

Representation Learning for Networks in Biology and Medicine: Advancements, Challenges, and Opportunities

Michelle M. Li^{1,3}, Kexin Huang², and Marinka Zitnik^{3,4,5,*}

¹Bioinformatics and Integrative Genomics Program, Harvard Medical School

²Health Data Science Program, Harvard T.H. Chan School of Public Health

³Department of Biomedical Informatics, Harvard Medical School

⁴Broad Institute of MIT and Harvard

⁵Harvard Data Science

*Correspondence: marinka@hms.harvard.edu

Abstract

With the remarkable success of representation learning in providing powerful predictions and data insights, we have witnessed a rapid expansion of representation learning techniques into modeling, analysis, and learning with networks. Biomedical networks are universal descriptors of systems of interacting elements, from protein interactions to disease networks, all the way to healthcare systems and scientific knowledge. In this review, we put forward an observation that long-standing principles of network biology and medicine—while often unspoken in machine learning research—can provide the conceptual grounding for representation learning, explain its current successes and limitations, and inform future advances. We synthesize a spectrum of algorithmic approaches that, at their core, leverage topological features to embed networks into compact vector spaces. We also provide a taxonomy of biomedical areas that are likely to benefit most from algorithmic innovation. Representation learning techniques are becoming essential for identifying causal variants underlying complex traits, disentangling behaviors of single cells and their impact on health, and diagnosing and treating diseases with safe and effective medicines.

1 Introduction

Networks, or graphs, are pervasive in biology and medicine, from molecular interaction maps to dependencies between diseases in a person, all the way to populations encompassing social and health interactions. Depending on the type of information encoded in networks, the implications of an “interaction” between two entities can differ. For instance, edges in a protein-protein interaction (PPI) network can indicate physical interactions measured by experiments, such as yeast two-hybrid screens and mass spectrometry (e.g., [148, 197]); edges in a regulatory network can indicate causal interactions between genes measured by dynamic single-cell expression (e.g., [174]); and edges in an electronic health record (EHR) network can indicate hierarchical relationships found in medical ontologies (e.g., [182, 190]). From molecular to healthcare systems, networks have emerged as a predominant paradigm for representing, learning, and reasoning about biomedical systems.

The case for representation learning on biomedical networks. Capturing interactions in a biomedical system gives rise to a bewildering degree of complexity that can likely only be fully understood through

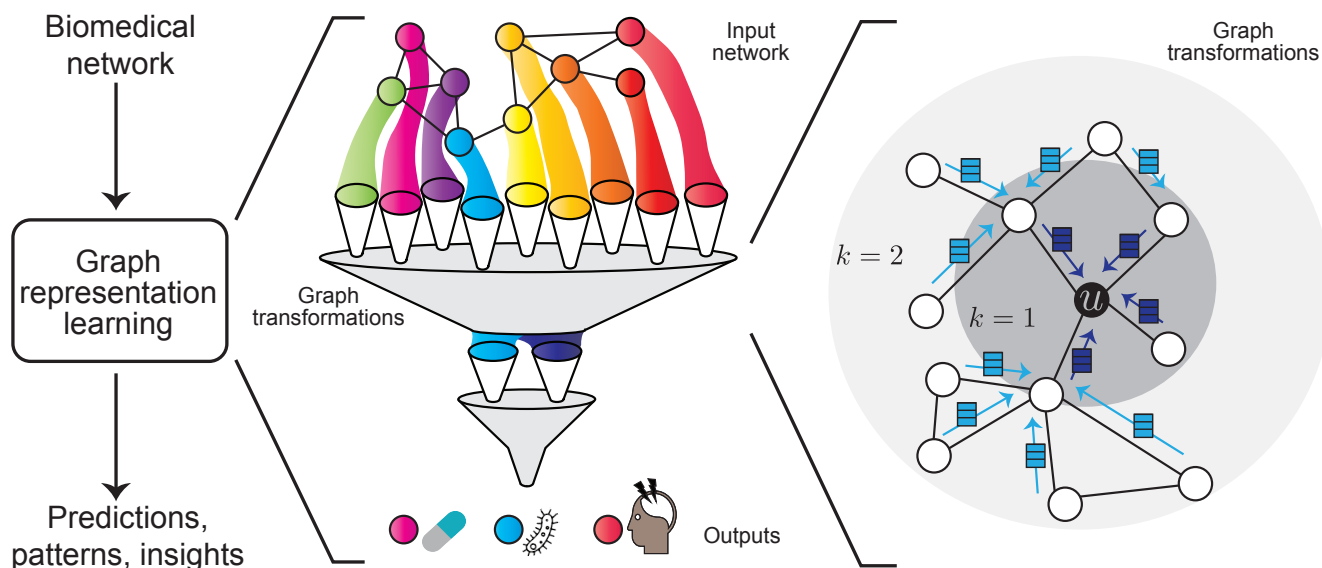


Figure 1: **Representation learning for networks in biology and medicine.** Given a rich biomedical network, a representation learning method transforms the graph (e.g., a graph neural network does so through multiple layers of graph convolutions, and network diffusion does so through a stochastic process that traces the flow of information through the graph) in order to extract patterns/representations (often referred to as *embeddings*) predictive of an endpoint of interest (e.g., assignment of protein nodes to biological processes, or patient nodes to outcomes). For example, the far right panel shows a local 2-hop neighborhood around node u . It illustrates how information (e.g., neural messages shown as blue vectors) can be propagated along edges in the neighborhood, transformed, and finally aggregated at node u to arrive at the u 's embedding.

a holistic and integrated systems view [17, 28, 164]. To this end, network biology and medicine have over the last two decades identified a series of *organizing principles* (e.g., [16, 86, 106, 262]) that govern biomedical networks. These principles *link network structure to molecular phenotypes, biological roles, disease, and health*. The long-standing principles—while often unspoken in machine learning research—provide the conceptual grounding that, we argue, can explain the successes (and limitations) of representation learning for modeling biomedical networks and inform future development of the field. In particular, while interpretation of an edge in a network depends on the context, interacting entities tend to be more similar than non-interacting entities. For example, disease ontologies are structured such that disease terms connected by an edge tend to be more alike than unconnected disease terms. In PPI networks, mutations in interacting proteins often lead to similar diseases. Conversely, proteins involved in the same disease have an increased tendency to interact with each other. In cellular networks, components associated with a specific phenotype tend to cluster in the same network neighborhood.

Representation learning to realize key principles of network biology and medicine. We posit that *representation learning can realize key principles of network biology and medicine*. A corollary of this hypothesis is that representation learning can be well-suited for analysis, learning, and reasoning about biomedical networks. At the core of representation learning is the notion of vector space *embeddings*. The idea is to learn how to represent nodes (or larger graph structures) in a network as points in a low-dimensional space, where the geometry of this space is optimized to reflect the structure of interactions between the nodes. Representation learning formalizes this idea by specifying (deep, non-linear) transformation functions that map nodes to points in a compact vector space, termed *embeddings*. These functions are optimized to embed the input graph so that performing algebraic operations in the learned space reflects the graph's topology. Nodes are mapped to embeddings such that nodes with similar network neighborhoods are embedded close together in the

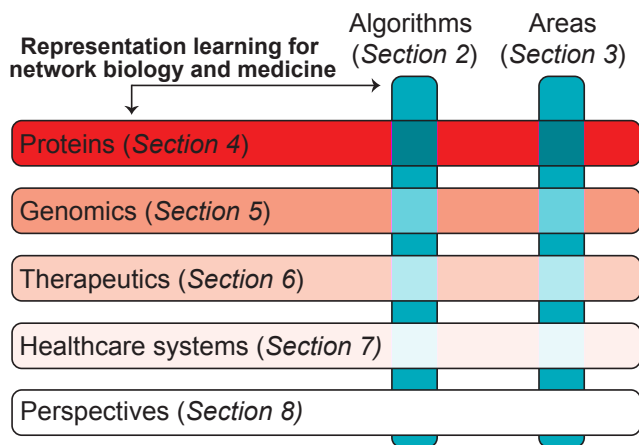


Figure 2: **Organization of this review.** Section 2 formulates a taxonomy of graph representation learning for biomedical networks. Section 3 provides an overview of their applications in biology and medicine. The following sections discuss advances enabled by graph representation learning to elucidate protein structure, interaction, and function (Section 4); genome-wide associations that drive disease progression (Section 5); small compound structure, interactions with proteins and other compounds, and indications (Section 6); precision medicine based on patients’ interactions with healthcare systems, represented by electronic health records and medical images (Section 7). Finally, Section 8 offers perspectives on machine learning for interconnected datasets.

embedding space. Notably, the implication of the embedding space for understanding biomedical networks (e.g., PPI networks) is that the proximity of points in the space (e.g., the distance between protein embeddings) naturally reflects the similarity of entities those points represent (e.g., the similarity of proteins’ phenotypes), suggesting that embeddings can be thought of as a differentiable manifestation of key principles in network biomedicine.

Algorithmic paradigms (Figure 1). Network science and graph theoretic techniques have fueled biomedical discoveries, from uncovering relationships between diseases [91, 135, 159, 200] to repurposing drugs [41, 42, 96]. Further algorithmic innovations, such as random walks [40, 229, 242], kernels [83], and network propagation [214], have also played crucial roles in capturing structural and neighborhood information from networks to generate embeddings for downstream predictions. Feature engineering is another commonly used paradigm for machine learning on biomedical networks, including but not limited to hard-coding network features (e.g., higher-order structures, network motifs, degree counts, and common neighbor statistics) and feeding the engineered feature vectors into predictive models. This strategy, while powerful, does not fully leverage network information and can fail to generalize to new network types and datasets [255].

Recently, graph representation learning methods have emerged as a leading paradigm for deep learning on biomedical networks. However, deep learning on graphs is challenging because graphs contain complex topographical structure, there is no fixed node ordering and no reference points, and they comprise of many different kinds of entities (nodes) and various types of interactions (edges) relating them to each other. Classic deep learning methods are unable to consider such diverse structural properties and rich interactions, which are the essence of biomedical networks. This is because classic deep models are designed primarily for fixed-size grids (e.g., images and tabular datasets) or optimized for text and sequences. As such, they have brought extraordinary gains in computer vision, natural language processing, speech, and robotics. Akin to how deep learning on images and sequences has revolutionized the image analysis and natural language processing fields, graph representation learning is poised to transform the study of complex systems in biology and medicine.

Scope of this review. Our focus is on representation learning, specifically manifold learning [27], graph transformer networks [250], differential geometric deep learning [25], topological data analysis (TDA) [34, 224], and graph neural networks (GNN) [125]. Figure 2 depicts the structure and organization of this review. We first provide a technical exposition of prevailing graph learning paradigms and describe their critical impact

in accelerating biomedical research. With every current application area of graph representation learning (Figure 4), we demonstrate the potential directions in which graph representation learning could take through four unique prospective studies, each addressing at least one of the following key prediction tasks in graph machine learning: node-, edge-, subgraph-, and graph-level prediction, continuous embedding, and generation.

Distinctive contributions of this review. Most existing reviews independently discuss deep learning on structured data [29, 255], graph neural networks [232], network representation learning for homogeneous and heterogeneous graphs [36, 138, 249], solely heterogeneous graphs [63], data fusion [265], dynamic graphs [119], network propagation [53], topological data analysis in biology [22], and biomedical networks [22, 28, 129, 144, 176]. Furthermore, many survey the impact of graph neural networks on molecular generation [55, 227], as well as drug discovery and repurposing [116, 201]. In contrast, our review aims to unify graph representation learning paradigms that have enabled molecular, genomic, and therapeutic discoveries, and precision medicine advancements.

2 Representation learning for graphs in biology and medicine

Machine learning on graphs has rapidly evolved to better analyze networks that are only becoming more comprehensive and complex. As a result, graph machine learning encompasses a wide range of methods, from graph theoretic techniques rooted in classic network science principles to graph neural networks and generative models (Figure 3). To provide an overview of such methods, we first describe key notations and definitions regarding graphs (Section 2.1). Next, we outline the three main types of graph machine learning tasks (Section 2.2). Finally, we summarize and categorize the seven major paradigms of machine learning for biomedical networks (Section 2.3).

2.1 Notation and definitions

We start with the notation, and proceed with an overview of key graph elements and a brief description of network types commonly encountered in biology and medicine.

Graph-theoretic elements. Graphs consist of the following key elements:

- *Node* v represents a biomedical entity, ranging from atoms to patients (Figure 4).
- *Edge* $e_{u,v}$ is a relation or link between node entities u and v , such as a bond between atoms, an affinity between molecules, a disease association between phenotypes, and a referral between a patient and a doctor. We denote the edge set in the graph as \mathcal{E} and the complementary non-edge set as \mathcal{E}^c . The edge can be directed, oriented such that it points from a source (or head node) to a destination (or tail node). The edge can also be undirected, where the nodes have two-way relations.
- *Graph* $G = (\mathcal{V}, \mathcal{E})$ consists of a collection of nodes \mathcal{V} that are connected by an edge set \mathcal{E} , such as a molecular graph or protein interaction network. Adjacency matrix \mathbf{A} is commonly used to represent a graph, where each entry $\mathbf{A}_{u,v}$ is 1 if nodes u, v are connected, and 0 otherwise. $\mathbf{A}_{u,v}$ can also be the edge weight between nodes u, v . We denote the number of the nodes $|\mathcal{V}| = n$ and the number of edges $|\mathcal{E}| = m$.
- *Subgraph* $S = (\mathcal{V}_S, \mathcal{E}_S)$ is a subset of a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}_S \subseteq \mathcal{V}, \mathcal{E}_S \subseteq \mathcal{E}$. Examples include disease modules in a protein interaction network, or communities in a patient-doctor referral network.
- *Node Feature* $\mathbf{x}_v \in \mathbb{R}^d$ describes attributes of node v . The node feature matrix is denoted as $\mathbf{X} \in \mathbb{R}^{n \times d}$. Similarly, we can have edge features $\mathbf{x}_{u,v}^e \in \mathbb{R}^c$ for edge $e_{u,v}$ collected together into edge feature matrix $\mathbf{X}^e \in \mathbb{R}^{m \times c}$.

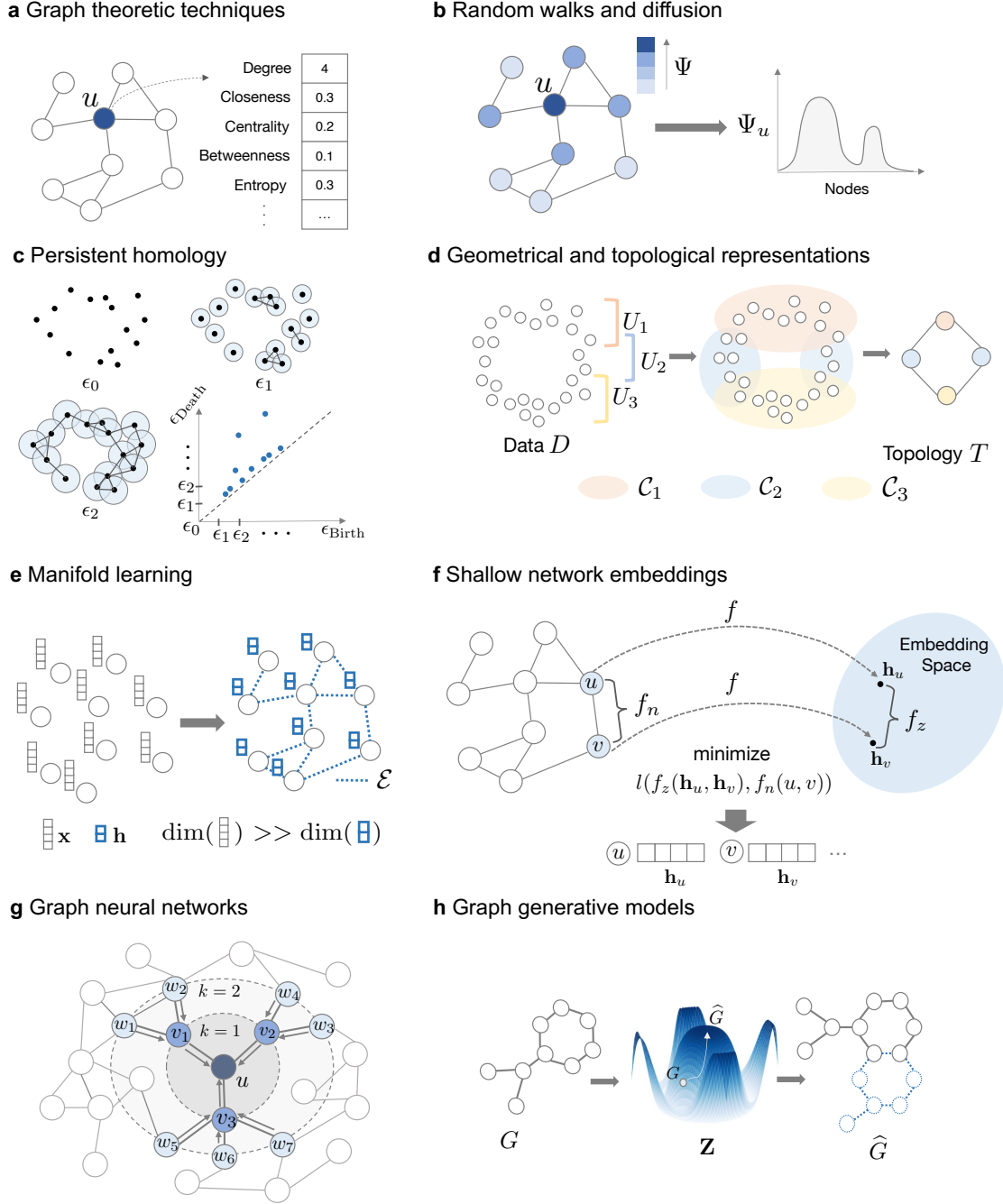


Figure 3: **Predominant graph learning paradigms.** Graph machine learning is a large field with a diverse set of methods. Here, we summarize and categorize seven major graph machine learning methods paradigms. **(a)** Graph theoretic techniques compute a deterministic value that describes patterns in the graph (Section 2.3.1); **(b)** diffusion process captures the importance and influence of nodes through network diffusion (Section 2.3.2); **(c-d)** topological data analysis provides summarized views of the shape of the data (Section 2.3.3); **(e)** manifold learning aims to obtain the underlying graph structure of data and a low-dimensional embedding (Section 2.3.4); **(f)** shallow network embeddings generate node representations through direct encoding of node similarities in the input graph (Section 2.3.5); **(g)** graph neural networks learn graph embeddings through supervised signals by neural networks (Section 2.3.6); **(h)** generative models generate novel graphs that have desirable properties (Section 2.3.7). Note that panels (c) and (d) are two major diagrams of TDA.

