nature biomedical engineering

Perspective

https://doi.org/10.1038/s41551-022-00942-x

Graph representation learning in biomedicine and healthcare

Received: 30 October 2021

Accepted: 9 August 2022

Published online: 31 October 2022

Check for updates

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

Networks—or graphs—are universal descriptors of systems of interacting elements. In biomedicine and healthcare, they can represent, for example, molecular interactions, signalling pathways, disease co-morbidities or healthcare systems. In this Perspective, we posit that representation learning can realize principles of network medicine, discuss successes and current limitations of the use of representation learning on graphs in biomedicine and healthcare, and outline algorithmic strategies that leverage the topology of graphs to embed them into compact vectorial spaces. We argue that graph representation learning will keep pushing forward machine learning for biomedicine and healthcare applications, including the identification of genetic variants underlying complex traits, the disentanglement of single-cell behaviours and their effects on health, the assistance of patients in diagnosis and treatment, and the development of safe and effective medicines.

Networks are pervasive in biology and medicine. They can represent molecular interaction maps or population-scale social and health interactions, for example. Because of the multitude of biological entities and associations that networks can describe, graph representations of biological organization and biomedical knowledge are prevalent. For instance, edges in a regulatory network can indicate 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 (EHRs) can indicate co-occurrences of medical codes across patients[3–5]. The ability to model biomedical discoveries and even overlay patient information in a unified data representation has driven the development of deep learning for networks. In fact, the data diversity and multimodality in networks not only boost the performance of predictive deep learning models, they enable their broad generalization to settings not seen during training[6] and improve model interpretability[7,8]. However, networks can give rise to a bewildering degree of complexity that can only be fully understood through a holistic and integrated view[9–11].

Fortunately, deep learning on graphs is rooted on organizing principles identified in the past two decades in systems biology and medicine[12–15]. These principles link network structure to molecular phenotypes, biological functions or disease states. Thus, we argue that they provide a conceptual grounding that explains the successes of representation learning on graphs—that is, of machine-learning techniques for the generation of optimized mathematical representations of data structured as graphs—and that informs its future developments. For instance, as defined by the local hypothesis, interacting entities are typically more similar than non-interacting entities[13]. The local hypothesis hence implicates that, in protein interaction networks, mutations in interacting proteins often lead to similar diseases[13]. According to the disease-module hypothesis[13], cellular components (such as genes, proteins or metabolites associated with a specific disease) tend to cluster in the same network neighbourhood[16]. According to the shared-components hypothesis, diseases driven by perturbations of the same components (or of closely associated components) are phenotypically similar and have similar responses when targeted by a therapeutic. Furthermore, essential genes are typically found in hubs of a molecular network, whereas non-essential genes (including genes associated with disease) are located at the network's periphery[13]. Moreover, the parsimony principle dictates that the shortest paths in a molecular network involving the fewest disease-associated components correlate with causal molecular pathways[13]. These hypotheses and principles continue to drive discoveries.

[1]Bioinformatics and Integrative Genomics Program, Harvard Medical School, Boston, MA, USA. [2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [3]Health Data Science Program, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Harvard Data Science Initiative, Cambridge, MA, USA. e-mail: marinka@hms.harvard.edu

https://doi.org/10.1038/s41551-022-00942-x

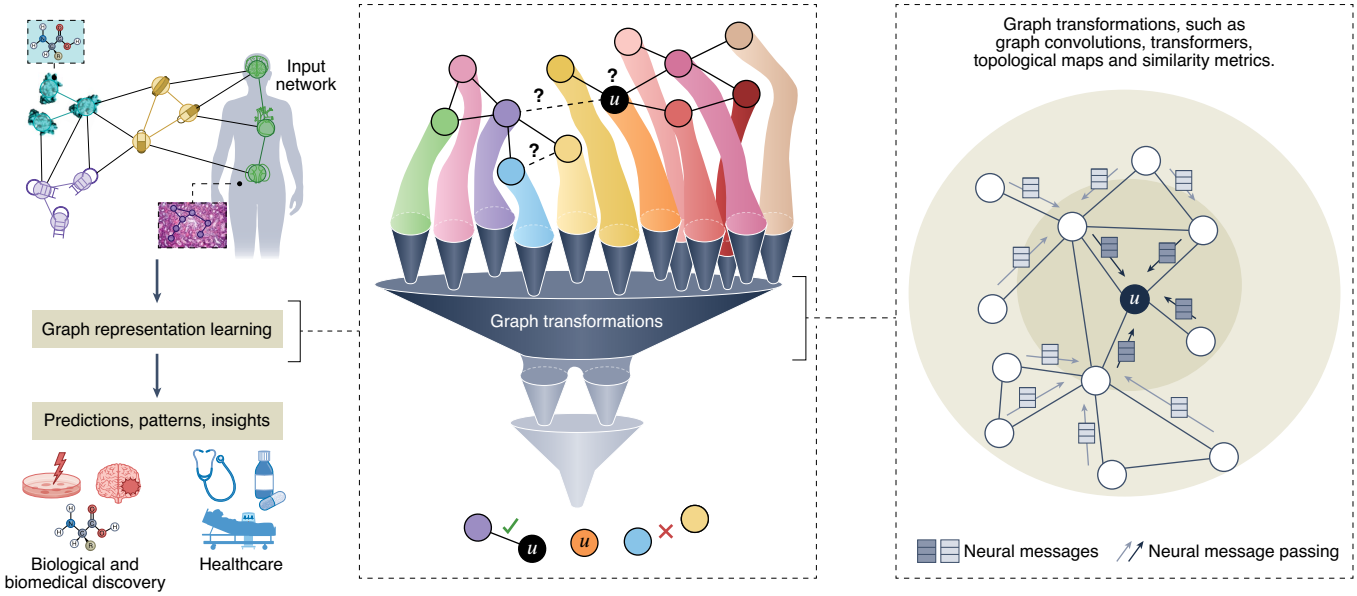


Fig. 1 | Representation learning for networks in biomedicine and healthcare. For any network, graph representation learning transforms the network to extract patterns, make predictions or gain insights, and leverages these to produce compact vectorial representations (denoted by the tube-like shapes) that can be optimized for the downstream task. The right-most schematic shows a local two-hop neighbourhood around node u; it illustrates how information (or 'neural messages') can be propagated along edges in the neighbourhood, transformed and then aggregated at node u to arrive at the embedding of u. The shaded concentric rings englobe the sets of one-hop neighbourhood and two-hop neighbourhood of u.

We argue that representation learning can realize principles of network medicine. The 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 the space is optimized to reflect the structure of interactions between nodes. Concretely, representation learning specifies nonlinear transformation functions that map nodes to points in a compact vectorial space (or embeddings). Such functions are optimized to embed the input network, so that nodes with similar network neighbourhoods are embedded closely in the vectorial space (and algebraic operations performed in this learned space reflect the network's topology). Hence, nodes in the same positional regions should have similar embeddings, owing to the local hypothesis (for example, highly similar pairs of protein embeddings suggest similar phenotypic consequences). Additionally, node embeddings can capture whether the nodes lie within a hub on the basis of their degree (that is, the number of connected nodes), which is an important aspect of local neighbourhood (for instance, strongly clustered gene embeddings indicate essential housekeeping roles). Because of the shared-components hypothesis, two nodes with significantly overlapping sets of neighbours should have similar embeddings, owing to shared message passing (for example, highly similar disease embeddings imply shared disease-associated cellular components).

In this Perspective, we survey the capabilities of graph representation learning and highlight notable applications in biomedicine and healthcare. Some aspects of graph representation learning have been covered extensively in the literature: deep learning on structured data[17,18]; graph neural networks[19–21] (GNNs); representation learning for homogeneous and heterogeneous graphs[22–24], solely heterogeneous graphs[25] and dynamic graphs[26]; data fusion[27]; network propagation[28]; topological data analysis[29] (TDA); and the creation of biomedical networks[9,29–32]. Biomedically focused review articles have surveyed the use of GNNs for molecular generation[33,34], single-cell biology[35], drug discovery and drug repurposing[36–40], and histopathology[41]. Other articles have focused on GNNs, excluding many approaches in graph representation learning, or have not considered patient-centric methods[42]. Here, we overview the uses of graph representation learning across a range of areas in biomedicine and healthcare.

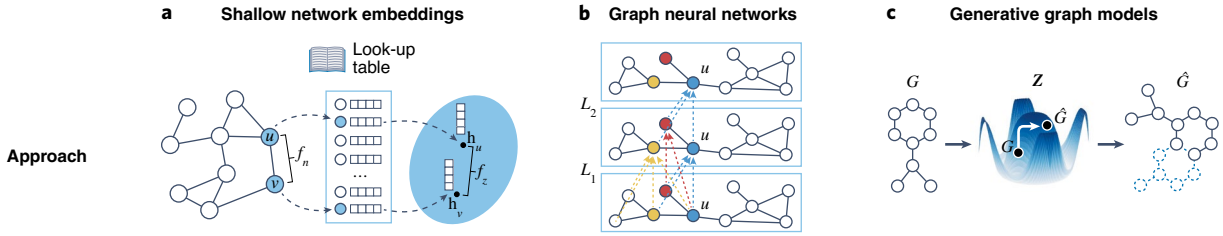
Graph representation learning

Graph theoretic techniques have fuelled many discoveries, from uncovering relationships between diseases[43–46] to repurposing drugs[6,47,48]. Algorithmic innovations, such as random walks[49–51], kernels[52] and network propagation[53], have played a role in capturing structural information from networks. Feature engineering—the process of extracting predetermined features from a network to suit a user-specified machine-learning method[54]—is also a common approach applied to machine learning on networks. It involves the hard-coding of network features (for example, higher order structures, network motifs, degree counts and common neighbour statistics) and the feeding of the engineered feature vectors into a machine-learning model. However, handcrafting optimally predictive features across diverse types of networks and applications can be challenging[18].

For these reasons, graph representation learning has emerged as a leading machine-learning approach for networks. However, its development is challenging because graphs comprise many kinds of entity (nodes) and various types of interaction (edges) among the entities, can be topographically complex and have no fixed node ordering or reference points. Classic deep-learning methods cannot handle such diverse structural properties and rich interactions (which are predominant in biomedical networks) because the methods are designed for fixed-size grids (such as matrices of pixels in 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, we anticipate that graph representation learning will transform the study of complex systems.

