Graph Representation Learning in Biomedicine and Healthcare

Michelle M. Li[1,3], **Kexin Huang**[2], **Marinka Zitnik**[3,4,5,*]

[1]Bioinformatics and Integrative Genomics Program, Harvard Medical School, Boston, MA 02115, USA

[2]Health Data Science Program, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

[3]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

[4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[5]Harvard Data Science Initiative, Cambridge, MA 02138, USA

Abstract

Biomedical networks (or graphs) are universal descriptors for systems of interacting elements, from molecular interactions and disease co-morbidity to healthcare systems and scientific knowledge. Advances in artificial intelligence, specifically deep learning, have enabled us to model, analyze, and learn with such networked data. In this review, we put forward an observation that long-standing principles of systems biology and medicine—while often unspoken in machine learning research—provide the conceptual grounding for representation learning on graphs, explain its current successes and limitations, and even inform future advancements. We synthesize a spectrum of algorithmic approaches that, at their core, leverage graph topology to embed networks into compact vector spaces. We also capture the breadth of ways in which representation learning has dramatically improved the state-of-the-art in biomedical machine learning. Exemplary domains covered include identifying variants underlying complex traits, disentangling behaviors of single cells and their effects on health, assisting in diagnosis and treatment of patients, and developing safe and effective medicines.

1 Introduction

Networks (or graphs) are pervasive in biology and medicine, from molecular interaction maps to population-scale social and health interactions. With the multitude of bioentities and associations that can be described by networks, they are prevailing representations of biological organization and biomedical knowledge. For instance, edges in a regulatory network can indicate causal activating and inhibitory relationships between genes [1]; edges between genes and diseases can indicate genes that are 'upregulated by', 'downregulated by', or 'associated with' a disease [2]; and edges in a knowledge network built from electronic health records (EHR) can indicate co-occurrences of medical codes across patients [3, 4, 5]. The ability to model all biomedical discoveries to date—even overlay

*Correspondence: marinka@hms.harvard.edu.

patient-specific information—in a unified data representation has driven the development of artificial intelligence, specifically deep learning, for networks. In fact, the diversity and multimodality in networks not only boost performance of predictive models, but importantly enable broad generalization to settings not seen during training [6] and improve model interpretability [7, 8]. Nevertheless, interactions in networks give rise to a bewildering degree of complexity that can likely only be fully understood through a holistic and integrated view [9, 10, 11]. As a result, systems biology and medicine—upon which deep learning on graphs is founded—have identified over the last two decades *organizing principles* that govern networks [12, 13, 14, 15].

The organizing principles governing networks *link network structure to molecular phenotypes, biological roles, disease, and health*, thus providing the conceptual grounding that, we argue, can explain the successes (and limitations) of graph representation learning and inform future development of the field. Here, we exemplify how a series of such principles has uncovered disease mechanisms. First, interacting entities are typically more similar than non-interacting entities, as implicated by the local hypothesis [13]. In protein interaction networks, for instance, mutations in interacting proteins often lead to similar diseases [13]. Given by the shared components and disease module hypotheses [13], cellular components associated with the same phenotype tend to cluster in the same network neighborhood [16]. Further, essential genes are often found in hubs of a molecular network whereas non-essential genes (e.g., those associated with disease) are located on the periphery [13]. Thus, the network parsimony principle dictates that shortest molecular paths between known disease-associated components tend to correlate with causal molecular pathways [13]. To this day, these hypotheses and principles continue to drive discoveries.

We posit that *representation learning can realize network biomedicine principles*. Its core 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 nodes. More concretely, representation learning specifies deep, non-linear transformation functions that map nodes to points in a compact vector space, termed *embeddings*. Such functions are optimized to embed the input network so that nodes with similar network neighborhoods are embedded close together in the embedding space, and algebraic operations performed in this learned space reflect the network's topology. To provide concrete connections between graph representation learning and systems biology and medicine: nodes in the same positional regions should have similar embeddings due to the local hypothesis (e.g., highly similar pairs of protein embeddings suggest similar phenotypic consequence); node embeddings can capture whether the nodes lie within a hub based on their degree, an important aspect of local neighborhood (e.g., strongly clustered gene embeddings indicate essential housekeeping roles); and given by the shared components hypothesis, two nodes with significantly overlapping sets of network neighbors should have similar embeddings due to shared message passing (e.g., highly similar disease embeddings imply shared disease-associated cellular components). Hence, artificial intelligence methods that produce representations can be thought of as differentiable engines of key network biomedicine principles.

Our survey provides an exposition of graph artificial intelligence capability and highlights important applications for deep learning on biomedical networks. Given the prominence of graph representation learning, specific aspects of it have been covered extensively. However, existing reviews independently discuss deep learning on structured data [17, 18]; graph neural networks [19, 20, 21]; representation learning for homogeneous and heterogeneous graphs [22, 23, 24], solely heterogeneous graphs [25], and dynamic graphs [26]; data fusion [27]; network propagation [28]; topological data analysis [29]; and creation of biomedical networks [9, 29, 30, 31, 32]. Biomedically-focused reviews survey graph neural networks exclusively on molecular generation [33, 34], single-cell biology [35], drug discovery and repurposing [36, 37, 38, 39, 40], or histopathology [41]. Other reviews tend to focus solely on graph neural networks, excluding other graph representation learning approaches or do not consider patient-centric methods [42]. In contrast, our survey unifies graph representation learning approaches across molecular, genomic, therapeutic, and precision medicine areas.

2 Graph representation learning

Graph theoretic techniques have fueled discoveries, from uncovering relationships between diseases [43, 44, 45, 46] to repurposing drugs [6, 47, 48]. Further algorithmic innovations, such as random walks [49, 50, 51], kernels [52], and network propagation [53], have also played a role in capturing structural information in networks. Feature engineering, the process of extracting predetermined features from a network to suit a user-specified machine learning method [54], is a common approach for machine learning on 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 a machine learning model. While powerful, it can be challenging to hand engineer optimally-predictive features across diverse types of networks and applications [18].

For these reasons, *graph representation learning*, the idea of automatically learning optimal features for networks, has emerged as a leading artificial intelligence approach for networks. Graph representation learning is challenging because graphs contain complex topographical structure, have no fixed node ordering or reference points, and are comprised 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, because classic deep methods are designed for fixed-size grids (e.g., images and tabular datasets) or optimized for text and sequences. Akin to how deep learning on images and sequences has revolutionized image analysis and natural language processing, graph representation learning is poised to transform the study of complex systems.

Graph representation learning methods generate vector representations for graph elements such that the learned representations, i.e., *embeddings*, capture the structure and semantics of networks, along with any downstream supervised task, if any (Box 1). Graph representation learning encompasses a wide range of methods, including manifold learning, topological data analysis, graph neural networks and generative graph models (Figure 2). We

next describe graph elements and outline main artificial intelligence tasks on graphs (Box 1). We then outline graph representation learning methods (Section 2.1-2.3).

2.1 Shallow graph embedding approaches

Shallow embedding methods optimize a compact vector space such that points that are close in the graph are mapped to nearby points in the embedding space, which is measured by a predefined distance function or an outer product. These approaches are transductive embedding methods where the encoder function is a simple embedding lookup (Figure 2). More concretely, the methods have three steps: (1) *Mapping to an embedding space*. Given a pair of nodes u and v in graph G, we specify an encoder, a learnable function f that maps nodes to embeddings $\mathbf{h}_u$ and $\mathbf{h}_v$. (2) *Defining graph similarity*. We define the graph similarity $f_n(u, v)$, for example, measured by distance between u and v in the graph, and the embedding similarity $f_z(\mathbf{h}_u, \mathbf{h}_v)$, for example, an Euclidean distance function or pairwise dot-product. (3) *Computing loss*. Then, we define the loss $\mathscr{L}(f_n(u, v), f_z(\mathbf{h}_u, \mathbf{h}_v))$, which quantifies how the resulting embeddings preserve the desired input graph similarity. Finally, we apply an optimization procedure to minimize the loss $\mathscr{L}(f_n(u, v), f_z(\mathbf{h}_u, \mathbf{h}_v))$. The resulting encoder f is a graph embedding method that serves as a shallow embedding lookup and considers the graph structure only in the loss function.

Shallow embedding methods vary given various definitions of similarities. For example, shortest path length between nodes is often used as the network similarity and dot product as the embedding similarity. Perozzi *et al.* [55] define similarity as co-occurrence in a series of random walks of length k. Unsupervised techniques that predict which node belongs to the walk, such as skip-gram [66], are then applied on the walks to generate embeddings. Grover *et al.* [56] propose an alternative way for walks on graphs, using a combination of depth-first search and breadth-first search. In heterogeneous graphs, information on the semantic meaning of edges, i.e., relation types, can be important. Knowledge graph methods expand similarity measures to consider relation types [58, 67, 68, 69, 70, 71]. When shallow embedding models are trained, the resulting embeddings can be fed into separate models to be optimized towards a specific classification or regression task.

2.2 Graph neural networks

Graph neural networks (GNNs) are a class of neural networks designed for graph-structured datasets (Figure 2). They learn compact representations of graph elements, their attributes, and supervised labels if any. A typical GNN consists of a series of propagation layers [72], where layer l carries out three operations: (1) *Passing neural messages*. The GNN computes a message $\mathbf{m}_{u,v}^{(l)} = \mathrm{MSG}(\mathbf{h}_u^{(l-1)}, \mathbf{h}_v^{(l-1)})$ 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) *Aggregating neighborhoods*. The messages between node u and its neighbors $\mathscr{N}_u$ are aggregated as $\widehat{\mathbf{m}}_u^{(l)} = \mathrm{AGG}(\mathbf{m}_{uv}^{(l)} \mid v \in \mathscr{N}_u)$. (3) *Updating representations*. Non-linear transformation is applied to update node embeddings as $\mathbf{h}_u^{(l)} = \mathrm{UPD}(\widehat{\mathbf{m}}_u^{(l)}, \mathbf{h}_u^{(l-1)})$ using the aggregated message and the embedding from the previous layer. In contrast to shallow embeddings, GNNs can capture higher-order and non-linear patterns through multi-hop propagation within several layers of neural message passing.

Additionally, GNNs can optimize supervised signals and graph structure simultaneously, whereas shallow embedding approaches require a two-stage approach to achieve the same.

A myriad of GNN architectures define different messages, aggregation, and update schemes to derive deep graph embeddings [59, 73, 74, 75]. For example, [61, 76, 77, 78, 79] assign importance scores for nodes during neighborhood aggregation such that more important nodes play a larger effect in the embeddings. [80, 81] improve GNNs' ability to capture graph structural information by posing structural priors, such as a higher-order adjacency matrix. Graph pooling techniques [82] learn abstract topological structures. GNNs designed for molecules [83, 84] inject physics-based scores and domain knowledge into propagation layers.

As biomedical networks can be multimodal and massive, special consideration is needed to scale GNNs to large and heterogeneous networks. To this end, [85, 86] developed sampling strategies to intelligently select small subsets of the whole local network neighborhoods and use them in training GNN models. To tackle heterogeneous relations, [76, 87, 88] designed aggregation transformations to fuse diverse types of relations and attributes. Recent architectures describe dynamic message passing [76, 89, 90] to deal with evolving and time-varying graphs and few-shot learning [91] or self-supervised strategies [92, 93] to deal with graphs that are poorly annotated and have limited label information.