- *Label* is a target value associated with either a node Y_v , an edge Y_e , a subgraph Y_S , or a graph Y_G .

Walks and paths in graphs. A walk of length l from node v_1 to node v_l is a sequence of nodes and edges $v_1 \xrightarrow{e_{1,2}} v_2 \cdots v_{l-1} \xrightarrow{e_{l-1,l}} v_l$. A (simple) path is a type of walks where all nodes in the walk is distinct. For every two nodes u, v in the graph G , we define the distance $d(u, v)$ as the length of the shortest path between them. In a heterogeneous graph, a meta-path is a sequence of node types V_i and their edge types $R_{i,j}$: $V_1 \xrightarrow{R_{1,2}} V_2 \cdots V_{l-1} \xrightarrow{R_{l-1,l}} V_l$.

Local neighborhoods. For a node v , we denote its neighborhood $\mathcal{N}(v)$ as the collection of nodes that are connected to v , and its degree is the size of $\mathcal{N}(v)$. The k hop neighborhood of node v is the set of nodes that are exactly k hops away from node v : $\mathcal{N}^k(v) = \{u | d(u, v) = k\}$.

Graph types. The following types of graphs are commonly considered to model complex biomedical systems.

- *Simple, weighted, and attributed graphs.* A simple graph $G = (\mathcal{V}, \mathcal{E})$ is fully described by nodes \mathcal{V} and edges \mathcal{E} . For example, a set of binary PPIs gives rise to an unweighted PPI network (Section 4.2). Further, nodes and edges can be accompanied by single-dimensional (e.g., edge weights) or multi-dimensional attribute vectors describing node and edge properties. For example, each node in a cell-cell interaction network can have a gene expression attribute vector encoding the gene's expression profile.
- *Multimodal or heterogeneous graphs.* Multimodal or heterogeneous graphs consist of nodes of different types (node type set \mathcal{A}) connected by diverse kinds of edges (edge type set \mathcal{R}). For example, in a drug-target-disease interaction network, nodes represent $\mathcal{A} = \{\text{drugs, proteins, diseases}\}$ and edges indicate $\mathcal{R} = \{\text{drug-target binding, disease-associated mutations, treatments}\}$ (Section 6.2).
- *Knowledge graphs.* Biomedical knowledge graph is a heterogeneous graph that captures knowledge retrieved from literature and biorepositories (Section 7.2). Knowledge graph is given by a set of triplets $(u, r, v) \in \mathcal{V} \times \mathcal{R} \times \mathcal{V}$, where nodes u, v belong to node types \mathcal{A} and are connected by edges with type $r \in \mathcal{R}$.
- *Multi-layer graphs.* Multi-layer graphs capture hierarchical relations by grouping individual networks into different layers. Formally, we have a set of networks G_1, \dots, G_n and l layers where each layer corresponds to a set of networks. Different layers can represent distinct contexts, such as tissues or diseases. Edges can also be added across layers. For example, each tissue can be represented with a tissue-specific PPI network, and PPIs for every tissue can be organized by tissue taxonomy, where each layer corresponds to a tissue taxonomy (Section 4). Inter-layer edges can be connected for the same proteins across tissues [263].
- *Temporal graphs.* Biomedical systems evolve over time. A temporal graph consists of a sequence of graphs G_1, \dots, G_T ordered by time, where at each time step t we observe a subset of all nodes and their activity. For example, a human brain can be modeled as a temporal graph of brain regions showing task-related increases in neural activity at time t (e.g., greater activity during an experimental task than during a baseline state) and linked based on functional connectivity at t .
- *Spatial graphs.* Nodes or edges in a spatial graph are spatial elements usually associated with coordinates in one, two or three dimensions, e.g., a spatial representation of cell-cell interactions in the 3-dimensional (3D) Euclidean tissue environment or a 3D point cloud of the protein's atomic coordinates. Spatial graphs are defined by having nodes or edges with spatial locations. This small modification to aspatial graphs has profound effects on how these graphs are used and interpreted because a spatial graph is a location map of points with the constraints of space rather than an abstract structure.

The graph types described above can be combined to give rise to new objects, such as multi-layer spatial

graphs or multimodal temporal graphs.

2.2 Graph machine learning tasks

We divide machine learning tasks on biomedical graphs into three broad categories: graph prediction, latent graph learning, and graph generation. Each category is associated with several individual graph machine learning tasks.

Canonical graph prediction. Graph prediction aims to predict a label in the graph. The label can be associated with any unit of the graph. There are four canonical graph learning tasks: (1) *Node classification/regression* aims to find a function $f : \mathcal{V} \rightarrow Y_{\mathcal{V}}$ that predicts the label of a node in the graph; (2) *Link prediction* aims to find a function $f : \mathcal{E} \cup \mathcal{E}^c \rightarrow \{0, 1\}$ to predict whether there exists a link between a given pair of nodes in the graph; (3) *Edge classification/regression* aims to find a function $f : \mathcal{E} \rightarrow Y_{\mathcal{E}}$ that predicts the label of an edge; (4) *Graph classification/regression* aims to find a function $f : \mathcal{G} \rightarrow Y_{\mathcal{G}}$ that maps each graph in a graph set to the correct label.

Other graph prediction tasks. In addition to the four standard prediction tasks on graphs, there are additional tasks that are particularly important for biomedical graphs: (1) *Module detection* aims to detect a subgraph module in the graph that contributes to a variable; (2) *Clustering or community detection* aims to partition the graphs into a set of subgraphs such that each subgraph contains similar nodes; (3) *Subgraph classification/regression* aims to predict a label for the subgraph or module; (4) *Dynamic graph prediction* aims to perform the above prediction tasks in a sequence of dynamic graphs.

Latent graph learning. While graph prediction tasks predict given the graph-structured data, latent graph learning aims to obtain a function $f : \mathcal{V}, \mathbf{X} \rightarrow \mathcal{E}$ to learn the underlying graph structure (e.g., edges) given only the nodes and their feature attributes. The learned graph can be used to (1) perform graph prediction tasks; (2) obtain the inherent topology of the data; and (3) generate latent low-dimensional representations of the feature attributes.

Graph generation. The objective of graph generation is to generate a never-before-seen graph G with some properties of interest. Given a set of training graphs \mathcal{G} with certain shared characteristics, the task is to learn a function $f : \mathcal{G} \rightarrow \mathcal{D}_{\mathcal{G}}$ to obtain a distribution $\mathcal{D}_{\mathcal{G}}$ that characterizes the training graphs. Then, the learned distribution can be used to generate a new graph G' , which has the same characteristics as or optimized properties compared to the training graphs.

2.3 Graph machine learning paradigms

Machine learning on graphs has been studied widely in both network science and machine learning communities over many years. There have been numerous proposed methods. To provide an overall perspective, we summarize and categorize them into seven distinct method paradigms. In the following, we introduce the method paradigms and provide several notable works for each.

2.3.1 Graph theoretic techniques

Networks model relations among real world subjects. Such relations form patterns of structures in the network. Among the network science community, many have studied these patterns of graph structures, and proposed graph statistics to measure their characteristics. We summarize such statistics into the following three categories: node-, link-, and subgraph-level statistics. See Figure 3a for an illustration of network statistics.

Node-level techniques. The goal of node-level statistics is to measure the role of a node u in a graph. For instance, betweenness [78] calculates the number of shortest paths that pass through the node. A node with high betweenness is called a bottleneck because it controls the information flow of the network. Various centrality statistics are proposed to measure the various roles of a node regarding its structure and function. For example, k-core [126] measures the position of the node in the graph.

Link-level techniques. Statistics have been demonstrated as a powerful tool for link prediction. They are used to represent the likelihood that a link exists between two nodes. Classic statistics include common neighbor index, Adamic–Adar index, resource allocation index, etc [81]. Most of them rely on the homophily principle. However, [130] recently showed that the protein-protein interaction (PPI) network does not follow homophily because interacting proteins are not necessarily similar, and similar proteins do not necessarily interact. They propose L3, which calculates the number of paths of length 3, and it shows strong performance in learning PPI networks. For a complete list of link-level graph statistics methods, we refer readers to [144].

Subgraph-level techniques. Many subnetwork patterns recur in the network. Such patterns are called motifs, and they are shown to be the basic blocks of complex networks. For instance, a feed-forward loop is an important three-node motif in gene regulatory networks [161]. They have functional roles, such as increasing the response to signals [155]. Individualized motifs for each network can also be computed through frequent subgraph mining algorithms [113].

Formulation 1: Graph theoretic techniques

Graph theoretic techniques are functions that map network components to real-values representing aspects of graph structure, such as node proximity [4, 19, 159] and node centrality [86]. We use betweenness as an example.

Example. For node u in graph G , the betweenness is calculated as: $B_u = \sum_{s \neq u \neq t} \frac{\sigma_{s,t}(u)}{\sigma_{s,t}}$, where $\sigma_{s,t}$ is the number of shortest paths between nodes s and t , and $\sigma_{s,t}(u)$ is the number of shortest paths that pass through node u . Basically, the larger the betweenness, the larger the influence of this node u on the network.

2.3.2 Network diffusion

Nodes in a graph influence each other along the paths. Diffusion measures these spreads of influences. The typical resulting outcomes of interest include a scalar ψ between every node v to the source node u that measures the influence, or a diffusion profile ψ_u for source node u , which captures the local connectivity patterns (Figure 3b). Many have studied the effect of diffusion in physics, economics, epidemiology and various formulations have been proposed [5, 52, 177]. On a very related line of work, label propagation leverages connected links to propagate labels [218].

Diffusion state distance. One effective method for biomedical networks is the diffusion state distance [30], which first calculates the number of times a random walk starting at source node will visit a destination node given a fixed number of steps, and iterates this process for every destination node in the graph.

Unsupervised extension. Recent efforts, such as GraphWave [64], adopt an unsupervised learning method to learn an embedding for each node by leveraging heat wavelet diffusion patterns. The resulting embeddings allow nodes residing in different parts of a graph to have similar structural roles within their local network topology.

Formulation 2: Network diffusion

Network diffusion computes the network influence signatures based on propagation on the networks. We use Diffusion State Distance (DSD) [30] as an example.

Example. For a node u , we calculate its diffusion distance $D(u, v_i)$ from every node $v_i \in \mathcal{V}$ as the expected number of times that p_u , a random walk of length k starting from u , will visit v_i . Formally, $D(u, v_i) = \mathbb{E}[\text{sign}(v_i, p_u)]$, where $\text{sign}(v_i, p_u)$ is 1 if node $v_i \in p_u$ and 0 otherwise. So, we obtain a vector $D(u) = (D(u, v_1), \dots, D(u, v_n))$. Then, the DSD between nodes u and v is defined as $\text{DSD}(u, v) = \|D(u) - D(v)\|_1$, the L_1 -norm of the diffusion vector difference. Intuitively, DSD measures the differences in the node influence from every other node to see if they have similar local connectivity.

2.3.3 Topological data analysis

For a large and high-dimensional dataset D , it is hard to directly gauge their characteristics or obtain a summary of the data. However, all data have an underlying shape or topology T , which can be considered as a network. Topological data analysis (TDA) analyzes the topology of the data to generate an underlying graph structure. There are two major diagrams in TDA: persistent homology (Figure 3c) and mapper (Figure 3d).

Persistent homology. Persistent homology [67] obtains a vector that quantifies various topological shapes at different spatial resolutions. As the resolution scale expands, the noise and artifacts would disappear while the important structure persists. Recent works have used neural networks on top of persistent diagrams to learn augmented topological features [33, 104]. [102] propose a differentiable persistent homology layer in the network to make any GNN topology-aware.

Formulation 3: Persistent homology

Given a dataset D , persistent homology generates a persistence diagram (often referred to as a barcode) that captures the significant topological features in D , such as connected components, holes, and cavities [26, 103]. It consists of the following steps:

1. **Construction of the Rips Complex.** Consider each data point in the original dataset D as a vertex. For each pair of vertices u, v , create an edge if the distance between them is at most ϵ , i.e., $\mathcal{E} = \{(u, v) | d(u, v) \leq \epsilon\}$, given the distance metric d . Consider a monotonically increasing sequence of $\epsilon_0, \dots, \epsilon_n$, we then generate a filtration of Rips complexes G_0, \dots, G_n .
2. **Homology.** Homology characterizes topological structures (e.g., 0-th order homology measures connected components, 1-st order homology measures holes, 2-nd order homology measures voids). A class in a k -th order homology is an instantiation of the homology. In each Rips complex, we can track the various homology classes, which capture various topological structures. For a rigorous definition of homology, we refer reader to [68].
3. **Persistence Diagrams.** At each ϵ , we denote the emergence of a new homology class a as its birth using the current ϵ . Similarly, we denote the disappearance of a previous homology class a as its death. Thus, for each homology class a , we can represent them as $(\epsilon_{\text{Birth}}, \epsilon_{\text{Death}})$. After ϵ reaches the end of the sequence, we have a set of 2-dimensional points (x, y) , each corresponding to the birth and death of a homology class. The persistence diagram is a plot of these points. The farther away from the diagonal, the longer the lifespan of the homology class, implying that the structure is an important shape of the data topology.

Mapper. While persistent homology provides a vector of topology, mapper generates a topology graph [169, 193]. This graph is obtained by first clustering topologically similar data into a node, where similarity is defined by a filter function, and then connecting the clusters as the backbone of the data topology. The resulting shape can be used to visualize the data and understand data subtypes [57] and trajectories of development [153]. Mapper is highly dependent on the filter function, and it is usually constructed with domain expertise. Recent works have integrated neural networks to automatically learn the filter function from the data [23].

Formulation 4: Geometric representations along preassigned guiding functions called filters

A mapper generates the data topology T from the data D [193, 221]. A standard mapper procedure consists of the following steps:

1. **Reference Map.** *Given a set of data D , we first define a continuous filter function $f : D \rightarrow Z$ that assigns every data point in D to a value in Z .*
2. **Construction of a Covering.** *A finite covering $\mathbb{U} = \{U_\alpha\}_{\alpha \in A}$ is constructed on Z . Each cover consists of a set of points D_α , and is the pre-image of the cover $D_\alpha = f^{(-1)}(U_\alpha)$.*
3. **Clustering.** *For each subset D_α , we apply a clustering algorithm C that generates N_α clusters.*
4. **Topology Graph.** *Each cluster forms a node. If two clusters share data points in D , then an edge is formed. The resulting graph is the topology graph T of data D . Formally, the topology graph is defined as the nerve of the cover by the path-connected components.*

2.3.4 Manifold learning

Real-world data is usually high-dimensional. To better interpret them, a mapping to find the low-dimensional characterization of the data is ideal. This mapping is called manifold learning, or non-linear dimensionality reduction (Figure 3e). Note that the underlying manifold can be considered as a weighted network such that higher weights are assigned to edges between data points that are closer in the manifold.

In a typical setting, we only have a set of data points, or nodes u_1, \dots, u_n , and their associated attributes $\mathbf{x}_1, \dots, \mathbf{x}_n$ without any connections among them. To learn the underlying manifold using graphs, the first step is to construct an edge set \mathcal{E} that connects nodes given some distance measure. With this connectivity graph, one approach is to apply a graph statistics operator to directly compute the low-dimensional embedding, such as laplacian eigenmap [20] and isomap [205]. Another approach is to optimize various kinds of cost functions to generate low-dimensional embeddings that preserve distance measures on the graph, such as t-SNE [151]. However, such methods are all multi-stage processes in which the outcome depends on the defined distance measures. Recently, a line of research emerged that can generate embeddings in an end-to-end learnable manner, where the manifold is learned by the signals from the downstream prediction task [97, 224].