In graph representation learning, learned vector representations (or embeddings) of graph elements are generated such that they capture the structure and semantics of the network along with any downstream supervised task (Fig. 1). There is a wide range of methods for graph representation learning, including manifold learning, TDA, GNNs and generative graph models (Fig. 2). Box 1 describes the elements of a graph and outlines the main tasks of machine learning on graphs. In what follows, we outline the main methods of graph representation learning (additional techniques are outlined in Supplementary Note 4).

https://doi.org/10.1038/s41551-022-00942-x



	Shallow network embeddings	Graph neural networks	Generative graph models
Input	Graph structure	Local graph neighbourhoods Node and edge attributes	Graph structure Node and edge attributes Graph attributes
Representation output	Node embeddings Edge embeddings	Node embeddings Edge embeddings Subgraph embeddings Graph embeddings	Graph structure
Learning task	Node-property prediction Link-property prediction	Node-property prediction Link-property prediction Graph-property prediction	Molecular graph generation Molecule optimization
Examples of implementation methods	DeepWalk, Node2vec, LINE, Metapath2vec	GCN, GIN, GAT, JK-Net	GCPN, JT-VAE, GraphRNN

Fig. 2 | Algorithmic paradigms in graph representation learning. a, Methods for shallow network embedding generate a dictionary of representations $\mathbf{h}_u$ for every node u that summarize graph topology surrounding every node in the graph. This is achieved by learning a function f_z that maps nodes into an embedding space such that nodes with similar graph neighbourhoods measured by function f_n get embedded closely. An independent decoder can optimize learned embeddings for downstream tasks, such as the prediction of the property of a node or a link. Example methods include DeepWalk[55], Node2vec[57], LINE[58] and Metapath2vec[59], which differ in how they define the similarity function f_n via graph-traversal techniques (unbiased, biased or typed random walks). **b**, In contrast with methods for shallow network embedding, GNNs can generate representations for any graph element by capturing both the network structure, the attributes, and node metadata. The embeddings are generated through a series of nonlinear transformations (that is, message-passing layers; L_k denotes transformations at layer k) that iteratively aggregate information from neighbouring nodes at the target node u. GNN models can be optimized for performance on a variety of downstream tasks. Examples of GNN methods include GCNs (an architecture for simple graphs with multiple message-passing layers[66]), GIN (an architecture that is probably the most expressive among the class of GNNs[67]), GAT (an architecture that stacks layers in which nodes are able to up-weight and down-weight other nodes in their neighbourhoods[71]) and JK-Net (a jumping-knowledge network that flexibly leverages, for each node, neighbourhoods of different size to enable better representations[215]. **c**, Generative graph models optimize a latent distribution ($\mathbf{Z}$) to capture the structure and properties of input graphs (G). The models use the optimized distribution to generate new graphs ($\hat{G}$) predicted to have the same desirable properties as input graphs (for example, a generated graph can represent a molecular graph of a drug candidate). Examples of these methods include GCPN (a graph convolutional policy network that produces molecular graphs with desired properties such as drug likeness and synthetic accessibility, while obeying physical laws such as chemical valency[98]), JT-VAE (a variational autoencoder that generates molecular graphs in two phases, by first generating a tree-structured scaffold over chemical substructures, and then combining them into a molecule with a message-passing network[93]) and GraphRNN (a deep autoregressive model that learns to generate graphs by training on a set of graphs and decomposing the graph-generation process into a sequence of node and edge formations[99]. Supplementary Fig. 1 and Supplementary Note 4 outline other representation-learning techniques.

Shallow graph embeddings

Shallow-embedding methods optimize a compact vectorial space such that points close in the graph are mapped to nearby points in the embedding space, measured by a predefined distance function or an outer product. These methods are transductive where the encoder function is a simple embedding lookup (Fig. 2). Concretely, t methods involve three steps: the mapping to an embedding space (given a pair of nodes u and v in a graph and a learnable function f that maps nodes to embeddings, the mapping specifies $\mathbf{h}_u$ and $\mathbf{h}_v$); the definition of graph similarity ($f_n(u,v)$; for example, measured by the distance between u and v in the graph) and of embedding similarity ($f_z(\mathbf{h}_v,\mathbf{h}_v)$; for example, a Euclidean distance function or pairwise dot-product); and the computation of a loss function ($\mathcal{L}(f_n(u,v),f_z(\mathbf{h}_u,\mathbf{h}_v))$), which quantifies how the resulting embeddings preserve the desired input-graph similarity). Then an optimization procedure to minimize the loss $\mathcal{L}(f_n(u,v),f_z(\mathbf{h}_u,\mathbf{h}_v))$ is applied. The resulting f serves as a shallow lookup of embeddings that considers the graph structure only in the loss function.

Shallow embedding methods vary according to various definitions of similarities. For example, the shortest path length between nodes is often used as the network similarity, and the dot-product as the embedding similarity. Similarity can also be defined as co-occurrence in a series of random walks of length k (ref. [55]). Unsupervised techniques that predict which node belongs to the walk, such as Skip-gram[56] (an unsupervised learning technique that identifies the nearby nodes, or context, of any given node to learn its most related nodes), are then applied on the walks to generate embeddings. Supervised techniques[57,58], such as Node2vec[57] (a semi-supervised learning technique that combines depth-first search and breadth-first search to capture a node's network neighbourhood), have been used similarly. In heterogeneous graphs, information on the semantic meaning of edges (that is, relation types) can be important. Knowledge graph methods expand similarity measures to consider relation types[59–64]. Once shallow embedding models are trained, the resulting embeddings can be fed into separate models optimized for downstream analyses, such as classification and regression.

Graph neural networks

GNNs are a class of neural networks designed for graph-structured datasets (Fig. 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[65], where layer l carries out three operations: the passing of neural messages (the GNN computes a message $\mathbf{m}_{u,v}^{(l)} = \mathrm{MSG}\left(\mathbf{h}_u^{(l-1)}, \mathbf{h}_v^{(l-1)}\right)$ for linked nodes u, v on the basis of their embeddings from the previous layer $\mathbf{h}_u^{(l-1)}$ and $\mathbf{h}_v^{(l-1)}$); the aggregation of neighbourhoods (the messages between node u and its neighbours $\mathcal{N}_u$ are aggregated as $\hat{\mathbf{m}}_u^{(l)} = \mathrm{AGG}(\mathbf{m}_{u,v}^{(l)} | v \in \mathcal{N}_u)$); and the

https://doi.org/10.1038/s41551-022-00942-x

BOX 1

Fundamentals of graph representation learning

Elements of graphs

A graph $G = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes $\mathcal{V}$ that are connected by a set of edges $\mathcal{E}$. A homogeneous graph has only one type of node and one type of edge, whereas a heterogeneous graph consists of nodes of different type connected by diverse types of edge. Each node in the graph describes real-world entities typically encoded as attribute vectors. Similarly, each edge has an attribute vector describing its associated information. An adjacency matrix **A** is used to represent a graph, where an entry in column *u* (representing node *u*) and row *v* (representing node *v*) is 1 if nodes *u* and *v* are connected and 0 otherwise. These entries can also be edge weights between nodes *u* and *v*. A path from a source node to a target node is given by an ordered sequence of edges connecting them. A subgraph $S = (\mathcal{V}_S, \mathcal{E}_S)$ is a subset of a graph *G*, where $\mathcal{V}_S$ is a subset of $\mathcal{V}$ and $\mathcal{E}_S$ is a subset of $\mathcal{E}$. For any node *u*, its neighbourhood is a subgraph composed of nodes that are directly linked to *u* (that is, there is a path of length 1 between *u* and any other nodes in the subgraph). Supplementary Note 1 provides additional information.

Machine-learning tasks on graphs

To extract information from networks, classic machine learning relies on summary statistics (that is, degrees or clustering coefficients) or carefully engineered features to measure network structures (such as network motifs). By contrast, representation learning automatically learns to encode networks into low-dimensional representations (or embeddings) using transformation techniques based on deep learning and nonlinear dimensionality reduction. The learned representations can be used in a myriad of tasks (Supplementary Note 2).

Prediction of the properties of nodes, links and graphs

The objective is to learn representations of graph elements, namely nodes, edges, subgraphs or 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 (a node-classification task), the binding affinity of a chemical compound to a target protein (a link-prediction task) and the toxicity profile of a candidate drug (a 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 of many biological problems, where graphs such as gene-regulation 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. Also, 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 drug-likeness. The model is given a set of graphs $\mathcal{G}$ with such a property, and is tasked with learning a nonlinear 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 that of the input graphs.

updating of representations. A nonlinear transformation is applied to update node embeddings as $\mathbf{h}_u^{(l)} = \mathrm{UPD}(\hat{\mathbf{m}}_u^{(l)}, \mathbf{h}_u^{(l-1)})$ using the aggregated message and the embedding from the previous layer. In contrast to shallow embeddings, GNNs can capture higher order and nonlinear patterns through multi-hop propagation within several layers of neural message passing. Additionally, GNNs can optimize supervised signals and the graph structure simultaneously, whereas a shallow embedding method requires 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[66–70]. For example, in refs. [71,72–75], the researchers assigned importance scores for nodes during neighbourhood aggregation such that more important nodes played a larger effect in the embeddings. In refs. [76,77], the ability of GNNs to capture structural information of a graph was improved by imposing structural priors, such as a higher order adjacency matrix. Graph-pooling techniques[78] learn abstract topological structures. And GNNs designed for molecules[79,80] inject physics-based scores and domain knowledge into propagation layers.

As biomedical networks can be large and multimodal, special consideration is needed to scale GNNs to large and heterogeneous networks. To this end, refs. [81,82] developed sampling strategies to intelligently select small subsets of the whole local network neighbourhoods, and used them to train GNN models. To tackle heterogeneous relations, in refs. [72,83,84] aggregation transformations were designed to fuse diverse types of relations and attributes. Recent architectures have leveraged dynamic message passing[72,85,86] to deal with evolving and time-varying graphs, as well as few-shot learning[87] or self-supervised strategies[88,89] to deal with graphs that are poorly annotated and that have limited information about labels.

Generative graph models

Generative graph models generate new structures of nodes and edges (and even entire graphs) that are likely to have desired properties, such as novel molecules with acceptable toxicity profiles (Fig. 2). Traditionally, network science models can generate graphs using deterministic or probabilistic rules. For instance, starting from an empty graph, the Erdös–Rényi model[90] iteratively adds random edges according to a predefined probability. The Barabási–Albert model[91] grows a graph by adding nodes and edges such that the degree of the resulting graph has a power-law distribution, which is often observed in real-world networks. The configuration model[92] adds edges on the basis of predefined node degree sequences to generate graphs with arbitrary degree distributions. Although they are powerful as random graph generators, such models cannot optimize graph structures according to properties of interest.

Deep generative models address the challenge by estimating distributional graph properties on the basis of a dataset of graphs $\mathcal{G}$ and by inferring graph structures using such optimized distributions. A generative graph model first learns a latent distribution $P(Z|\mathcal{G})$ that characterizes the input graph set $\mathcal{G}$. Then, conditioned on this distribution, it decodes a new graph (that is, it generates a new graph $\hat{G}$). There are different ways to encode the input graphs and to learn the

