Published in "Physics Reports 846: 1–66, 2020"
which should be cited to refer to this work.

Computational network biology: Data, models, and applications

Chuang Liu [a], Yifang Ma [b], Jing Zhao [c], Ruth Nussinov [d,e], Yi-Cheng Zhang [i,a,*], Feixiong Cheng [f,g,h,*], Zi-Ke Zhang [j,a,*]

[a] *Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, China*
[b] *Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, China*
[c] *Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai, China*
[d] *Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute at Frederick, Frederick, MD 21702, USA*
[e] *Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel*
[f] *Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA*
[g] *Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA*
[h] *Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA*
[i] *Department of Physics, University of Fribourg, Fribourg, CH 1700, Switzerland*
[j] *College of Media and International Culture, Zhejiang University, Hangzhou 310028, China*

Biological entities are involved in intricate and complex interactions, in which uncovering the biological information from the network concepts are of great significance. Benefiting from the advances of network science and high-throughput biomedical technologies, studying the biological systems from network biology has attracted much attention in recent years, and networks have long been central to our understanding of biological systems, in the form of linkage maps among genotypes, phenotypes, and the corresponding environmental factors. In this review, we summarize the recent developments of computational network biology, first introducing various types of biological networks and network structural properties. We then review the network-based approaches, ranging from some network metrics to the complicated machine-learning methods, and emphasize how to use these algorithms to gain new biological insights. Furthermore, we highlight the application in neuroscience, human disease, and drug developments from the perspectives of network science, and we discuss some major challenges and future directions. We hope that this review will draw increasing interdisciplinary attention from physicists, computer scientists, and biologists.

Keywords:
Complex networks
Network biology
Disease module
Machine learning

Contents

* Corresponding authors.
E-mail addresses: yi-cheng.zhang@unifr.ch (Y.-C. Zhang), chengf@ccf.org (F. Cheng), zhangzike@gmail.com (Z.-K. Zhang).

1. Introduction

The fast growth of biomedical data, especially data from individual cells, pose concrete questions that we can undertake, such as how to interpret disease mutations, what their potential network consequences are, and implications for, disease diagnosis and drug discovery, and how can these help in treatment decisions. To date, assessments of human diseases are generally made based on statistics and correlations. Yet, statistics do not always permit mechanistic understanding. Statistics reflect the normalized sum totals; it does not account for the distinct, personal environments. Thus, often, it is not only the most frequently observed disease-correlated mutation (driver) that matters; other protein residue substitutions can take place, as can differ cellular networks in cells, tissues, and individuals (see Fig. 1). These can affect the molecular network landscape, and the consequences, such as disease emergence or progression. Can we forecast which substitutions can collaborate to lead to certain network-level outcomes? This argues for a detailed approach, and raises the question of not only which protocols to consider, but also which factors to scrutinize, and, broadly, how can we then integrate the disciplines. Therefore, on the one hand, we are overwhelmed by the rapidly accumulating data from next-generation sequencing technologies and thus statistics; on the other hand, their interpretation, in terms of considerations of dynamic network perturbation, which translate genomics data into function, and dysfunction, lag behind. Genetics, large data screenings, and pharmacology are insufficient; to be actionable, the data must merge with the network basis of the protein (node) behavior, and latch onto the human cellular network.

It is widely accepted that bio-molecules do not perform their functions alone, but do so interactively with one another to form so-called bio-molecular networks. For instance, a disease is rarely a consequence of an abnormality in a single gene, but reflects the perturbations or malfunctions of the complex biological networks that link tissues and organ systems [1]. Genes associated with similar diseases are likely to interact and have similar expression, forming the "disease module" [2]. The terminology of "network biology" can be described from different standpoints, ranging from molecule–molecule interaction to cell–cell interaction, from unipartite network to bipartite network (e.g., drug–target interaction), or even tri-partite network (e.g., drug–disease–protein interaction). It is clear that networks are being generated in ever increasing sizes by exploiting the success of experimental techniques. The availability of large network datasets and the affordable computing capability have driven the development of bioinformatics algorithms and computational biology approaches to analyze data to offer biological insights. Benefiting from the advances of network science and high-throughput biomedical technologies, studying the biological systems from the network perspective has attracted much attention recently [3–5] and networks have long been central to our understanding of biological systems, in the form of linkage maps among proteins (genes or neurons), their associated phenotypes (diseases), and the corresponding environmental factors (i.e., drugs) [6].