Formulation 5: Manifold learning

The goal of manifold learning is to learn a low-dimensional embedding that captures the data manifold from high-dimensional data [27]. In the following, we use isomap [205] as an example.

Example. *First, given a set of data points with high-dimensional feature vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$, we apply a k -nearest neighbor algorithm and connect the nearest neighbors to form a neighborhood graph G . Next, we calculate the distance matrix \mathbf{D} , where each entry $\mathbf{d}_{i,j}$ is the length of the shortest path between nodes i and j in*

the neighborhood graph. Then, we apply an optimization algorithm to obtain a set of low-dimensional vectors $\mathbf{h}_1, \dots, \mathbf{h}_n$ that minimizes $\sum_{i < j} (\|\mathbf{h}_i - \mathbf{h}_j\| - d_{i,j})^2$.

2.3.5 Shallow network embeddings

The goal of graph representation learning is to generate embeddings for each node that capture the graph information. In other words, it is to encode nodes such that similarity in the embedding space approximates the similarity in the network (Figure 3f). Thus, we need to define the similarity function and the encoder.

General networks. The direct implementation uses shortest path length as the network similarity between two nodes and dot product as the encoder. To capture higher-order network similarity, one method is to define similarity as the co-occurrence in a walk of length k . Unsupervised learning techniques that predict which node belongs to the walk, such as skipgram [160], can then be applied on sampled walks to generate embeddings. Different ways to sample walks are proposed, such as random walk [173], a mixture of depths, and breath-first search walks [89].

Heterogeneous networks. In heterogeneous graphs, such as knowledge graphs (KGs), [62] propose meta-path-based sampling to circumvent the problem of bias towards dominant types of nodes. The relation between two node types in a KG is very important. Various KG embeddings have been proposed with varying similarity metrics of head, tail and relation embeddings (e.g., distance) [24, 168, 202, 209, 239].

Formulation 6: Shallow network embedding

Shallow network embedding generates a mapping that preserves the similarity in the network [105]. Formally, typical shallow network embeddings are learned in the following three steps:

1. **Mapping to an embedding space.** Given a pair of nodes u, v in network G , we obtain a function f to map these nodes to an embedding space to generate \mathbf{h}_u and \mathbf{h}_v .
2. **Defining network similarity.** We next define the network similarity as $f_n(u, v)$, and the embedding similarity as $f_z(\mathbf{h}_u, \mathbf{h}_v)$.
3. **Computing loss.** Then, we define the loss $\mathcal{L}(f_n(u, v), f_z(\mathbf{h}_u, \mathbf{h}_v))$, which measures whether the embedding preserves the distance in the original networks. Finally, we apply an optimization procedure to minimize the loss $\mathcal{L}(f_n(u, v), f_z(\mathbf{h}_u, \mathbf{h}_v))$.

2.3.6 Graph neural networks

Graph neural networks (GNNs) are a type of neural network that models graphs through a series of local message aggregation and propagation steps (Figure 3g). They can return vector representations of graph components that capture both the graph network structure and the attributes without any feature engineering.

Deep network representation learning. [58, 66, 216] laid the foundational works in applying operations on graph structured data, and GNN was first popularized by [125] with the formulation of Graph Convolutional Networks (GCN). Since then, numerous variants to improve the representation have emerged. For instance, GAT [213] assigns an attention for each node during aggregation. Recent works apply a transformer self-attention mechanism [212] to GNNs [46, 108, 238, 250] and show improved representation learning. Many works have aimed to improve GNNs' ability to capture graph structural information. For example, JK-Net [236] includes skip connections, and MixHop [1] considers a higher-order adjacency matrix to capture higher-order local structures. Graph pooling techniques, such as DiffPool [245], learn the topological structures of a graph.

Applying GNNs on molecular graphs have shown promising results. Special GNNs designed for molecules, such as DimeNet [127] and SchNet [191], inject chemical knowledge and domain assumptions, which improves results. These advances are also accompanied by rich theoretical analyses to understand GNNs, such as connecting their expressive powers to the Weisfeiler-Lehman test [235], label propagation [219], and universal invariance [79, 121].

GNN extensions to large, dynamic, and heterogeneous graphs. As real-world networks are enormous, various sampling strategies to improve the scalability of GNNs are proposed [44, 251]. Standard GNNs consider only homogenous graphs. However, many biomedical networks, such as biomedical knowledge graphs, are heterogeneous. [108, 189, 223] have designed new aggregation mechanisms to consider the heterogeneity of various node and relation types in realistic networks. Standard GNNs operate on static graphs, and do not work with dynamic graphs. Recently, [108, 171, 185] propose various forms of dynamic update mechanisms that can take in a sequence of graphs.

Graph transfer learning. As labels for biomedical problems are scarce, applying transfer learning to enable fast adaptation from a large pretrained GNN model is important. [107] designs strategies to learn improved node representations through unsupervised pretraining, such as local graph structure context prediction. Self-supervised approaches [247] are also designed to leverage unsupervised network statistics prediction to facilitate downstream task prediction.

Explainability in graph neural networks. GNNs are neural networks that are parameterized by a huge number of parameters. Thus, they are not interpretable out-of-the-box. GNNExplainer [244] uses an information theoretical approach to identify relevant subgraphs and their attributes for graph-level prediction.

Formulation 7: Graph neural networks

GNNs learn compact representations or embeddings that capture network structure and node features [99, 111, 233]. A GNN generates outputs through a series of propagation layers [85], where propagation at layer l consists of the following three steps:

1. **Neural message passing.** *The GNN computes a message $\mathbf{m}_{u,v}^{(l)} = \text{MSG}(\mathbf{h}_u^{(l-1)}, \mathbf{h}_v^{(l-1)})$ for every linked nodes u, v based on their embeddings from the previous layer $\mathbf{h}_u^{(l-1)}$ and $\mathbf{h}_v^{(l-1)}$.*
2. **Neighborhood aggregation.** *The messages between node u and its neighbors \mathcal{N}_u are aggregated as $\hat{\mathbf{m}}_u^{(l)} = \text{AGG}(\mathbf{m}_{uv}^{(l)} | v \in \mathcal{N}_u)$.*
3. **Update.** *The GNN applies a non-linear function to update node embeddings as $\mathbf{h}_u^{(l)} = \text{UPD}(\hat{\mathbf{m}}_u^{(l)}, \mathbf{h}_u^{(l-1)})$ using the aggregated message and the embedding from the previous layer.*

2.3.7 Generative models

Generative modeling aims to generate a novel graph with the desired property (Figure 3h).

Network science approaches. Traditionally, many network science models generate new graphs based on a process in which new nodes or edges are added or deleted following some deterministic or probabilistic rules. For instance, the Erdős-Rényi model [69] (e.g., random graph) adds edges with a fixed probability; the Barabási-Albert model [7] adds edges based on the node degree to reflect the real-world power law distribution; to generalize, the configuration model [15] adds edges based on a predefined sequence of degrees to allow arbitrary degree distributions.

Deep generative models. While powerful, previous works are restricted to generalize over arbitrary graphs, and cannot optimize based on the defined properties of the graph. Deep generative models can tackle the challenge present in network science approaches by automatically learning from a set of graphs. They first learn a latent distribution $P(Z|\mathcal{G})$ that characterizes the input graph set \mathcal{G} . Then, conditioned on this distribution, they can decode to generate a new graph \hat{G} . There are different ways to encode and learn the latent distribution, such as through variational autoencoders [87, 117, 124] or generative adversarial networks [220]. In addition to using a straightforward neural network, we can apply other methods to decode from the latent representation. For example, auto-regressive decoders generate edges sequentially conditioned on the edges just generated [142, 246].

Formulation 8: Generative modeling

Generative models optimize and learn data distributions in order to generate graphs with desirable properties. In the following, we use VGAE [124] as an example.

Example. VGAE is a graph extension of variational autoencoders [123]. Given a graph G with adjacency matrix \mathbf{A} and node features \mathbf{X} , we use two GNNs to encode the graph and generate the latent mean vector $\mu_{\mathbf{Z}}^u$ and the log-variance $\log(\sigma_{\mathbf{Z}}^u)$ parameters for each node u . Formally, $\mu_{\mathbf{Z}} = \text{GNN}_{\mu}(\mathbf{A}, \mathbf{X}) \in \mathbb{R}^{|\mathcal{V}| \times d}$, and $\log(\sigma_{\mathbf{Z}}) = \text{GNN}_{\sigma}(\mathbf{A}, \mathbf{X}) \in \mathbb{R}^{|\mathcal{V}| \times d}$. We can then obtain the latent distribution $q(\mathbf{Z}|G) \sim \mathcal{N}(\mu_{\mathbf{Z}}, (\log(\sigma_{\mathbf{Z}})))$, where we can draw a latent embedding sample $\mathbf{Z} \in \mathbb{R}^{|\mathcal{V}| \times d}$ from the latent distribution. Then, given the latent embedding, we feed into a probabilistic decoder $p_{\theta}(\hat{\mathbf{A}}|\mathbf{Z})$ to generate a network \hat{G} with adjacency matrix $\hat{\mathbf{A}}$. In VGAE, the decoder is a dot product, i.e., $p_{\theta}(\hat{\mathbf{A}}_{uv} = 1|\mathbf{Z}) = \text{sigmoid}(\mathbf{z}_u^T \mathbf{z}_v)$. This way, we obtain a newly generated graph \hat{G} . Finally, we optimize the weight using the variational reconstruction loss $\mathcal{L} = \mathbb{E}_{q(\mathbf{Z}|G)}[p(G|\mathbf{Z})] - \text{KL}(q(\mathbf{Z}|G)||p(\mathbf{Z}))$.

3 Application areas in biology and medicine

Biomedical data involve rich multimodal and heterogeneous interactions that span from the molecular scale to whole ecosystems—from the molecular level to the level of connections between diseases in a person, all the way to the societal level encompassing human interactions (Figure 4). These interactions at different levels give rise to a bewildering degree of complexity which is only likely to be fully understood through a holistic and integrated systems view and the study of combined, multi-level networks. In this review, we focus on the following areas of biology and medicine through which graph representation learning has permeated: molecules, genomics, therapeutics, and healthcare systems.

Molecular-level applications (Section 4). Among the most commonly modeled bioentities are proteins. Molecular structure can be translated from atoms and bonds into nodes and edges, respectively. Protein interactions naturally form a network based on the existence of a physical interaction or functional relationship. For instance, a protein can bind to, upregulate, or downregulate another protein in a biological pathway. Further, with the ability to learn molecular representations of proteins and their physical interactions, graph ML methods are well-suited to predict protein function.

Genomic-level applications (Section 5). Genetic elements are frequently incorporated into biomedical networks. The interaction of genes can be extracted from transcriptomic data to improve predictions on disease classification. In addition to coding genes, noncoding RNA elements have been used to construct more comprehensive molecular association networks. As a result, with the help of representation learning, disease predictions based on genome-wide interactions are possible.

Therapeutics-level applications (Section 6). Zooming out from the genomics level to that of small compounds, biomedical networks can be composed of drugs (drug-drug interaction network), proteins (drug-protein interaction network), and diseases (drug-protein-disease interaction network). As a result, graph representation learning models are able to leverage these networks to predict drug-drug, drug-target, and drug-disease associations. For instance, leveraging affinity information between drugs and proteins can improve predictions of candidate drugs for treating a disease, potential off-target effects, etc.

Healthcare-level applications (Section 7). Patient records, such as medical images and electronic health records (EHRs), can be represented as networks. For instance, cell spatial graphs are well-suited to capture the structural information of the nuclei in histopathology slides. Given the advances in predicting protein features, genome-wide associations, and therapeutics, graph representation learning methods have been successful in integrating patient records with molecular (including proteins and small compounds), genomic, and disease networks to make personalized predictions for patients.

In the following sections, we survey representation learning techniques for each of the above areas (Sections 4-7) and provide examples of possible future studies enabled by these techniques (Applications 1-4).

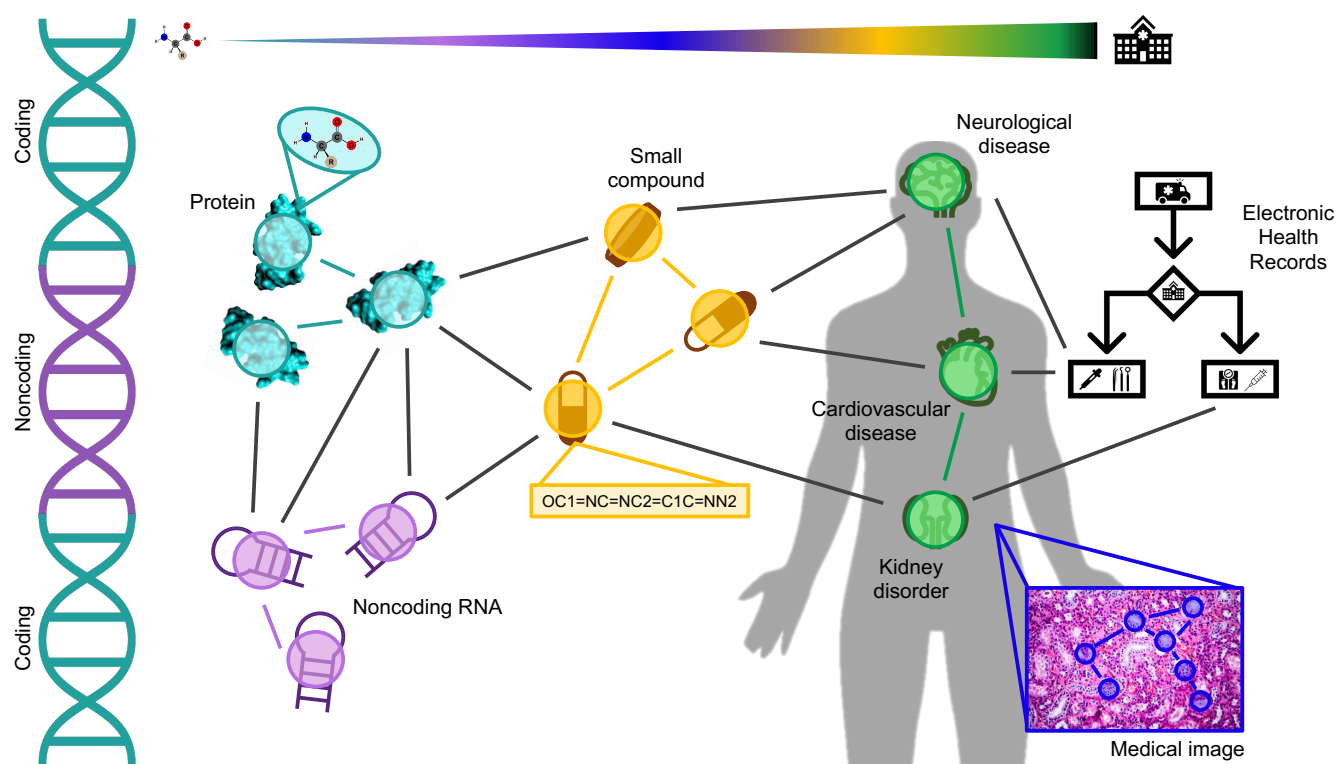


Figure 4: **Overview of biomedical applications areas.** Networks are prevalent across biomedical areas, from the molecular level to the healthcare systems level. Protein structures and interactions can be represented as molecular level networks (Section 4). Coding and noncoding regions of the genome can be modeled by protein, miRNA, and lncRNA networks to depict genome-wide associations (Section 5). Small compounds' structure, interactions with protein targets and noncoding RNAs, and indications can be represented by therapeutics level networks (Section 6). Patient-specific data, such as medical images and electronic health records, can be integrated into a cross-domain knowledge graph of proteins, noncoding RNA, drugs, diseases, and more to enable healthcare systems level applications (Section 7).

Area	Data Type	Example Data Source(s)
Molecules (Section 4)	Protein structure	ProteinNet [8]
	Physical interactions	BioGRID [197], HuRI [148], STRING [204]
	Functional interactions	STRING [204]
	Protein function	Gene Ontology [32], ProteomeHD [131], STRING [204]
Genomics (Section 5)	Gene expression data	Connectivity Map [133], Monarch Initiative [164]
	Disease mechanisms	Disease Maps [157], HENA [199]
	Noncod. RNA interactions	iMDA-BN [258], Molecular Association Network [92]
	Disease ontologies	Disease Ontology [190], HetNet [101]
	Genome-wide interactions	GIANT [88], HetNet [101], Molecular Association Network [92]
Therapeutics (Section 6)	Small compounds structure	Therapeutics Data Commons (TDC) [110], Drug Repur. Hub [56]
	Drug-target interactions	DrugBank [228], Comp. Toxicogenom. DB (CTD) [56], TDC [110]
	Drug-drug interactions	DrugCombDB [145], TDC [110]
	Drug-disease associations	CTD [51], Drug Repurposing Hub [56], TDC [110]
Healthcare (Section 7)	Medical images	The Cancer Imaging Archive [48], Osteoarthritis Initiative [167]
	Clinical knowledge graphs	SPOKE [166], Pubmed Knowledge Graph [234]

Table 1: **Biomedical networks and databases for each application area.** From molecules to patients, existing data types are listed alongside example commonly used data sources.