https://doi.org/10.1038/s41551-022-00942-x

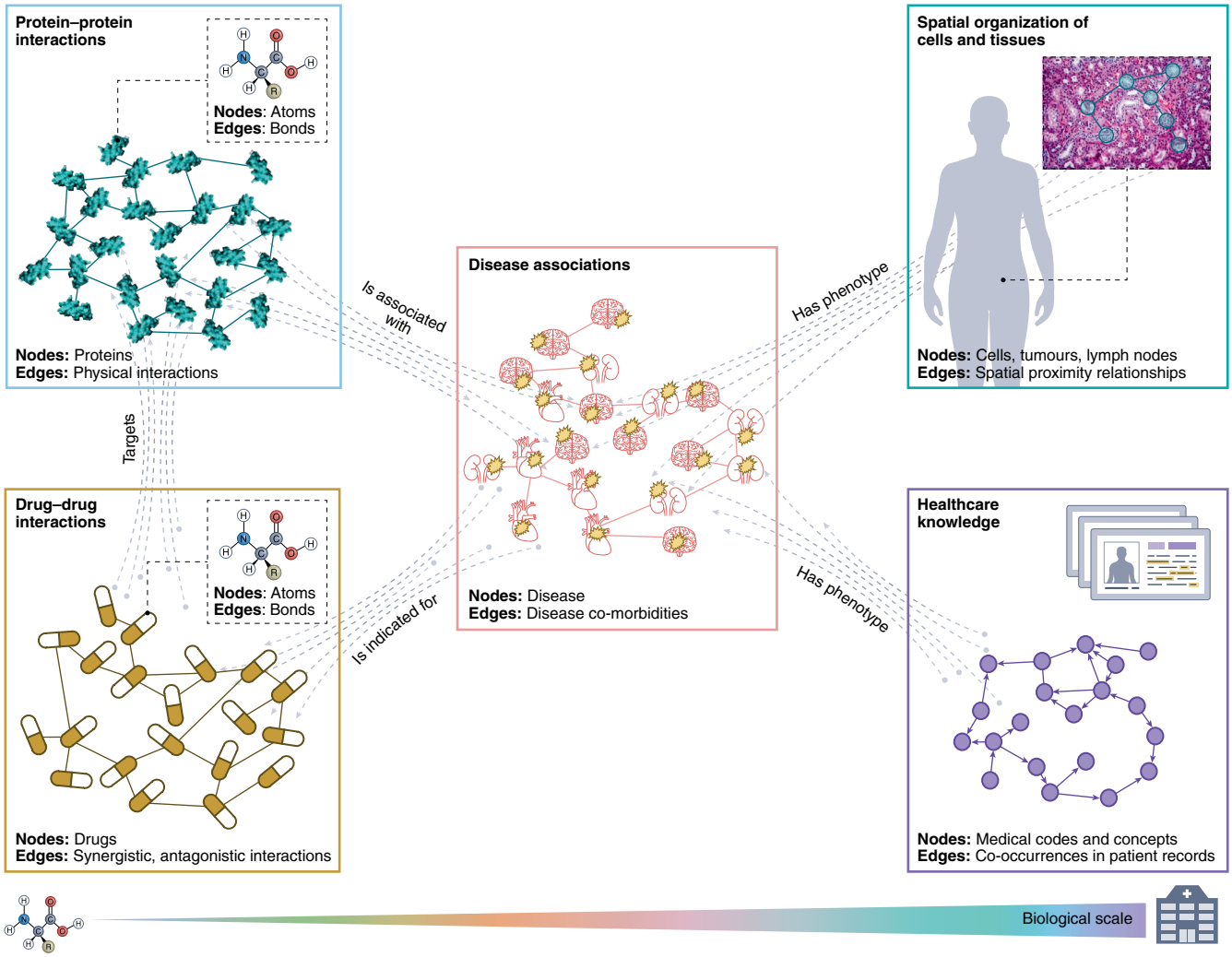


Fig. 3 | Biomedical applications of graph representation learning. Networks are prevalent across biomedical areas. Protein structures and chemical compounds can be modelled as a network in which nodes represent atoms and edges indicate a bond between pairs of atoms. In protein–protein interaction networks, the nodes represent proteins and the edges indicate physical interactions (top left). In drug–drug interaction networks, the nodes are drugs, and are connected by synergistic or antagonistic relationships (bottom left). In networks of protein–drug interactions, edges indicate that a drug binds to a protein target. Edges between proteins and diseases indicate proteins (or genes) associated with a disease, and edges between drugs and diseases represent drugs that are indicated for the disease. Patient information, such as medical images (modelled as spatial networks of cells, tumours and lymph nodes, for example; top right) and EHRs (modelled as networks of medical codes and of concepts generated by co-occurrences in the patient records; bottom right), are often integrated into a cross-domain knowledge graph of proteins, drugs and diseases (centre). Such disease-association networks often represent diseases as nodes and co-morbidities as edges. Edge relations can also mean 'targets', 'is associated with', 'is indicated for' or 'has phenotype', for example.

latent distribution (in particular, through variational autoencoders[93–95] or generative adversarial networks[96]). Decoding a new graph is more difficult than decoding an image or text because a graph is discrete and unbounded in structure and size, and the nodes in it have no particular order. Common practices to generate new graphs include the prediction of a probabilistic fully connected graph followed by the use of graph matching to find the optimal subgraph[97]; the decomposition of a graph into a tree of subgraph structures and the generation of a tree structure instead, followed by the generation of assemblies of subgraphs[93]; and the sequential sampling of new nodes and edges[98,99].

Applications in biomedicine

Biomedical datasets involve rich multimodal and heterogeneous types of data, such as molecular interactions and healthcare systems (Fig. 3). Methods of graph representation learning are suited to leverage structural information in such multimodal datasets[100].

For instance, at the molecular level, atoms and bonds can be represented as nodes and edges, respectively. Physical interactions or functional relationships between proteins also naturally form a network. Whether an unknown protein clusters in a particular neighbourhood of known proteins and shares direct neighbours with them is informative of the binding affinity and function of the unknown protein[101]. Hence, by learning molecular representations of proteins and their physical interactions, graph representation learning can be applied to predicting protein function.

At the genomic level, genetic elements can be incorporated into networks by extracting the co-expression information of coding genes from transcriptomic data. Because spatial molecular profiling at the single-cell level has enabled the mapping of genetic interactions at the cellular and tissue levels, investigating the cellular circuitry of molecular functions through gene co-expression data can help uncover disease mechanisms. For instance, as implicated by the network parsimony principle[13], the shortest path in a molecular network between

disease-associated genes often correlates with causal molecular pathways[16,45]. Also, learned embeddings that capture genome-wide interactions can enhance disease predictions at the resolutions of single cells and tissues.

Moreover, networks composed of small molecule drugs, proteins and diseases can be used to model drug–drug interactions, the binding of drugs to target proteins, and the identification of drug-repurposing opportunities. For example, according to a corollary of the local hypothesis[13], the topology of drug combinations is indicative of synergistic or antagonistic relationships[102]. Learning the topology of graphs with nodes representing drugs, proteins and diseases can improve predictions of candidate drugs, the identification of potential off-target effects, and the prioritization of novel drug combinations.

Proteins

Graph representation learning has been widely used to model proteins and produce new protein designs by optimizing over the input space (such as amino acid sequences) of a predictive model, and to find proteins that satisfy the design criteria (such as having specific protein functions[103,104]). Specifically, the inductive ability of graph convolutional networks (GCNs) to generalize to data points unseen during model training, and to generate new data points from scratch by decoding latent representation from the embedding space, has enabled the discovery of new molecules, interactions and functions[100,105,106].

Computationally elucidating protein structure has been an ongoing challenge[33]. Because proteins are folded into complex 3D structures, they can be represented as graphs. For example, a contact distance graph can be constructed where the nodes are individual residues and the edges are determined by a physical distance threshold[107]. Edges can also be defined by the ordering of amino acids in the primary sequence[107]. Additionally, spatial relationships between residues (such as distances and angles) may be used as features for edges[108].

Protein structures can be modelled by capturing dependencies in their sequences of amino acids (for example, by applying GNNs to learn the local neighbourhood structure of each node) to generate protein embeddings[108,109]. Concretely, protein embeddings can be learned by identifying short- and long-range dependencies across sequences corresponding to their 3D structures, and then used to predict primary sequences from 3D structures[109]. Alternatively, one can use a hierarchical process of learning atom-connectivity motifs to capture molecular structure at varying levels of granularity (at the levels of the motif, connectivity and atoms) in the protein embeddings, with which new 3D structures can be generated. This is a difficult task, owing to the computational constraints of generalizability across different classes of molecules and of flexibility for a wide range of sizes[110]. Recent review articles have covered machine learning for molecular design[33,111], graph generation[112], the prediction of molecular properties[33,34], and therapeutic-compound design and generation.

Protein interactions

Various data modalities, including chemical structure, binding affinities, physical and chemical principles, and amino acid sequences, have been integrated to improve the quantification of protein interactions[33]. GNNs are commonly used to generate representations of proteins on the basis of chemical features (for example, the locations of free electrons donors and of proton donors) and of geometric features (such as distance-dependent curvature) to predict protein-pocket–ligand interactions and protein–protein interactions[113]; to generate intramolecular and intermolecular residue contact graphs to predict intramolecular and intermolecular energies, binding affinities and quality measures for a pair of molecular complexes[114]; and to generate ligand–protein and receptor–protein graphs to predict whether a pair of residues from the ligand and receptor proteins belongs to an interface[108]. Combining evolutionary, topological and energetic information about molecules enables the scoring of docked conformations on the basis of the similarity of random walks simulated on a pair of protein graphs (Supplementary Note 3)[52].

Owing to experimental and resource constraints, the most updated networks of protein–protein interactions are limited in their number of nodes (proteins) and edges (physical interactions)[115]. Yet topology-based methods can capture and leverage the dynamics of biological systems to enrich existing protein–protein interaction networks[116]. Some of these methods first apply graph convolutions to aggregate structural information in the graphs of interest (such as protein–protein interaction networks and ligand–receptor networks), use sequence modelling to learn the dependencies in amino acid sequences, and then concatenate the two outputs to predict the presence of physical interactions[100,117]. Interestingly, such concatenated outputs have been treated as 'image' inputs to convolutional neural networks[117]. Similar graph convolution methods can also be used to remove less credible interactions, thereby constructing a more reliable protein–protein interaction network[118].

Protein functions and cellular phenotypes

Characterizing a protein's function in specific biological contexts is a challenging and experimentally intensive task[119,120]. However, innovations in techniques for the representation of protein structures and interactions have facilitated the prediction of protein function[121], especially when leveraging gene ontologies and transcriptomic data.

Gene ontology terms[122] are a standardized vocabulary for describing molecular functions, biological processes and cellular locations of gene products[123]. They have been built as a hierarchical graph that GNNs can leverage to learn dependencies of the terms[123], and can also be directly used as protein-function labels[103,124]. In the latter case, sequence-similarity networks are typically constructed and combined with protein–protein interaction networks, and then protein features (such as amino acid sequence, protein domains, subcellular location or gene expression profiles) are integrated to predict protein function[103,124]. Additionally, gene-interaction networks that leverage transcriptomic data[105,125] can capture context-specific interactions between genes (Fig. 4a).

Other methods of graph representation learning for the prediction of protein function involve defining diffusion-based distance metrics on protein–protein networks for predicting protein function[126]; the use of the theory of topological persistence to compute signatures of a protein on the basis of its 3D structure[127]; and the application of TDA to extract features from protein-contact networks created from 3D coordinates[128] (Supplementary Note 3). Additionally, an attention mechanism for protein-sequence embeddings generated by the language model BERT (for 'bidirectional encoder representations from transformers') has facilitated the interpretability of the predictions of such networks[129,130].

Gene expression

Diseases can be classified according to symptoms, and these can sometimes be caused by molecular dysfunction resulting from genetic mutations. Hence, diagnosing many diseases requires knowledge of alterations in the transcription of coding genes, so as to capture genome-wide associations driving disease onset and progression. Methods of graph representation learning allow for the analysis of heterogeneous networks of multimodal information, from genomic data to pathophysiology (Fig. 4b).

Approaches that rely solely on gene expression data typically transform the co-expression matrix into a more topologically meaningful form[131–133]. Gene-expression data can be transformed into a coloured graph that captures the shape of the data (by using TDA[133] Supplementary Note 3), which then enables downstream analyses through network-science metrics and graph machine learning. Topological landscapes present in gene-expression data can be vectorized