Network biology corresponds to a theoretical framework for representing the biological structure, termed a graphical model, and the functional flow of information through this structure. Understanding and modeling the network structure would lead to a better knowledge of its evolutionary mechanisms, and to a better acceptance of its dynamical and functional behaviors. In this review, we will describe some common properties observed in the topology of biological networks as well as their measures, with a special emphasis on how to model the network structure [7,8] and how the topology structure affects the biological functional behaviors [9,10]. For instance, the biological networks are always scale-free [11,12], and the network hubs tend to be the essential parts [13], which would be critically important in disease progression, cellular dysfunction, and so on. The efficient and functional flow of information in the biological network would be related to the small-world properties [14], and the network nodes cluster into the module to exert its biological functions [15]. The biological networks are always heterogeneous, whereas the disease can be considered as the increase of network heterogeneity [16]. In addition, the highly resilient nature of the network failures would help the life system to retain biological functionality in the face of the physical damage or environmental variation.

As graphical models are essential representations that characterize many different domains of physics and engineering, they can significantly promote the development of computational network biology approaches and their applications across these domains. Therefore, we review a list of network-based computational algorithms, such as network centrality [17], network propagation [18], network controllability [19], and module detection [20]. Going beyond the simple summarization of the computational methods, we emphasize how to use these approaches to mine biological information, e.g., to identify the disease associated genes using the network propagation approaches. Importantly, machine learning, especially the deep-learning-based methods, show a strong capacity in extracting hidden information and new insights into network biology [21,22]. Based on these computational network-based approaches, we introduce how to use the knowledge of the human interactome networks to interpret associations between genetic variants and diseases [23,24] and discuss how to use computational network models for drug development [25–27]. We will highlight the possibilities of network-based methodologies for accelerating novel therapeutic development and personalized treatment [28,29]. In summary, human cellular networks are governed by universal laws and offer a cutting-edge technology that would revolutionize structural view of biology, disease, and therapeutics in the new genomic era [30].

The rest of this article is organized as follows. First, we introduce several important types of biological networks, including the protein–protein interaction, isoform–isoform networks, genetic interactions, metabolic associations, neuron–neuron interactions, and drug–target interactions, and we also provide some data resources in Section 2. In Section 3, we review the network structure in biological systems, focusing on two points: biological network structural properties and the mapping between the cell function and topological structure of biological networks. In Section 4, we introduce the

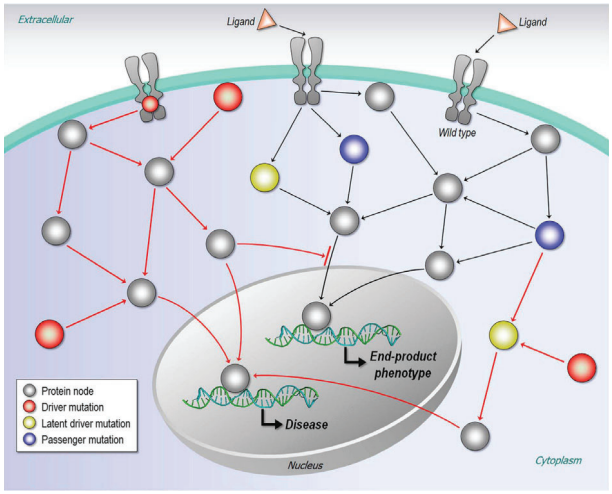

Fig. 1. Illustration of human interactome network perturbation altered by disease mutations. A protein in a signaling pathway is typically modulated by upstream activation (or inhibition), with effects (transmission of the signals or blocking them) propagating via interactors to pathways downstream, thereby affecting the network.
Source: Figure from Ref. [28].

network-based methods in network biology, which can be roughly divided into network-structure-oriented methods [31] and machine-learning-based methods. In Section 5, we introduce the application of the computational network biology, concentrating on discuss how to use the computational network biology to interpret human disease, neuroscience, and drug development. Finally, we conclude this review in Section 6 with several major challenges and future directions of computational network biology.

2. Types of biological networks

Simplifying sophisticated macromolecules as "nodes", and the large number of interactions (physical, biochemical, or functional) between them as "edges", offers a first-order understanding of sub-cellular systems at a global scale for nearly the entire human interactome. In this review, we focus on several types of biological networks, including human protein–protein interactions, isoform–isoform networks, genetic interactions, metabolic associations, neuron–neuron correlations, and drug–target interactions.