4 Graph representation learning for molecules

We begin our deep dive into the applications of graph representation learning at the molecular level. Representation learning methods for molecules have been deviating from implementing network-based metrics and random walk algorithms towards building graph machine learning methods [120]. Prior non-graph-based machine learning approaches to predict protein-protein interactions and protein function include sequence-based deep learning methods [241]. However, these do not take into account structural information in the protein-protein interaction network, such as degrees, neighborhood, etc [241]. As a result, due to their ability to capture structural information (i.e. by neighborhood aggregation), graph representation learning methods have proven to be quite successful in modeling protein interactions and predicting protein function [74, 98, 120]. Specifically, graph convolution neural networks’ inductiveness (Section 2.3.6) and graph auto-encoders’ generative ability (Section 2.3.7) enable the discovery of new molecular structure, interaction, and even function [65, 98, 241].

Generating protein structure. Computationally elucidating protein structure has been an ongoing challenge for decades [55]. Experimental methods have been created and new metrics have been defined to predict protein structure based on amino acid sequences, physical and chemical principles, etc [55]. Graph representation learning has played a role in leveraging such information to generate molecular structures. Here, our focus is on studies aimed to better model interactions between proteins and even with small compounds.

Characterizing protein interactions. Because characterizing protein interactions requires time-consuming and expensive experiments, domain experts are motivated to develop machine learning models to create a more complete PPI network. Current PPI networks are built using methods ranging from experimentally intensive processes, i.e. yeast two-hybrid screens, to computationally inferred methods, i.e. natural language processing of published articles [184, 187]. Experimental methods result in the most accurate measurements of protein interaction. However, due to the experimental constraints of the method, the most updated PPI is still

limited in its number of nodes (proteins) and edges (physical interactions) [148]. Computational methods that scrape publications for evidence of protein interactions are much more comprehensive, but quite noisy. As a result, graph representation learning methods, which we review next, have been shown—using various data modalities, including chemical structure, binding affinities, and amino acid sequences—to improve protein interaction quantification by measuring the likelihood of an edge between any two given proteins.

Predicting protein phenotypes. Identifying the roles and functions of protein specific biological contexts is a challenging and experimentally intensive task [162, 264]. Applying and innovating graph representation learning techniques to integrate gene ontology terms and gene expression data into protein interaction networks have demonstrated success in understanding multicellular functions [77, 88, 225, 263].

4.1 Generating protein molecular graphs

Here, we briefly discuss the impact that graph representation learning has had on modeling protein structure and interactions.

Modeling 3D protein structure. As protein structures are folded into complex 3D structures, it is natural to model them in graphs. For example, many such as [109] construct a contact distance graph where nodes are individual residues and edges are constructed by a physical distance threshold. [111] uses a modified GAT (Section 2.3.6) for protein design, aiming to generate good primary sequences from the given 3D structures.

Leveraging structure to predict interactions. iScore [84] develops a graph kernel metric (Section 2.3.1) on the intermolecular protein interface graph to predict protein-protein interaction. [76] applies graph convolutions (Section 2.3.6) to ligand-protein graphs and protein receptor graphs to perform protein interface prediction. MASIF [80] uses multiple geodesic convolutional layers to generate fingerprints for 3D protein molecular surfaces patches, which have been shown to improve performance in protein pocket-ligand prediction and protein-protein interaction site prediction. [114] uses GNN layers (e.g., GCN or GAT) to model the contact graph generated by a predictor and predict drug-target interactions.

4.2 Quantifying protein interactions

In addition to molecular maps, protein sequences and interaction networks constructed from *in vitro* experimental results enable accurate prediction of interacting proteins using graph representation learning.

Utilizing protein sequences and structures. In addition to protein structure, many methods utilize amino acid sequences to predict protein-protein interactions. S-VGAE [241] (Section 2.3.7) first learns protein embeddings based on their sequences using GCN, and then concatenates a pair of proteins' embeddings to predict whether there exists an interaction. EGCN [31] applies a multi-head attention mechanism (Section 2.3.6) to predict intra- and inter-molecular energies, binding affinities, and quality measures for a pair of molecular complexes.

Leveraging protein interaction networks. Topology-based methods for physical interaction networks, e.g., PPI, are able to capture and leverage the dynamics of biological systems [139]. [147] applies GNN layers (Section 2.3.6) to aggregate structural information in the graphs of ligand and receptor proteins, uses sequence modeling to restore the sequential information in their amino acid sequences, and concatenates the resulting matrices as input for CNN layers to predict whether there is an interaction. [243] applies GCN to remove less credible protein-protein interactions in order to construct a more reliable PPI network.

4.3 Interpreting protein functions and cellular phenotypes

The ability to represent protein structures and interactions facilitates downstream protein function prediction. This can be done by leveraging existing gene ontologies and transcriptomic data, or even deriving it from sequence and structural information via attention mechanisms.

Integrating hierarchical gene ontology. Gene Ontology (GO) terms [50] are commonly used as labels for protein function because they are a standardized vocabulary for describing molecular functions, biological processes, and cellular locations of gene products [259]. As such, [259] applies GCN (Section 2.3.6) to a hierarchical graph built from GO terms. To leverage physical interaction networks, Graph2GO [73] integrates protein features into a PPI network, and applies VGAE (Section 2.3.7) and GCN to predict protein functions on gene ontology (GO). [65] demonstrates the value of generating gene interaction networks based on transcriptomic data. So, after building a graph based on DNA sequence similarity, where the node features are composed of miRNA, gene, and protein interactions and gene expression profiles, Pseudo2GO [74] applies GCN to compute probability scores for GO terms. [100] introduces a GCN framework to learn protein representations from transcription and PPI networks in order to predict gene expression.

Deriving function from sequences and structures. Alternative graph representation learning methods for predicting protein function include TDA and Transformers. [30] defines a novel diffusion-based distance metric to enable protein function prediction in protein interaction networks (Section 2.3.2). [60] uses the theory of topological persistence (Section 2.3.3) to compute signatures able to discriminate structural information for classifying protein function. [156] applies TDA to extract features from protein contact networks created from 3D coordinates. [165] adapts Bidirectional Encoder Representations from Transformers (BERT) to embed protein sequences and classify protein families. Similarly, [215] applies an attention mechanism to the BERT model in order to enable interpretability to the protein sequences.

Application 1: Cell-type aware protein representation learning

Motivation. *Gene expression is not homogeneous across cells. Single cell transcriptomic data is able to capture the heterogeneity of gene expression across a collection of cells [128]. By leveraging cell type specific expression information as labels, we can use graph representation learning to generate protein embeddings that are cell-type aware.*

Network Representation. *We can use any form of protein interaction network. Examples include PPI networks in which edges represent a physical interaction or co-expression.*

Learning Task. *With the protein interaction network, we can perform multilabel node classification to predict whether a protein's corresponding gene is differentially expressed in a cell type based on single cell RNA sequencing (scRNA-seq) experiments (Figure 5). If there are N cell types identified in scRNA-seq analysis for a given experiment, each protein is assigned a vector of length N . Each element of the vector corresponds to a cell type, where 1 indicates that the protein/gene is differentially expressed in that cell type and 0 otherwise.*

Impact. *By generating protein embeddings that consider their differential expression at the cell type level, we can begin to make predictions at a single cell resolution. We can consider factors including disease and cell states, and temporal and spatial dependencies. Implications of such cell-type aware protein embeddings extend to cellular function prediction and cell-type-specific disease classification.*

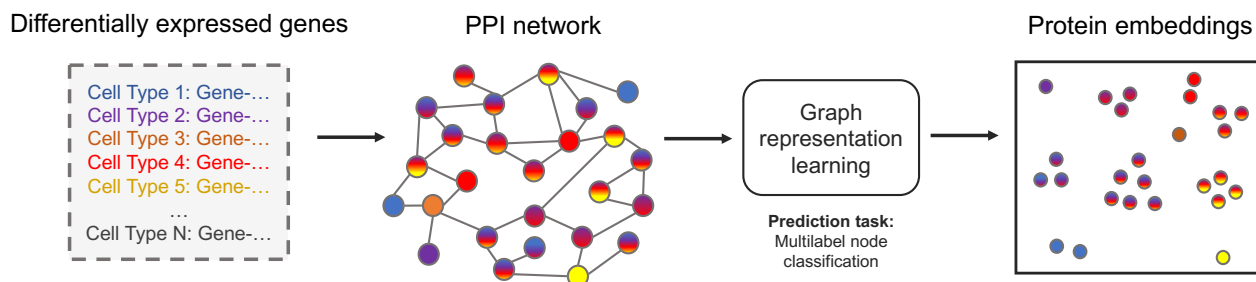


Figure 5: **Cell-type aware protein representation learning.** Given differentially expressed genes generated from scRNA-seq data, we perform multilabel node classification on a PPI network. The resulting protein embeddings should cluster such that cell type specificity is reflected.

5 Graph representation learning for genomics

We next zoom out from modeling protein structure and interactions to capturing the relationships between genome-wide elements for disease diagnosis. Diseases are classified based on the presenting symptoms of patients, which are often caused by internal dysfunctions, such as genetic mutations. As a result, diagnosing diseases requires knowledge about alterations in the transcription of coding genes and the interactions between non-coding genes to capture genome-wide associations driving disease acquisition and progression [118, 196, 252]. Advances in graph representation learning methods to analyze heterogeneous networks of multimodal data allow us to integrate and make predictions on data across domains, from genomic level data (e.g., gene expression and copy number information) to clinically relevant data (e.g., pathophysiology and tissue localization).

Integrating gene expression data. Comparing transcriptomic profiles from healthy individuals to those of patients with a specific disease can inform clinicians of its causal gene(s). By injecting gene expression data into PPI networks, we are able to leverage information regarding differential expression of genes-of-interest to (1) identify markers that are disease-specific and (2) utilize such markers to more accurately classify diseases of interest.

Leveraging non-coding genetic elements. Coding genes are not the only bioentities driving disease progression. In fact, they only account for approximately 2% of the genome [70]. Non-coding genes, e.g., long non-coding RNA and microRNA, play a significant role as well [12]. Experimental and computational methods are still being developed to discover ways in which non-coding RNA fundamentally contribute to human disease. As such, interaction networks of non-coding RNAs modeling crosstalk between the two types of RNA have accelerated the discovery of their associations with diseases [12, 196, 252].

Fusing genome-wide data. Ultimately, to improve disease classification models, we must consider the interactions of both coding and non-coding genes. Drawing genome-wide associations is key to capturing the complete genetic picture of a disease and a patient’s risk [118, 196]. To this end, graph representation learning methods have been successful in (1) integrating cross domain knowledge relevant to diseases and (2) learning their importance (weights) for predicting specific diseases.

5.1 Leveraging gene expression measurements

Gene expression is the direct readout of the effects of a perturbation. As such, we can model important interactions between genes based on changes in their expression.

Transforming co-expression matrices into graphs. Methods that rely solely on gene expression data typically transform the co-expression matrix into a more topologically meaningful form [99, 154, 169]. [169] transforms gene expression data into a colored graph that captures the shape of the data (Section 2.3.3), which enables downstream analysis using network science metrics and graph representation learning (Section 2.3.1). [99] combines matrix factorization with GCN to draw disease-gene associations, akin to a recommendation task (Section 2.3.6). [154] vectorizes the topological landscapes present in gene expression data (Section 2.3.3), and feeds them into a GCN to classify the disease type. [240] initiates a gene correlation network using a subset of gene expression matrices, and apply CONDGEN, a joint GCN, VAE, and GAN framework, to generate the desired final graph (Section 2.3.7).

Enhancing existing networks with gene expression data. Because gene expression data can be noisy and variational, recent advances include fusing the co-expression matrices with existing biomedical networks, such as Gene Ontology (GO) annotations and PPI, and feeding the resulting graph into graph convolution layers (Section 2.3.6) [43, 54, 180]. [54] incorporates associations between gene ontology terms and differentially expressed genes in a given sample to create a factor graph neural network, resulting in a relatively transparent and interpretable disease classification model (Section 2.3.6). [43, 180] integrate gene expression data into PPI networks, which are then used as input for their GCN models. However, [180] additionally defines a relation network to prioritize the edges weighted by the graph convolution layers, representing the relevant gene sets for the classification task.

Limiting aspects of PPI networks. Alternative to fusing gene expression data with the PPI network, [178] compares the performance of GCN (Section 2.3.6) on four separate graphs: a co-expression graph generated from RNA sequencing data given a threshold for correlation between a pair of genes, a PPI network, a co-expression graph with singleton nodes, and a PPI network with singleton nodes. As a result, [178] reports that the model trained solely on the PPI network performs the worst compared to the other graphs, suggesting that the PPI network is insufficient for accurately classifying diseases. The PPI network does not capture all gene regulatory activities due to experimental constraints, and it lacks information about non-coding genes. To this end, we need to model interactions between non-coding genes as well as coding genes.

5.2 Incorporating non-coding RNA interactions

As non-coding regions of a genome play a significant role in disease progression, it is important to model their interactions in a network along with coding genes.

Characterizing non-coding interactions. Predicting associations between non-coding RNA and diseases has relied heavily on heterogeneous graphs modeling interactions between lncRNAs, miRNAs, and diseases [152, 237, 253]. [253] vectorizes the lncRNA, miRNA, and disease nodes using DeepWalk (Section 2.3.5). [152] generates embeddings of lncRNAs and diseases using node2vec[89] (Section 2.3.5), and feeds them into their deep belief network. [231, 237] adopt a standard GCN framework, and [248] combines inductive matrix completion and GCNs to address the issue of incompleteness in experimentally-derived datasets (Section 2.3.6).

Classifying diseases across the genome. Incorporating both coding and non-coding gene associations can elucidate genome-wide interactions. [257] applies GCN on a gene interaction network to prioritize target coding genes for lncRNAs (Section 2.3.6). [82] integrates gene expression, copy number alteration, and clinical data in a unified GNN framework to make predictions on short- and long-term survival for a cancer patient. [94] constructs a molecular association network comprised of mRNA, protein, miRNA, lncRNA, circRNA, drug, disease, and microbe interactions, and applies DeepWalk (Section 2.3.5) to generate node embeddings of such bioentities for downstream predictions. Clearly, considering genome-wide interactions broadens the

range of biomedical questions that can be answered as well, including predictions regarding drug-disease and drug-target associations and even patient-centric prognoses.

Application 2: Disease classification using subgraphs

Motivation. Phenotypes are observable characteristics resulting from interactions between genotypes, as well as environment. Physicians utilize a standardized vocabulary of phenotypes to describe human diseases. By modeling diseases or disorders as collections of associated phenotypes, we can diagnose patients based on their presenting symptoms.

Network Representation. Consider a graph $G = (V, E)$ built from the standardized vocabulary of phenotypes, such as the Human Phenotype Ontology (HPO). HPO forms a directed acyclic graph with nodes $v \in V$ representing phenotypes and edges $e \in E$ indicating a hierarchical relationship between the pair of phenotypes [182]. We can represent a disease or disorder as a set of nodes, formally defined as a subgraph $S = (V', E')$ where $V' \subseteq V$ and $E' \subseteq E$, in the HPO graph.

Learning Task. With subgraphs representing diseases or disorders, we can train subgraph embeddings and predict the disease or disorder most consistent with the set of phenotypes that the embedding represents [9] (Figure 6).

Impact. Modeling a disease or disorder as a subgraph enables a more flexible representation of diseases than relying on individual nodes or edges. As a result, we can better resolve complex phenotypic relationships to improve differentiation of diseases or disorders.

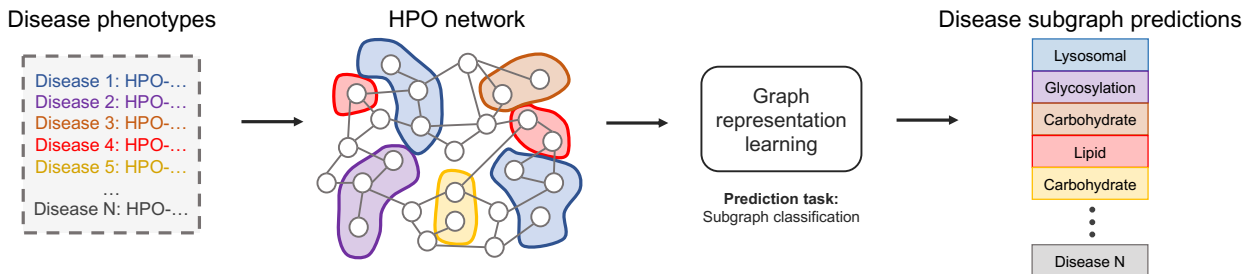


Figure 6: **Disease classification using subgraphs.** Given a list of phenotypes associated with a disease and a network of the Human Phenotype Ontology (HPO), we feed into our graph representation learning model the disease-specific characteristics as subgraphs and the underlying network. The model then classifies each subgraph by its disease type.