https://doi.org/10.1038/s41551-022-00942-x

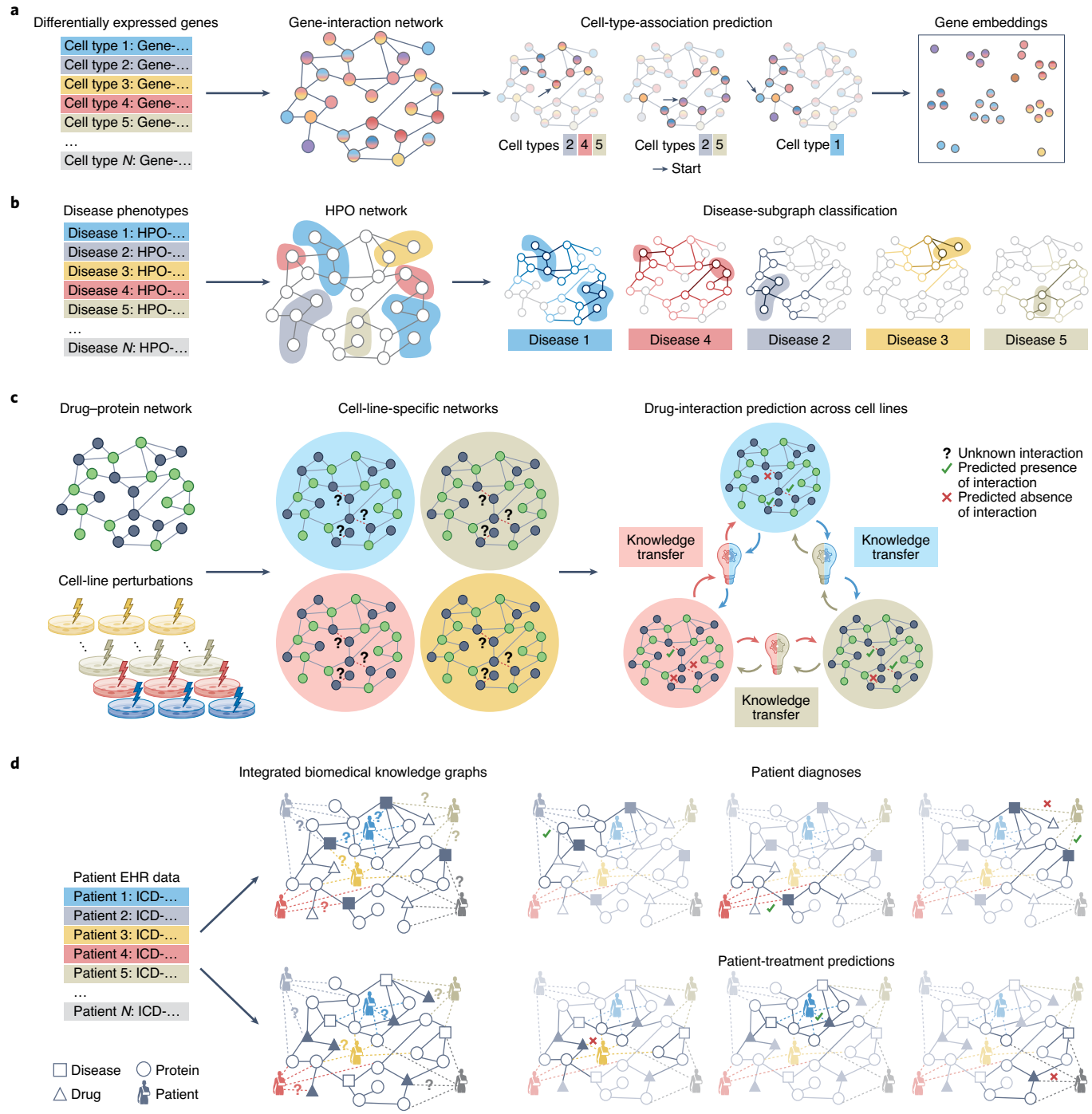


Fig. 4 | Representation learning in four areas of biomedicine and healthcare. **a**, Cell-type-aware protein-representation learning via multilabel node classification. **b**, Disease classification using subgraphs. **c**, Cell-line-specific prediction of interacting drug pairs via edge regression with transfer learning across cell lines. **d**, Integration of health data into knowledge graphs to predict patient diagnoses or treatments via edge regression. Box 2 provides context and details for each panel. HPO, Human Phenotype Ontology.

and fed into a GCN to classify the disease type[132]. Alternatively, gene expression data can be used directly to construct networks of genes and diseases that are then input into a joint matrix factorization and a GCN to draw disease–gene associations, akin to a recommendation task[131]. Additionally, applying GCNs, variational autoencoders and generative adversarial networks jointly to gene-correlation networks (initialized with a subset of gene-expression matrices) can generate disease networks with the desired properties[134].

Because gene-expression data can be noisy and variational, the co-expression matrices can be fused with existing biomedical networks (for example, networks of gene-ontology annotations and of protein–protein interactions), and the resulting graph fed into graph convolutional layers[135–137]. Doing so has enabled more interpretable disease-classification models (such as models weighting gene interactions on the basis of existing biological knowledge). However, models trained solely on gene-interaction networks are unable to capture all gene-regulation activities[138]. To this end, methods of graph representation learning, such as GNNs, can learn robust and meaningful representations of molecules (even with an incomplete interactome[101]) and inductively infer new edges between pairs of nodes[139].

BOX 2

Learning multiscale representations with graphs

Proteins and cell types

Dataset

Single-cell transcriptomic and proteomic data capture the heterogeneity of gene expression across diverse types of cells[216,217]. GNNs can help inject cell-type-specific gene-expression information into cell-type-specific gene-interaction networks[140,218,219]. To do so, a global protein-interaction network[115,220] is needed.

Learning task

On a global gene-interaction network, multilabel node classification can be performed to predict whether a gene is activated in a specific cell type on the basis of scRNA-seq experiments. If *N* cell types are identified in each 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 (Fig. 4a). To enable inductive learning, nodes (that is, genes) are split into training, validation and test sets such that the model can generalize to genes that it has not seen.

Impact

Generating gene embeddings that consider differential expression at the cell-type level can enable predictions at single-cell resolution, with considerations for factors including disease and cell states, and temporal and spatial dependencies[140,142]. The implications of such cell-type-aware gene embeddings extend to the prediction of cellular function and to the identification of cell-type-specific disease features[140]. For example, quantifying ligand–receptor interactions using single-cell-expression data has predicted intercellular interactions in tumour microenvironments (in particular, via CellPhoneDB[221] or NicheNet[222]). Experimental validation of the predicted cell–cell interactions in distinct spatial regions of tissues and tumours showed the importance of spatial heterogeneity in tumours[223]. Unlike methods for standard representation learning, GNNs can explicitly model dependencies (such as physical interactions) between proteins as well as single-cell gene expression[224,225].

Diseases and phenotypes

Dataset

Physicians use a standardized vocabulary of symptoms (that is, phenotypes) to describe human diseases. Hence, diseases can be modelled as collections of associated phenotypes and used to diagnose patients on the basis of the symptoms that they present. In a graph built from the standardized vocabulary of phenotypes (the Human Phenotype Ontology[3]), the nodes represent phenotypes and the edges indicate hierarchical relationships between them. A disease described by a set of its phenotypes thus corresponds to a subset of nodes in the ontology, and thus forms a subgraph of it (a subgraph can contain many disconnect components dispersed across the entire graph[169]).

Learning task

Given a dataset of subgraphs and disease labels for a select number of them, the task is to generate an embedding for every subgraph and to use the learned subgraph embeddings to predict the disease most consistent with the set of phenotypes that the embedding represents[169] (Fig. 4b).

Impact

Modelling diseases as rich graph structures (such as subgraphs) enables a more flexible representation of diseases than relying on individual nodes or edges. Graph structures can better resolve complex phenotypic relationships and improve the differentiation of related diseases or disorders.

Drugs and drug combinations

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 whether the combination leads to effects that are different from the additive effects of the individual drugs. Graph representation learning can leverage perturbation experiments performed across cell lines to predict the responses, to drug combinations, of unseen cell lines with mutations of interest (in particular, disease-causing mutations). 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 (such as synergistic effects, where the effects of the combination are different from the contributions of the effects of each drug)[226,227] can be constructed for every cell line, yielding a collection of cell-line-specific networks[226] (Fig. 4c).

Learning task

From the drug–protein network of a single cell line, one can predict whether two or more drugs are interacting[226]. Concretely, nodes of a drug–protein network are embedded into a compact space such that distances between node embeddings correspond to the similarities of the local neighbourhoods of the nodes. The learned embeddings can then be used to decode drug–drug edges and to predict the probabilities of two drugs interacting. Transfer learning can then be applied to leverage the knowledge gained from one cell-line-specific network so as to accelerate the training and to improve the accuracy of the model across other cell-line-specific networks[228] (Fig. 4c). Specifically, a model can be developed by using a drug–protein network for one cell line, and reused on the drug–protein network of any other cell line.

Impact

Standard methods are unable to capture topological dependencies between drugs and targets, and most predictive models for drug combinations do not consider the tissue specificity or cell-line specificity of drugs. Because the effects of drugs on the human body are not uniform, it is crucial to account for such anatomical differences. Additionally, the ability to prioritize candidate drug combinations in silico could reduce the cost of developing and testing them experimentally.

Personalized health information fused with knowledge graphs

Dataset

Robust methods that can inject biomedical knowledge into patient-specific information are needed to produce actionable and trustworthy predictions[229]. Because EHRs can also be represented

(continued from previous page)

by networks, networks of EHRs can be fused with biomedical networks, thus enabling graph representation learning to make predictions on patient-specific features. An example is a knowledge graph, where nodes and edges represent different types of biological entities and their various relationships. Examples of such relations are 'upregulate / downregulate', 'treats', binds', 'encodes' and 'localizes'[7]. To integrate patient data into a network, a distinct metanode is created to represent each patient, and edges are added between the patient's metanode and its associated biomedical-entity nodes (Fig. 4d).

Learning tasks

Node embeddings for each patient can be learned while predicting (via edge regression) the probability of a patient developing a specific disease or of a drug effectively treating the patient[7] (Fig. 4d).

Impact

Most networks do not consider patient data, which can prevent robust predictions of a patient's conditions and their potential responsiveness to particular drugs. The ability to integrate patient data with biomedical knowledge may address this.

Single-cell transcriptomics

Single-cell RNA sequencing (scRNA-seq) data lend themselves to graph representation learning for the modelling of cellular differential processes[140,141] and disease states[142]. A predominant approach to analyse scRNA-seq datasets is to transform them into gene-similarity networks, (such as gene co-expression networks) or into cell-similarity networks (by correlating gene-expression readouts across individual cells). Applied to such networks, graph representation learning can, for instance, impute scRNA-seq data[143,144] and predict cell clusters[144,145]. 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[144]. 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[146]. Beyond GNNs and graph autoencoders, learning a manifold over a cell-state space can quantify the effects of experimental perturbations[141]. To this end, cell-similarity graphs are constructed for control samples and treated samples, and used to estimate the likelihood of a cell population being observed under a given perturbation[141].

Spatial molecular profiling can measure both gene expression at the cellular level and the location of cells in tissue[147]. As a result, spatial transcriptomics data can be used to construct cell graphs[148], spatial gene-expression graphs[149], gene-co-expression networks or molecular-similarity graphs[35]. Creating graphs of cell neighbourhood and of spatial gene expression requires a distance metric, as edges are determined on the basis of spatial proximity, whereas graphs of gene co-expression and molecular similarity need a threshold applied on the gene-expression data[35]. From such networks, methods of graph representation learning produce embeddings that capture the network topology and that can be further optimized for downstream tasks. For instance, a cell-neighbourhood graph and a gene-pair expression matrix enable GNNs to predict ligand–receptor interactions[148]. In fact, because these interactions are directed, they could be used to infer causal interactions of previously unknown ligand–receptor pairs[148,150].

Small-molecule drugs

Modern drug discovery requires elucidating the chemical structure of a candidate drug, identifying its drug targets, quantifying its efficacy and toxicity, and detecting its potential side effects[13,14,32,151]. Because such processes are costly and time consuming, drug-discovery pipelines leverage in silico approaches. However, cross-domain expertise is necessary to develop a drug with optimal binding affinity and optimal specificity to biomarkers, maximal therapy efficacy, and minimal adverse effects. Therefore, it is critical to integrate chemical-structure information, protein interactions and clinically relevant data (such as indications and reported side effects) into predictive models for drug discovery and drug repurposing. Graph representation learning can be used to characterize drugs at the systems level without patient data to make predictions about interactions with other drugs, protein targets, side effects and diseases[6,38–40,48,152].

As with proteins, small molecules are modelled 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) that are added to the model[79,153]. Edges can also be added to indicate the spatial distance between each two atoms[65] or information on bond angles and rotations can be incorporated into the molecular graph[80].

Representing molecules as graphs has improved predictions on various quantum-chemistry properties. Simplistically, GNNs aggregate information from neighbouring atoms and bonds to learn the local chemistry of each atom[153]. For example, generating representations of the atoms, distances and angles has allowed the identification of the angles and directions of the interactions between atoms[80]. Producing atom-centred representations based on a weighted combination of their neighbours' features (via an attention mechanism) can be used to model interactions among reactants and to predict the outcomes of the reactions[154].