2.3   Generative graph models

Generative graph models generate new node and edge structures—even entire graphs —that are likely to have desired properties, such as novel molecules with acceptable toxicity profiles (Figure 2). Traditionally, network science models can generate graphs using deterministic or probabilistic rules. For instance, the Erdös-Rényi model [94] keeps iteratively adding random edges according to a predefined probability starting from an empty graph; the Barabási-Albert model [95] grows a graph by adding nodes and edges such that the resulting graph has a power-law degree distribution often observed in real-world networks; the configuration model [96] adds edges based on a predefined node degree sequences to generate graphs with arbitrary degree distributions. While powerful as random graph generators, such models cannot optimize graph structure based on properties of interest.

Deep generative models address the challenge by estimating distributional graph properties based on a dataset of graphs $\mathcal{G}$ and inferring graph structures using such optimized distributions. In particular, a generation graph model first learns a latent distribution $P(Z \mid \mathcal{G})$ that characterizes the input graph set $\mathcal{G}$. Then, conditioned on this distribution, it decodes a new graph, i.e., generates a new graph $\hat{G}$. There are different ways to encode the input graphs and learn the latent distribution, such as through variational autoencoders [64, 97, 98] and generative adversarial networks [99]. Decoding to generate a novel graph presents a unique challenge compared to an image or text since a graph is discrete, is unbounded in structure and size, and has no particular ordering of nodes. Common practices to generate new graphs include (1) predicting a probabilistic fully-connected graph and then using graph matching to find the optimal subgraph [100]; (2) decomposing a graph into a tree

of subgraphs structure and generating a tree structure instead, followed by assembles of subgraphs) [64]; and (3) sequentially sampling new nodes and edges [63, 65].

3 Application areas in biology and medicine

Biomedical data involve rich multimodal and heterogeneous interactions that span from molecular to societal levels (Figure 3). Unlike machine learning approaches designed to analyze data modalities like medical images and biological sequences, graph representation learning methods are uniquely able to leverage structural information in multimodal datasets [101].

Starting at the **molecular level** (Section 4), molecular structure is translated from atoms and bonds into nodes and edges, respectively. Physical interactions or functional relationships between proteins also naturally form a network. Given by the key organizing principles that govern network medicine—for instance, the local hypothesis and the shared components hypothesis [13]—whether an unknown protein clusters in a particular neighborhood of and shares direct neighbors with known proteins are informative of its binding affinity, function, etc [102]. Grounded in network medicine principles, graph machine learning is commonly applied to learn molecular representations of proteins and their physical interactions for predicting protein function.

At the **genomic level** (Section 5), genetic elements are incorporated into networks by extracting coding genes' co-expression information from transcriptomic data. Single-cell and spatial molecular profiling have further enabled the mapping of genetic interactions at the cellular and tissue level. Investigating the cellular circuitry of molecular functions in the resulting gene co-expression or regulatory networks can uncover disease mechanisms. For instance, as implicated by the network parsimony principle [13], shortest path length in a molecular network between disease-associated components often correlates with causal molecular pathways [16, 45]. Learned embeddings—using graph representation learning methods—that capture genome-wide interactions has enhanced disease predictions, even at the resolution of tissues and single cells.

At the **therapeutics level** (Section 6), networks are composed of drugs (e.g., small compounds), proteins, and diseases to allow the modeling of drug-drug interactions, binding of drugs to target proteins, and identification of drug repurposing opportunities. For example, by the corollary of the local hypothesis [13], the topology of drug combinations are indicative of synergistic or antagonistic relationships [103]. Learning the topology of graphs containing drug, protein, and disease nodes has improved predictions of candidate drugs for treating a disease, identification of potential off-target effects, and prioritization of novel drug combinations.

Finally, at the **healthcare level** (Section 7), patient records, such as medical images and EHRs, can be represented as networks and incorporated into protein, disease, and drug networks. As an example, given by three network medicine principles, the local hypothesis, the shared components hypothesis, and the disease module hypothesis [13], rare disease patients with common neighbors and topology likely have similar phenotypes and even

disease mechanisms [104, 105]. Graph representation learning methods have been successful in integrating patient records with molecular, genomic, and disease networks to personalize patient predictions.

4   Graph representation learning for molecules

Graph representation learning has been widely used for predicting protein interactions and function [106, 107]. Specifically, the inductive ability of graph convolution neural networks to generalize to data points unseen during model training (Section 2.2) and even generate new data points from scratch by decoding latent representation from the embedding space (Section 2.3) has enabled the discovery of new molecules, interactions, and functions [101, 108, 109].

4.1   Modeling protein molecular graphs

Computationally elucidating protein structure has been an ongoing challenge for decades [33]. Since protein structures are folded into complex 3D structures, it is natural to represent them as graphs. For example, we construct a contact distance graph where nodes are individual residues and edges are determined by a physical distance threshold [110]. Edges can also be defined by the ordering of amino acids in the primary sequence [110]. Spatial relationships between residues (e.g., distances, angles) may be used as features for edges [111].

We then model the 3D protein structures by capturing dependencies in their sequences of amino acids (e.g., applying GNNs (Section 2.2) to learn each node's local neighborhood structure) in order to generate protein embeddings [111, 112]. After learning short- and long-range dependencies across sequences corresponding to their 3D structures to produce embeddings of proteins, we can predict primary sequences from 3D structures [112]. Alternatively, we can use a hierarchical process of learning atom connectivity motifs to capture molecular structure at varying levels of granularity (e.g., at the motif-, connectivity-, and atomic-levels) in our protein embeddings, with which we can generate new 3D structures—a difficult task due to computational constraints of being both generalizable across different classes of molecules and flexible to a wide range of sizes [113]. As the field of machine learning for molecules is vast, we refer readers to existing reviews on molecular design [33, 114], graph generation [115], and molecular properties prediction [33, 34], and Section 6.1 on therapeutic compound design and generation.

4.2   Quantifying protein interactions

Many have integrated various data modalities, including chemical structure, binding affinities, physical and chemical principles, and amino acid sequences, to improve protein interaction quantification [33]. GNNs (as described in Section 4.1) are commonly used to generate representations of proteins based on chemical (e.g., free electrons' and protons donors' locations) and geometric (e.g., distance-dependent curvature) features to predict protein pocket-ligand and protein-protein interactions sites [116]; intra- and inter-molecule residue contact graphs to predict intra- and inter-molecular energies, binding affinities, and quality measures for a pair of molecular complexes [117]; and ligand- and receptor-protein

graphs to predict whether a pair of residues from the ligand and receptor proteins is a part of an interface [111]. Combining evolutionary, topological, and energetic information about molecules enables the scoring of docked conformations based on the similarity of random walks simulated on a pair of protein graphs (refer to graph kernel metrics in SI Note 3) [52].

Due to experimental and resource constraints, the most updated PPI networks are still limited in their number of nodes (proteins) and edges (physical interactions) [118]. Topology-based methods have been shown to capture and leverage the dynamics of biological systems to enrich existing PPI networks [119]. Predominant methods first apply graph convolutions (Section 2.2) to aggregate structural information in the graphs of interest (e.g., PPI networks, ligand-receptor networks), then use sequence modeling to learn the dependencies in amino acid sequences, and finally concatenate the two outputs for predicting the presence of physical interactions [101, 120]. Interestingly, such concatenated outputs have been treated as "image" inputs to CNNs [120], demonstrating the synergy of graph- and non-graph based machine learning methods. Similar graph convolution methods are also used to remove less credible interactions, thereby constructing a more reliable PPI network [121].

4.3 Interpreting protein functions and cellular phenotypes

Characterizing a protein's function in specific biological contexts is a challenging and experimentally intensive task [122, 123]. However, innovating graph representation learning techniques to represent protein structures and interactions has facilitated protein function prediction [124], especially when we leverage existing gene ontologies and transcriptomic data.

Gene Ontology (GO) terms [125] are a standardized vocabulary for describing molecular functions, biological processes, and cellular locations of gene products [126]. They have been built as a hierarchical graph that GNNs then leverage to learn dependencies of the terms [126], or directly used as protein function labels [106, 127]. In the latter case, we typically construct sequence similarity networks, combine them with PPI networks, and integrate protein features (e.g. amino acid sequence, protein domains, subcellular location, gene expression profiles) to predict protein function [106, 127]. Others have even created gene interaction networks using transcriptomic data [108, 128] to capture context-specific interactions between genes, which PPI networks lack.

Alternative graph representation learning methods for predicting protein function include defining diffusion-based distance metrics on PPI networks for predicting protein function [129]; using the theory of topological persistence to compute signatures of a protein based on its 3D structure [130]; and applying topological data analysis to extract features from protein contact networks created from 3D coordinates [131] (SI Note 3). Many have also adopted an attention mechanism for protein sequence embeddings generated by a Bidirectional Encoder Representations from Transformers (BERT) model to enable interpretability [132, 133], showcasing the synergy of graph-based and language models.

5 Graph representation learning for genomics

Diseases are classified based on the presenting symptoms of patients, which can be caused by molecular dysfunctions, such as genetic mutations. As a result, diagnosing diseases requires knowledge about alterations in the transcription of coding genes to capture genome-wide associations driving disease acquisition and progression. Graph representation learning methods allow us to analyze heterogeneous networks of multimodal data and make predictions across domains, from genomic level data (e.g., gene expression, copy number information) to clinically relevant data (e.g., pathophysiology, tissue localization).

5.1 Leveraging gene expression measurements

Comparing transcriptomic profiles from healthy individuals to those of patients with a specific disease informs clinicians of its causal genes. As gene expression is the direct readout of perturbation effects, changes in gene expression are often used to model disease-specific co-expression or regulatory interactions between genes. Further, injecting gene expression data into PPI networks has identified disease biomarkers, which are then used to more accurately classify diseases of interest.

Methods that rely solely on gene expression data typically transform the co-expression matrix into a more topologically meaningful form [146, 147, 148]. Gene expression data can be transformed into a colored graph that captures the shape of the data (e.g., using TDA [148]; refer to SI Note 3), which then enables downstream analysis using network science metrics and graph machine learning. Topological landscapes present in gene expression data can be vectorized and fed into a GCN to classify the disease type [147]. Alternatively, gene expression data can directly be used to construct disease and gene networks that are then input into a joint matrix factorization and GCN method to draw disease-gene associations, akin to a recommendation task (Section 2.2) [146]. Further, applying a joint GCN, VAE, and GAN framework (Section 2.3) to gene correlation networks—initialized with a subset of gene expression matrices—can generate disease networks with the desired properties [149].

Because gene expression data can be noisy and variational, recent advances include fusing the co-expression matrices with existing biomedical networks, such as GO annotations and PPI, and feeding the resulting graph into graph convolution layers (Section 2.2) [150, 151, 152]. Doing so has enabled more interpretable disease classification models (e.g., weighting gene interactions based on existing biological knowledge). However, despite the utility of PPI networks, they have been reported to limit models trained solely on PPI networks because they are unable to capture all gene regulatory activities [153]. To this end, graph representation learning methods, such as GNNs, have been developed to learn robust and meaningful representations of molecules despite the incomplete interactome [102] and to inductively infer new edges between pairs of nodes [154].

5.2 Injecting single cell and spatial information into molecular networks

Single-cell RNA sequencing (scRNA-seq) data lend themselves to graph representation learning to model cellular differential processes [136, 155] and disease states [140]. In particular, a predominant approach to analyze scRNA-seq datasets is to transform them into

gene similarity networks, such as gene co-expression networks, or cell similarity networks by correlating gene expression readouts across individual cells. Applied to such networks, graph representation learning can impute scRNA-seq data [156, 157], predict cell clusters [157, 158], etc. Cell similarity graphs have also been created using autoencoders by first embedding gene expression readouts and then connecting genes based on how similar their embeddings are [157]. Alternatively, variational graph autoencoders produce cell embeddings and interpretable attention weights indicating what genes the model attends to when deriving an embedding for a given cell [159]. Beyond GNNs and graph autoencoders, learning a manifold over a cell state space can quantify the effects of experimental perturbations [155]. To this end, cell similarity graphs are constructed for control and treated samples and used to estimate the likelihood of a cell population observed under a given perturbation [155].