6 Graph representation learning for therapeutics

After exploring the role of networks in interrogating the underlying genetic mechanisms of disease, we now discuss ways in which graph representation learning has accelerated the developmental pipeline for therapeutics. Modern drug discovery requires elucidating a candidate drug’s chemical structure, identifying its drug targets, quantifying its efficacy and toxicity, and detecting its potential side effects [16, 90, 106, 176]. Because such processes are costly and time-consuming, *in silico* approaches have been adopted into the drug discovery pipeline. However, cross-domain expertise is necessary to develop a drug with the optimal binding affinity and specificity to biomarkers, maximal therapy efficacy, and minimal adverse effects. As a result, it is critical to integrate chemical structure information, protein interactions, and clinical relevant data (i.e.

indications and reported side effects) into predictive models for drug discovery and repurposing. Graph representation learning has been successful in characterizing drugs at the systems level without patient data to make predictions about interactions with other drugs, protein targets, side effects, diseases, etc.

Generating drug structure. Similar to proteins, small compounds can easily be modeled as 2D and 3D molecular graphs. In the simplest 2D case, nodes and edges of a graph can represent atoms and bonds of a small compound’s chemical structure, respectively. However, this does not consider the 3D structure of the molecule. So, we can include node and edge features like chirality and bond length, respectively [55]. As such, graph representation learning methods are well-suited to generate small compound molecular graphs.

Characterizing drug interactions. Corresponding to molecular structure is binding affinity and specificity to biomarkers. Such measurements are important for ensuring that a drug is effective in treating its intended disease, and does not have significant off-target effects [227]. However, quantifying these metrics requires labor- and cost-intensive experiments [55, 227]. Graph representation learning has been widely used to determine the existence, category, and extent of interaction between a given drug and protein target.

Quantifying drug efficacy. Part of the drug discovery pipeline is minimizing adverse drug events [55, 227]. But, in addition to high financial cost, the experiments required to measure drug-drug interactions and toxicity face a combinatorial explosion problem [55]. Hence, graph representation learning methods have played a major role in prioritizing candidate drugs that have little to no side effects with existing drugs on the market. In the same vein, graph representation learning has been shown to be effective in ranking candidate drugs for repurposing by considering gene expression data, gene ontologies, drug similarity, and other clinically relevant data regarding side effects and indications.

6.1 Modeling compound molecular graphs

Graph representation learning has improved the characterization of a drug’s structure and its potential targets, which is a key step in the standard therapeutics development pipeline. We first discuss representation learning and generative modeling methods to capture drugs’ chemical structure.

Integrating chemical properties of drugs. A drug compound is naturally represented as a 2D molecular graph, where nodes are atoms and edges are bonds. It can also be modeled in 3D where edges contain weights about the physical distance between the atoms. Recently, GNN-based computational modeling of molecules has shown significant performance gains by obtaining a stronger representation of the molecular structure. For example, enn-s2s [85], SchNet [191], DimeNet [127] integrate chemical knowledge into the GNN architecture design (Section 2.3.6), showing that their methods improve predictions on various quantum chemistry properties.

Generating molecules and characterizing interactions. Modeling the chemical graph *in silico* enables numerous downstream tasks, such as predicting molecule compounds’ drug-likeness properties, estimating a compound’s binding affinity to its target protein, and generating reaction outcomes. JT-VAE [117] applies a GNN to generate latent representations of graphs, and uses a junction tree based encoder-decoder approach to generate a new molecule that has desirable properties (Section 2.3.7). [49] builds a GNN to model the interactions among reactants and predict the organic reaction outcomes (Section 2.3.6).

6.2 Quantifying drug-drug and drug-target interactions

Next, we demonstrate the role of graph representation learning in precisely measuring the interactions between a drug with candidate targets and other drugs.

Modeling drug and target interactions. Graph representation learning methods, such as TDA (Section 2.3.3) and node2vec [89], have been used primarily to learn latent representations of drugs and targets. [6] adopts TDA to generate a graph where nodes represent compounds and edges indicate a level of similarity (Section 2.3.3). [206] generates embeddings for drugs and targets using node2vec to compute drug-drug, drug-target, and target-target similarities (Section 2.3.5), and denoise the original drug-target interaction network to be fed into downstream machine learning models.

Fusing compound sequence, structure, and clinical implications. More recently, graph neural networks are being used to enable more accurate drug-drug and drug-target interaction predictions. [150] applies GCN with an attention mechanism on drug graphs (Section 2.3.6), with chemical structures and side effects as features, to predict drug-drug interactions. [261] uses GCN to predict side effects on a drug-protein network. [114] learns representations of protein and small molecule graphs in two separate GNNs in order to predict drug-target affinity. Additionally, several methods leverage amino acid sequence information to build more robust predictive models. [38, 175, 210] apply GCN to learn protein structure representations, which are then combined with protein sequence representations generated by word2vec or CNNs, to predict the probability of compound-protein interaction. [143] follows a similar framework, but applies GAT instead of GCN to learn protein target representations.

6.3 Identifying drug-disease associations and biomarkers for complex disease

Ultimately, we can leverage our knowledge of diseases and drugs in our graph representation learning models to more accurately predict the efficacy of a drug for treating a disease.

Constructing drug-disease-protein networks from medical ontology. Drug and disease representations can be learned on homogeneous graphs of drugs, diseases, or targets. [93] constructs a drug-disease graph using Medical Subject Headings (MeSH) terms, and learns latent representations of drugs and diseases using various graph embedding algorithms, including DeepWalk and LINE (Section 2.3.5). [221] applies TDA (Section 2.3.3) to construct graphs of drugs, targets, and diseases separately to learn representations of such entities for downstream prediction.

Integrating multimodal drug- and disease-related datasets. To emphasize the systems-level complexity of diseases, recent methods fuse multimodal data to generate a heterogeneous graph, enabling methods to potentially elucidate drug action. [217] aggregates neighborhood information from heterogeneous networks comprised of drug, target, and disease information to predict drug-target interactions (Section 2.3.6). [233] is a GNN that combines PPI network with genomic features to predict drug sensitivity. [194] applies a graph attention propagation mechanism to predict the therapeutic efficacy of kinase inhibitors across tumors.

Application 3: Cell-line specific prediction of interacting drug pairs

Motivation. *Drug combinations are commonly used to treat complex diseases and co-morbidities. However, evaluating the interactions of candidate drug combinations is experimentally intensive and costly. Using graph representation learning, we can leverage perturbation experiments on cell lines, and predict cell-line specific drug combinations.*

Network Representation. *Consider a multimodal network with protein-protein, protein-drug, and drug-drug interactions $G = (V, E)$ where nodes $v \in V$ are proteins and drugs, and edges $e \in E$ indicate interactions between protein pairs, drug pairs, or protein-drug pairs. Given a score s_i indicating an interaction between a drug pair d_i tested in cell line c , we create new $G'_c = (V'_c, E'_c)$ such that edges are retained if s_i is above a*

certain threshold [115].

Learning Task #1: Single Cell Line Network. From a single cell line’s drug-protein network G_c , we can determine whether a drug pair is interacting in that specific cell line through edge regression (Figure 7).

Learning Task #2: Collection of Cell Line Networks. We can apply transfer learning to Learning Task #1 to train a model that can generate predictions of drug interactions across many cell types. Intuitively, we develop a model using one cell line’s drug-protein network G_c , “reuse” the model on the next cell line’s drug-protein network, and repeat until we have trained on all cell line specific drug-protein networks. The utility of transfer learning here is to leverage the knowledge gained from one network to accelerate the training and improve the accuracy across all networks (Figure 8).

Impact. Most predictive models for drug combinations do not consider tissue or cell-line specificity of drugs. Because drugs’ effects on the body are not uniform, it is crucial to account for such anatomical differences. Further, the ability to prioritize candidate drug combinations in silico could reduce the cost of developing and testing them experimentally, thereby enabling more robust evaluation of the high potential candidates.

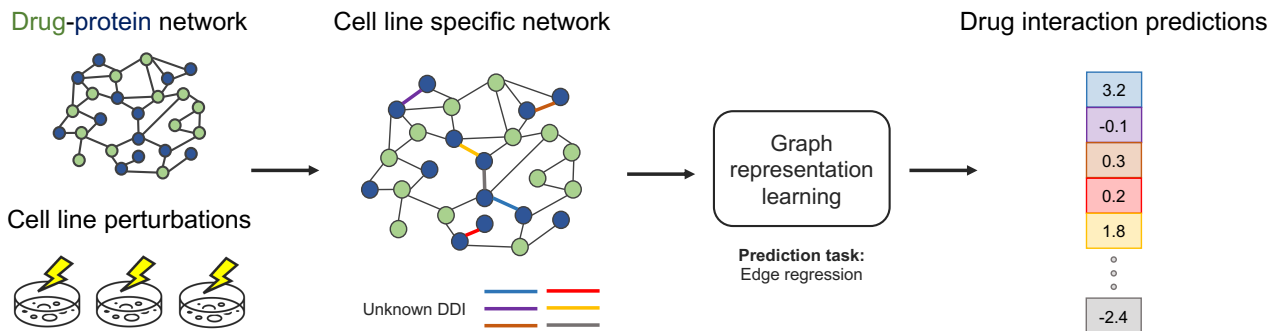


Figure 7: **Cell-line specific prediction of interacting drug pairs.** Given a drug-protein network and cell line drug perturbations data, we construct a cell line specific PPI network based on interaction scores computed from the cell line drug perturbations. Next, we feed the network into our graph representation learning model, which predicts drug-drug interactions (DDI) in the cell line specific network.

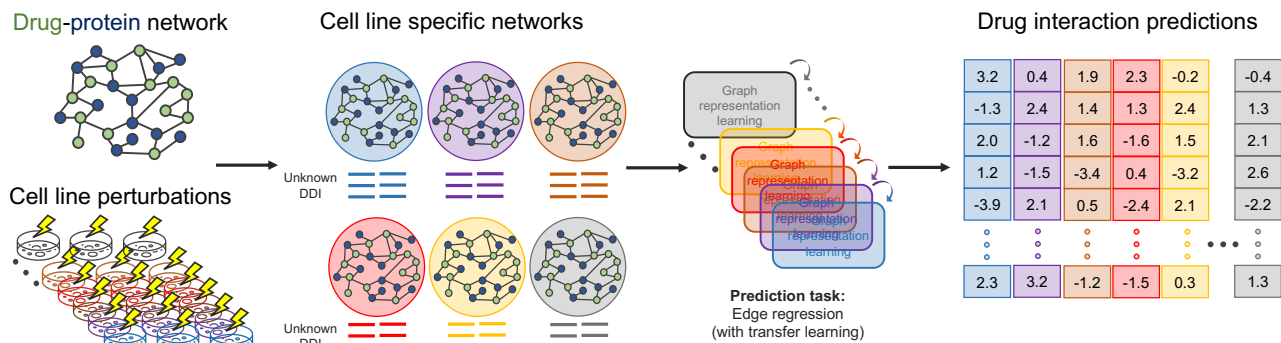


Figure 8: **Multimodal cell-line specific prediction of interacting drug pairs.** Given a drug-protein network and cell line drug perturbations data, we construct cell line specific PPI networks based on interaction scores computed from the cell line drug perturbations. Next, we feed the networks into our graph representation learning models (with transfer learning), which predict drug-drug interactions (DDI) in each cell line specific network.

7 Graph representation learning for healthcare systems

Beyond generating predictions of drug discovery and drug-disease associations at the systems level, graph representation learning has been used to fuse multimodal biomedical knowledge with patient records to better enable precision medicine. Two modes of patient data that have been especially successfully integrated into graph representation learning models and networks are histopathological images [18, 39] and EHRs [45, 141]. By representing these data modalities as networks, graph ML has improved diagnostic imaging and personalized medicine powered by EHRs.

Characterizing diseases through medical imaging. Medical images of patients, including histopathology slides, enable clinicians to more comprehensively observe the effects of a disease on the patient’s body [95]. Numerous deep learning tools exist for detecting subtle signs of disease progression in the images [61]. To enable graph representation learning, GNNs have been developed to convert medical images into cell spatial graphs, where nodes represent cells in the image and edges indicate that a pair of cells are adjacent in space. Moreover, recent methods have integrated other modalities, (e.g., tissue localization [172] and genomic features [39]), to improve the accuracy of predictions on medical images.

Integrating electronic health records for personalized medicine. Electronic health records are typically represented by ICD (International Classification of Disease) codes [45, 141]. The hierarchical information inherent to ICD codes (medical ontologies) naturally lend itself to creating a rich network of medical knowledge. In addition to ICD codes, medical knowledge can take the form of other data types, including presenting symptoms, molecular data, drug interactions, and side effects. As demonstrated in prior sections, graph representation learning is well-suited to heterogeneous networks (e.g., knowledge graphs) to make accurate predictions about diseases and drugs. By integrating patient records into our networks, we can use graph representation learning to advance precision medicine, generating predictions tailored to individual patients.

7.1 Leveraging networks for diagnostic imaging

By transforming images into networks, we can take advantage of graph representation learning to improve disease diagnostics.

Translating patient histopathological images into graphs. Cell-tissue graphs generated from histopathological images are able to encode the spatial context of cells and tissues for a given patient. [2, 11, 39, 260] apply GNNs to generate cell graphs, aggregating cell morphology and tissue micro-architecture information, for grading cancer histology images (Section 2.3.6). Additionally, [2] introduces an attention mechanism via pooling to infer relevant patches in the image. [172] is a hierarchical GNN that generates a cell-to-tissue graph representation capable of capturing cell morphology and interactions, tissue morphology and spatial distribution, cell-to-tissue hierarchies, and spatial distribution of cells with respect to tissues. Because interpretability is critical for models aimed to generate patient-specific predictions, CGEXPLAINER [112] performs post-hoc graph pruning optimization on a cell graph generated from a histopathology image to define a subgraph explaining the original cell graph analysis.

Constructing networks from CT and MRI images. Graph representation learning methods have also been proven successful for classifying other types of medical images. [35] applies a GNN to model relationships between lymph nodes to compute the spread of lymph node gross tumor volume based on radiotherapy CT images (Section 2.3.6). [10, 195, 226] convert MRI images into graphs and apply GCN to classify the progression of Alzheimer’s Disease. [137] uses TDA to generate graphs of whole-slide images, which include tissues from various patient sources (Section 2.3.3), and applies a GNN to classify the stage of colon cancer.

Integrating multimodal systems- and patient-level data. Finally, since multimodal data enables more robust predictions, [39] applies a GCN (Section 2.3.6) to generate cell spatial graphs from histopathology images and then fuses genomic and transcriptomic data to predict treatment response and resistance, histopathology grading, and patient survival.

7.2 Personalizing medical knowledge networks with patient records

Integrating patient-specific data into networks of well-established biomedical knowledge can improve precision medicine. Here, we discuss ways in which representing medical ontologies and temporal data has been effective in enabling more accurate diagnoses.

Leveraging hierarchical electronic health records. Methods that embed medical entities, including EHRs and medical ontologies, leverage the inherently hierarchical structure in the medical concepts KG [186]. ME2VEC [230] generates low dimensional embeddings of EHR data by separately considering medical services (via WORD2VEC), doctors (via GAT), and patients (via LINE) (Sections 2.3.5 and 2.3.6). GRAM [45], KAME [149], and [203] apply an attention mechanism on EHR data and medical ontologies to capture the parent-child relationships. Rather than assuming a certain structure in the EHRs, [46] is a Graph Convolution Transformer that learns the hidden EHR structure.

Modeling spatial and temporal dependencies. EHRs also have underlying spatial and/or temporal dependencies [37] that many methods have recently taken advantage of to perform time-dependent prediction tasks. [136, 146] combine an LSTM and GNN to represent patient status sequences and temporal medical event graphs, respectively, to predict future prescriptions or disease codes (Section 2.3.6). GNDP [141] designs a ST-GCN [238] to generate patient diagnoses to utilize the underlying spatial and temporal dependence of EHR data. MPVAA [47] is a mixed pooling multi-view self-attention autoencoder that generates patient representations in order to predict either a patient’s risk of developing a disease in a future visit, or the diagnostic codes of the next visit. [183] constructs a patient graph based on the similarity of patients, and applies an LSTM-GNN architecture to learn patient embeddings.

Integrating multimodal datasets. EHRs can be used with other modalities, such as diseases, symptoms,

molecular data, drug interactions, etc [37, 132, 166, 256]. [140] utilizes a probabilistic knowledge graph of EHR data, which can include medical history, drug prescriptions, and laboratory examination results, to consider the semantic relations between EHR entities (Section 2.3.5). [132] initializes node features for drugs and diseases using Skipgram, and applies GCN to predict adverse drug events. GAMENET [192] integrates drug and disease interactions with EHR data, and combines both RNNs and GCNs to recommend medication combinations. [105] exploits meta-paths in the EHR-derived KG to leverage higher order, semantically important relations for disease classification.