Alternatively, molecular graphs have been decomposed into a 'junction tree', where each node represents a substructure in the molecule. This aims to learn representations of both the molecular graph and the junction tree, for the generation of new molecules with desirable properties[93]. In fact, iteratively editing fragments of a molecular graph during training has improved predictions of high-quality drug candidates targeting a protein of interest[155].

Drug–drug and drug–target interactions

A drug's binding affinity and specificity to its target determine the drug's effectiveness and potential for off-target effects[34]. However, quantifying these metrics requires labour-intensive and costly experiments[33,34]. Modelling the molecular structure of the protein targets of small molecules as well as their binding affinities and specificities by using graph representation learning has accelerated the study of drug–target interactions.

Topological data analysis[156] and shallow network embedding[57] have been used to learn representations of drugs and targets. Concretely, TDA transforms experimental data into a graph where nodes represent compounds and edges indicate a level of similarity between them[156] (Supplementary Note 3). Methods of shallow network embedding can also be used to generate embeddings for drugs and targets by computing drug–drug, drug–target and target–target similarities[157]. Non-graph methods have also been used to create graphs that are then fed into a graph model to generate embeddings. For instance, the *k*-nearest-neighbours algorithm is commonly used to construct drug-similarity and target-similarity networks[158]. The resulting embeddings are fed into downstream machine-learning models.

Predictions of drug–drug and drug–target interactions have been improved by fusing chemical structures, target sequences and clinical implications. 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[159].

Additionally, two separate GNNs may be used to learn representations of protein graphs and small-molecule graphs, to predict drug–target affinity[160]. And protein-structure representations generated by graph convolutions have been combined with protein-sequence representations (using shallow network embedding or convolutional neural networks) to predict the probability of small-molecule–protein interactions[161–164].

Drug–disease associations and disease biomarkers

Part of the drug-discovery pipeline involves minimizing any adverse events[33,34]. However, the experiments required to measure drug–drug interactions and toxicity are costly and face a combinatorial explosion problem[33]. By considering gene-expression data, gene ontologies, drug similarity and other clinically relevant data regarding side effects and indications, methods of graph representation learning enable the in silico modelling of drug action, allowing for a more efficient ranking of candidate drugs for repurposing.

Drug and disease representations have been learned on homogeneous graphs of drugs, diseases or targets. For instance, medical terms in subject headings may be used to construct a drug–disease graph, from which latent representations of drugs and diseases are learned using various graph-embedding algorithms (such as DeepWalk and LINE; ref. [165]). TDA (Supplementary Note 3) has also been applied for the separate construction of graphs of drugs, targets and diseases; representations of such entities are learned and optimized for downstream prediction[166].

Recent methods have fused multimodal data to create heterogeneous graphs. For example, neighbourhood information can be aggregated from heterogeneous networks of drugs, targets and diseases, to predict drug–target interactions[167]. Protein–protein interaction networks have also been combined with genomic features to predict drug sensitivity using GNNs[168]. Overall, approaches integrating cross-domain knowledge as a vast heterogeneous network or into the model's architecture may better predict drug action (Fig. 4c).

Applications in healthcare

Patient records, such as medical images and EHRs, can be represented as networks, and can be incorporated into networks of proteins, diseases and drugs. For example, following the local hypothesis, the shared-components hypothesis and the disease-module hypothesis[13], patients with rare diseases probably have similar phenotypes and even share disease mechanisms if they are represented by nodes that have common neighbours and topology[169,170].

Methods of graph representation learning can, in principle, integrate patient records with molecular, genomic and disease networks for personalized predictions. Graph representation learning has also been used to fuse multimodal knowledge with patient records. Here we highlight two types of patient data that have been successfully integrated using deep graph learning: histopathology images[8,171,172] and EHRs[173,174].

Histopathology images

Whole histopathology slides and other medical images can typically be converted into spatial graphs, where nodes represent the cells in the image and edges indicate that a pair of cells are adjacent in space. Deep graph learning can then detect subtle signs of disease progression in the images, also by integrating other modalities (such as tissue localization[175] and genomic features[8]).

Cell–tissue graphs generated from histopathology images can encode the spatial context of cells and tissues for a given patient. Information on cell morphology and tissue microarchitecture can be aggregated into cell graphs to, for instance, grade cancer histology images (for example, using GNNs)[8,176–178]. An example aggregation method involves pooling with an attention mechanism to infer relevant patches in the image[176]. A hierarchical GNN can then learn relevant representations of cell morphology and cell–cell interactions, tissue morphology and the spatial distribution of cells, cell-to-tissue hierarchies, and the spatial distribution of cells in the tissue, as all of these can be captured in a cell-to-tissue graph[175]. Because interpretability is critical for models that generate patient predictions, post-hoc graph-pruning optimization may be performed on a cell graph generated from a histopathology image, to define subgraphs that explain the original cell-graph analysis[179].

Methods of graph representation learning can also be used for classifying other types of medical images. For instance, GNNs can model relationships between lymph nodes to compute the spread of lymph-node gross tumour volume on the basis of computed-tomography images[180]. GNNs have been used to classify the progression of Alzheimer's disease from magnetic resonance images that are converted into graphs[181–183]. GNNs can also leverage relational structures, such as similarities among chest X-rays, to improve downstream tasks, such as disease diagnosis and localization[184]. Moreover, GNNs have been used to classify patients according to colon-cancer stage[185] after TDA had been applied to generate graphs from whole-slide images of tissues from various sources (Supplementary Note 3).

With spatial gene-expression graphs (weighted and undirected) and corresponding histopathology images, gene-expression information can be aggregated to generate embeddings of genes that could then be used to investigate spatial domains (to differentiate between cancer and noncancer regions in tissues, for instance)[186]. Because multimodal data enables more robust predictions, GNNs have been applied to spatial graphs of cells from histopathology images alongside genomic and transcriptomic data, to predict treatment responses and resistance, histopathology grading, and patient survival[8].

Patient records

EHRs are typically represented by ICD (International Classification of Disease) codes[173,174]. The hierarchical information inherent to ICD codes (medical ontologies) lends itself to the creation of a rich network of medical knowledge. In addition to ICD codes, medical knowledge can take the form of many data types, including symptoms, molecular data, drug interactions and side effects. By integrating patient records into networks, graph representation learning can generate predictions tailored to individual patients.

Methods that embed medical entities, including EHRs and medical ontologies, leverage the inherently hierarchical structure of knowledge graphs of medical concepts[187]. For example, low-dimensional embeddings of EHR data can be generated by separately considering medical services, doctors and patients in shallow network embeddings and GNNs[188,189]. Alternatively, attention mechanisms may be applied on EHR data and medical ontologies to capture parent–child relationships[173,190,191]. Rather than assuming a certain structure in the EHRs, a graph convolution transformer can learn hidden EHR structures[75].

EHRs also have underlying spatial and temporal dependencies[192] that many recent methods have leveraged to perform time-dependent prediction tasks. A mixed pooling multi-view self-attention autoencoder can generate 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[193]. Combined long short-term and GNN models have been used to represent patient-status sequences and temporal-medical-event graphs, respectively, to predict future prescriptions or disease codes[194,195]. Alternatively, a patient graph can be constructed on the basis of patient similarities, and patient embeddings learned by a long short-term model with a GNN architecture are then optimized to predict patient outcomes[196]. Furthermore, a short-term GCN[74] has been designed to leverage the underlying spatial and temporal dependencies of EHR data to generate patient diagnoses[174].

EHRs are often supplemented with other modalities, such as diseases, symptoms, molecular data and drug interactions[7,192,197,198] (Fig. 4d).

A probabilistic knowledge graph of EHR data, which include medical history, drug prescriptions and laboratory examination results, has been used to explore semantic relations between EHR entities in a shallow network embedding method[199]. Meta-paths may alternatively be exploited in an EHR-derived knowledge graph to leverage higher-order and semantically important relations for disease classification[200]. Node features for drugs and diseases can be initialized using Skip-gram and then a GNN leveraging multilayer message passing can be applied to predict adverse drug events[197]. Moreover, models combining recurrent neural networks and GNNs have been applied to EHR data integrated with drug–disease interactions to better recommend combinations of medications[201].

Outlook

Powered by network principles founded on decades of research, deep learning on graphs is poised to address major gaps in biology and medicine. Graph representation learning has provided insights into the structure and function of proteins and small molecules, captured disease-associated transcriptional changes (at single-cell resolution and considering the spatial context), and enabled new analyses via the fusion of biomedical knowledge and patient information.

As graph representation learning has aided the mapping of genotypes to phenotypes, leveraging it for fine-scale mapping of genetic variants appears promising[202]. By re-imagining genome-wide-association studies and expression-quantitative-trait-loci studies[203] as networks, biologically meaningful modules can be discovered that highlight key genes involved in the underlying mechanisms of a disease[204]. Alternatively, network propagation can be seeded with quantitative-trait-locus candidate genes[202]. Because graphs can model long-range dependencies or interactions, they can also model chromatin elements and the effects of their binding to regions across the genome[205,206]. Three-dimensional chromosomal structures could be reconstructed by predicting the 3D coordinates of nodes derived from a Hi-C contact map[207]. Graph representation learning for the analysis of spatial molecular-profiling data will continue to expand. For instance, with causal GNNs, one may be able to better capture changes in expression levels observed in scRNA-seq data over time or as a result of a perturbation[141,208].

Effective integration of healthcare data with molecular, genomic, disease and drug data can help generate more accurate and interpretable predictions about biological systems underlying health and disease[209]. Because of the utility of graphs, there has been a major push to generate knowledge graphs that synthesize and model multiscale and multimodal data, from genotype–phenotype associations to population-scale epidemiological dynamics. In public health, spatial and temporal networks could model space-dependent and time-dependent observations (such as disease states or susceptibility to infection[210]) to spot trends, detect anomalies and interpret temporal dynamics.

Importantly, as algorithms for graph representation learning are increasingly employed in biomedicine and healthcare, it is essential to ensure that the representations are explainable[211], fair[212] and robust[213], and that the algorithms are revisited to minimize health disparities[214] and to take into account new information on algorithmic biases.

