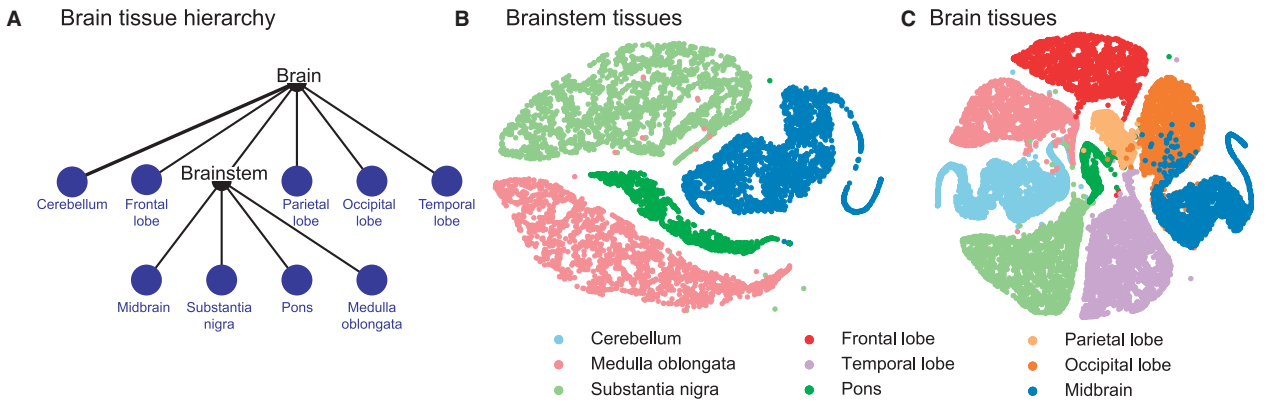

Fig. 5. Visualization of the brain tissue-specific protein interaction networks. (**A**) The two-level brain tissue hierarchy as specified by the BRENDA Tissue Ontology (Chang *et al.*, 2014) and used in the case study in Section 5.3. Leaves of the hierarchy (in blue) represent nine brain tissues each of which is associated with a tissue-specific protein interaction network. (**B**) Visualization of the brainstem-specific networks. The proteins are mapped to the 2D space using the t-SNE package with learned features as input. Color of a node indicates the tissue of the protein. (**C**) Visualization of the brain-specific networks. The proteins are mapped and colored using the same procedure as in **B**

that considering the relationships between tissues when learning features for proteins has a significant impact on transfer performance.

In general, we observed that the transferability of classifiers decreased when the tree-based distance between the source and the target tissue in the tissue hierarchy increased, which is consistent with the empirical evidence in transfer learning (Yosinski *et al.*, 2014). This also matches our intuition that a source tissue should be most informative for predicting cellular functions in an anatomically close target tissue (e.g. source and target tissues are both part of the same organ).

5.3 The multiscale model of brain tissues

We have seen in Section 4.1 that human tissues have a multi-level hierarchical organization. The tissue hierarchy categorizes tissues into: cell types, groups of cells with similar structure and function; organs, groups of tissues that work together to perform a specific activity; and organ systems, groups of two or more tissues that work together for the good of the entire body. We now aim to empirically demonstrate this fact and show that *OhmNet* in fact can discover embeddings that obey this organization.

We first construct a multi-layer brain network by integrating nine brain-specific protein interaction networks (e.g. the cerebellum, frontal lobe, brainstem and other brain tissues). Each of nine brain-specific networks is one layer in the multi-layer network. The layers are organized according to a two-level hierarchy (Fig. 5A). We run *OhmNet* on this multi-layer network to find node features in a purely unsupervised way. We then map the nodes to the 2D space based on the learned features. This way we assign every node in every layer to a point in the two-dimensional space based solely on the node's learned features. We then visualize the points and color them based on the layer they belong to.

Figure 5B shows the example for the brainstem tissues: substantia nigra, pons, midbrain and medulla oblongata. Laying out these tissue-specific networks is very challenging as the four brainstem tissues are very closely related to each other in the human body. However, the visualization using *OhmNet* performs quite well. Notice how points of the same color are closely distributed, and how well regions of the same color are separated from each other. In the brainstem example, this means that *OhmNet* generates a meaningful layout of the brainstem tissue-specific networks, in which proteins belonging to the same tissues are clustered together.

Figure 5C shows the example for the brain, which is located one level up from the brainstem in the tissue hierarchy. Again, *OhmNet* produces a meaningful layout of the nine brain tissue-specific networks.