Application 4: Integration of health data into knowledge graphs to predict patient outcomes

Motivation. While rich biomedical networks have been constructed and utilized for advancing our understanding of proteins, genomics, diseases, and therapeutics, precision medicine is limited due to the lack of robust methods for integrating patient-specific information with basic science. Since EHRs can also be represented by networks, we can fuse patients' EHR networks with cross-domain biomedical networks, thus enabling graph representation learning to make predictions on patient-specific features.

Network Representation. Consider a knowledge graph for biomedical data $G = (V, E)$, where nodes $v \in V$ and edges $e \in E$ represent different kinds of bioentities and their various relationships, respectively. Examples of relations may include "up-/down-regulate," "treats," "binds," "encodes," and "localizes" [166]. To integrate patients into the network, we can create a meta-node that represents a patient, and add edges between the patient meta-node and its associated bioentity nodes.

Learning Task #1. We can learn node embeddings for each patient, as well as predict the probability of a patient with a disease, using edge regression (Figure 9).

Learning Task #2. We can learn node embeddings for each patient, as well as predict the probability of a drug effectively treating the patient, using edge regression (Figure 10).

Impact. Precision medicine requires an understanding of patient-specific data as well as the underlying biological mechanisms of diseases and drug action. Most biomedical networks do not consider patient data, which can prevent robust predictions of patients' conditions and potential responsiveness to drugs. The ability to integrate patient data with biomedical knowledge can address such issues.

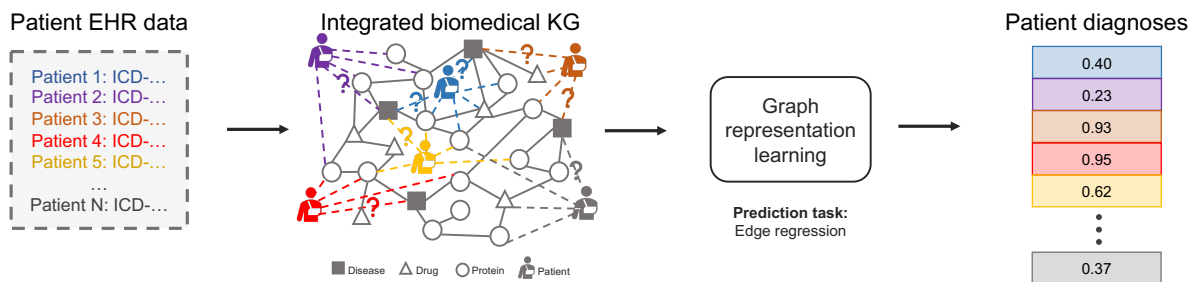


Figure 9: **Integration of health data into knowledge graphs to predict patient diagnoses.** Given a list of ICD codes associated with a patient, we create new "patient" nodes with edges connecting to the ICD codes in a multi-modal biomedical knowledge graph. We feed the resulting integrated biomedical KG into our graph representation learning model, which then predicts the probability of an edge between a patient and a disease.

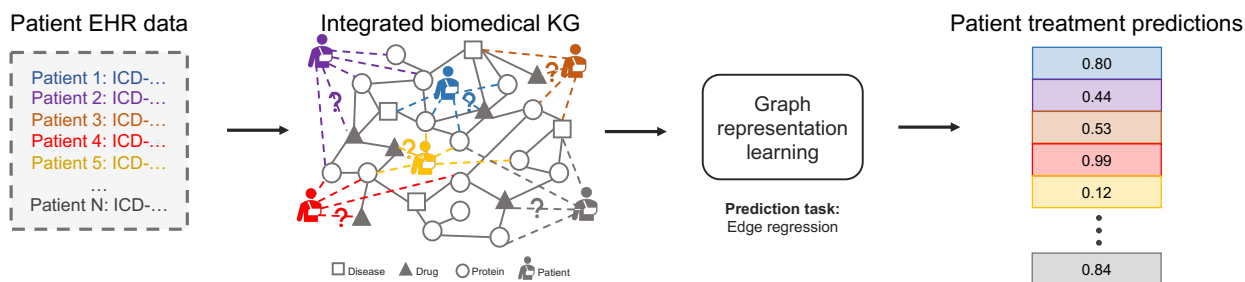


Figure 10: **Integration of health data into knowledge graphs to predict patient treatments.** Given a list of ICD codes associated with a patient, we create new “patient” nodes with edges connecting to the ICD codes in a multi-modal biomedical knowledge graph. We feed the resulting integrated biomedical KG into our graph representation learning model, which then predicts the probability of an edge between a patient and a drug.

8 Perspectives and Conclusions

Without question, multi-omics data integration has accelerated biological research by highlighting important, often even overlooked, relationships between bioentities involved in complex biological processes [198]. As demonstrated throughout this review, biomedical knowledge graphs enable more comprehensive and interpretable investigation of the underlying molecular mechanisms of a disease or drug. Given the utility of graphs in both the biological and biomedical domains, there has been a major push to further generate biomedical knowledge graphs that integrate multimodal data, from genotype-phenotype associations to population-scale epidemiological dynamics.

Identifying causal variants impacting complex traits. As graph representation learning has been applied quite extensively to aid in the mapping of genotypes to phenotypes, leveraging graph representation learning for fine-scale mapping of variants will be a promising new direction. Genome-wide association studies (GWAS) reveal relationships not only between a variant and a disease, but between variants or loci as well [72, 207]. By integrating networks from GWAS, expression Quantitative Trait Loci (eQTL) studies [211], and the human interactome, we can already begin to discover biologically meaningful modules to highlight key genes involved in the underlying mechanisms of a disease [222]. Additionally, because graphs can model long-range dependencies or interactions, we can model chromatin elements and the effects of their binding to regions across the genome as a network, allowing us to take advantage of graph representation learning [59, 134]. Thus, by re-imagining GWAS and eQTL studies as networks, we could utilize graph representation learning to improve predictions regarding the significance and impact of a potentially causal variant on a disease.

Classifying diseases at single-cell resolution. Single cell sequencing techniques’ ability to capture the heterogeneity of responses to perturbations aids in the discovery of complex genotype-phenotype relationships [13, 71]. Using manifold learning, we can begin to model the cellular state space as a graph, thereby enabling us to study the cellular differential process [27, 181]. Graph attention neural networks have also allowed us to predict disease state from scRNA-seq data [179]. With dynamic GNNs, we could even better capture changes in expression levels observed in scRNA-seq data over time or as a result of a perturbation. In addition to resources such as the Human Cell Atlas [13] and the Single-Cell Atlas of Early Human Brain Development [71], more comprehensive biomedical networks that integrate cell-type, transcriptomic, tissue-level, cell-states data would only amplify the predictive power of graph representation learning models aimed to address biomedical problems at a single cell resolution.

Differentiating disease-specific features in the human microbiome. Constructing a network of the

human microbiome allows us to highlight key interactions between microbes and uncover their roles in the host’s health [14]. Using temporal graph representation learning methods, we can capture the dynamics of cross-species communication to accurately predict disease features and states [158]. We can also leverage the structure of phylogenetic trees in formulating our GNN architecture to better discriminate disease-specific signals [122]. Furthermore, the human microbiome has been shown to be modulated more so by environmental factors than host genetics to affect human health [188]. Hence, modeling interspecies communication within and across individual microbiota could provide valuable insights into both genomic- and environmental-level influences on patient health.

Diagnosing and treating patients through precision medicine. Seamless fusion of data modalities will ultimately enhance diagnostic and therapeutic efforts for individual patients. Effective integration of health-care data with fundamental knowledge about molecular, genomic, disease-level, and drug-level data can help generate more accurate and interpretable predictions about the biological systems of health and disease [75]. Constructing networks of interhospital communication can potentially improve risk assessments for transferring patients between hospitals to increase their survival [163]. The ability to draw from cross-domain knowledge and regularly update our biomedical knowledge base will empower us to narrow the gap between fundamental biological and computational research and bedside care.

Modeling spatial and temporal dynamics of complex systems. In public health, modeling the dynamics of disease propagation [208] at the population scale requires constructing and analyzing a multitude of rich and higher-order networks, including tensors, simplicial complexes, and hypergraphs. Many important problems can be modeled as a system of interconnected entities (e.g., individuals, municipalities), where each entity is spatially positioned and is recording time-dependent observations (e.g., disease states). To spot trends, detect anomalies, and interpret the temporal dynamics of such data, it is essential to understand the relationships between the different entities and how they evolve over time. Graph representation learning methods will soon be able to reason about questions [21], such as whether a pair of individuals has ever been close enough to infect each other (in this case, a simple undirected graph—the traditional setting, wherein each edge records a possible route of infection—is appropriate); or whether a set of individuals has ever formed all or part of a group whose members came close enough to infect each other (in this case, a simplicial complex is appropriate because it allows for downward closure—any subset of nodes within a simplex also forms a simplex identifying a set of people who might be infected simultaneously). New methodologies can also feature the behavior of processes, such as random walks, on top of the interconnected data. Thus, we would be able to harness graph representation learning to model biological and biomedical systems at the molecular scale as well as the population scale.

Towards learning fair and unbiased representations of interconnected data in medicine. As embeddings output by graph representation learning models are increasingly employed in real-world applications, it has become essential to ensure that the representations are fair and robust, and that, if necessary, existing algorithms are reformulated to address important sources of algorithmic bias and health disparities [170]. In the near future, we envision methods that will address the myriad of problems in learning graph representations that are fair and stable (e.g., [3]) and can defend models against adversarial attacks (e.g., [254]).

Acknowledgements

M.M.L. is supported by T32HG002295 from the National Human Genome Research Institute and a National Science Foundation Graduate Research Fellowship. We also gratefully acknowledge the support of NSF under nos. IIS-2030459 and IIS-2033384, the Harvard Data Science Initiative, the Amazon Research Award, and the Bayer Early Excellence in Science Award.

References

- [1] S. Abu-El-Haija, B. Perozzi, A. Kapoor, N. Alipourfard, K. Lerman, H. Harutyunyan, G. V. Steeg, and A. Galstyan. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In *ICML*, 2019.
- [2] M. Adnan, S. Kalra, and H. R. Tizhoosh. Representation learning of histopathology images using graph neural networks. In *CVPR Workshop*, 2020.
- [3] C. Agarwal, H. Lakkaraju, and M. Zitnik. Towards a unified framework for fair and stable graph representation learning. *arXiv:2102.13186*, 2021.
- [4] M. Agrawal, M. Zitnik, J. Leskovec, et al. Large-scale analysis of disease pathways in the human interactome. In *PSB*, 2018.
- [5] M. Akbarpour and M. O. Jackson. Diffusion in networks and the virtue of burstiness. *PNAS*, 2018.
- [6] M. Alagappan, D. Jiang, N. Denko, and A. C. Koong. A multimodal data analysis approach for targeted drug discovery involving topological data analysis (tda). In *Tumor Microenvironment*. 2016.
- [7] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 2002.
- [8] M. AlQuraishi. Proteinnet: a standardized data set for machine learning of protein structure. *BMC Bioinformatics*, 2019.
- [9] E. Alsentzer, S. G. Finlayson, M. M. Li, and M. Zitnik. Subgraph neural networks. In *NeurIPS*, 2020.
- [10] X. An, Y. Zhou, Y. Di, and D. Ming. Dynamic functional connectivity and graph convolution network for alzheimer’s disease classification. *arXiv:2006.13510*, 2020.
- [11] D. Anand, S. Gadiya, and A. Sethi. Histograms: graphs in histopathology. In *Medical Imaging 2020: Digital Pathology*, 2020.
- [12] T. S. Assmann, F. I. Milagro, and J. A. Martínez. Crosstalk between microRNAs, the putative target genes and the lncRNA network in metabolic diseases. *Molecular Medicine Reports*, 2019.
- [13] R. Aviv, S. A. Teichmann, E. S. Lander, A. Ido, B. Christophe, B. Ewan, B. Bernd, P. Campbell, C. Piero, C. Menna, et al. The human cell atlas. *Elife*, 2017.
- [14] V. D. Badal, D. Wright, Y. Katsis, H.-C. Kim, A. D. Swafford, R. Knight, and C.-N. Hsu. Challenges in the construction of knowledge bases for human microbiome-disease associations. *Microbiome*, 2019.
- [15] A.-L. Barabási et al. *Network Science*. 2016.
- [16] A.-L. Barabási, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 2011.
- [17] A.-L. Barabási. Network medicine — from obesity to the “diseasome”. *New England Journal of Medicine*, 2007.
- [18] L. Barisoni, K. J. Lafata, S. M. Hewitt, A. Madabhushi, and U. G. Balis. Digital pathology and computational image analysis in nephropathology. *Nature Reviews Nephrology*, 2020.

- [19] A. Baryshnikova. Systematic functional annotation and visualization of biological networks. *Cell Systems*, 2016.
- [20] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NeurIPS*, 2001.
- [21] A. R. Benson, D. F. Gleich, and D. J. Higham. Higher-order network analysis takes off, fueled by classical ideas and new data. *arXiv:2103.05031*, 2021.
- [22] A. S. Blevins and D. S. Bassett. Topology in biology. *Handbook of the Mathematics of the Arts and Sciences*, 2020.
- [23] C. Bodnar, C. Cangea, and P. Liò. Deep graph mapper: Seeing graphs through the neural lens. *arXiv:2002.03864*, 2020.
- [24] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NeurIPS*, 2013.
- [25] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 2017.
- [26] P. Bubenik. Statistical topological data analysis using persistence landscapes. *JMLR*, 2015.
- [27] D. B. Burkhardt, J. S. Stanley, A. L. Perdigoto, S. A. Gigante, K. C. Herold, G. Wolf, A. J. Giraldez, D. van Dijk, and S. Krishnaswamy. Quantifying the effect of experimental perturbations in single-cell rna-sequencing data using graph signal processing. *bioRxiv*, 2019.
- [28] T. J. Callahan, I. J. Tripodi, H. Pielke-Lombardo, and L. E. Hunter. Knowledge-based biomedical data science. *Annual Review of Biomedical Data Science*, 2020.
- [29] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins. Next-generation machine learning for biological networks. *Cell*, 2018.
- [30] M. Cao, H. Zhang, J. Park, N. M. Daniels, M. E. Crovella, L. J. Cowen, and B. Hescott. Going the distance for protein function prediction: a new distance metric for protein interaction networks. *PLOS One*, 2013.
- [31] Y. Cao and Y. Shen. Energy-based graph convolutional networks for scoring protein docking models. *Proteins: Structure, Function, and Bioinformatics*, 2020.
- [32] S. Carbon, E. Douglass, B. M. Good, D. R. Unni, N. L. Harris, C. J. Mungall, S. Basu, R. L. Chisholm, R. J. Dodson, E. Hartline, et al. The gene ontology resource: enriching a gold mine. *Nucleic Acids Research*, 2021.
- [33] M. Carrière, F. Chazal, Y. Ike, T. Lacombe, M. Royer, and Y. Umeda. Perslay: A neural network layer for persistence diagrams and new graph topological signatures. In *AISTATS*, 2020.
- [34] E. Cawi, P. S. La Rosa, and A. Nehorai. Designing machine learning workflows with an application to topological data analysis. *PLOS One*, 2019.
- [35] C.-H. Chao, Z. Zhu, D. Guo, K. Yan, T.-Y. Ho, J. Cai, A. P. Harrison, X. Ye, J. Xiao, A. Yuille, et al. Lymph node gross tumor volume detection in oncology imaging via relationship learning using graph neural network. In *MICCAI*, 2020.

- [36] F. Chen, Y.-C. Wang, B. Wang, and C.-C. J. Kuo. Graph representation learning: a survey. *APSIPA Transactions on Signal and Information Processing*, 2020.
- [37] I. Y. Chen, M. Agrawal, S. Horng, and D. Sontag. Robustly extracting medical knowledge from ehrrs: A case study of learning a health knowledge graph. In *PSB*, 2020.
- [38] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, and M. Zheng. Transformerpci: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 2020.
- [39] R. J. Chen, M. Y. Lu, J. Wang, D. F. Williamson, S. J. Rodig, N. I. Lindeman, and F. Mahmood. Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 2020.
- [40] Z.-H. Chen, Z.-H. You, Z.-H. Guo, H.-C. Yi, G.-X. Luo, and Y.-B. Wang. Prediction of drug–target interactions from multi-molecular network based on deep walk embedding model. *Frontiers in Bioengineering and Biotechnology*, 2020.
- [41] F. Cheng, I. A. Kovács, and A.-L. Barabási. Network-based prediction of drug combinations. *Nature Communications*, 2019.
- [42] F. Cheng, W. Lu, C. Liu, J. Fang, Y. Hou, D. E. Handy, R. Wang, Y. Zhao, Y. Yang, J. Huang, et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nature Communications*, 2019.
- [43] H. Chereda, A. Bleckmann, F. Kramer, A. Leha, and T. Beissbarth. Utilizing molecular network information via graph convolutional neural networks to predict metastatic event in breast cancer. In *GMDS*, 2019.
- [44] W. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C. Hsieh. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *KDD*, 2019.
- [45] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun. GRAM: graph-based attention model for healthcare representation learning. In *KDD*, 2017.
- [46] E. Choi, Z. Xu, Y. Li, M. Dusenberry, G. Flores, E. Xue, and A. M. Dai. Learning the graphical structure of electronic health records with graph convolutional transformer. In *AAAI*, 2020.
- [47] S. Chowdhury, C. Zhang, P. S. Yu, and Y. Luo. Mixed pooling multi-view attention autoencoder for representation learning in healthcare. *arXiv:1910.06456*, 2019.
- [48] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of Digital Imaging*, 2013.
- [49] C. W. Coley, W. Jin, L. Rogers, T. F. Jamison, T. S. Jaakkola, W. H. Green, R. Barzilay, and K. F. Jensen. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chemical Science*, 2019.
- [50] G. O. Consortium. The Gene Ontology resource: 20 years and still GOing strong. *Nucleic Acids Research*, 2019.