References

1. Qiu, X. et al. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Syst* **10**, 265–274.e11 (2020).
2. Nicholson, D. N. & Greene, C. S. Constructing knowledge graphs and their biomedical applications. *Comput. Struct. Biotechnol. J.* **18**, 1414–1428 (2020).
3. Robinson, P. N. et al. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* **83**, 610–615 (2008).
4. Schriml, L. M. et al. Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* **40**, D940–D946 (2012).
5. Hong, C. et al. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. *npj Digital Med* **4**, 151 (2021).
6. Gysi, D. M. et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc. Natl Acad. Sci. USA* **118**, e2025581118 (2021).
7. Nelson, C. A., Butte, A. J. & Baranzini, S. E. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nat. Commun.* **10**, 3045 (2019).
8. Chen, R. J. et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. In *IEEE Transactions on Medical Imaging* Vol. 41, 757–770 (IEEE, 2022).
9. Callahan, T. J., Tripodi, I. J., Pielke-Lombardo, H. & Hunter, L. E. Knowledge-based biomedical data science. *Annu. Rev. Biomed. Data Sci.* **3**, 23–41 (2020).
10. Barabási, A.-L. Network medicine — from obesity to the "diseasome". *N. Engl. J. Med.* **357**, 404–407 (2007).
11. Mungall, C. J. et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res* **45**, D712–D722 (2017).
12. Goh, K.-I. et al. The human disease network. *Proc. Natl Acad. Sci. USA* **104**, 8685–8690 (2007).
13. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
14. Hu, J. X., Thomas, C. E. & Brunak, S. Network biology concepts in complex disease comorbidities. *Nat. Rev. Genet.* **17**, 615–629 (2016).
15. Zitnik, M. et al. Evolution of resilience in protein interactomes across the tree of life. *Proc. Natl Acad. Sci. USA* **116**, 4426–4433 (2019).
16. Agrawal, M., Zitnik, M. & Leskovec, J. Large-scale analysis of disease pathways in the human interactome. *In Pac. Symp. Biocomput.* **23**, 111–122 (2018).
17. Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C. & Collins, J. J. Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592 (2018).
18. Zhang, Z., Cui, P. & Zhu, W. Deep learning on graphs: a survey. In *IEEE Transactions on Knowledge and Data Engineering* Vol. 34, 249–270 (IEEE, 2020).
19. Hamilton, W. L., Ying, R. & Leskovec, J. Representation learning on graphs: methods and applications. *IEEE Data Eng. Bull* **40**, 52–74 (2017).
20. Hamilton, W. L. in *Synthesis Lectures on Artificial Intelligence and Machine Learning* Vol. 14, 1–159 (Morgan and Claypool, 2020).
21. Wu, Z. et al. A comprehensive survey on graph neural networks. In *IEEE Transactions on Neural Networks and Learning Systems* Vol. 32, 4–24 (IEEE, 2020).
22. Chen, F., Wang, Y.-C., Wang, B. & Kuo, C.-C. J. Graph representation learning: a survey. In *APSIPA Transactions on Signal and Information Processing* Vol. 9, E15 (Cambridge Univ. Press, 2020).
23. Li, B. & Pi., D. Network representation learning: a systematic literature review. *Neural Comput. Appl.* **34**, 16647–16679 (2020).
24. Yue, X. et al. Graph embedding on biomedical networks: methods, applications and evaluations. *Bioinformatics* **36**, 1241–1251 (2020).
25. Dong, Y., Hu, Z., Wang, K., Sun, Y. & Tang, J. Heterogeneous network representation learning. In *Proc. 29th International Joint Conference on Artificial Intelligence* 4861–4867 (IJCAI, 2020).

26. Kazemi, S. M. et al. Representation learning for dynamic graphs: a survey. *J Mach. Learn. Res.* **21**, 1–73 (2020).
27. Zitnik, M. et al. Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf. Fusion* **50**, 71–91 (2019).
28. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).
29. Blevins, A. S. & Bassett, D. S. in *Handbook of the Mathematics of the Arts and Sciences* (ed. Sriraman, B.) 2073–2095 (Springer, 2020).
30. Koutrouli, M., Karatzas, E., Paez-Espino, D. & Pavlopoulos, G. A. A guide to conquer the biological network era using graph theory. *Front. Bioeng. Biotechnol.* **8**, 34 (2020).
31. Liu, C. et al. Computational network biology: data, models, and applications. *Phys. Rep.* **846**, 1–66 (2020).
32. Rai, A., Shinde, P. & Jalan, S. Network spectra for drug-target identification in complex diseases: new guns against old foes. *Appl. Netw. Sci.* **3**, 51 (2018).
33. David, L., Thakkar, A., Mercado, R. & Engkvist, O. Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminformatics* **12**, 56 (2020).
34. Wieder, O. et al. A compact review of molecular property prediction with graph neural networks. *Drug Discov. Today. Technol.* **37**, 1–12 (2020).
35. Hetzel, L., Fischer, D. S., Günnemann, S. & Theis, F. J. Graph representation learning for single cell biology. *Curr. Opin. Syst. Biol.* **28**, 100347 (2021).
36. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2**, 573–584 (2020).
37. Sun, M. et al. Graph convolutional networks for computational drug development and discovery. *Brief. Bioinform.* **21**, 919–935 (2020).
38. Gaudelet, T. et al. Utilizing graph machine learning within drug discovery and development. *Brief. Bioinform.* **22**, bbab159 (2021).
39. MacLean, F. Knowledge graphs and their applications in drug discovery. *Expert Opin. Drug Discov* **16**, 1057–1069 (2021).
40. Zeng, X., Tu, X., Liu, Y., Fu, X. & Su, Y. Toward better drug discovery with knowledge graph. *Curr. Opin. Struct. Biol.* **72**, 114–126 (2022).
41. Ahmedt-Aristizabal, D., Armin, M. A., Denman, S., Fookes, C. & Petersson, L. A survey on graph-based deep learning for computational histopathology. *Comput. Med. Imaging Graph.* **95**, 102027 (2021).
42. Muzio, G., O'Bray, L. & Borgwardt, K. Biological network analysis with deep learning. *Brief. Bioinform.* **22**, 1515–1530 (2021).
43. Guo, M. et al. Analysis of disease comorbidity patterns in a large-scale China population. *BMC Med. Genomics* **12**, 177 (2019).
44. Le, D.-H. & Dang, V.-T. Ontology-based disease similarity network for disease gene prediction. *Vietnam J. Comput. Sci.* **3**, 197–205 (2016).
45. Menche, J. et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
46. Sumathipala, M., Maiorino, E., Weiss, S. T. & Sharma, A. Network diffusion approach to predict lncrna disease associations using multi-type biological networks: Lion. *Front. Physiol.* **10**, 888 (2019).
47. Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1197 (2019).
48. Cheng, F. et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat. Commun.* **10**, 3476 (2019).
49. Chen, Z.-H. et al. Prediction of drug–target interactions from multi-molecular network based on deep walk embedding model. *Front. Bioeng. Biotechnol.* **8**, 338 (2020).
50. Wong, L. et al. MIPDH: a novel computational model for predicting microRNA–mRNA interactions by DeepWalk on a heterogeneous network. *ACS Omega* **5**, 17022–17032 (2020).
51. Yang, K. et al. HerGePred: heterogeneous network embedding representation for disease gene prediction. *IEEE J. Biomed. Health Inform.* **23**, 1805–1815 (2018).
52. Geng, C. et al. iScore: a novel graph kernel-based function for scoring protein–protein docking models. *Bioinformatics* **36**, 112–121 (2020).
53. Veselkov, K. et al. HyperFoods: machine intelligent mapping of cancer-beating molecules in foods. *Sci. Rep.* **9**, 9237 (2019).
54. Zheng, A. & Casari, A. Distributed multi-task classification: a decentralized online learning approach. *Mach. Learn.* **107**, 727–747 (2018).
55. Perozzi, B., Al-Rfou, R. & Skiena, S. DeepWalk: online learning of social representations. In *Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 701–710 (ACM, 2014).
56. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed representations of words and phrases and their compositionality. In *Proc. 26th International Conference on Neural Information Processing Systems* Vol. 2 (eds Burges, C. J. et al.) 3111–3119 (Curran Associates, 2013).
57. Grover, A. & Leskovec, J. Node2vec: scalable feature learning for networks. In *Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 855–864 (ACM, 2016).
58. Tang, J. et al. LINE: Large-scale information network embedding. In *Proc. ACM Web Conference* 1067–1077 (ACM, 2015).
59. Dong, Y., Chawla, N. V. & Swami, A. metapath2vec: scalable representation learning for heterogeneous networks. In *Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 135–144 (ACM, 2017).
60. Bordes, A., Usunier, N., García-Durán, A., Weston, J. & Yakhnenko, O. Translating embeddings for modeling multi-relational data. In *Proc. 26th International Conference on Neural Information Processing Systems* Vol. 2 (eds Burges, C. J. et al.) 2787–2795 (Curran Associates, 2013).
61. Nickel, M., Tresp, V. & Kriegel, H. A three-way model for collective learning on multi-relational data. In *Proc. 28th International Conference on International Conference on Machine Learning* 809–816 (PMLR, 2011).
62. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É. & Bouchard, G. Complex embeddings for simple link prediction. In *Proc. 33rd International Conference on International Conference on Machine Learning* Vol. 48 (eds Balcan, M. F. & Weinberger, K. Q.) 2071–2080 (PMLR, 2016).
63. Sun, Z., Deng, Z., Nie, J. & Tang, J. RotatE: knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations* (2019).
64. Yang, B., Yih, W., He, X., Gao, J. & Deng, L. Embedding entities and relations for learning and inference in knowledge bases. In *International Conference on Learning Representations* (2015).
65. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *Proc. 34th International Conference on Machine Learning* Vol. 70, 1263–1272 (PMLR, 2017).
66. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations* (2017).
67. Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? In *International Conference on Learning Representations* (2019).
68. Duvenaud, D. et al. Convolutional networks on graphs for learning molecular fingerprints. In *Proc. 28th International Conference on Neural Information Processing Systems* Vol. 2 (eds Cortes, C. et al.) 2224–2232 (ACM, 2015).