2.1. Protein–protein interaction network

Protein–protein interactions (PPIs) are the physical links between two or more protein macromolecules characterized by biochemical contacts in cells. There are several experimental strategies for mapping human PPIs, such as yeast two-hybrid assay (Y2H), which measures direct physical interactions in cells, and affinity-purification-mass spectrometry, which measures the composition of protein complexes. Specifically, we can group PPIs into six different experimental evidentiary types by assembling various publicly available PPI data: (i) Binary, physical PPIs tested by high-throughput Y2H systems from publicly available high-quality Y2H datasets [32,33]; (ii) three-dimensional structure-solved PPIs built from the published protein structure databases, such as the Protein Data Bank [34], Interactome3D [35], Instruct [36], and Interactome INSIDER [37]; (iii) kinase-substrate interactions by literature-derived low- and high-throughput experiments from KinomeNetworkX [38], PhosphoNetworks [39], PhosphositePlus [40], DbPTM 3.0 [41], and Phospho. ELM [42]; (iv) signaling networks by literature-derived low-throughput experiments as annotated in SignaLink2.0 [43]; (v) protein complexes data identified by a robust affinity-purification-mass spectrometry methodology collected from BioPlex V2.0 [24]; and (vi) carefully literature-curated PPIs identified by affinity purification followed by mass spectrometry (AP-MS) and by literature-derived low-throughput experiments from Human Protein Resource Database (HPRD) [44], BioGRID [45], PINA [46], MINT [47], IntAct [48], and InnateDB [49]. Fig. 2 shows the most comprehensive human protein–protein interactome network which contains 351,444 PPIs connecting 17,706 proteins [50], and also provides the degree distribution and several topological characteristics. The detailed bioinformatics resources for curating human PPIs are provided in Table 1.

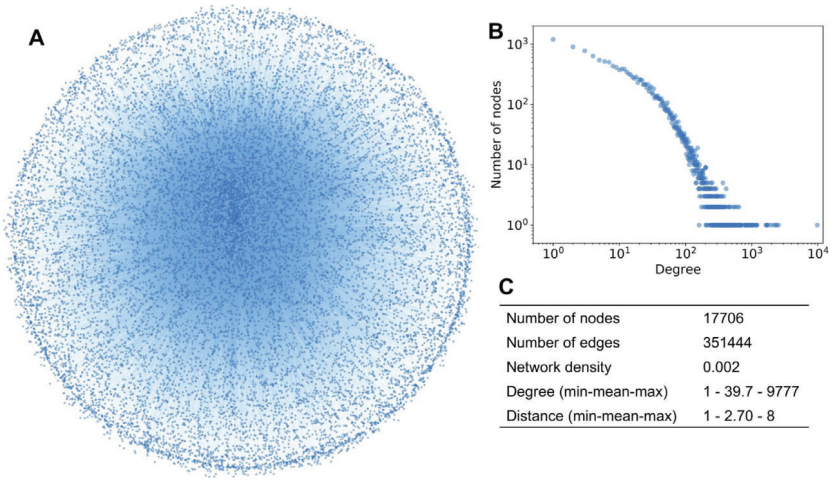

Fig. 2. Network picture of human protein–protein interactome network. This interactome contains 17,706 human proteins and 351,444 interactions that have been curated by assembling six types of experimental evidence. (A) Overall complex network of human interactome. (B) Degree (connectivity) distribution of proteins by following a power-law tail. (C) Several selected network topological characteristics of the interactome.

Table 1
Summary of representative bioinformatics resources for curating human protein–protein interaction networks.

Databases	Description	Website	Refs.
BioGRID	Integrated protein–protein interaction data.	http://thebiogrid.org	[45]
HPRD	Human protein–protein interaction data.	http://www.hprd.org	[51]
Interactome3D	Manually curated PPIs with known three-dimensional structure information.	http://interactome3d.irbbarcelona.org	[35]
STRING	Functional protein association networks database.	http://string-db.org	[52]
MINT	Protein–protein interactions in refereed journals.	http://mint.bio.uniroma2.it/mint	[47]
KinomeNetworkX	Integrative kinase-substrate database.		[38]
PhosphoNetworks	High-resolution phosphorylation network connecting the specific phosphorylation sites present in substrates with their upstream kinases.	http://www.phosphonetworks.org/	[39,53]
PhosphositePlus	Database and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, acetylation, and more.	https://www.phosphosite.org/homeAction.action	[40]