In addition, we repeated this analysis by visualizing protein features learned by running principal component analysis (PCA) or non-negative matrix factorization (NMF) algorithm on the brain-specific PPI networks. Acknowledging the subjective nature of this analysis, we observed that visualizations using PCA or NMF were not very meaningful, as proteins belonging to the same tissue were not clustered together (data not shown).

OhmNet's result in Figure 5 is especially appealing because of two reasons. First, it shows that *OhmNet* can learn node features that adhere to a given hierarchy of layers. In the brain example, *OhmNet* learns the protein features that expose the multiscale tissue hierarchy. Second, it shows that *OhmNet* can generate meaningful visualizations of network embeddings despite the fact that *OhmNet*'s objective is independent of the visualization task.

6 Conclusion

We presented *OhmNet*, an approach for unsupervised feature learning in multi-layer networks. We use *OhmNet* to learn state-of-the-art task-independent protein features on a multi-layer network with 107 tissues. *OhmNet* models tissue interdependence up and down a tissue hierarchy spanning dozens of biological scales. The learned features achieve excellent accuracy on the cellular function prediction task, allow us to transfer functions to unannotated tissues, and provide insights into tissues.

There are several directions for future work. Our approach assumes the dependencies between layers are given in the form of a hierarchy. In several biological scenarios, the dependencies are given in the form of a graph, and we hope to extend the approach to handle graph-based dependencies. As the learned protein features are independent of any downstream task, it would be interesting to see whether our approach performs equally well for gene–disease association prediction and disease pathway detection.

Funding

This research has been supported in part by NSF IIS-1149837, NIH BD2K U54EB020405, DARPA SIMPLEX N66001 and Chan Zuckerberg Biohub.

Conflict of Interest: none declared.

References

Antanaviciute,A. *et al*. (2015) GeneTIER: prioritization of candidate disease genes using tissue-specific gene expression profiles. *Bioinformatics*, **31**, 2728–2735.

Ashburner,M. *et al*. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet*., **25**, 25–29.

Barutcuoglu,Z. *et al*. (2006) Hierarchical multi-label prediction of gene function. *Bioinformatics*, **22**, 830–836.

Belkin,M., and Niyogi,P. (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, vol. 14, MIT Press, Cambridge, pp. 585–591.

Cannistraci,C.V. *et al*. (2013) Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics*, **29**, i199–i209.

Carvunis,A.-R., and Ideker,T. (2014) Siri of the cell: what biology could learn from the iPhone. *Cell*, **157**, 534–538.

Chang,A. *et al*. (2014) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res*. **43**, D439–D446.

Chatr-Aryamontri,A. *et al*. (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res*., **43**, D470–D478.

Costanzo,M. *et al*. (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science*, **353**, aaf1420.

De Domenico,M. *et al*. (2014) Navigability of interconnected networks under random failures. *PNAS*, **111**, 8351–8356.

De Domenico,M. *et al*. (2015) Ranking in interconnected multilayer networks reveals versatile nodes. *Nat. Commun*., **6**, 6868.

De Domenico,M. *et al*. (2016) The physics of spreading processes in multilayer networks. *Nat. Phys*., **12**, 901–906.

Dutkowski,J. *et al*. (2012) A gene ontology inferred from molecular networks. *Nat. Biotechnol*., **31**, 38–45.

Fagerberg,L. *et al*. (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteom*., **13**, 397–406.

Ganegoda,G.U. *et al*. (2014) Prediction of disease genes using tissue-specified gene-gene network. *BMC Syst. Biol*., **8**, S3.

Greene,C.S. *et al*. (2015) Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet*., **47**, 569–576.

Grover,A., and Leskovec,J. (2016) Node2vec: scalable feature learning for networks. In *KDD*, pp. 855–864.

GTEx,C. *et al*. (2015) The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.

Guan,Y. *et al*. (2012) Tissue-specific functional networks for prioritizing phenotype and disease genes. *PLoS Comput. Biol*., **8**, e1002694.

Hayes,W. *et al*. (2013) Graphlet-based measures are suitable for biological network comparison. *Bioinformatics*, **29**, 483–491.

Hou,C. *et al*. (2014) Joint embedding learning and sparse regression: a framework for unsupervised feature selection. *IEEE Trans. Cybernet*., **44**, 793–804.

Hu,J.X. *et al*. (2016) Network biology concepts in complex disease comorbidities. *Nat. Rev. Genet*., **17**, 615–629.

Kitsak,M. *et al*. (2016) Tissue specificity of human disease module. *Sci. Rep*., **6**, 35241.