Spatial molecular profiling can measure both gene expression at the cellular level and location of cells in a tissue [160]. As a result, spatial transcriptomics data can be used to construct cell graphs [161], spatial gene expression graphs [162], gene co-expression networks, or molecular similarity graphs [35]. Creating cell neighborhood and spatial gene expression graphs require a distance metric, as edges are determined based on spatial proximity, while gene co-expression and molecular similarity graphs need a threshold applied on the gene expression data [35]. From such networks, graph representation learning methods produce embeddings that capture the network topology and that can be further optimized for downstream tasks. For instance, a cell neighborhood graph and a gene pair expression matrix enable classic GNNs to predict ligand-receptor interactions [161]. In fact, as ligand-receptor interactions are directed, they could be used to infer causal interactions of previously unknown ligand-receptor pairs [161, 163].

6 Graph representation learning 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 [13, 14, 32, 164]. 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 clinically relevant data (e.g., 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 [6, 38, 39, 40, 48, 165].

6.1 Modeling compound molecular graphs

Similar to proteins, small compounds are modeled as 2D and 3D molecular graphs such that nodes are atoms and edges are bonds. Each atom and bond may include features, such as atomic mass, atomic number, and bond type, to be included in the model [83, 166]. Edges

can also be added to indicate pairwise spatial distance between two atoms [72], or directed with information on bond angles and rotations incorporated into the molecular graph [84].

Representing molecules as graphs has improved predictions on various quantum chemistry properties. Intuitively, message passing steps (i.e., in GNNs (Section 2.2)) aggregate information from neighboring atoms and bonds to learn the local chemistry of each atom [166]. For example, generating representations of the atoms, distances, and angles to be propagated along the molecular graph has allowed us to identify the angle and direction of interactions between atoms [84]. Producing atom-centered representations based on a weighted combination of their neighbors' features (i.e., using an attention mechanism) is able to model interactions among reactants for predicting organic reaction outcomes [167]. Alternatively, molecular graphs have been decomposed into a "junction tree," where each node represents a substructure in the molecule, to learn representations of both the molecular graph and the junction tree for generating new molecules with desirable properties (Section 2.3) [64]. Due to the major challenge of finding novel and diverse compounds with specific chemical properties, iteratively editing fragments of a molecular graph during training has improved predictions for high-quality candidates targeting our proteins of interest [168].

6.2 Quantifying drug-drug and drug-target 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 [34]. However, quantifying these metrics requires labor- and cost-intensive experiments [33, 34]. Modeling small compounds' and protein targets' molecular structure as well as their binding affinities and specificity, for instance, using graph representation learning has enabled accelerated investigation of interactions between a given drug and protein target.

First, we learn representations of drugs and targets using graph-based methods, such as TDA [169] or shallow network embedding approaches [56]. Concretely, TDA (refer to SI Note 3) transforms experimental data into a graph where nodes represent compounds and edges indicate a level of similarity between them [169]. Shallow network embedding techniques are also used to generate embeddings for drugs and targets by computing drug-drug, drug-target, and target-target similarities (Section 2.1) [170]. Non-graph based methods have also been used to construct graphs that are then fed into a graph representation learning model to generate embeddings. K-nearest neighbors, for instance, is a common used method to construct drug and target similarity networks [171]. The resulting embeddings are fed into downstream machine learning models.

Fusing compound sequence, structure, and clinical implications has significantly improved drug-drug and drug-target interaction predictions. For example, attention mechanisms have been applied on drug graphs, with chemical structures and side effects as features, to generate interpretable predictions of drug-drug interactions [172]. Also, two separate GNNs may be used to learn representations of protein and small molecule graphs for predicting drug-target affinity [173]. To be flexible with other graph- and non-graph-based methods, protein structure representations generated by graph convolutions has been combined with

protein sequence representations (e.g., shallow network embedding methods or CNNs) to predict the probability of compound-protein interactions [174, 175, 176, 177].

6.3 Identifying drug-disease associations and biomarkers for complex disease

Part of the drug discovery pipeline is minimizing adverse drug events [33, 34]. But, in addition to high financial cost, the experiments required to measure drug-drug interactions and toxicity face a combinatorial explosion problem [33]. Graph representation learning methods enables *in silico* modeling of drug action, which allows for more efficient ranking of candidate drugs for repurposing, such as by considering gene expression data, gene ontologies, drug similarity, and other clinically relevant data regarding side effects and indications.

Drug and disease representations have been learned on homogeneous graphs of drugs, diseases, or targets. For instance, Medical Subject Headings (MeSH) terms may be used to construct a drug-disease graph, from which latent representations of drugs and diseases are learned using various graph embedding algorithms, including DeepWalk and LINE (Section 2.1) [178]. TDA (refer to SI Note 3) has also been applied to construct graphs of drugs, targets, and diseases separately, from which representations of such entities are learned and optimized for downstream prediction [179].

To emphasize the systems-level complexity of diseases, recent methods fuse multimodal data to generate heterogeneous graphs. For example, neighborhood information are aggregated from heterogeneous networks comprised of drug, target, and disease information to predict drug-target interactions (Section 2.2) [180]. In other instances, PPI networks are combined with genomic features to predict drug sensitivity using GNNs [181]. As a result, approaches that integrate cross-domain knowledge as a vast heterogeneous network and/or into the model's architecture seem better equipped to elucidate drug action.

7 Graph representation learning for healthcare

Graph representation learning has been used to fuse multimodal knowledge with patient records to better enable precision medicine. Two modes of patient data successfully integrated using deep graph learning are histopathological images [8, 185] and EHRs [186, 187].

7.1 Leveraging networks for diagnostic imaging

Medical images of patients, including histopathology slides, enable clinicians to comprehensively observe the effects of a disease on the patient's body [188]. Medical images, such as large whole histopathology slides, are typically converted into cell spatial graphs, where nodes represent cells in the image and edges indicate that a pair of cells is adjacent in space. Deep graph learning has been shown to detect subtle signs of disease progression in the images while integrating other modalities (e.g., tissue localization [189] and genomic features [8]) to improve medical image processing.

Cell-tissue graphs generated from histopathological images are able to encode the spatial context of cells and tissues for a given patient. Cell morphology and tissue micro-

architecture information can be aggregated from cell graphs to grade cancer histology images (e.g., using GNNs (Section 2.2)) [8, 190, 191, 192]. An example aggregation method is pooling with an attention mechanism to infer relevant patches in the image [190]. Further, cell morphology and interactions, tissue morphology and spatial distribution, cell-to-tissue hierarchies, and spatial distribution of cells with respect to tissues can be captured in a cell-to-tissue graph, upon which a hierarchical GNN can learn representations using these different data modalities [189]. Because interpretability is critical for models aimed to generate patient-specific predictions, post-hoc graph pruning optimization may be performed on a cell graph generated from a histopathology image to define subgraphs explaining the original cell graph analysis [193].

Graph representation learning methods have also been proven successful for classifying other types of medical images. GNNs are able to model relationships between lymph nodes to compute the spread of lymph node gross tumor volume based on radiotherapy CT images (Section 2.2) [194]. MRI images can be converted into graphs that GNNs are applied to for classifying the progression of Alzheimer's Disease [195, 196, 197]. GNNs are also shown to leverage relational structures like similarities among chest X-rays to improve downstream tasks, such as disease diagnosis and localization [198]. Alternatively, TDA can generate graphs of whole-slide images, which include tissues from various patient sources (SI Note 3), and GNNs are then used to classify the stage of colon cancer [199].

Further, spatial molecular profiling benefits from methodological advancements made for medical images. With spatial gene expression graphs (weighted and undirected) and corresponding histopathology images, gene expression information are aggregated to generate embeddings of genes that could then be used to investigate spatial domains (e.g., differentiate between cancer and noncancer regions in tissues) [200]. Since multimodal data enables more robust predictions, GNNs (Section 2.2) have been applied to generate cell spatial graphs from histopathology images and then fuse genomic and transcriptomic data for predicting treatment response and resistance, histopathology grading, and patient survival [8].

7.2 Personalizing medical knowledge networks with patient records

Electronic health records are typically represented by ICD (International Classification of Disease) codes [186, 187]. 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. By integrating patient records into our networks, graph representation learning is well-equipped to advance precision medicine by generating predictions tailored to individual patients.

Methods that embed medical entities, including EHRs and medical ontologies, leverage the inherently hierarchical structure in the medical concepts KG [201]. Low dimensional embeddings of EHR data can be generated by separately considering medical services, doctors, and patients in shallow network embeddings (Section 2.1) and graph neural networks (Section 2.2) [202, 203]. Alternatively, attention mechanisms may be applied on EHR data and medical ontologies to capture the parent-child relationships [186, 204, 205].

Rather than assuming a certain structure in the EHRs, a Graph Convolution Transformer can even learn the hidden EHR structure [79].

EHRs also have underlying spatial and/or temporal dependencies [206] that many methods have recently taken advantage of to perform time-dependent prediction tasks. A mixed pooling multi-view self-attention autoencoder has generated patient representations for predicting either a patient's risk of developing a disease in a future visit, or the diagnostic codes of the next visit [207]. A combined LSTM and GNN model has also been used to represent patient status sequences and temporal medical event graphs, respectively, to predict future prescriptions or disease codes (Section 2.2) [208, 209]. Alternatively, a patient graph may be constructed based on the similarity of patients, and patient embeddings learned by an LSTM-GNN architecture are optimized to predict patient outcomes [210]. An ST-GCN [78] is designed to utilize the underlying spatial and temporal dependencies of EHR data for generating patient diagnoses [187].

EHRs are often supplemented with other modalities, such as diseases, symptoms, molecular data, drug interactions, etc [7, 206, 211, 212]. A probabilistic KG of EHR data, which include medical history, drug prescriptions, and laboratory examination results, has been used to consider the semantic relations between EHR entities in a shallow network embedding method (Section 2.1) [213]. Meta-paths may alternatively be exploited in an EHR-derived KG to leverage higher order, semantically important relations for disease classification [214]. Node features for drugs and diseases can be initialized using Skipgram and then a GNN leveraging multi-layer message-passing can be applied to predict adverse drug events [211]. Moreover, combined RNNs and GNNs models have been applied to EHR data integrated with drug and disease interactions to better recommend medication combinations [215].

8   Outlook

As graph representation learning has aided in the mapping of genotypes to phenotypes, leveraging graph representation learning for fine-scale mapping of variants is a promising new direction [217]. By re-imagining GWAS and expression Quantitative Trait Loci (eQTL) studies [218] as networks, we can already begin to discover biologically meaningful modules to highlight key genes involved in the underlying mechanisms of a disease [219]. We can alternatively seed network propagation with QTL candidate genes [217]. 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 [220, 221]. We could even reconstruct 3D chromosomal structures by predicting 3D coordinates for nodes derived from a Hi-C contact map [222]. Further, as spatial molecular profiling has enabled profound discoveries for diseases, graph representation learning repertoire to analyze such datasets will continue to expand. For instance, with dynamic GNNs, we may be able to better capture changes in expression levels observed in single cell RNA sequencing data over time or as a result of a perturbation [155, 223].

Effective integration of healthcare data with 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 [224]. Given the utility of graphs in both the biological and medical domains, there has been a major push to generate knowledge graphs that synthesize and model multi-scaled, multi-modal data, from genotype-phenotype associations to population-scale epidemiological dynamics. In public health, spatial and temporal networks can model space- and time-dependent observations (e.g., disease states, susceptibility to infection [225]) to spot trends, detect anomalies, and interpret temporal dynamics.