2.2. Isoform–isoform network

The newest human GENCODE project has revealed ~20,000 protein-coding genes in the human genome; however, full protein coding capacity of the genome and the full extent to which different isoforms are differentially expressed remains underexplored. Human genes can encode multiple protein "forms" via alternative transcription, splicing, 3'-end formation, translation, and post-translational modification [54]. Previous study has suggested a central role for alternative splicing in network organization, function, and cross-tissue dynamics, demonstrating the importance of an isoform-resolved global view of interactome networks [55]. Compared to PPIs, isoform–isoform networks are significantly larger by possible post-translational modification of products of all of the possible transcript informs. For example, it is estimated that as many as ~100,000 distinct isoform transcripts could be produced from the ~20,000 human protein-coding genes, which may lead to perhaps over a million distinct isoform–isoform interactions [54,56]. Binary PPI mapping has allowed us to discriminate isoform-specific interactions [55]. In a global analysis of interactome network maps, alternative isoforms are more likely to have as distinct proteins rather than minor variants of each other. Interaction partners specific to alternative isoform tend to be expressed in a highly tissue-specific manner and form distinct functional modules [55]. Mathematically, the isoform–isoform interaction network could be formatted as a network-of-networks [57]. In this context, a protein, which is a single node in the traditional protein interactome, becomes a sub-network of interaction between the different isoforms encoded by the underlying gene (see Fig. 3). This network-of-networks framework may offer potential application

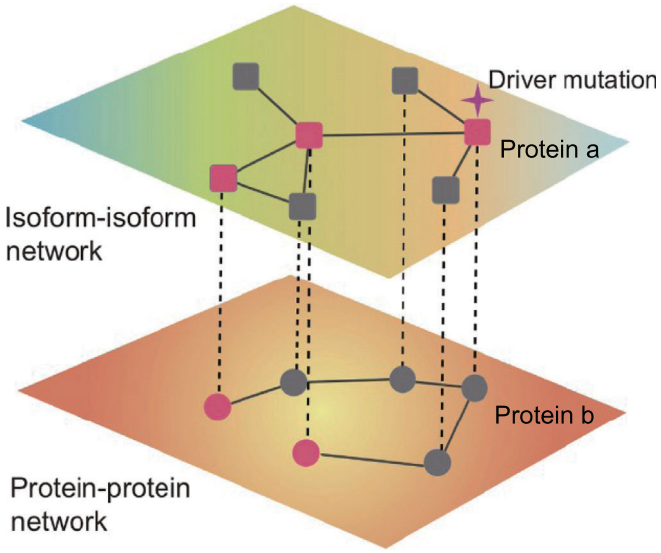

Fig. 3. A diagram illustrating isoform-level network-of-networks framework to model the effect of specific disease (cancer) variants. In a traditional view, proteins interact in the protein–protein interactions (PPIs) without discernment of the underlying gene alternative splicing. Different isoforms from the same gene can have different interacting partners. When a mutation (e.g., cancer driver mutations) affects a particular isoform, it only affects its interacting partners. The square nodes in the above layer denote different isoforms of protein *a*, while the cycle nodes in the below layer denote different isoforms of protein *b*. Links (edges) denote isoform–isoform interactions among isoforms with the same protein and dash lines denote isoform–isoform interactions among different protein isoforms.

in human disease, e.g., cancer. For example, blocking isoform-specific cancer mutations on the RAS family has shown potential in treating cancer [29].

2.3. Genetic interaction network

In the context of genetic studies, a genetic interaction (i.e., synthetic lethal) involves two genes: the cell is viable upon perturbation of either gene alone, but simultaneous perturbation of both genes by genetic or genomic alterations will result in cell growth or death [58]. A genetic interaction occurring between a tumor-specific somatic mutation and a gene that drives tumorigenesis and tumor progression offers an ideal therapeutic target in effective anticancer therapeutic development [59]. Furthermore, discovery of genetic interactions through identification of a second-site synthetic lethal druggable target facilitates indirect targeting of tumor alterations of undruggable proteins (e.g., KRAS or TP53) [58,60,61]. Recent advances in functional genomic technologies, such as RNA interference (RNAi) or CRISPR-Cas9 assays, have offered innovative tools with which to comprehensively screen human cancer cells for genetic interactions [62,63]. By application of CRISPR-based screens, Wang et al. uncovered PREX1, a key synthetic lethal interactor of oncogenic RAS in human acute myeloid leukemia cell lines [61]. However, measurements of cell proliferation in genome-scale CRISPR-Cas9 loss-of-function screens have a potentially high false-positive rate in copy-number-amplified regions [64,65]. Furthermore, large-scale experimental assays are expensive and time-consuming. Fig. 4 provides a landscape of genetic interactions in cells [66].