Kotlyar,M. *et al*. (2015) Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res*., **44**, D536–D541.

Kramer,M. *et al*. (2014) Inferring gene ontologies from pairwise similarity data. *Bioinformatics*, **30**, i34–i42.

Li,Y. *et al*. (2015) Gated graph sequence neural networks. *arXiv:1511.05493*.

Lois,C. *et al*. (2002) Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science*, **295**, 868–872.

Lopes,T.J. *et al*. (2011) Tissue-specific subnetworks and characteristics of publicly available human protein interaction databases. *Bioinformatics*, **27**, 2414–2421.

Magger,O. *et al*. (2012) Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput. Biol*., **8**, e1002690.

Menche,J. *et al*. (2015) Uncovering disease-disease relationships through the incomplete interactome. *Science*, **347**, 1257601.

Mikolov,T. *et al*. (2013) Efficient estimation of word representations in vector space. *arXiv:1301.3781*.

Mostafavi,S., and Morris,Q. (2009) Using the gene ontology hierarchy when predicting gene function. In *UAI*, AUAI Press, Corvallis, pp. 419–427.

Mostafavi,S. *et al*. (2008) GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*., **9**, 1.

Nickel,M. *et al*. (2011) A three-way model for collective learning on multi-relational data. In *ICML*, ACM, Bellevue, pp. 809–816.

Okabe,Y., and Medzhitov,R. (2014) Tissue-specific signals control reversible program of localization and functional polarization of macrophages. *Cell*, **157**, 832–844.

Orchard,S. *et al*. (2013) The MIntAct projectintact as a common curation platform for 11 molecular interaction databases., *Nucleic Acids Res*. **42**, D358–D363.

Perozzi,B. *et al*. (2014) Deepwalk: online learning of social representations. In *KDD*, ACM, pp. 701–710.

Prasad,T.K. *et al*. (2009) Human protein reference database-2009 update. *Nucleic Acids Res*. **37**, D767–D772.

Pržulj,N. (2007) Biological network comparison using graphlet degree distribution. *Bioinformatics*, **23**, e177–e183.

Radivojac,P. *et al*. (2013) A large-scale evaluation of computational protein function prediction. *Nat. Methods*, **10**, 221–227.

Rakyan,V.K. *et al*. (2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tdmrs). *Genome Res*., **18**, 1518–1529.

Rolland,T. *et al*. (2014) A proteome-scale map of the human interactome network. *Cell*, **159**, 1212–1226.

Ruepp,A. *et al*. (2010) CORUM: the comprehensive resource of mammalian protein complexes-2009. *Nucleic Acids Res*. **38**, D497–D501.

Stojanova,D. *et al*. (2013) Using PPI network autocorrelation in hierarchical multi-label classification trees for gene function prediction. *BMC Bioinformatics*, **14**, 1.

Tang,J. *et al*. (2015) Line: Large-scale information network embedding. In *WWW*, pp. 1067–1077.

Tang,L. *et al*. (2012) Scalable learning of collective behavior. *IEEE Trans. Knowl. Data Eng*., **24**, 1080–1091.

Tenenbaum,J.B. *et al*. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319–2323.

Vidulin,V. *et al*. (2016) Extensive complementarity between gene function prediction methods. *Bioinformatics*, **32**, 3645–3653.

Wang,D. *et al*. (2016a) Structural deep network embedding. In *KDD*, ACM, pp. 1225–1234.

Wang,W. *et al*. (2016b) Tissue-specific pathway association analysis using genome-wide association study summaries. *Bioinformatics*, **33**, 243–247.

Xiaoyi,L., D., Nan,L.H. *et al*. (2014) A deep learning approach to link prediction in dynamic networks. In *SDM*, 289–297

Yeger-Lotem,E., and Sharan,R. (2015) Human protein interaction networks across tissues and diseases. *Front. Genet*., **6**, 257.

Yosinski,J. *et al*. (2014) How transferable are features in deep neural networks? In *NIPS*, Curran Associates, pp. 3320–3328.

Yu,M. *et al*. (2016) Translation of genotype to phenotype by a hierarchy of cell systems. *Cell Syst*., **2**, 77–88.

Zhai,S., and Zhang,Z. (2015) Dropout training of matrix factorization and autoencoder for link prediction in sparse graphs. In *SDM*, 451–459.

Žitnik,M., and Zupan,B. (2015) Data fusion by matrix factorization. *IEEE Trans. Pattern Anal. Mach. Intell*., **37**, 41–53.

Zuberi,K. *et al*. (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res*., **41**, W115–W122.