As artificial intelligence tools implementing graph representation learning algorithms are increasingly employed in clinical applications, it is essential to ensure that representations are explainable [226], fair [227], and robust [228], and that existing tools are revisited in light of algorithmic bias and health disparities [229].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

[1]. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C, and Kannan S. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. Cell Systems, 2020.

[2]. Nicholson DN and Greene CS. Constructing knowledge graphs and their biomedical applications. Computational and Structural Biotechnology Journal, 2020.

[3]. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, and Mundlos S. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. The American Journal of Human Genetics, 2008.

[4]. Schriml LM, Arze C, Nadendla S, Chang Y-WW, Mazaitis M, Felix V, Feng G, and Kibbe WA. Disease ontology: a backbone for disease semantic integration. Nucleic Acids Research, 2012.

[5]. Hong C, Rush E, Liu M, Zhou D, Sun J, Sonabend A, Castro VM, Schubert P, Panickan VA, Cai T, et al. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. medRxiv, 2021.

[6]. Gysi DM, Do Valle Í, Zitnik M, Ameli A, Gan X, Varol O, Ghiassian SD, Patten J, Davey RA, Loscalzo J, et al. Network medicine framework for identifying drug-repurposing opportunities for covid-19. PNAS, 2021.

[7]. Nelson CA, Butte AJ, and Baranzini SE. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. Nature Communications, 2019.

[8]. Chen RJ, Lu MY, Wang J, Williamson DF, Rodig SJ, Lindeman NI, and Mahmood F. Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging, 2020.

[9]. Callahan TJ, Tripodi IJ, Pielke-Lombardo H, and Hunter LE. Knowledge-based biomedical data science. Annual Review of Biomedical Data Science, 2020.

[10]. Barabási A-L. Network medicine — from obesity to the "diseasome". New England Journal of Medicine, 2007.

[11]. Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, Carbon S, Conlin T, Dunn N, Engelstad M, et al. The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Research, 2017.

[12]. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, and Barabási A-L. The human disease network. PNAS, 2007.

[13]. Barabási A-L, Gulbahce N, and Loscalzo J. Network medicine: a network-based approach to human disease. Nature Reviews Genetics, 2011.

[14]. Hu JX, Thomas CE, and Brunak S. Network biology concepts in complex disease comorbidities. Nature Reviews Genetics, 2016.

[15]. Zitnik M, Feldman MW, Leskovec J, et al. Evolution of resilience in protein interactomes across the tree of life. PNAS, 2019.

[16]. Agrawal M, Zitnik M, and Leskovec J. Large-scale analysis of disease pathways in the human interactome. In Pacific Symposium on Biocomputing, 2018.

[17]. Camacho DM, Collins KM, Powers RK, Costello JC, and Collins JJ. Next-generation machine learning for biological networks. Cell, 2018.

[18]. Zhang Z, Cui P, and Zhu W. Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering, 2020.

[19]. Hamilton WL, Ying R, and Leskovec J. Representation learning on graphs: Methods and applications. IEEE Data Engineering Bulletin, 2017.

[20]. Hamilton WL. Graph representation learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2020.

[21]. Wu Z, Pan S, Chen F, Long G, Zhang C, and Philip SY. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2020.

[22]. Chen F, Wang Y-C, Wang B, and Kuo C-CJ. Graph representation learning: a survey. APSIPA Transactions on Signal and Information Processing, 2020.

[23]. Li B and Pi D. Network representation learning: a systematic literature review. Neural Computing and Applications, 2020.

[24]. Yue X, Wang Z, Huang J, Parthasarathy S, Moosavinasab S, Huang Y, Lin SM, Zhang W, Zhang P, and Sun H. Graph embedding on biomedical networks: methods, applications and evaluations. Bioinformatics, 2020.

[25]. Dong Y, Hu Z, Wang K, Sun Y, and Tang J. Heterogeneous network representation learning. In IJCAI, 2020.

[26]. Kazemi SM, Goel R, Jain K, Kobyzev I, Sethi A, Forsyth P, and Poupart P. Representation learning for dynamic graphs: A survey. JMLR, 2020.

[27]. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, and Hoffman MM. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. Information Fusion, 2019.

[28]. Cowen L, Ideker T, Raphael BJ, and Sharan R. Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics, 2017.

[29]. Blevins AS and Bassett DS. Topology in biology. Handbook of the Mathematics of the Arts and Sciences, 2020.

[30]. Koutrouli M, Karatzas E, Paez-Espino D, and Pavlopoulos GA. A guide to conquer the biological network era using graph theory. Frontiers in Bioengineering and Biotechnology, 2020.

[31]. Liu C, Ma Y, Zhao J, Nussinov R, Zhang Y-C, Cheng F, and Zhang Z-K. Computational network biology: Data, models, and applications. Physics Reports, 2020.

[32]. Rai A, Shinde P, and Jalan S. Network spectra for drug-target identification in complex diseases: new guns against old foes. Applied Network Science, 2018.

[33]. David L, Thakkar A, Mercado R, and Engkvist O. Molecular representations in ai-driven drug discovery: a review and practical guide. Journal of Cheminformatics, 2020.

[34]. Wieder O, Kohlbacher S, Kuenemann M, Garon A, Ducrot P, Seidel T, and Langer T. A compact review of molecular property prediction with graph neural networks. Drug Discovery Today: Technologies, 2020.

[35]. Hetzel L, Fischer DS, Günnemann S, and Theis FJ. Graph representation learning for single cell biology. Current Opinion in Systems Biology, 2021.

[36]. Jiménez-Luna J, Grisoni F, and Schneider G. Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2020.

[37]. Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, and Wang F. Graph convolutional networks for computational drug development and discovery. Briefings in Bioinformatics, 2020.

[38]. Gaudelet T, Day B, Jamasb AR, Soman J, Regep C, Liu G, Hayter JB, Vickers R, Roberts C, Tang J, et al. Utilizing graph machine learning within drug discovery and development. Briefings in Bioinformatics, 2021.

[39]. MacLean F. Knowledge graphs and their applications in drug discovery. Expert Opinion on Drug Discovery, 2021.

[40]. Zeng X, Tu X, Liu Y, Fu X, and Su Y. Toward better drug discovery with knowledge graph. Current Opinion in Structural Biology, 2022.

[41]. Ahmedt-Aristizabal D, Armin MA, Denman S, Fookes C, and Petersson L. A survey on graph-based deep learning for computational histopathology. Computerized Medical Imaging and Graphics, 2021.

[42]. Muzio G, O'Bray L, and Borgwardt K. Biological network analysis with deep learning. Briefings in Bioinformatics, 2021.

[43]. Guo M, Yu Y, Wen T, Zhang X, Liu B, Zhang J, Zhang R, Zhang Y, and Zhou X. Analysis of disease comorbidity patterns in a large-scale china population. BMC Medical Genomics, 2019.

[44]. Le D-H and Dang V-T. Ontology-based disease similarity network for disease gene prediction. Vietnam Journal of Computer Science, 2016.

[45]. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, and Barabási A-L. Uncovering disease-disease relationships through the incomplete interactome. Science, 2015.

[46]. Sumathipala M, Maiorino E, Weiss ST, and Sharma A. Network diffusion approach to predict lncrna disease associations using multi-type biological networks: Lion. Frontiers in Physiology, 2019.

[47]. Cheng F, Kovács IA, and Barabási A-L. Network-based prediction of drug combinations. Nature Communications, 2019.

[48]. Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, Wang R, Zhao Y, Yang Y, Huang J, et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. Nature Communications, 2019.

[49]. Chen Z-H, You Z-H, Guo Z-H, Yi H-C, Luo G-X, and Wang Y-B. Prediction of drug–target interactions from multi-molecular network based on deep walk embedding model. Frontiers in Bioengineering and Biotechnology, 2020.

[50]. Wong L, You Z-H, Guo Z-H, Yi H-C, Chen Z-H, and Cao M-Y. Mipdh: A novel computational model for predicting microRNA–mRNA interactions by DeepWalk on a heterogeneous network. ACS Omega, 2020.

[51]. Yang K, Wang R, Liu G, Shu Z, Wang N, Zhang R, Yu J, Chen J, Li X, and Zhou X. Hergepred: heterogeneous network embedding representation for disease gene prediction. IEEE Journal of Biomedical and Health Informatics, 2018.

[52]. Geng C, Jung Y, Renaud N, Honavar V, Bonvin AM, and Xue LC. iscore: a novel graph kernel-based function for scoring protein–protein docking models. Bioinformatics, 2020.

[53]. Veselkov K, Gonzalez G, Aljifri S, Galea D, Mirnezami R, Youssef J, Bronstein M, and Laponogov I. Hyperfoods: Machine intelligent mapping of cancer-beating molecules in foods. Scientific Reports, 2019.

[54]. Zheng A and Casari A. Feature engineering for machine learning: principles and techniques for data scientists. "O'Reilly Media, Inc." 2018.

[55]. Perozzi B, Al-Rfou R, and Skiena S. Deepwalk: online learning of social representations. In KDD, 2014.

[56]. Grover A and Leskovec J. node2vec: Scalable feature learning for networks. In KDD, 2016.

[57]. Tang J, Qu M, Wang M, Zhang M, Yan J, and Mei Q. LINE: Large-scale information network embedding. In International Conference on World Wide Web, pages 1067–1077, 2015.

[58]. Dong Y, Chawla NV, and Swami A. metapath2vec: Scalable representation learning for heterogeneous networks. In KDD, 2017.

[59]. Kipf TN and Welling M. Semi-supervised classification with graph convolutional networks. In ICLR, 2017.

[60]. Xu K, Hu W, Leskovec J, and Jegelka S. How powerful are graph neural networks? In ICLR, 2019.

[61]. Velickovic P, Cucurull G, Casanova A, Romero A, Liò P, and Bengio Y. Graph attention networks. In ICLR, 2018.

[62]. Xu K, Li C, Tian Y, Sonobe T, Kawarabayashi K.-i., and Jegelka S. Representation learning on graphs with jumping knowledge networks. In International Conference on Machine Learning, pages 5453–5462, 2018.

[63]. You J, Liu B, Ying R, Pande V, and Leskovec J. Graph convolutional policy network for goal-directed molecular graph generation. NeurIPS, 2018.

[64]. Jin W, Barzilay R, and Jaakkola TS. Junction tree variational autoencoder for molecular graph generation. In ICML, 2018.

[65]. You J, Ying R, Ren X, Hamilton WL, and Leskovec J. Graphrnn: Generating realistic graphs with deep auto-regressive models. In ICML, 2018.

[66]. Mikolov T, Sutskever I, Chen K, Corrado GS, and Dean J. Distributed representations of words and phrases and their compositionality. In NeurIPS, 2013.

[67]. Bordes A, Usunier N, García-Durán A, Weston J, and Yakhnenko O. Translating embeddings for modeling multi-relational data. In NeurIPS, 2013.

[68]. Nickel M, Tresp V, and Kriegel H. A three-way model for collective learning on multi-relational data. In ICML, 2011.

[69]. Trouillon T, Welbl J, Riedel S, Gaussier É, and Bouchard G. Complex embeddings for simple link prediction. In ICLR, 2016.

[70]. Sun Z, Deng Z, Nie J, and Tang J. Rotate: Knowledge graph embedding by relational rotation in complex space. In ICLR, 2019.

[71]. Yang B, Yih W, He X, Gao J, and Deng L. Embedding entities and relations for learning and inference in knowledge bases. In ICLR, 2015.

[72]. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, and Dahl GE. Neural message passing for quantum chemistry. In ICML, 2017.

[73]. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, Aspuru-Guzik A, and Adams RP. Convolutional networks on graphs for learning molecular fingerprints. In NeurIPS, 2015.