- [51] S. M. Corsello, J. A. Bittker, Z. Liu, J. Gould, P. McCarren, J. E. Hirschman, S. E. Johnston, A. Vrcic, B. Wong, M. Khan, et al. The drug repurposing hub: a next-generation drug library and information resource. *Nature Medicine*, 2017.
- [52] R. Cowan and N. Jonard. Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, 2004.
- [53] L. Cowen, T. Ideker, B. J. Raphael, and R. Sharan. Network propagation: a universal amplifier of genetic associations. *Nature Reviews Genetics*, 2017.
- [54] J. Crawford and C. S. Greene. Incorporating biological structure into machine learning models in biomedicine. *Current Opinion in Biotechnology*, 2020.
- [55] L. David, A. Thakkar, R. Mercado, and O. Engkvist. Molecular representations in ai-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 2020.
- [56] A. P. Davis, C. J. Grondin, R. J. Johnson, D. Sciaky, R. McMorran, J. Wiegiers, T. C. Wiegiers, and C. J. Mattingly. The comparative toxicogenomics database: update 2019. *Nucleic Acids Research*, 2019.
- [57] L. De Cecco, M. Nicolau, M. Giannoccaro, M. G. Daidone, P. Bossi, L. Locati, L. Licitra, and S. Canevari. Head and neck cancer subtypes with biological and clinical relevance: Meta-analysis of gene-expression data. *Oncotarget*, 2015.
- [58] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *NeurIPS*, 2016.
- [59] J. Dekker and T. Misteli. Long-range chromatin interactions. *Cold Spring Harbor Perspectives in Biology*, 2015.
- [60] T. K. Dey and S. Mandal. Protein classification with improved topological data analysis. In *WABI*, 2018.
- [61] N. Dimitriou, O. Arandjelović, and P. D. Caie. Deep learning for whole slide image analysis: an overview. *Frontiers in Medicine*, 2019.
- [62] Y. Dong, N. V. Chawla, and A. Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*, 2017.
- [63] Y. Dong, Z. Hu, K. Wang, Y. Sun, and J. Tang. Heterogeneous network representation learning. In *IJCAI*, 2020.
- [64] C. Donnat, M. Zitnik, D. Hallac, and J. Leskovec. Learning structural node embeddings via diffusion wavelets. In *KDD*, 2018.
- [65] F. Dutil, J. P. Cohen, M. Weiss, G. Derevyanko, and Y. Bengio. Towards gene expression convolutions using gene interaction graphs. *ICML Workshop on Computational Biology*, 2018.
- [66] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *NeurIPS*, 2015.
- [67] H. Edelsbrunner and J. Harer. Persistent homology-a survey. *Contemporary Mathematics*, 2008.
- [68] H. Edelsbrunner and J. Harer. *Computational topology: an introduction*. 2010.

- [69] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 1960.
- [70] M. Esteller. Non-coding rnas in human disease. *Nature Reviews Genetics*, 2011.
- [71] U. C. Eze, A. Bhaduri, M. Haeussler, T. J. Nowakowski, and A. R. Kriegstein. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nature Neuroscience*, 2021.
- [72] L. Fachal, H. Aschard, J. Beesley, D. R. Barnes, J. Allen, S. Kar, K. A. Pooley, J. Dennis, K. Michailidou, C. Turman, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nature Genetics*, 2020.
- [73] K. Fan, Y. Guan, and Y. Zhang. Graph2go: a multi-modal attributed network embedding method for inferring protein functions. *GigaScience*, 2020.
- [74] K. Fan and Y. Zhang. Pseudo2go: A graph-based deep learning method for pseudogene function prediction by borrowing information from coding genes. *Frontiers in Genetics*, 2020.
- [75] N. Fortelny and C. Bock. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biology*, 2020.
- [76] A. Fout, J. Byrd, B. Shariat, and A. Ben-Hur. Protein interface prediction using graph convolutional networks. In *NeurIPS*, 2017.
- [77] M. Franz, H. Rodriguez, C. Lopes, K. Zuberi, J. Montojo, G. D. Bader, and Q. Morris. GeneMANIA update 2018. *Nucleic Acids Research*, 2018.
- [78] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 1977.
- [79] F. Fuchs, D. E. Worrall, V. Fischer, and M. Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *NeurIPS*, 2020.
- [80] P. Gainza, F. Sverrisson, F. Monti, E. Rodola, D. Boscaini, M. Bronstein, and B. Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 2020.
- [81] F. Gao, K. Musial, C. Cooper, and S. Tsoka. Link prediction methods and their accuracy for different social networks and network metrics. *Scientific Programming*, 2015.
- [82] J. Gao, T. Lyu, F. Xiong, J. Wang, W. Ke, and Z. Li. MGNN: A multimodal graph neural network for predicting the survival of cancer patients. In *SIGIR*, 2020.
- [83] C. Geng, Y. Jung, N. Renaud, V. Honavar, A. M. Bonvin, and L. C. Xue. iscore: a novel graph kernel-based function for scoring protein–protein docking models. *Bioinformatics*, 2020.
- [84] C. Geng, Y. Jung, N. Renaud, V. Honavar, A. M. Bonvin, and L. C. Xue. iscore: a novel graph kernel-based function for scoring protein–protein docking models. *Bioinformatics*, 2020.
- [85] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *ICML*, 2017.
- [86] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási. The human disease network. *PNAS*, 2007.

- [87] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 2018.
- [88] C. S. Greene, A. Krishnan, A. K. Wong, E. Ricciotti, R. A. Zelaya, D. S. Himmelstein, R. Zhang, B. M. Hartmann, E. Zaslavsky, S. C. Sealfon, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nature Genetics*, 2015.
- [89] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *KDD*, 2016.
- [90] E. Guney, J. Menche, M. Vidal, and A.-L. Barábasi. Network-based in silico drug efficacy screening. *Nature Communications*, 2016.
- [91] M. Guo, Y. Yu, T. Wen, X. Zhang, B. Liu, J. Zhang, R. Zhang, Y. Zhang, and X. Zhou. Analysis of disease comorbidity patterns in a large-scale china population. *BMC Medical Genomics*, 2019.
- [92] Z.-H. Guo, H.-C. Yi, and Z.-H. You. Construction and comprehensive analysis of a molecular association network via lncrna–mirna–disease–drug–protein graph. *Cells*, 2019.
- [93] Z.-H. Guo, Z.-H. You, D.-S. Huang, H.-C. Yi, K. Zheng, Z.-H. Chen, and Y.-B. Wang. Meshheading2vec: a new method for representing mesh headings as vectors based on graph embedding algorithm. *Briefings in Bioinformatics*, 2020.
- [94] Z.-H. Guo, Z.-H. You, Y.-B. Wang, D.-S. Huang, H.-C. Yi, and Z.-H. Chen. Bioentity2vec: Attribute-and behavior-driven representation for predicting multi-type relationships between bioentities. *GigaScience*, 2020.
- [95] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, 2009.
- [96] D. M. Gysi, Í. D. Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, H. Sanchez, R. M. Baron, D. Ghiassian, J. Loscalzo, et al. Network medicine framework for identifying drug repurposing opportunities for COVID-19. *arXiv:2004.07229*, 2020.
- [97] J. Halcrow, A. Mosoi, S. Ruth, and B. Perozzi. Grale: Designing networks for graph learning. In *KDD*, 2020.
- [98] W. L. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *NeurIPS*, 2017.
- [99] P. Han, P. Yang, P. Zhao, S. Shang, Y. Liu, J. Zhou, X. Gao, and P. Kalnis. GCN-MF: disease-gene association identification by graph convolutional networks and matrix factorization. In *KDD*, 2019.
- [100] R. Hasibi and T. Michoel. Predicting gene expression from network topology using graph neural networks. *arXiv:2005.03961*, 2020.
- [101] D. S. Himmelstein and S. E. Baranzini. Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLOS Computational Biology*, 2015.
- [102] C. D. Hofer, F. Graf, B. Rieck, M. Niethammer, and R. Kwitt. Graph filtration learning. In *ICML*, 2020.
- [103] C. D. Hofer, R. Kwitt, and M. Niethammer. Learning representations of persistence barcodes. *JMLR*, 2019.

- [104] C. D. Hofer, R. Kwitt, M. Niethammer, and A. Uhl. Deep learning with topological signatures. In *NeurIPS*, 2017.
- [105] A. Hosseini, T. Chen, W. Wu, Y. Sun, and M. Sarrafzadeh. Heteromed: Heterogeneous information network for medical diagnosis. In *CIKM*, 2018.
- [106] J. X. Hu, C. E. Thomas, and S. Brunak. Network biology concepts in complex disease comorbidities. *Nature Reviews Genetics*, 2016.
- [107] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V. S. Pande, and J. Leskovec. Strategies for pre-training graph neural networks. In *ICLR*, 2020.
- [108] Z. Hu, Y. Dong, K. Wang, and Y. Sun. Heterogeneous graph transformer. In *WWW*, 2020.
- [109] J. Huan, D. Bandyopadhyay, W. Wang, J. Snoeyink, J. Prins, and A. Tropsha. Comparing graph representations of protein structure for mining family-specific residue-based packing motifs. *Journal of Computational Biology*, 2005.
- [110] K. Huang, T. Fu, W. Gao, Y. Zhao, Y. Roohani, J. Leskovec, C. W. Coley, C. Xiao, J. Sun, and M. Zitnik. Therapeutics data commons: Machine learning datasets and tasks for therapeutics. *arXiv:2102.09548*, 2021.
- [111] J. Ingraham, V. K. Garg, R. Barzilay, and T. S. Jaakkola. Generative models for graph-based protein design. In *NeurIPS*, 2019.
- [112] G. Jaume, P. Pati, A. Foncubierta-Rodriguez, F. Feroce, G. Scognamiglio, A. M. Anniciello, J.-P. Thiran, O. Goksel, and M. Gabrani. Towards explainable graph representations in digital pathology. *ICML Workshop on Computational Biology*, 2020.
- [113] C. Jiang, F. Coenen, and M. Zito. A survey of frequent subgraph mining algorithms. *Knowledge Engineering Review*, 2013.
- [114] M. Jiang, Z. Li, S. Zhang, S. Wang, X. Wang, Q. Yuan, and Z. Wei. Drug–target affinity prediction using graph neural network and contact maps. *RSC Advances*, 2020.
- [115] P. Jiang, S. Huang, Z. Fu, Z. Sun, T. M. Lakowski, and P. Hu. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Computational and Structural Biotechnology Journal*, 2020.
- [116] J. Jiménez-Luna, F. Grisoni, and G. Schneider. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2020.
- [117] W. Jin, R. Barzilay, and T. S. Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *ICML*, 2018.
- [118] H. Kagami, T. Akutsu, S. Maegawa, H. Hosokawa, and J. C. Nacher. Determining associations between human diseases and non-coding rnas with critical roles in network control. *Scientific Reports*, 2015.
- [119] S. M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart. Representation learning for dynamic graphs: A survey. *JMLR*, 2020.
- [120] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of Computer-Aided Molecular Design*, 2016.

- [121] N. Keriven and G. Peyré. Universal invariant and equivariant graph neural networks. In *NeurIPS*, 2019.
- [122] S. Khan and L. Kelly. Multiclass disease classification from microbial whole-community metagenomes. In *PSB*, 2020.
- [123] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *ICLR*, 2014.
- [124] T. N. Kipf and M. Welling. Variational graph auto-encoders. *NeurIPS Workshop*, 2016.
- [125] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- [126] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 2010.
- [127] J. Klicpera, J. Groß, and S. Günnemann. Directional message passing for molecular graphs. In *ICLR*, 2020.
- [128] M. Kondratova, U. Czerwinska, N. Sompairac, S. D. Amigorena, V. Soumelis, E. Barillot, A. Zinovyev, and I. Kuperstein. A multiscale signalling network map of innate immune response in cancer reveals cell heterogeneity signatures. *Nature Communications*, 2019.
- [129] M. Koutrouli, E. Karatzas, D. Paez-Espino, and G. A. Pavlopoulos. A guide to conquer the biological network era using graph theory. *Frontiers in Bioengineering and Biotechnology*, 2020.
- [130] I. A. Kovács, K. Luck, K. Spirohn, Y. Wang, C. Pollis, S. Schlabach, W. Bian, D.-K. Kim, N. Kishore, T. Hao, et al. Network-based prediction of protein interactions. *Nature Communications*, 2019.
- [131] G. Kustatscher, P. Grabowski, T. A. Schrader, J. B. Passmore, M. Schrader, and J. Rappsilber. Co-regulation map of the human proteome enables identification of protein functions. *Nature Biotechnology*, 2019.
- [132] H. Kwak, M. Lee, S. Yoon, J. Chang, S. Park, and K. Jung. Drug-disease graph: Predicting adverse drug reaction signals via graph neural network with clinical data. In *PAKDD*, 2020.
- [133] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 2006.
- [134] J. Lanchantin and Y. Qi. Graph convolutional networks for epigenetic state prediction using both sequence and 3d genome data. *Bioinformatics*, 2020.
- [135] D.-H. Le and V.-T. Dang. Ontology-based disease similarity network for disease gene prediction. *Vietnam Journal of Computer Science*, 2016.
- [136] D. Lee, X. Jiang, and H. Yu. Harmonized representation learning on dynamic ehr graphs. *Journal of Biomedical Informatics*, 2020.
- [137] J. Levy, C. Haudenschild, C. Bar, B. Christensen, and L. Vaickus. Topological feature extraction and visualization of whole slide images using graph neural networks. *PSB*, 2020.
- [138] B. Li and D. Pi. Network representation learning: a systematic literature review. *Neural Computing and Applications*, 2020.

- [139] D. Li and J. Gao. Towards perturbation prediction of biological networks using deep learning. *Scientific Reports*, 2019.
- [140] L. Li, P. Wang, Y. Wang, S. Wang, J. Yan, J. Jiang, B. Tang, C. Wang, and Y. Liu. A method to learn embedding of a probabilistic medical knowledge graph: Algorithm development. *JMIR Medical Informatics*, 2020.
- [141] Y. Li, B. Qian, X. Zhang, and H. Liu. Graph neural network-based diagnosis prediction. *Big Data*, 2020.
- [142] R. Liao, Y. Li, Y. Song, S. Wang, W. L. Hamilton, D. Duvenaud, R. Urtasun, and R. S. Zemel. Efficient graph generation with graph recurrent attention networks. In *NeurIPS*, 2019.
- [143] X. Lin, K. Zhao, T. Xiao, Z. Quan, Z.-J. Wang, and P. S. Yu. Deepgs: Deep representation learning of graphs and sequences for drug-target binding affinity prediction. *ECAI*, 2020.
- [144] C. Liu, Y. Ma, J. Zhao, R. Nussinov, Y.-C. Zhang, F. Cheng, and Z.-K. Zhang. Computational network biology: Data, models, and applications. *Physics Reports*, 2020.
- [145] H. Liu, W. Zhang, B. Zou, J. Wang, Y. Deng, and L. Deng. Drugcombdb: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Research*, 2020.
- [146] S. Liu, T. Li, H. Ding, B. Tang, X. Wang, Q. Chen, J. Yan, and Y. Zhou. A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction. *IJMLC*, 2020.
- [147] Y. Liu, H. Yuan, L. Cai, and S. Ji. Deep learning of high-order interactions for protein interface prediction. In *KDD*, 2020.
- [148] K. Luck, D.-K. Kim, L. Lambourne, K. Spirohn, B. E. Begg, W. Bian, R. Brignall, T. Cafarelli, F. J. Campos-Laborie, B. Charlotiaux, et al. A reference map of the human binary protein interactome. *Nature*, 2020.
- [149] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao. KAME: knowledge-based attention model for diagnosis prediction in healthcare. In *CIKM*, 2018.
- [150] T. Ma, C. Xiao, J. Zhou, and F. Wang. Drug similarity integration through attentive multi-view graph auto-encoders. In *IJCAI*, 2018.
- [151] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *JMLR*, 2008.
- [152] M. Madhavan et al. Deep belief network based representation learning for lncrna-disease association prediction. *arXiv:2006.12534*, 2020.
- [153] K. F. Madhobi, M. Kamruzzaman, A. Kalyanaraman, E. Lofgren, R. Moehring, and B. Krishnamoorthy. A visual analytics framework for analysis of patient trajectories. In *ACM-BCB*, 2019.
- [154] S. Mandal, A. Guzmán-Sáenz, N. Haiminen, S. Basu, and L. Parida. A topological data analysis approach on predicting phenotypes from gene expression data. In *AICoB*, 2020.
- [155] O. C. Martin, A. Krzywicki, and M. Zagorski. Drivers of structural features in gene regulatory networks: From biophysical constraints to biological function. *Physics of Life Reviews*, 2016.
- [156] A. Martino, A. Rizzi, and F. M. F. Mascioli. Supervised approaches for protein function prediction by topological data analysis. In *IJCNN*, 2018.