69. Vinyals, O., Bengio, S. & Kudlur, M. Order matters: sequence to sequence for sets. In *International Conference on Learning Representations* (2016).
70. Defferrard, M., Bresson, X. & Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In *Proc. International Conference on Neural Information Processing Systems* (eds Lee, D. et al.) 3844–3852 (Curran Associates, 2016).
71. Velickovic, P. et al. Graph attention networks. In *International Conference on Learning Representations* (2018).
72. Hu, Z., Dong, Y., Wang, K. & Sun, Y. Heterogeneous graph transformer. In *Proc. ACM Web Conference* 2704–2710 (ACM, 2020).
73. Yun, S., Jeong, M., Kim, R., Kang, J. & Kim, H. J. Graph transformer networks. In *Proc. 33rd Conference on Neural Information Processing Systems* (eds Wallach, H. et al.) (Curran Associates, 2019).
74. Yan, S., Xiong, Y. & Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proc. AAAI Conference on Artificial Intelligence* (eds McIlraith, S. A. & Weinberger, K. Q.) 7444–7452 (AAAI, 2018).
75. Choi, E. et al. Learning the graphical structure of electronic health records with graph convolutional transformer. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 34, 606–613 (AAAI, 2020).
76. Xu, K. et al. Representation learning on graphs with jumping knowledge networks. In *Proc. 35th International Conference on MachineLearning* 5482–5493 (PMLR, 2018).
77. Abu-El-Haija, S. et al. MixHop: higher-order graph convolutional architectures via sparsified neighborhood mixing. In *Proc. 36th International Conference on MachineLearning* (PMLR, 2019).
78. Ying, Z. et al. Hierarchical graph representation learning with differentiable pooling. In *Proc. 32nd International Conference on Neural Information Processing Systems* (eds Bengio, S et al.) 4805–4815 (Curran Associates, 2018).
79. Schütt, K. et al. Schnet: a continuous-filter convolutional neural network for modeling quantum interactions. In *31st Conference on Neural Information Processing Systems* (eds Guyon, I. et al.) (Curran Associates, 2017).
80. Klicpera, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations* (2020).
81. Chiang, W. et al. Cluster-GCN: an efficient algorithm for training deep and large graph convolutional networks. In *Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 257–266 (ACM, 2019).
82. Zeng, H., Zhou, H., Srivastava, A., Kannan, R. & Prasanna, V. K. GraphSAINT: graph sampling based inductive learning method. In *International Conference on Learning Representations* (2020).
83. Schlichtkrull, M. et al. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference* (eds Gangemi, A. et al.) 593–607 (Springer, 2018).
84. Wang, X. et al. Heterogeneous graph attention network. In *Proc. ACM Web Conference* (eds Liu, L. & White, R.) 2022–2032 (ACM, 2019).
85. Pareja, A. et al. EvolveGCN: evolving graph convolutional networks for dynamic graphs. In *Proc. 34th AAAI Conference on Artificial Intelligence* Vol. 34 5363–5370 (AAAI, 2020).
86. Rossi, E. et al. Temporal graph networks for deep learning on dynamic graphs. In *International Conference on Machine Learning Workshop on Graph Representation Learning and Beyond* (2020).
87. Huang, K. & Zitnik, M. Graph meta learning via local subgraphs. In *Proc. International Conference on Neural Information Processing Systems* (eds Larochelle, H. et al.) 5862–5874 (Curran Associates, 2020).
88. Hu, W. et al. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations* (2020).
89. You, Y., Chen, T., Wang, Z. & Shen, Y. When does self-supervision help graph convolutional networks? In *International Conference on Machine Learning* (eds Daumé, H. & Singh, A.) 10871–10880 (JMLR, 2020).
90. Erdös, P. & Rényi, A. *On the Rvolution of Random Graphs* (Mathematical Institute of the Hungarian Academy of Sciences, 1960).
91. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
92. Barabási A.-L. et al. *Network Science* (Cambridge Univ. Press, 2016).
93. Jin, W., Barzilay, R. & Jaakkola, T. S. Junction tree variational autoencoder for molecular graph generation. In *International Conference on Machine Learning* (PMLR, 2018).
94. Kipf, T. N. & Welling, M. Variational graph auto-encoders. In *Advances in Neural Information Processing Systems Bayesian Deep Learning Workshop* (2016).
95. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
96. Wang, H. et al. GraphGAN: Graph representation learning with generative adversarial nets. In *Proc. 32nd AAAI Conference on Artificial Intelligence* (AAAI, 2018).
97. Simonovsky, M. & Komodakis, N. GraphVAE: towards generation of small graphs using variational autoencoders. In *International Conference on Artificial Neural Networks* (Springer, 2018).
98. You, J., Liu, B., Ying, R., Pande, V. & Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *Proc. International Conference on Neural Information Processing Systems* (eds Bengio S. et al.) (Curran Associates, 2018).
99. You, J., Ying, R., Ren, X., Hamilton, W. L. & Leskovec, J. GraphRNN: generating realistic graphs with deep auto-regressive models. In *International Conference on Machine Learning* (2018).
100. Yang, F., Fan, K., Song, D. & Lin, H. Graph-based prediction of protein–protein interactions with attributed signed graph embedding. *BMC Bioinf.* **21**, 323 (2020).
101. Huang, K., Xiao, C., Glass, L. M., Zitnik, M. & Sun, J. SkipGNN: predicting molecular interactions with skip-graph networks. *Sci. Rep.* **10**, 21092 (2020).
102. Yin, N. et al. Synergistic and antagonistic drug combinations depend on network topology. *PLoS ONE* **9**, e93960 (2014).
103. Fan, K. & Zhang, Y. Pseudo2GO: A graph-based deep learning method for pseudogene function prediction by borrowing information from coding genes. *Front. Genet.* **11**, 807 (2020).
104. Kearnes, S., McCloskey, K., Berndl, M., Pande, V. & Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.* **30**, 595–608 (2016).
105. Dutil, F., Cohen, J. P., Weiss, M., Derevyanko, G. & Bengio, Y. Towards gene expression convolutions using gene interaction graphs. In *International Conference on Machine Learning Workshop on Computational Biology* (2018).
106. Hamilton, W. L., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. In *Proc. International Conference on Neural Information Processing Systems* (eds Guyon, I. et al.) (Curran Associates, 2017).
107. Huan, J. et al. Comparing graph representations of protein structure for mining family-specific residue-based packing motifs. *J. Comput. Biol.* **12**, 657–671 (2005).
108. Fout, A., Byrd, J., Shariat, B. & Ben-Hur, A. Protein interface prediction using graph convolutional networks. In *Proc. International Conference on Neural Information Processing Systems* (eds Guyon, I. et al.) (Curran Associates, 2019).
109. Ingraham, J., Garg, V. K., Barzilay, R. & Jaakkola, T. S. Generative models for graph-based protein design. In *Proc. International*

Conference on Neural Information Processing Systems (eds Wallach, H. et al.) (Curran Associates, 2019).
110. Jin, W., Barzilay, R. & Jaakkola, T. Hierarchical generation of molecular graphs using structural motifs. In *International Conference on Machine Learning* 4552–4561 (2020).
111. Elton, D. C., Boukouvalas, Z., Fuge, M. D. & Chung, P. W. Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.* **4**, 828–849 (2019).
112. Guo, X. & Zhao, L. A systematic survey on deep generative models for graph generation. In *Association for Computing Machinery* (ACM, 2020).
113. Gainza, P. et al. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
114. Cao, Y. & Shen, Y. Energy-based graph convolutional networks for scoring protein docking models. *Proteins Struct. Funct. Bioinf.* **88**, 1091–1099 (2020).
115. Luck, K. et al. A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
116. Li, D. & Gao, J. Towards perturbation prediction of biological networks using deep learning. *Sci. Rep.* **9**, 11941 (2019).
117. Liu, Y., Yuan, H., Cai, L., & Ji, S. Deep learning of high-order interactions for protein interface prediction. In *Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 679–687 (ACM, 2020).
118. Yao, H., Guan, J. & Liu, T. Denoising protein-protein interaction network via variational graph auto-encoder for protein complex detection. *J. Bioinform. Comput. Biol.* **18**, 2040010 (2020).
119. Moreau, Y. & Tranchevent, L.-C. Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nat. Rev. Genet.* **13**, 523–536 (2012).
120. Zitnik, M. et al. Gene prioritization by compressive data fusion and chaining. *PLoS Comput. Biol.* **11**, 1004552 (2015).
121. Zhou, H., Beltran, F. & Brito, I. L. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Sci. Adv.* **7**, eabj5056 2021.
122. The Gene Ontology Consortium. The Gene Ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330–D338 (2019).
123. Zhou, G., Wang, J., Zhang, X. & Yu, G. DeepGOA: predicting gene ontology annotations of proteins via graph convolutional network. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 1836–1841 (IEEE, 2019).
124. Fan, K., Guan, Y. & Zhang, Y. Graph2GO: a multi-modal attributed network embedding method for inferring protein functions. *GigaScience* **9**, giaa081 (2020).
125. Hasibi, R. & Michoel, T. A Graph feature auto-encoder for the prediction of unobserved node features on biological networks. *BMC Bioinform* **22**, 525 (2021).
126. Cao, M. et al. Going the distance for protein function prediction: a new distance metric for protein interaction networks. *PLoS ONE* **8**, e76339 (2013).
127. Dey, T. K. & Mandal, S. Protein classification with improved topological data analysis. In *18th International Workshop on Algorithms in Bioinformatics (WABI)* (eds Parida, L. & Ukkonen, E.) 6:1–6:13 (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018).
128. Martino, A., Rizzi, A. & Mascioli, F. M. F. Supervised approaches for protein function prediction by topological data analysis. In *2018 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2018).
129. Nambiar, A. et al. Transforming the language of life: transformer neural networks for protein prediction tasks. In *11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (ACM, 2020).
130. Vig, J. et al. Bertology meets biology: interpreting attention in protein language models. In *International Conference on Learning Representations* (2021).
131. Han, P. et al. GCN-MF: disease–gene association identification by graph convolutional networks and matrix factorization. In *Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 705–713 (ACM, 2019).
132. Mandal, S., Guzmán-Sáenz, A., Haiminen, N., Basu, S. & Parida, L. A topological data analysis approach on predicting phenotypes from gene expression data. In *International Conference on Algorithms for Computational Biology* (eds Martín-Vide, C. et al.) 178–187 (Springer, 2020).
133. Nicolau, M., Levine, A. J. & Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc. Natl Acad. Sci. USA* **108**, 7265–7270 (2011).
134. Yang, C., Zhuang, P., Shi, W., Luu, A. & Li, P. Conditional structure generation through graph variational generative adversarial nets. In *Proc. International Conference on Neural Information Processing Systems* (eds Wallach, H. et al.) 124632 (Curran Associates, 2019).
135. Chereda, H., Bleckmann, A., Kramer, F., Leha, A. & Beissbarth, T. in *German Medical Data Sciences: Shaping Change—Creative Solutions for Innovative Medicine* (eds Röhrig, R. et al) 181–186 (GMDS, 2019).
136. Crawford, J. & Greene, C. S. Incorporating biological structure into machine learning models in biomedicine. *Curr. Opin. Biotechnol.* **63**, 126–134 (2020).
137. Rhee, S., Seo, S. & Kim, S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. In *Proc. Twenty-Seventh International Joint Conference on Artificial Intelligence* 3527–3534 (International Joint Conferences on Artificial Intelligence Organization, 2018).
138. Ramirez, R. et al. Classification of cancer types using graph convolutional neural networks. *Front. Phys.* **8**, 203 (2020).
139. Liu, S., Grau, B., Horrocks, I. & Kostylev, E. INDIGO: GNN-based inductive knowledge graph completion using pair-wise encoding. In *Proc. International Conference on Neural Information Processing Systems* (eds Ranzato, S. et al.) 2034–2045 (Curran Associates, 2021).
140. Rizvi, A. H. et al. Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* **35**, 551–560 (2017).
141. Burkhardt, D. B. et al. Quantifying the effect of experimental perturbations at single-cell resolution. *Nat. Biotechnol.* **39**, 619–629 (2021).
142. Ravindra, N., Sehanobish, A., Pappalardo, J. L., Hafler, D. A. & van Dijk, D. Disease state prediction from single-cell data using graph attention networks. In *Conference on Health, Inference, and Learning* 121–130 (ACM, 2020).
143. Huang, K. scGNN: scRNA-seq dropout imputation via induced hierarchical cell similarity graph. In *International Conference on Machine Learning Workshop on Computational Biology* (2020).
144. Wang, J. et al. scGNN is a novel graph neural network framework for single-cell RNA-seq analyses. *Nat. Commun.* **12**, 1882 (2021).
145. Chen, H., Ryu, J., Vinyard, M. E., Lerer, A. & Pinello, L. SIMBA: Single-cell embedding along with features. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.17.464750 (2021).
146. Buterez, D., Bica, I., Tariq, I., Andrés-Terré, H. & Liò, P. CellVGAE: an unsupervised scRNA-seq analysis workflow with graph attention networks. *Bioinformatics* **38**, 1277–1286 (2021).
147. Marx, V. Method of the year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14 (2021).
148. Yuan, Y. & Bar-Joseph, Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol* **21**, 300 (2020).