[74]. Vinyals O, Bengio S, and Kudlur M. Order matters: Sequence to sequence for sets. In ICLR, 2016.

[75]. Defferrard M, Bresson X, and Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In NeurIPS, 2016.

[76]. Hu Z, Dong Y, Wang K, and Sun Y. Heterogeneous graph transformer. In WWW, 2020.

[77]. Yun S, Jeong M, Kim R, Kang J, and Kim HJ. Graph transformer networks. In NeurIPS, 2019.

[78]. Yan S, Xiong Y, and Lin D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In AAAI, 2018.

[79]. Choi E, Xu Z, Li Y, Dusenberry M, Flores G, Xue E, and Dai AM. Learning the graphical structure of electronic health records with graph convolutional transformer. In AAAI, 2020.

[80]. Xu K, Li C, Tian Y, Sonobe T, Kawarabayashi K, and Jegelka S. Representation learning on graphs with jumping knowledge networks. In ICML, 2018.

[81]. Abu-El-Haija S, Perozzi B, Kapoor A, Alipourfard N, Lerman K, Harutyunyan H, Steeg GV, and Galstyan A. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In ICML, 2019.

[82]. Ying Z, You J, Morris C, Ren X, Hamilton WL, and Leskovec J. Hierarchical graph representation learning with differentiable pooling. In NeurIPS, 2018.

[83]. Schütt K, Kindermans P, Felix HES, Chmiela S, Tkatchenko A, and Müller K. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In NeurIPS, 2017.

[84]. Klicpera J, Groß J, and Günnemann S. Directional message passing for molecular graphs. In ICLR, 2020.

[85]. Chiang W, Liu X, Si S, Li Y, Bengio S, and Hsieh C. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In KDD, 2019.

[86]. Zeng H, Zhou H, Srivastava A, Kannan R, and Prasanna VK. Graphsaint: Graph sampling based inductive learning method. In ICLR, 2020.

[87]. Schlichtkrull M, Kipf TN, Bloem P, Van Den Berg R, Titov I, and Welling M. Modeling relational data with graph convolutional networks. In ESWC, 2018.

[88]. Wang X, Ji H, Shi C, Wang B, Ye Y, Cui P, and Yu PS. Heterogeneous graph attention network. In WWW, 2019.

[89]. Pareja A, Domeniconi G, Chen J, Ma T, Suzumura T, Kanezashi H, Kaler T, Schardl TB, and Leiserson CE. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In AAAI, 2020.

[90]. Rossi E, Chamberlain B, Frasca F, Eynard D, Monti F, and Bronstein M. Temporal graph networks for deep learning on dynamic graphs. ICML Workshop on Graph Representation Learning and Beyond, 2020.

[91]. Huang K and Zitnik M. Graph meta learning via local subgraphs. NeurIPS, 33, 2020.

[92]. Hu W, Liu B, Gomes J, Zitnik M, Liang P, Pande VS, and Leskovec J. Strategies for pre-training graph neural networks. In ICLR, 2020.

[93]. You Y, Chen T, Wang Z, and Shen Y. When does self-supervision help graph convolutional networks? In ICML, 2020.

[94]. Erdös P and Rényi A. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci, 1960.

[95]. Albert R and Barabási A-L. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002.

[96]. Barabási A-L et al. Network Science. 2016.

[97]. Kipf TN and Welling M. Variational graph auto-encoders. NeurIPS Workshop, 2016.

[98]. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, Aguilera-Iparraguirre J, Hirzel TD, Adams RP, and Aspuru-Guzik A. Automatic chemical design using a data-driven continuous representation of molecules. ACS Central Science, 2018.

[99]. Wang H, Wang J, Wang J, Zhao M, Zhang W, Zhang F, Xie X, and Guo M. Graphgan: Graph representation learning with generative adversarial nets. In AAAI, 2018.

[100]. Simonovsky M and Komodakis N. Graphvae: Towards generation of small graphs using variational autoencoders. In ICANN, pages 412–422. Springer, 2018.

[101]. Yang F, Fan K, Song D, and Lin H. Graph-based prediction of protein-protein interactions with attributed signed graph embedding. BMC Bioinformatics, 2020.

[102]. Huang K, Xiao C, Glass LM, Zitnik M, and Sun J. SkipGNN: predicting molecular interactions with skip-graph networks. Scientific Reports, 2020.

[103]. Yin N, Ma W, Pei J, Ouyang Q, Tang C, and Lai L. Synergistic and antagonistic drug combinations depend on network topology. PloS One, 2014.

[104]. Alsentzer E, Finlayson SG, Li MM, and Zitnik M. Subgraph neural networks. In NeurIPS, 2020.

[105]. Buphamalai P, Kokotovic T, Nagy V, and Menche J. Network analysis reveals rare disease signatures across multiple levels of biological organization. Nature Communications, 2021.

[106]. Fan K and Zhang Y. Pseudo2go: A graph-based deep learning method for pseudogene function prediction by borrowing information from coding genes. Frontiers in Genetics, 2020.

[107]. Kearnes S, McCloskey K, Berndl M, Pande V, and Riley P. Molecular graph convolutions: moving beyond fingerprints. Journal of Computer-Aided Molecular Design, 2016.

[108]. Dutil F, Cohen JP, Weiss M, Derevyanko G, and Bengio Y. Towards gene expression convolutions using gene interaction graphs. ICML Workshop on Computational Biology, 2018.

[109]. Hamilton WL, Ying Z, and Leskovec J. Inductive representation learning on large graphs. In NeurIPS, 2017.

[110]. Huan J, Bandyopadhyay D, Wang W, Snoeyink J, Prins J, and Tropsha A. Comparing graph representations of protein structure for mining family-specific residue-based packing motifs. Journal of Computational Biology, 2005.

[111]. Fout A, Byrd J, Shariat B, and Ben-Hur A. Protein interface prediction using graph convolutional networks. In NeurIPS, 2017.

[112]. Ingraham J, Garg VK, Barzilay R, and Jaakkola TS. Generative models for graph-based protein design. In NeurIPS, 2019.

[113]. Jin W, Barzilay R, and Jaakkola T. Hierarchical generation of molecular graphs using structural motifs. In ICML, 2020.

[114]. Elton DC, Boukouvalas Z, Fuge MD, and Chung PW. Deep learning for molecular design—a review of the state of the art. Molecular Systems Design & Engineering, 2019.

[115]. Guo X and Zhao L. A systematic survey on deep generative models for graph generation. ACM, 2020.

[116]. Gainza P, Sverrisson F, Monti F, Rodola E, Boscaini D, Bronstein M, and Correia B. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. Nature Methods, 2020.

[117]. Cao Y and Shen Y. Energy-based graph convolutional networks for scoring protein docking models. Proteins: Structure, Function, and Bioinformatics, 2020.

[118]. Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, et al. A reference map of the human binary protein interactome. Nature, 2020.

[119]. Li D and Gao J. Towards perturbation prediction of biological networks using deep learning. Scientific Reports, 2019.

[120]. Liu Y, Yuan H, Cai L, and Ji S. Deep learning of high-order interactions for protein interface prediction. In KDD, 2020.

[121]. Yao H, Guan J, and Liu T. Denoising protein-protein interaction network via variational graph auto-encoder for protein complex detection. Journal of Bioinformatics and Computational Biology, 2020.

[122]. Moreau Y and Tranchevent L-C. Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nature Reviews Genetics, 2012.

[123]. Zitnik M, Nam EA, Dinh C, Kuspa A, Shaulsky G, and Zupan B. Gene prioritization by compressive data fusion and chaining. PLOS Computational Biology, 2015.

[124]. Zhou H, Beltran F, and Brito IL. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. Science Advances, 7, 2021.

[125]. G. O. Consortium. The Gene Ontology resource: 20 years and still GOing strong. Nucleic Acids Research, 2019.

[126]. Zhou G, Wang J, Zhang X, and Yu G. Deepgoa: Predicting gene ontology annotations of proteins via graph convolutional network. In BIBM, 2019.

[127]. Fan K, Guan Y, and Zhang Y. Graph2go: a multi-modal attributed network embedding method for inferring protein functions. GigaScience, 2020.

[128]. Hasibi R and Michoel T. Predicting gene expression from network topology using graph neural networks. arXiv:2005.03961, 2020.

[129]. Cao M, Zhang H, Park J, Daniels NM, Crovella ME, Cowen LJ, and Hescott B. Going the distance for protein function prediction: a new distance metric for protein interaction networks. PLOS One, 2013.

[130]. Dey TK and Mandal S. Protein classification with improved topological data analysis. In WABI, 2018.

[131]. Martino A, Rizzi A, and Mascioli FMF. Supervised approaches for protein function prediction by topological data analysis. In IJCNN, 2018.

[132]. Nambiar A, Heflin M, Liu S, Maslov S, Hopkins M, and Ritz A. Transforming the language of life: Transformer neural networks for protein prediction tasks. In Proceedings of the 11th ACM-BCB, 2020.

[133]. Vig J, Madani A, Varshney LR, Xiong C, Socher R, and Rajani NF. Bertology meets biology: Interpreting attention in protein language models. ICLR, 2021.

[134]. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, and Smibert P. Simultaneous epitope and transcriptome measurement in single cells. Nature Methods, 2017.

[135]. Kondratova M, Czerwinska U, Sompairac N, Amigorena SD, Soumelis V, Barillot E, Zinovyev A, and Kuperstein I. A multiscale signalling network map of innate immune response in cancer reveals cell heterogeneity signatures. Nature Communications, 2019.

[136]. Rizvi AH, Camara PG, Kandror EK, Roberts TJ, Schieren I, Maniatis T, and Rabadan R. Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development. Nature Biotechnology, 2017.

[137]. Mohammadi S, Davila-Velderrain J, and Kellis M. Reconstruction of cell-type-specific interactomes at single-cell resolution. Cell Systems, 2019.

[138]. Li MM and Zitnik M. Deep contextual learners for protein networks. ICML Workshop on Computational Biology, 2021.

[139]. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, and Tyers M. Biogrid: a general repository for interaction datasets. Nucleic Acids Research, 2006.

[140]. Ravindra N, Sehanobish A, Pappalardo JL, Hafler DA, and van Dijk D. Disease state prediction from single-cell data using graph attention networks. In CHIL, 2020.

[141]. Efremova M, Vento-Tormo M, Teichmann SA, and Vento-Tormo R. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nature Protocols, 2020.

[142]. Browaeys R, Saelens W, and Saeys Y. Nichenet: modeling intercellular communication by linking ligands to target genes. Nature Methods, 2020.

[143]. Zhang Y and Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. Cellular & Molecular Immunology, 2020.

[144]. Alessandri L, Cordero F, Beccuti M, Licheri N, Arigoni M, Olivero M, Di Renzo MF, Sapino A, and Calogero R. Sparsely-connected autoencoder (sca) for single cell rnaseq data mining. NPJ Systems Biology and Applications, 2021.

[145]. Tran D, Nguyen H, Tran B, La Vecchia C, Luu HN, and Nguyen T. Fast and precise single-cell data analysis using a hierarchical autoencoder. Nature Communications, 2021.

[146]. Han P, Yang P, Zhao P, Shang S, Liu Y, Zhou J, Gao X, and Kalnis P. GCN-MF: disease-gene association identification by graph convolutional networks and matrix factorization. In KDD, 2019.

[147]. Mandal S, Guzmán-Sáenz A, Haiminen N, Basu S, and Parida L. A topological data analysis approach on predicting phenotypes from gene expression data. In AlCoB, 2020.

[148]. Nicolau M, Levine AJ, and Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. PNAS, 2011.