- [157] A. Mazein, M. Ostaszewski, I. Kuperstein, S. Watterson, N. Le Novère, D. Lefaudeux, B. De Meulder, J. Pellet, I. Balaur, M. Saqi, et al. Systems medicine disease maps: community-driven comprehensive representation of disease mechanisms. *npj Systems Biology and Applications*, 2018.
- [158] K. Melnyk, S. Klus, G. Montavon, and T. O. Conrad. Graphkke: graph kernel koopman embedding for human microbiome analysis. *Applied Network Science*, 2020.
- [159] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, and A.-L. Barabási. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 2015.
- [160] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NeurIPS*, 2013.
- [161] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 2002.
- [162] Y. Moreau and L.-C. Tranchevent. Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nature Reviews Genetics*, 2012.
- [163] S. K. Mueller, J. Fiskio, and J. Schnipper. Interhospital transfer: Transfer processes and patient outcomes. *Journal of Hospital Medicine*, 2019.
- [164] C. J. Mungall, J. A. McMurry, S. Köhler, J. P. Balhoff, C. Borromeo, M. Brush, S. Carbon, T. Conlin, N. Dunn, M. Engelstad, et al. The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Research*, 2017.
- [165] A. Nambiar, M. Heflin, S. Liu, S. Maslov, M. Hopkins, and A. Ritz. Transforming the language of life: Transformer neural networks for protein prediction tasks. In *Proceedings of the 11th ACM-BCB*, 2020.
- [166] C. A. Nelson, A. J. Butte, and S. E. Baranzini. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nature Communications*, 2019.
- [167] M. Nevitt, D. Felson, and G. Lester. The osteoarthritis initiative. *Protocol for the Cohort Study*, 2006.
- [168] M. Nickel, V. Tresp, and H. Kriegel. A three-way model for collective learning on multi-relational data. In *ICML*, 2011.
- [169] M. Nicolau, A. J. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *PNAS*, 2011.
- [170] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 2019.
- [171] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. B. Schardl, and C. E. Leiserson. Evolvegc: Evolving graph convolutional networks for dynamic graphs. In *AAAI*, 2020.
- [172] P. Pati, G. Jaume, L. A. Fernandes, A. Foncubierta-Rodríguez, F. Feroce, A. M. Anniciello, G. Scognamiglio, N. Brancati, D. Riccio, M. Di Bonito, et al. Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis*. 2020.
- [173] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: online learning of social representations. In *KDD*, 2014.

- [174] X. Qiu, A. Rahimzamani, L. Wang, B. Ren, Q. Mao, T. Durham, J. L. McFaline-Figueroa, L. Saunders, C. Trapnell, and S. Kannan. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Systems*, 2020.
- [175] Z. Quan, Y. Guo, X. Lin, Z.-J. Wang, and X. Zeng. Graphcpi: Graph neural representation learning for compound-protein interaction. In *BIBM*, 2019.
- [176] A. Rai, P. Shinde, and S. Jalan. Network spectra for drug-target identification in complex diseases: new guns against old foes. *Applied Network Science*, 2018.
- [177] A. Raj, A. Kuceyeski, and M. Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 2012.
- [178] R. Ramirez, Y.-C. Chiu, A. Herrera, M. Mostavi, J. Ramirez, Y. Chen, Y. Huang, and Y.-F. Jin. Classification of cancer types using graph convolutional neural networks. *Frontiers in Physics*, 2020.
- [179] N. Ravindra, A. Sehanobish, J. L. Pappalardo, D. A. Hafler, and D. van Dijk. Disease state prediction from single-cell data using graph attention networks. In *CHIL*, 2020.
- [180] S. Rhee, S. Seo, and S. Kim. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. In *IJCAI*, 2018.
- [181] A. H. Rizvi, P. G. Camara, E. K. Kandror, T. J. Roberts, I. Schieren, T. Maniatis, and R. Rabadan. Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development. *Nature Biotechnology*, 2017.
- [182] P. N. Robinson, S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics*, 2008.
- [183] E. Rocheteau, C. Tong, P. Veličković, N. Lane, and P. Liò. Predicting patient outcomes with graph representation learning. *AAAI*, 2021.
- [184] T. Rolland, M. Taşan, B. Charlotiaux, S. J. Pevzner, Q. Zhong, N. Sahni, S. Yi, I. Lemmens, C. Fontanillo, R. Mosca, et al. A proteome-scale map of the human interactome network. *Cell*, 2014.
- [185] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein. Temporal graph networks for deep learning on dynamic graphs. *ICML Workshop on Graph Representation Learning and Beyond*, 2020.
- [186] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag. Learning a health knowledge graph from electronic medical records. *Scientific Reports*, 2017.
- [187] J.-F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 2005.
- [188] P. Scepanovic, F. Hodel, S. Mondot, V. Partula, A. Byrd, C. Hammer, C. Alanio, J. Bergstedt, E. Patin, M. Touvier, et al. A comprehensive assessment of demographic, environmental, and host genetic associations with gut microbiome diversity in healthy individuals. *Microbiome*, 2019.
- [189] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. In *ESWC*, 2018.

- [190] L. M. Schriml, C. Arze, S. Nadendla, Y.-W. W. Chang, M. Mazaitis, V. Felix, G. Feng, and W. A. Kibbe. Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Research*, 2012.
- [191] K. Schütt, P. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K. Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NeurIPS*, 2017.
- [192] J. Shang, C. Xiao, T. Ma, H. Li, and J. Sun. Gamenet: Graph augmented memory networks for recommending medication combination. In *AAAI*, 2019.
- [193] G. Singh, F. Mémoli, and G. E. Carlsson. Topological methods for the analysis of high dimensional data sets and 3d object recognition. *SPBG*, 2007.
- [194] M. Singha, L. Pu, K. Busch, H.-C. Wu, J. Ramanujam, M. Brylinski, et al. Graphgr: A graph neural network to predict the effect of pharmacotherapy on the cancer cell growth. *bioRxiv*, 2020.
- [195] T.-A. Song, S. R. Chowdhury, F. Yang, H. Jacobs, G. El Fakhri, Q. Li, K. Johnson, and J. Dutta. Graph convolutional neural networks for alzheimer’s disease classification. In *ISBI*, 2019.
- [196] M. Spielmann and S. Mundlos. Looking beyond the genes: the role of non-coding variants in human disease. *Human Molecular Genetics*, 2016.
- [197] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. Biogrid: a general repository for interaction datasets. *Nucleic Acids Research*, 2006.
- [198] I. Subramanian, S. Verma, S. Kumar, A. Jere, and K. Anamika. Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights*, 2020.
- [199] E. Sügis, J. Dauvillier, A. Leontjeva, P. Adler, V. Hindie, T. Moncion, V. Collura, R. Daudin, Y. Loe-Mie, Y. Herault, et al. Hena, heterogeneous network-based data set for alzheimer’s disease. *Scientific Data*, 2019.
- [200] M. Sumathipala, E. Maiorino, S. T. Weiss, and A. Sharma. Network diffusion approach to predict lncrna disease associations using multi-type biological networks: Lion. *Frontiers in Physiology*, 2019.
- [201] M. Sun, S. Zhao, C. Gilvary, O. Elemento, J. Zhou, and F. Wang. Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics*, 2020.
- [202] Z. Sun, Z. Deng, J. Nie, and J. Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *ICLR*, 2019.
- [203] Z. Sun, H. Yin, H. Chen, T. Chen, L. Cui, and F. Yang. Disease prediction via graph neural networks. *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [204] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, et al. String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 2019.
- [205] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000.
- [206] M. A. Thafar, R. S. Olayan, H. Ashoor, S. Albaradei, V. B. Bajic, X. Gao, T. Gojobori, and M. Essack. Dtigems+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. *Journal of Cheminformatics*, 2020.

- [207] C. Tian, B. S. Hromatka, A. K. Kiefer, N. Eriksson, S. M. Noble, J. Y. Tung, and D. A. Hinds. Genome-wide association and hla region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature Communications*, 2017.
- [208] L. Torres, A. S. Blevins, D. S. Bassett, and T. Eliassi-Rad. The why, how, and when of representations for complex systems. *arXiv:2006.02870*, 2020.
- [209] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *ICLR*, 2016.
- [210] M. Tsubaki, K. Tomii, and J. Sese. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, 2019.
- [211] B. D. Umans, A. Battle, and Y. Gilad. Where are the disease-associated eqtls? *Trends in Genetics*, 2020.
- [212] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *NeurIPS*, 2017.
- [213] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *ICLR*, 2018.
- [214] K. Veselkov, G. Gonzalez, S. Aljifri, D. Galea, R. Mirnezami, J. Youssef, M. Bronstein, and I. Laponogov. Hyperfoods: Machine intelligent mapping of cancer-beating molecules in foods. *Scientific Reports*, 2019.
- [215] J. Vig, A. Madani, L. R. Varshney, C. Xiong, R. Socher, and N. F. Rajani. Bertology meets biology: Interpreting attention in protein language models. *ICLR*, 2021.
- [216] O. Vinyals, S. Bengio, and M. Kudlur. Order matters: Sequence to sequence for sets. In *ICLR*, 2016.
- [217] F. Wan, L. Hong, A. Xiao, T. Jiang, and J. Zeng. Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics*, 2019.
- [218] F. Wang and C. Zhang. Label propagation through linear neighborhoods. In *ICML*, 2006.
- [219] H. Wang and J. Leskovec. Unifying graph convolutional neural networks and label propagation. *arXiv:2002.06755*, 2020.
- [220] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo. Graphgan: Graph representation learning with generative adversarial nets. In *AAAI*, 2018.
- [221] R. Wang, S. Li, L. Cheng, M. H. Wong, and K. S. Leung. Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning. *BMC Bioinformatics*, 2019.
- [222] T. Wang, Q. Peng, B. Liu, Y. Liu, and Y. Wang. Disease module identification based on representation learning of complex networks integrated from gwas, eqtl summaries, and human interactome. *Frontiers in Bioengineering and Biotechnology*, 2020.
- [223] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu. Heterogeneous graph attention network. In *WWW*, 2019.
- [224] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph CNN for learning on point clouds. *ACM Transactions On Graphics*, 2019.

- [225] D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 2010.
- [226] C.-Y. Wee, C. Liu, A. Lee, J. S. Poh, H. Ji, A. Qiu, A. D. N. Initiative, et al. Cortical graph neural network for ad and mci diagnosis and transfer learning across populations. *NeuroImage: Clinical*, 2019.
- [227] O. Wieder, S. Kohlbacher, M. Kuenemann, A. Garon, P. Ducrot, T. Seidel, and T. Langer. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 2020.
- [228] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 2018.
- [229] L. Wong, Z.-H. You, Z.-H. Guo, H.-C. Yi, Z.-H. Chen, and M.-Y. Cao. Mipdh: A novel computational model for predicting microRNA–mRNA interactions by DeepWalk on a heterogeneous network. *ACS Omega*, 2020.
- [230] T. Wu, Y. Wang, Y. Wang, E. Zhao, Y. Yuan, and Z. Yang. Representation learning of ehr data via graph-based medical entity embedding. *NeurIPS Graph Representation Learning Workshop*, 2019.
- [231] X. Wu, W. Lan, Q. Chen, Y. Dong, J. Liu, and W. Peng. Inferring lncrna-disease associations based on graph autoencoder matrix completion. *Computational Biology and Chemistry*, 2020.
- [232] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
- [233] Y. Xie, J. Peng, and Y. Zhou. Integrating protein-protein interaction information into drug response prediction by graph neural encoding. 2019.
- [234] J. Xu, S. Kim, M. Song, M. Jeong, D. Kim, J. Kang, J. F. Rousseau, X. Li, W. Xu, V. I. Torvik, et al. Building a pubmed knowledge graph. *Scientific Data*, 2020.
- [235] K. Xu, W. Hu, J. Leskovec, and S. Jegelka. How powerful are graph neural networks? In *ICLR*, 2019.
- [236] K. Xu, C. Li, Y. Tian, T. Sonobe, K. Kawarabayashi, and S. Jegelka. Representation learning on graphs with jumping knowledge networks. In *ICML*, 2018.
- [237] P. Xuan, S. Pan, T. Zhang, Y. Liu, and H. Sun. Graph convolutional network and convolutional neural network based method for predicting lncrna-disease associations. *Cells*, 2019.
- [238] S. Yan, Y. Xiong, and D. Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *AAAI*, 2018.
- [239] B. Yang, W. Yih, X. He, J. Gao, and L. Deng. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*, 2015.
- [240] C. Yang, P. Zhuang, W. Shi, A. Luu, and P. Li. Conditional structure generation through graph variational generative adversarial nets. In *NeurIPS*, 2019.
- [241] F. Yang, K. Fan, D. Song, and H. Lin. Graph-based prediction of protein-protein interactions with attributed signed graph embedding. *BMC Bioinformatics*, 2020.

- [242] K. Yang, R. Wang, G. Liu, Z. Shu, N. Wang, R. Zhang, J. Yu, J. Chen, X. Li, and X. Zhou. Hergepred: heterogeneous network embedding representation for disease gene prediction. *IEEE Journal of Biomedical and Health Informatics*, 2018.
- [243] H. Yao, J. Guan, and T. Liu. Denoising protein-protein interaction network via variational graph auto-encoder for protein complex detection. *Journal of Bioinformatics and Computational Biology*, 2020.
- [244] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec. Gnnexplainer: Generating explanations for graph neural networks. In *NeurIPS*, 2019.
- [245] Z. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec. Hierarchical graph representation learning with differentiable pooling. In *NeurIPS*, 2018.
- [246] J. You, R. Ying, X. Ren, W. L. Hamilton, and J. Leskovec. Graphrnn: Generating realistic graphs with deep auto-regressive models. In *ICML*, 2018.
- [247] Y. You, T. Chen, Z. Wang, and Y. Shen. When does self-supervision help graph convolutional networks? In *ICML*, 2020.
- [248] L. Yuan, J. Zhao, T. Sun, X.-S. Jiang, Z.-Y. Yang, X.-G. Wang, and Y.-S. Geng. Lncrna-disease association prediction based on graph neural networks and inductive matrix completion. In *ICIC*, 2020.
- [249] X. Yue, Z. Wang, J. Huang, S. Parthasarathy, S. Moosavinasab, Y. Huang, S. M. Lin, W. Zhang, P. Zhang, and H. Sun. Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics*, 2020.
- [250] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim. Graph transformer networks. In *NeurIPS*, 2019.
- [251] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. K. Prasanna. Graphsaint: Graph sampling based inductive learning method. In *ICLR*, 2020.
- [252] F. Zhang and J. R. Lupski. Non-coding genetic variants in human disease. *Human Molecular Genetics*, 2015.
- [253] H. Zhang, Y. Liang, C. Peng, S. Han, W. Du, and Y. Li. Predicting lncrna-disease associations using network topological similarity based on deep mining heterogeneous networks. *Mathematical Biosciences*, 2019.
- [254] X. Zhang and M. Zitnik. GNNGuard: Defending graph neural networks against adversarial attacks. *NeurIPS*, 2020.
- [255] Z. Zhang, P. Cui, and W. Zhu. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2020.
- [256] C. Zhao, J. Jiang, Y. Guan, X. Guo, and B. He. Emr-based medical knowledge representation and inference via markov random fields and distributed representation learning. *Artificial Intelligence in Medicine*, 2018.
- [257] T. Zhao, Y. Hu, J. Peng, and L. Cheng. Deeplgp: a novel deep learning method for prioritizing lncrna target genes. *Bioinformatics*, 2020.

- [258] K. Zheng, Z.-H. You, L. Wang, and Z.-H. Guo. imda-bn: Identification of mirna-disease associations based on the biological network and graph embedding algorithm. *Computational and Structural Biotechnology Journal*, 2020.
- [259] G. Zhou, J. Wang, X. Zhang, and G. Yu. Deepgoa: Predicting gene ontology annotations of proteins via graph convolutional network. In *BIBM*, 2019.
- [260] Y. Zhou, S. Graham, N. Alemi Koohbanani, M. Shaban, P.-A. Heng, and N. Rajpoot. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *ICCV Workshops*, 2019.
- [261] M. Zitnik, M. Agrawal, and J. Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 2018.
- [262] M. Zitnik, M. W. Feldman, J. Leskovec, et al. Evolution of resilience in protein interactomes across the tree of life. *PNAS*, 2019.
- [263] M. Zitnik and J. Leskovec. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*, 2017.
- [264] M. Zitnik, E. A. Nam, C. Dinh, A. Kuspa, G. Shaulsky, and B. Zupan. Gene prioritization by compressive data fusion and chaining. *PLOS Computational Biology*, 2015.
- [265] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, 2019.