149. Partel, G. & Wählby, C. Spage2vec: Unsupervised representation of localized spatial gene expression signatures. *FEBS J* **288**, 1859–1870 (2021).
150. Meinshausen, N. et al. Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl Acad. Sci. USA* **113**, 7361–7368 (2016).
151. Guney, E., Menche, J., Vidal, M. & Barábasi, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 10331 (2016).
152. Cheng, F. et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* **9**, 2691 (2018).
153. Stokes, J. M. et al. A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.e13 (2020).
154. Coley, C. W. et al. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **10**, 370–377 (2019).
155. Xie, Y. et al. MARS: Markov molecular sampling for multi-objective drug discovery. In *International Conference on Learning Representations* (2021).
156. Alagappan, M., Jiang, D., Denko, N. & Koong, A. C. in *Tumor Microenvironment* (eds Koumenis, C. et al.) 253–268 (Springer, 2016).
157. Thafar, M. A. et al. DTiGEMS+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. *J. Cheminformatics* **12**, 44 (2020).
158. Thafar, M. A. et al. DTi2Vec: Drug–target interaction prediction using network embedding and ensemble learning. *J. Cheminformatics* **13**, 71 (2021).
159. Ma, T., Xiao, C., Zhou, J. & Wang, F. Drug similarity integration through attentive multi-view graph auto-encoders. In *Proc. Twenty-Seventh International Joint Conference on Artificial Intelligence* (ed. Lang, J.) 3477–3483 (AAAI, 2018).
160. Jiang, M. et al. Drug–target affinity prediction using graph neural network and contact maps. *RSC Adv* **10**, 20701–20712 (2020).
161. Chen, L. et al. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics* **36**, 4406–4414 (2020).
162. Quan, Z., Guo, Y., Lin, X., Wang, Z.-J. & Zeng, X. GraphCPI: Graph neural representation learning for compound-protein interaction. In *BIBM* 717–722 (2019).
163. Tsubaki, M., Tomii, K. & Sese, J. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics* **35**, 309–318 (2019).
164. Lin, X. et al. DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction. *ECAI* (2020).
165. Guo, Z.-H. et al. MeSHHeading2vec: a new method for representing mesh headings as vectors based on graph embedding algorithm. *Brief. Bioinform.* **21**, 1641–1662 (2020).
166. Wang, R., Li, S., Cheng, L., Wong, M. H. & Leung, K. S. Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning. *BMC Bioinf.* **20**, 628 (2019).
167. Wan, F., Hong, L., Xiao, A., Jiang, T. & Zeng, J. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics* **35**, 104–111 (2019).
168. Xie, Y., Peng, J. & Zhou, Y. Integrating protein–protein interaction information into drug response prediction by graph neural encoding. Preprint at *Research Square* https://doi.org/10.21203/rs.2.18936/v1 (2019).
169. Alsentzer, E., Finlayson, S. G., Li, M. M. & Zitnik, M. Subgraph neural networks. In *Proc. International Conference on Neural Information Processing Systems* (eds Larochelle, H. et al.) 8017–8029 (Curran Associates, 2020).
170. Buphamalai, P., Kokotovic, T., Nagy, V. & Menche, J. Network analysis reveals rare disease signatures across multiple levels of biological organization. *Nat. Commun.* **12**, 6306 (2021).
171. Barisoni, L., Lafata, K. J., Hewitt, S. M., Madabhushi, A. & Balis, U. G. Digital pathology and computational image analysis in nephropathology. *Nat. Rev. Nephrol.* **16**, 669–685 (2020).
172. Gurcan, M. N. et al. Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* **2**, 147–171 (2009).
173. Choi, E., Bahadori, M. T., Song, L., Stewart, W. F. & Sun, J. GRAM: graph-based attention model for healthcare representation learning. In *Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 787–795 (ACM, 2017).
174. Li, Y., Qian, B., Zhang, X. & Liu, H. Graph neural network-based diagnosis prediction. *Big Data* **8**, 379–390 (2020).
175. Bonito, P. et al. in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis* (eds Sudre, C. H. et al.) 208–219 (Springer, 2020).
176. Adnan, M., Kalra, S. & Tizhoosh, H. R. Representation learning of histopathology images using graph neural networks. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 4254–4261 (IEEE, 2020).
177. Anand, D., Gadiya, S. & Sethi, A. Histographs: graphs in histopathology. In *Medical Imaging 2020: Digital Pathology* (eds Tomaszewski, J. E. & Ward, A. D.) (SPIE, 2020).
178. Zhou, Y. et al. CGC-Net: cell graph convolutional network for grading of colorectal cancer histology images. In *IEEE International Conference on Computer Vision Workshops* (IEEE, 2019).
179. Jaume, G. et al. Towards explainable graph representations in digital pathology. In *International Conference on Machine Learning Workshop on Computational Biology* (2020).
180. Chao, C.-H. et al. Lymph node gross tumor volume detection in oncology imaging via relationship learning using graph neural network. In *23rd International Conference on Medical Image Computing and Computer Assisted Intervention* (eds Martel, A. L. et al.) 772–782 (Springer-Verlag, 2020).
181. An, X., Zhou, Y., Di, Y. & Ming, D. Dynamic functional connectivity and graph convolution network for Alzheimer's disease classification. In *7th International Conference on Biomedical and Bioinformatics Engineering* 1–4 (ACM, 2020).
182. Song, T.-A. et al. Graph convolutional neural networks for Alzheimer's disease classification. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 414–417 (IEEE, 2019).
183. Wee, C.-Y. et al. Cortical graph neural network for ad and mci diagnosis and transfer learning across populations. *NeuroImage Clin* **23**, 101929 (2019).
184. Mao, C., Yao, L. & Luo, Y. ImageGCN: Multi-relational image graph convolutional networks for disease identification with chest x-rays. *IEEE Trans. Med. Imaging* **41**, 1990–2003 (2022).
185. Levy, J., Haudenschild, C., Bar, C., Christensen, B. & Vaickus, L. Topological feature extraction and visualization of whole slide images using graph neural networks. *Pac. Symp. Biocomput.* **26**, 285–296 (2021).
186. Hu, J. et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat. Methods* **18**, 1342–1351 (2021).
187. Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S. & Sontag, D. Learning a health knowledge graph from electronic medical records. *Sci. Rep.* **7**, 5994 (2017).
188. Wu, T. et al. Representation learning of EHR data via graph-based medical entity embedding. In *Advances in Neural Information Processing Systems Graph Representation Learning Workshop* (2019).

189. Mao, C., Yao, L. & Luo, Y. MedGCN: medication recommendation and lab test imputation via graph convolutional networks. *J. Biomed. Inform.* **127**, 104000 (2022).
190. Ma, F. et al. KAME: knowledge-based attention model for diagnosis prediction in healthcare. In *27th ACM International Conference on Information and Knowledge Management* 743–752 (ACM, 2018).
191. Sun, Z. et al. Disease prediction via graph neural networks. *IEEE J. Biomed. Health Inform.* **25**, 818–826 (2020).
192. Chen, I. Y., Agrawal, M., Horng, S. & Sontag, D. Robustly extracting medical knowledge from EHRs: a case study of learning a health knowledge graph. In *Pac. Symp. Biocomput.* **25**, 19–30 (2020).
193. Chowdhury, S., Zhang, C., Yu, P. S. & Luo, Y. Mixed pooling multi-view attention autoencoder for representation learning in healthcare. Preprint at https://arxiv.org/abs/1910.06456 (2019).
194. Liu, S. et al. A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction. *Int. J. Mach. Learn. Cybern.* **11**, 2849–2856 (2020).
195. Lee, D., Jiang, X. & Yu, H. Harmonized representation learning on dynamic her graphs. *J. Biomed. Inform.* **106**, 103426 (2020).
196. Tong, C., Rocheteau, E., Veličković, P., Lane, N. & Liò, P. in *AI for DiseaseSurveillance and Pandemic Intelligence. W3PHAI 2021. Studies in Computational Intelligence* Vol. 1013 (eds Shaban-Nejad, A. et al.) 281–293 (Springer, 2022).
197. Kwak, H. et al. Drug-disease graph: predicting adverse drug reaction signals via graph neural network with clinical data. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* 633–644 (Springer-Verlag, 2020).
198. Zhao, C., Jiang, J., Guan, Y., Guo, X. & He, B. EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning. *Artif. Intell. Med.* **87**, 49–59 (2018).
199. Li, L. et al. A method to learn embedding of a probabilistic medical knowledge graph: algorithm development. *JMIR Med. Inf.* **8**, e17645 (2020).
200. Hosseini, A., Chen, T., Wu, W., Sun, Y. & Sarrafzadeh, M. HeteroMed: heterogeneous information network for medical diagnosis. In *International Conference on Information and Knowledge Management* 763–772 (ACM, 2018).
201. Shang, J., Xiao, C., Ma, T., Li, H. & Sun, J. GameNet: graph augmented memory networks for recommending medication combination. In *Proc. AAAI Conference on Artificial Intelligence* 1126–1133 (AAAI, 2019).
202. Wu, S., Chen, D. & Snyder, M. P. Network biology bridges the gaps between quantitative genetics and multi-omics to map complex diseases. *Curr. Opin. Chem. Biol.* **66**, 102101 (2022).
203. Umans, B. D., Battle, A. & Gilad, Y. Where are the disease-associated eQTLs? *Trends Genet* **37**, 109–124 (2020).
204. Wang, T., Peng, Q., Liu, B., Liu, Y. & Wang, Y. Disease module identification based on representation learning of complex networks integrated from GWAS, eQTL summaries, and human interactome. *Front. Bioeng. Biotechnol.* **8**, 418 (2020).
205. Dekker, J. & Misteli, T. Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.* **7**, a019356 (2015).
206. Lanchantin, J. & Qi, Y. Graph convolutional networks for epigenetic state prediction using both sequence and 3D genome data. *Bioinformatics* **36**, i659–i667 (2020).
207. Hovenga, V., Oluwadare, O. & Kalita, J. Hic-GNN: a generalizable model for 3D chromosome reconstruction using graph convolutional neural networks. Preprint at *bioRxiv* https://doi.org/10.1101/2021.11.29.470405 (2021).
208. Ding, J., Sharon, N. & Bar-Joseph, Z. Temporal modelling using single-cell transcriptomics. *Nat. Rev. Genet.* **23**, 355–368 (2022).
209. Fortelny, N. & Bock, C. Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biol* **21**, 190 (2020).
210. Machens, A. et al. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infect. Dis.* **13**, 185 (2013).
211. Ying, Z., Bourgeois, D., You, J., Zitnik, M. & Leskovec, J. GNNExplainer: generating explanations for graph neural networks. In *Proc. International Conference on Neural Information Processing Systems* (ed Wallach, H. et al.) (Curran Associates, 2019).
212. Agarwal, C., Lakkaraju, H. & Zitnik, M. Towards a unified framework for fair and stable graph representation learning. In *Proc. Machine Learning Research* (eds de Campos, C. & Maathuis, M. H.) 2114–2124 (PMLR, 2021).
213. Zhang, X. & Zitnik, M. GNNGuard: defending graph neural networks against adversarial attacks. In *Proc. International Conference on Neural Information Processing Systems* (eds Larochelle, H. et al.) 9263–9275 (Curran Associates, 2020).
214. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
215. Xu, K. et al. Representation learning on graphs with jumping knowledge networks. In *Proc. 35th International Conference on MachineLearning* (PMLR, 2018).
216. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
217. Kondratova, M. et al. A multiscale signalling network map of innate immune response in cancer reveals cell heterogeneity signatures. *Nat. Commun.* **10**, 4808 (2019).
218. Mohammadi, S., Davila-Velderrain, J. & Kellis, M. Reconstruction of cell-type-specific interactomes at single-cell resolution. *Cell Syst* **9**, 559–568.e4 (2019).
219. Li, M. M. & Zitnik, M. Deep contextual learners for protein networks. In *International Conference on Machine Learning Workshop on Computational Biology* (2021).
220. Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535–D539 (2006).
221. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
222. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
223. Zhang, Y. & Zhang, Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cell. Mol. Immunol.* **17**, 807–821 (2020).
224. Alessandri, L. et al. Sparsely-connected autoencoder (SCA) for single cell RNAseq data mining. *npj Syst. Biol. Appl.* **7**, 1 (2021).
225. Tran, D. et al. Fast and precise single-cell data analysis using a hierarchical autoencoder. *Nat. Commun.* **12**, 1029 (2021).
226. Zitnik, M., Agrawal, M. & Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **34**, i457–i466 (2018).
227. Jiang, P. et al. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Comput. Struct. Biotechnol. J.* **18**, 427–438 (2020).
228. Kim, Y. et al. Anticancer drug synergy prediction in understudied tissues using transfer learning. *J. Am. Med. Iinf. Assoc.* **28**, 42–51 (2021).

229. National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a new Taxonomy of Disease* (National Academies Press, 2011).

Acknowledgements

We gratefully acknowledge the support of the National Science Foundation, under grants IIS-2030459 and IIS-2033384, the US Air Force Contract No. FA8702-15-D-0001, the Harvard Data Science Initiative, and awards from Amazon Research, Bayer Early Excellence in Science, AstraZeneca Research, and Roche Alliance with Distinguished Scientists. M.M.L. is supported by T32HG002295 from the National Human Genome Research Institute and a National Science Foundation Graduate Research Fellowship. Any opinions, findings, conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the funders.

Author contributions

M.M.L. and M.Z. conceived the work and shaped its framing. M.M.L. performed background research and wrote the manuscript together with K.H. and M.Z. All authors discussed the content, and reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41551-022-00942-x.

Correspondence should be addressed to Marinka Zitnik.

Peer review information *Nature Biomedical Engineering* thanks Feixiong Cheng, Fabian Theis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.