[149]. Yang C, Zhuang P, Shi W, Luu A, and Li P. Conditional structure generation through graph variational generative adversarial nets. In NeurIPS, 2019.

[150]. Chereda H, Bleckmann A, Kramer F, Leha A, and Beissbarth T. Utilizing molecular network information via graph convolutional neural networks to predict metastatic event in breast cancer. In GMDS, 2019.

[151]. Crawford J and Greene CS. Incorporating biological structure into machine learning models in biomedicine. Current Opinion in Biotechnology, 2020.

[152]. Rhee S, Seo S, and Kim S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. In IJCAI, 2018.

[153]. Ramirez R, Chiu Y-C, Hererra A, Mostavi M, Ramirez J, Chen Y, Huang Y, and Jin Y-F. Classification of cancer types using graph convolutional neural networks. Frontiers in Physics, 2020.

[154]. Liu S, Grau B, Horrocks I, and Kostylev E. Indigo: Gnn-based inductive knowledge graph completion using pair-wise encoding. NeurIPS, 34, 2021.

[155]. Burkhardt DB, Stanley JS, Tong A, Perdigoto AL, Gigante SA, Herold KC, Wolf G, Giraldez AJ, van Dijk D, and Krishnaswamy S. Quantifying the effect of experimental perturbations at single-cell resolution. Nature Biotechnology, 2021.

[156]. Huang K. scGNN: scrna-seq dropout imputation via induced hierarchical cell similarity graph. ICML Workshop on Computational Biology, 2020.

[157]. Wang J, Ma A, Chang Y, et al. scgnn is a novel graph neural network framework for single-cell rna-seq analyses. Nature Communications, 2021.

[158]. Chen H, Ryu J, Vinyard ME, Lerer A, and Pinello L. Simba: Single-cell embedding along with features. bioRxiv, 2021.

[159]. Buterez D, Bica I, Tariq I, Andrés-Terré H, and Liò P. Cellvgae: an unsupervised scrna-seq analysis workflow with graph attention networks. Bioinformatics, 2021.

[160]. Marx V. Method of the year: spatially resolved transcriptomics. Nature Methods, 2021.

[161]. Yuan Y and Bar-Joseph Z. Gcng: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. Genome Biology, 2020.

[162]. Partel G and Wählby C. Spage2vec: Unsupervised representation of localized spatial gene expression signatures. The FEBS journal, 2021.

[163]. Meinshausen N, Hauser A, Mooij JM, Peters J, Versteeg P, and Bühlmann P. Methods for causal inference from gene perturbation experiments and validation. PNAS, 2016.

[164]. Guney E, Menche J, Vidal M, and Barábasi A-L. Network-based in silico drug efficacy screening. Nature Communications, 2016.

[165]. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási A-L, and Loscalzo J. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nature Communications, 2018.

[166]. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, MacNair CR, French S, Carfrae LA, Bloom-Ackermann Z, et al. A deep learning approach to antibiotic discovery. Cell, 2020.

[167]. Coley CW, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, Barzilay R, and Jensen KF. A graph-convolutional neural network model for the prediction of chemical reactivity. Chemical Science, 2019.

[168]. Xie Y, Shi C, Zhou H, Yang Y, Zhang W, Yu Y, and Li L. Mars: Markov molecular sampling for multi-objective drug discovery. ICLR, 2021.

[169]. Alagappan M, Jiang D, Denko N, and Koong AC. A multimodal data analysis approach for targeted drug discovery involving topological data analysis (tda). In Tumor Microenvironment. 2016.

[170]. Thafar MA, Olayan RS, Ashoor H, Albaradei S, Bajic VB, Gao X, Gojobori T, and Essack M. Dtigems+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. Journal of Cheminformatics, 2020.

[171]. Thafar MA, Olayan RS, Albaradei S, Bajic VB, Gojobori T, Essack M, and Gao X. Dti2vec: Drug–target interaction prediction using network embedding and ensemble learning. Journal of Cheminformatics, 2021.

[172]. Ma T, Xiao C, Zhou J, and Wang F. Drug similarity integration through attentive multi-view graph auto-encoders. In IJCAI, 2018.

[173]. Jiang M, Li Z, Zhang S, Wang S, Wang X, Yuan Q, and Wei Z. Drug–target affinity prediction using graph neural network and contact maps. RSC Advances, 2020.

[174]. Chen L, Tan X, Wang D, Zhong F, Liu X, Yang T, Luo X, Chen K, Jiang H, and Zheng M. Transformercpi: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. Bioinformatics, 2020.

[175]. Quan Z, Guo Y, Lin X, Wang Z-J, and Zeng X. Graphcpi: Graph neural representation learning for compound-protein interaction. In BIBM, 2019.

[176]. Tsubaki M, Tomii K, and Sese J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics, 2019.

[177]. Lin X, Zhao K, Xiao T, Quan Z, Wang Z-J, and Yu PS. Deepgs: Deep representation learning of graphs and sequences for drug-target binding affinity prediction. ECAI, 2020.

[178]. Guo Z-H, You Z-H, Huang D-S, Yi H-C, Zheng K, Chen Z-H, and Wang Y-B. Meshheading2vec: a new method for representing mesh headings as vectors based on graph embedding algorithm. Briefings in Bioinformatics, 2020.

[179]. Wang R, Li S, Cheng L, Wong MH, and Leung KS. Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning. BMC Bioinformatics, 2019.

[180]. Wan F, Hong L, Xiao A, Jiang T, and Zeng J. Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. Bioinformatics, 2019.

[181]. Xie Y, Peng J, and Zhou Y. Integrating protein-protein interaction information into drug response prediction by graph neural encoding. 2019.

[182]. Zitnik M, Agrawal M, and Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 2018.

[183]. Jiang P, Huang S, Fu Z, Sun Z, Lakowski TM, and Hu P. Deep graph embedding for prioritizing synergistic anticancer drug combinations. Computational and Structural Biotechnology Journal, 2020.

[184]. Kim Y, Zheng S, Tang J, Jim Zheng W, Li Z, and Jiang X. Anticancer drug synergy prediction in understudied tissues using transfer learning. JAMIA, 2021.

[185]. Barisoni L, Lafata KJ, Hewitt SM, Madabhushi A, and Balis UG. Digital pathology and computational image analysis in nephropathology. Nature Reviews Nephrology, 2020.

[186]. Choi E, Bahadori MT, Song L, Stewart WF, and Sun J. GRAM: graph-based attention model for healthcare representation learning. In KDD, 2017.

[187]. Li Y, Qian B, Zhang X, and Liu H. Graph neural network-based diagnosis prediction. Big Data, 2020.

[188]. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, and Yener B. Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering, 2009.

[189]. Pati P, Jaume G, Fernandes LA, Foncubierta-Rodríguez A, Feroce F, Anniciello AM, Scognamiglio G, Brancati N, Riccio D, Di Bonito M, 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.

[190]. Adnan M, Kalra S, and Tizhoosh HR. Representation learning of histopathology images using graph neural networks. In CVPR Workshop, 2020.

[191]. Anand D, Gadiya S, and Sethi A. Histographs: graphs in histopathology. In Medical Imaging 2020: Digital Pathology, 2020.

[192]. Zhou Y, Graham S, Alemi Koohbanani N, Shaban M, Heng P-A, and Rajpoot N. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In ICCV Workshops, 2019.

[193]. Jaume G, Pati P, Foncubierta-Rodriguez A, Feroce F, Scognamiglio G, Anniciello AM, Thiran J-P, Goksel O, and Gabrani M. Towards explainable graph representations in digital pathology. ICML Workshop on Computational Biology, 2020.

[194]. Chao C-H, Zhu Z, Guo D, Yan K, Ho T-Y, Cai J, Harrison AP, Ye X, Xiao J, Yuille A, et al. Lymph node gross tumor volume detection in oncology imaging via relationship learning using graph neural network. In MICCAI, 2020.

[195]. An X, Zhou Y, Di Y, and Ming D. Dynamic functional connectivity and graph convolution network for alzheimer's disease classification. arXiv:2006.13510, 2020.

[196]. Song T-A, Chowdhury SR, Yang F, Jacobs H, El Fakhri G, Li Q, Johnson K, and Dutta J. Graph convolutional neural networks for alzheimer's disease classification. In ISBI, 2019.

[197]. Wee C-Y, Liu C, Lee A, Poh JS, Ji H, Qiu A, Initiative ADN, et al. Cortical graph neural network for ad and mci diagnosis and transfer learning across populations. NeuroImage: Clinical, 2019.

[198]. Mao C, Yao L, and Luo Y. ImageGCN: Multi-relational image graph convolutional networks for disease identification with chest x-rays. IEEE Transactions on Medical Imaging, 2022.

[199]. Levy J, Haudenschild C, Bar C, Christensen B, and Vaickus L. Topological feature extraction and visualization of whole slide images using graph neural networks. PSB, 2020.

[200]. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, Lee EB, Shinohara RT, and Li M. Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nature Methods, 2021.

[201]. Rotmensch M, Halpern Y, Tlimat A, Horng S, and Sontag D. Learning a health knowledge graph from electronic medical records. Scientific Reports, 2017.

[202]. Wu T, Wang Y, Wang Y, Zhao E, Yuan Y, and Yang Z. Representation learning of ehr data via graph-based medical entity embedding. NeurIPS Graph Representation Learning Workshop, 2019.

[203]. Mao C, Yao L, and Luo Y. Medgcn: Medication recommendation and lab test imputation via graph convolutional networks. Journal of Biomedical Informatics, 2022.

[204]. Ma F, You Q, Xiao H, Chitta R, Zhou J, and Gao J. KAME: knowledge-based attention model for diagnosis prediction in healthcare. In CIKM, 2018.

[205]. Sun Z, Yin H, Chen H, Chen T, Cui L, and Yang F. Disease prediction via graph neural networks. IEEE Journal of Biomedical and Health Informatics, 2020.

[206]. Chen IY, Agrawal M, Horng S, and Sontag D. Robustly extracting medical knowledge from ehrs: A case study of learning a health knowledge graph. In PSB, 2020.

[207]. Chowdhury S, Zhang C, Yu PS, and Luo Y. Mixed pooling multi-view attention autoencoder for representation learning in healthcare. arXiv:1910.06456, 2019.

[208]. Liu S, Li T, Ding H, Tang B, Wang X, Chen Q, Yan J, and Zhou Y. A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction. IJMLC, 2020.

[209]. Lee D, Jiang X, and Yu H. Harmonized representation learning on dynamic ehr graphs. Journal of Biomedical Informatics, 2020.

[210]. Rocheteau E, Tong C, Velickovic P, Lane N, and Liò P. Predicting patient outcomes with graph representation learning. AAAI, 2021.

[211]. Kwak H, Lee M, Yoon S, Chang J, Park S, and Jung K. Drug-disease graph: Predicting adverse drug reaction signals via graph neural network with clinical data. In PAKDD, 2020.

[212]. Zhao C, Jiang J, Guan Y, Guo X, and He B. Emr-based medical knowledge representation and inference via markov random fields and distributed representation learning. Artificial Intelligence in Medicine, 2018.

[213]. Li L, Wang P, Wang Y, Wang S, Yan J, Jiang J, Tang B, Wang C, and Liu Y. A method to learn embedding of a probabilistic medical knowledge graph: Algorithm development. JMIR Medical Informatics, 2020.

[214]. Hosseini A, Chen T, Wu W, Sun Y, and Sarrafzadeh M. Heteromed: Heterogeneous information network for medical diagnosis. In CIKM, 2018.

[215]. Shang J, Xiao C, Ma T, Li H, and Sun J. Gamenet: Graph augmented memory networks for recommending medication combination. In AAAI, 2019.

[216]. Council NR et al. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. 2011.

[217]. Wu S, Chen D, and Snyder MP. Network biology bridges the gaps between quantitative genetics and multi-omics to map complex diseases. Current Opinion in Chemical Biology, 2022.

[218]. Umans BD, Battle A, and Gilad Y. Where are the disease-associated eqtls? Trends in Genetics, 2020.

[219]. Wang T, Peng Q, Liu B, Liu Y, and Wang Y. Disease module identification based on representation learning of complex networks integrated from gwas, eqtl summaries, and human interactome. Frontiers in Bioengineering and Biotechnology, 2020.

[220]. Dekker J and Misteli T. Long-range chromatin interactions. Cold Spring Harbor Perspectives in Biology, 2015.

[221]. Lanchantin J and Qi Y. Graph convolutional networks for epigenetic state prediction using both sequence and 3d genome data. Bioinformatics, 2020.

[222]. Hovenga V, Oluwadare O, and Kalita J. Hic-gnn: A generalizable model for 3d chromosome reconstruction using graph convolutional neural networks. bioRxiv, 2021.

[223]. Ding J, Sharon N, and Bar-Joseph Z. Temporal modelling using single-cell transcriptomics. Nature Review Genetics, 2022.

[224]. Fortelny N and Bock C. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. Genome Biology, 2020.

[225]. Machens A, Gesualdo F, Rizzo C, Tozzi AE, Barrat A, and Cattuto C. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. BMC Infectious Diseases, 2013.

[226]. Ying Z, Bourgeois D, You J, Zitnik M, and Leskovec J. Gnnexplainer: Generating explanations for graph neural networks. In NeurIPS, 2019.

[227]. Agarwal C, Lakkaraju H, and Zitnik M. Towards a unified framework for fair and stable graph representation learning. In Uncertainty in Artificial Intelligence, pages 2114–2124, 2021.

[228]. Zhang X and Zitnik M. GNNGuard: Defending graph neural networks against adversarial attacks. NeurIPS, 2020.

[229]. Obermeyer Z, Powers B, Vogeli C, and Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science, 2019.

Box 1:

Fundamentals of graph representation learning

Elements of graphs.

Graph $G = (\mathcal{V}, \mathcal{E})$ consists of nodes $v \in \mathcal{V}$ and edges or relations $e_{u,v}^r = \mathcal{E}$ connecting nodes u and v via a relationship of type r. Subgraph $S = (\mathcal{V}_S, \mathcal{E}_S)$ is a subset of a graph G, where $\mathcal{V}_S \subseteq \mathcal{V}$ and $\mathcal{E}_S \subseteq \mathcal{E}$. Adjacency matrix $\mathbf{A}$ is used to represent a graph, where each entry $\mathbf{A}_{u,v}$ is 1 if nodes u, v are connected, and o otherwise. $\mathbf{A}_{u,v}$ can also be the edge weight between nodes u, v. Homogeneous graph is a graph with a single node and edge type. In contrast, heterogeneous graph consists of nodes of different types (node type set $\mathcal{A}$) connected by diverse kinds of edges (edge type set $\mathcal{R}$). Node attribute vector $\mathbf{x}_u \in \mathbb{R}^d$ describes side information and metadata of node u. The node attribute matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ brings together attribute vectors for all nodes in the graph. Similarly, edge attributes $\mathbf{x}_{u,v}^e \in \mathbb{R}^c$ for edge $e_{u,v}$ can be taken together to form an edge attribute matrix $\mathbf{X}^e \in \mathbb{R}^{m \times c}$. A path from node u_1 to node u_k is given by a sequence of edges $u_1 \xrightarrow{e_{1,2}} u_1 \dots u_{k-1} \xrightarrow{e_{k-1,k}} u_k$. For node u, we denote its neighborhood $\mathcal{N}(u)$ as nodes directly connected to u in G, and the node degree is the size of $\mathcal{N}(u)$. The k-hop neighborhood of node u is the set of nodes that are exactly k hops away from node u, that is, $\mathcal{N}^k(u) = \{v \mid d(u,v) = k\}$ where d denotes the shortest path distance (SI Note 1).

Artificial intelligence tasks on graphs.

To extract this information from networks, classic machine learning approaches rely on summary statistics (e.g., degrees or clustering coefficients) or carefully engineered features to measure network structures (e.g., network motifs). In contrast, representation learning approaches automatically learn to encode networks into low-dimensional representations (i.e., embeddings) using transformation techniques based on deep learning and nonlinear dimensionality reduction. The flexibility of learned representations shows in a myriad of tasks that representations can be used for (SI Note 2):

- **Node, link, and graph property prediction:** The objective is to learn representations of graph elements, such as nodes, edges, subgraphs, and entire graphs. Representations are optimized so that performing algebraic operations in the embedding space reflects the graph's topology. Optimized representations can be fed into models to predict properties of graph elements, such as the function of proteins in an interactome network (i.e., node classification task), the binding affinity of a chemical compound to a target protein (i.e., link prediction task), and the toxicity profile of a candidate drug (i.e., graph classification task).

- **Latent graph learning:** Graph representation learning exploits relational inductive biases for data that come in the form of graphs. In some settings, however, the graphs are not readily available for learning. This is typical for many biological problems, where graphs such as gene regulatory networks

are only partially known. Latent graph learning is concerned with inferring the graph from the data. The latent graph can be application-specific and optimized for the downstream task. Further, such a graph might be as important as the task itself, as it can convey insights about the data and offer a way to interpret the results.

- **Graph generation:** The objective is to generate a graph G representing a biomedical entity that is likely to have a property of interest, such as high druglikeness. The model is given a set of graphs $\mathcal{G}$ with such a property and is tasked with learning a non-linear mapping function characterizing the distribution of graphs in $\mathcal{G}$. The learned distribution is used to optimize a new graph G with the same property as input graphs.

Box 2:

Learning multi-scale representations of proteins and cell types (Figure 4a)

Graph dataset.

Activation of gene products can vary considerably across cells. Single-cell transcriptomic and proteomic data captures the heterogeneity of gene expression across diverse types of cells [134, 135]. With the help of GNNs, we inject cell type specific expression information into our construction of cell type specific gene interaction networks [136, 137, 138]. To do so, we need a global protein interaction network [118, 139].

Learning task.

On a global gene interaction network, we perform multilabel node classification to predict whether a gene is activated in a specific cell type based on single-cell RNA sequencing (scRNA-seq) experiments. In particular, if there are N cell types identified in a given experiment, each gene is associated with a vector of length N. Given the gene interaction network and label vectors for a select number of genes, the task is to train a model that predicts every element of the vector for a new gene such that predicted values indicate the probabilities of gene activation in various cell types (Figure 4a). To enable inductive learning, we split our nodes (i.e., genes) into train, validation, and test sets such that we can generalize to never-before-seen genes.

Impact.

Generating gene embeddings that consider differential expression at the cell type level enables predictions at a single cell resolution, with considerations for factors including disease/cell states and temporal/spatial dependencies [136, 140]. Implications of such cell-type aware gene embeddings extend to cellular function prediction and identification of cell-type-specific disease features [138]. For example, quantifying ligand-receptor interactions using single cell expression data has elucidated intercellular interactions in tumor microenvironments (e.g., via CellPhoneDB [141] or NicheNet [142]). In fact, upon experimental validation of the predicted cell-cell interactions in distinct spatial regions of tissues and/or tumors, these studies have demonstrated the importance of spatial heterogeneity in tumors [143]. Further, unlike most non-graph based methods, like autoencoders, GNNs are able to model dependencies (e.g., physical interactions) between proteins as well as single-cell expression [144, 145].

Box 3:

Learning representations of diseases and phenotypes (Figure 4b)

Graph dataset.

Symptoms are observable characteristics that typically result from interactions between genotypes. Physicians utilize a standardized vocabulary of symptoms, i.e., phenotypes, to describe human diseases. Thus, we model diseases as collections of associated phenotypes to diagnose patients based on their presenting symptoms. Consider a graph built from the standardized vocabulary of phenotypes, e.g., the Human Phenotype Ontology [3] (HPO). The HPO forms a directed acyclic graph with nodes representing phenotypes and edges indicating hierarchical relationships between them; however, it is typically treated as an undirected graph in most implementations of GNNs. A disease described by a set of its phenotypes corresponds to a subset of nodes in the HPO, forming a subgraph of the HPO. Note that a subgraph can contain many disconnect components dispersed across the entire graph [104].

Learning task.

Given a dataset of HPO subgraphs and disease labels for a select number of them, the task is to generate an embedding for every subgraph and use the learned subgraph embeddings to predict the disease most consistent with the set of phenotypes that the embedding represents [104] (Figure 4b).

Impact.

Modeling diseases as rich graph structures, such as subgraphs, enables a more flexible representation of diseases than relying on individual nodes or edges. As a result, we can better resolve complex phenotypic relationships and improve differentiation of diseases or disorders.

Box 4:

Learning representations of drugs and drug combinations (Figure 4c)

Graph dataset.

Combination therapies are increasingly used to treat complex and chronic diseases. However, it is experimentally intensive and costly to evaluate whether two or more drugs interact with each other and lead to effects that go beyond additive effects of individual drugs in the combination. Graph representation learning has been equipped to leverage perturbation experiments performed across cell lines to predict the response of never-before-seen cell lines with mutation(s) of interest (e.g., disease-causing) to drug combinations. Consider a multimodal network of protein-protein, protein-drug, and drug-drug interactions where nodes are proteins and drugs, and edges of different types indicate physical contacts between proteins, the binding of drugs to their target proteins, and interactions between drugs (e.g., synergistic effects, where the effects of the combination are larger than the sum of each drug's individual effect) [182, 183]. Such a multimodal drug-protein network is constructed for every cell line, yielding a collection of cell line specific networks (Figure 4c) [183].

Learning task.

From a single cell line's drug-protein network, we predict whether two or more drugs are interacting in the cell line [183]. Concretely, we embed nodes of a drug-protein network into a compact embedding space such that distances between node embeddings correspond to similarities of nodes' local neighborhoods in the network. We then use the learned embeddings to decode drug-drug edges, and predict probabilities of two drugs interacting based on their embeddings. Next, we apply transfer learning to leverage the knowledge gained from one cell line specific network to accelerate the training and improve the accuracy across other cell line specific networks (Figure 4c) [184]. Specifically, we develop a model using one cell line's drug-protein network, "reuse" the model on the next cell line's drug-protein network, and repeat until we have trained on drug-protein networks from all cell lines.

Impact.

Not only are non-graph based methods unsuited to capture topological dependencies between drugs and targets, 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 robust evaluation of the most promising combinatorial therapies.

Box 5:

Fusing personalized health information with knowledge graphs (Figure 4d)

Graph dataset.

To realize precision medicine, we need robust methods that can inject biomedical knowledge into patient-specific information to produce actionable and trustworthy predictions [216]. Since EHRs can also be represented by networks, we are able to fuse patients' EHR networks with biomedical networks, thus enabling graph representation learning to make predictions on patient-specific features. Consider a knowledge graph, where nodes and edges represent different types of bioentities and their various relationships, respectively. Examples of relations may include "up-/down-regulate," "treats," "binds," "encodes," and "localizes" [7]. To integrate patients into the network, we create a distinct meta node to represent each patient, and add edges between the patient's meta node and its associated bioentity nodes (Figure 4d).

Learning tasks.

We learn node embeddings for each patient while predicting (via edge regression) the probability of a patient developing a specific disease or of a drug effectively treating the patient (Figure 4d) [7].

Impact.

Precision medicine requires an understanding of patient-specific data as well as the underlying biological mechanisms of disease and healthy states. Most 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.
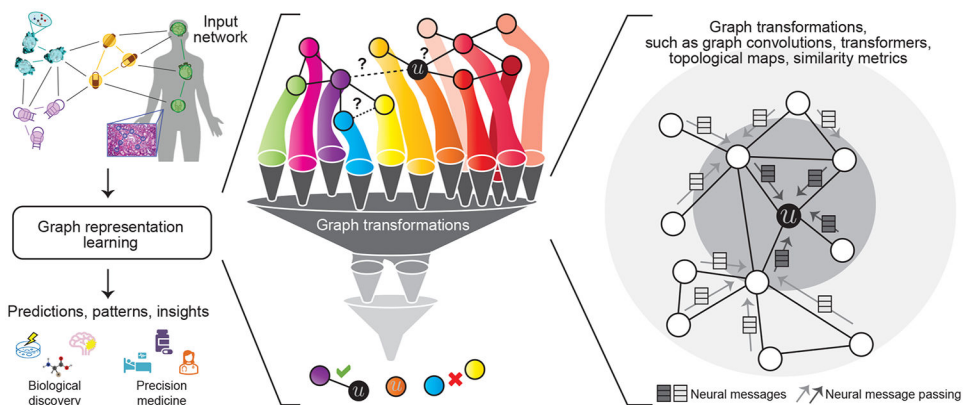
Figure 1: Representation learning for networks in biology and medicine.

Given a biomedical network, a representation learning method transforms the graph to extract patterns and leverage them to produce compact vector representations that can be optimized for the downstream task. The far right panel shows a local 2-hop neighborhood around node *u*, illustrating how information (e.g., neural messages) can be propagated along edges in the neighborhood, transformed, and finally aggregated at node *u* to arrive at the *u*'s embedding.
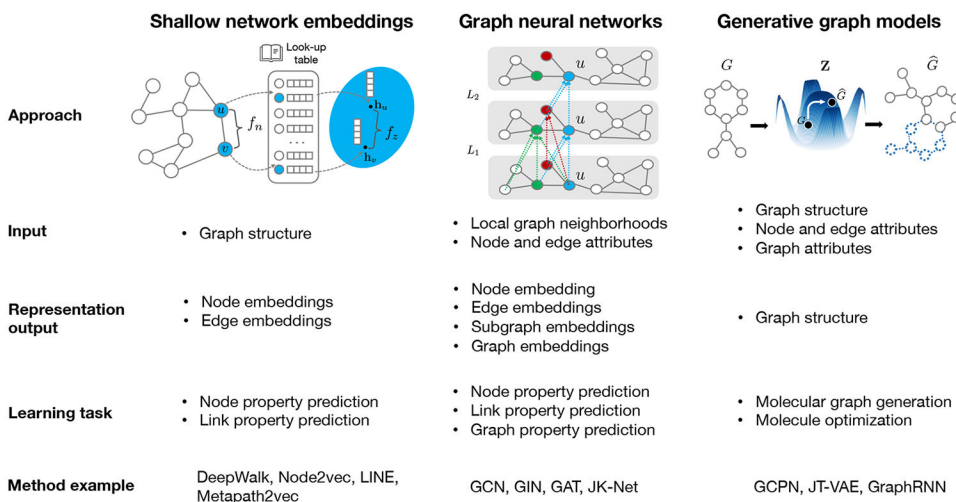
Figure 2: Predominant paradigms in graph representation learning.
(a) Shallow network embedding methods generate a dictionary of representations $\mathbf{h}_u$ for every node u that preserves the input graph structure information. This is achieved by learning a mapping function f_z that maps nodes into an embedding space such that nodes with similar graph neighborhoods measured by function f_n get embedded closer together (Section 2.1). Given the learned embeddings, an independent decoder method can optimize embeddings for downstream tasks, such as node or link property prediction. Method examples include Deep Walk [55], Node2vec [56], LINE [57], and Metapath2vec [58]. **(b)** In contrast with shallow network embedding methods, graph neural networks can generate representations for any graph element by capturing both network structure and node attributes and metadata. The embeddings are generated through a series of non-linear transformations, i.e., message-passing layers (L_k denotes transformations at layer k), that iteratively aggregate information from neighboring nodes at the target node u. GNN models can be optimized for performance on a variety of downstream tasks (Section 2.2). Method examples include GCN [59], GIN [60], GAT [61], and JK-Net [62]. **(c)** Generative graph models estimate a distribution landscape $\mathbf{Z}$ to characterize a collection of distinct input graphs. They use the optimized distribution to generate novel graphs $\widehat{G}$ that are predicted to have desirable properties, e.g., a generated graph can be represent a molecular graph of a drug candidate. Generative graph models use graph neural networks as encoders and produce graph representations that capture both network structure and attributes (Section 2.3). Method examples include GCPN [63], JT-VAE [64], and GraphRNN [65]. SI Figure 1 and SI Note 3 outline other representation learning techniques.
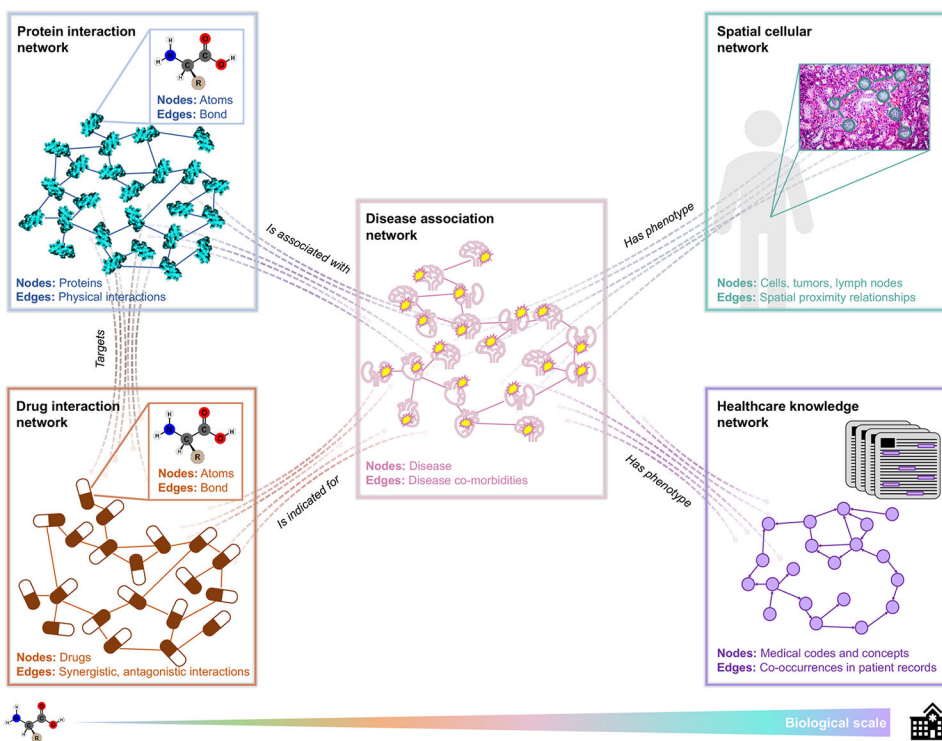
Figure 3: Overview of biomedical applications areas.
Networks are prevalent across biomedical areas, from the molecular level to the healthcare systems level. Protein structures and therapeutic compounds can be modeled as a network where nodes represent atoms and edges indicate a bond between pairs of atoms. Protein interaction networks contain nodes that represent proteins and edges that indicate physical interactions (top left). Drug interaction networks are comprised of drug nodes connected by synergistic or antagonistic relationships (bottom left). Protein- and drug-interaction networks can be combined using an edge type that signifies a protein being a "target" of a drug (left). Disease association networks often contain disease nodes with edges representing co-morbidity (middle). Edges exist between proteins and diseases to indicate proteins (or genes) associated with a disease (top middle). Edges exist between drugs and diseases to signify drugs that are indicated for a disease (bottom middle). Patient-specific data, such as medical images (e.g., spatial networks of cells, tumors, and lymph nodes) and EHRs (e.g., networks of medical codes and concepts generated by co-occurrences in patients' records), are often integrated into a cross-domain knowledge graph of proteins, drugs, and diseases (right). With such vast and diverse biomedical networks, we can derive fundamental insights about biology and medicine while enabling personalized representations of patients for precision medicine. Note that there are many other types of edge relations; "targets," "is associated with," "is indicated for," and "has phenotype" are a few examples.
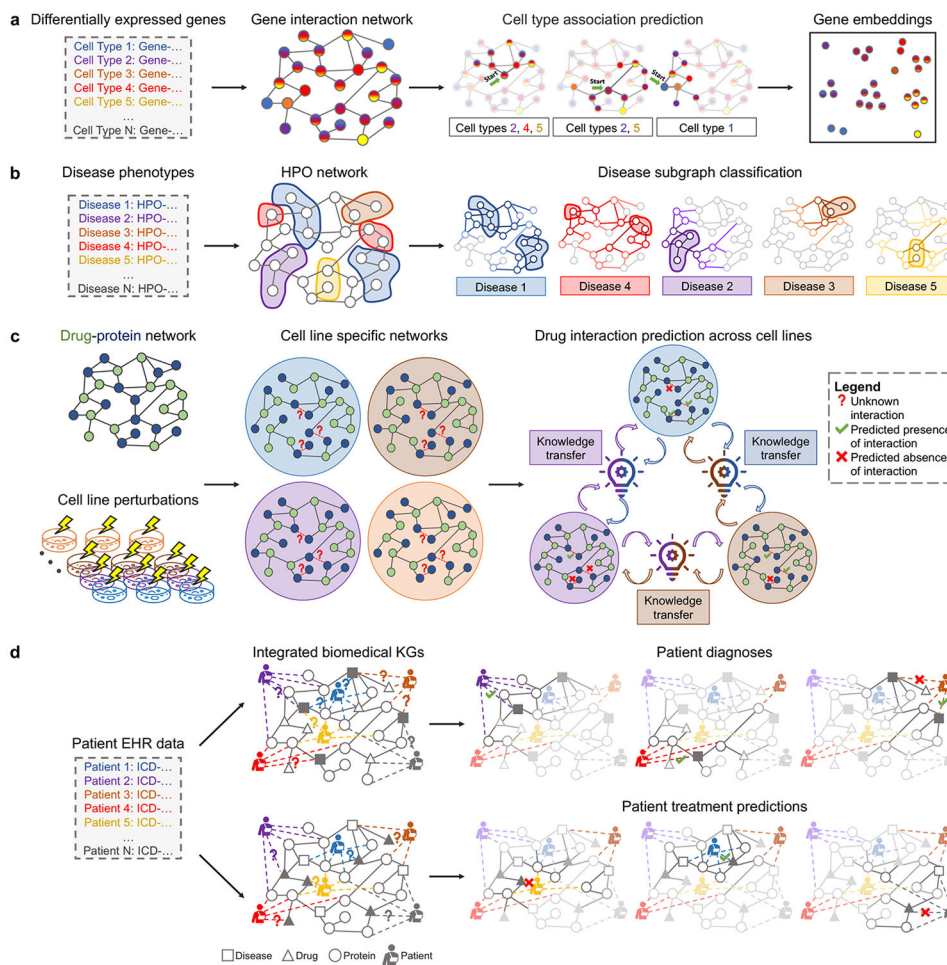
Figure 4: Representation learning in four areas of biology and medicine.
We present a case study on **(a)** cell-type aware protein representation learning via multilabel node classification (details in Box 2), **(b)** disease classification using subgraphs (details in Box 3), **(c)** cell-line specific prediction of interacting drug pairs via edge regression with transfer learning across cell lines (details in Box 4), and **(d)** integration of health data into knowledge graphs to predict patient diagnoses or treatments via edge regression (details in Box 5).