Computational approaches with low cost and high efficiency offer new tools for genome-wide identification of cancer genetic interactions and for inferring tumor evolution through analyzing publicly available large-scale tumor exome/genome sequencing data [67–69]. Recent efforts to map genetic interactions in tumor cells have suggested that tumor vulnerabilities can be exploited for the development of novel targeted therapies. Tumor-specific genomic alterations derived from multi-center cancer genome projects allow identification of genetic interactions that promote tumor vulnerabilities, offering novel strategies for development of targeted cancer therapies. Cheng et al. [69] proposed a mathematical model, termed the gene gravity model, derived from Newton's law of gravitation to systematic assessment of genetic interactions from cancer genomes. Specifically, the gene gravity model detects a gene-gene pair in which two genes are co-mutated and highly co-expressed simultaneously in a specific cancer type. By applying the gene gravity model to approximately 3000 cancer exomes across nine cancer types, Cheng et al. identified multiple potential cancer genes by altering genetic interactions and cancer genome evolutions [69].

2.4. Metabolic networks

A metabolic network is a large system containing a series of chemical reactions in organisms that determines the biochemical and physiological functions of the cell and maintains the balance and regularities of the entire system [70,71].

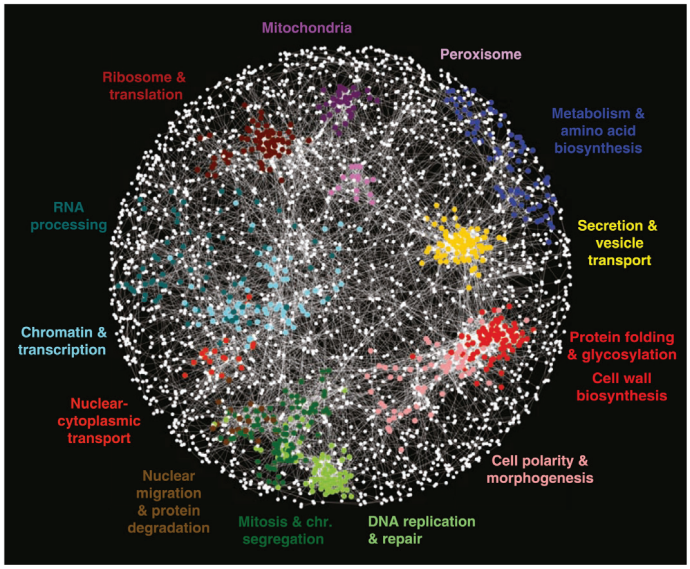

Fig. 4. Landscape of genetic interactions in cells. Edges between genes denote Pearson correlation coefficients (*PCC* > 0.2) calculated from the complete genetic interaction matrix. The network was laid out using an edge-weighted, spring-embedded, network layout algorithm. Colored regions indicate sets of genes enriched for GO biological processes summarized by the indicated terms. Genes sharing similar patterns of genetic interactions are proximal to each other; less similar genes are positioned farther apart.
Source: Figure from Ref. [66].

Table 2
List of recent published models in genome scale metabolic network reconstructions. The basic statistic of the corresponding networks are provided, and # indicates the missing corresponding data in the work.

Organism	Organism genes	Model genes	Metabolites	Reactions	Refs.
Akkermansia muciniphila	#	588	540	744	[73]
Bordetella pertussis	3464	762	#	1675	[74]
Corynebacterium glutamicum	#	773	950	1207	[75]
Geobacillus thermoglucosidasius	#	736	1163	1159	[76]
Oenococcus oeni	1864	454	536	660	[77]
Piscirickettsia salmonis	3074	584	801	1323	[78]
Streptomyces clavuligerus	7281	1021	1360	1494	[79]

This system can convert food (or any other kind of energy source) into the energy forms that an organism can use directly, convert fuel to building blocks for cell growth, and integrate small molecules, including proteins, lipids, nucleic acids, and carbohydrates. Work on the structure, function, and evolution of metabolic networks has been developed for decades, and metabolic networks are evolving from small-scale systems to the large-genome-scale metabolic networks in systems biology.

To construct a sub-graph of a metabolic network, at the genomic scale [72], in which nodes are products and links are the underlying biochemical reactions, we must answer the following questions: (i) what products does an enzyme act on, (ii) what are the stoichiometric coefficients for each metabolite in the reaction, (iii) which reactions are reversible or not, and (iv) what are the locations of the reactions? The construction processes are also complex, which contains four steps: the draft reconstruction phase, curated reconstruction, genome-scale metabolic model, and platform for design and discovery. With an increasing number of scientists around the world working on these reconstruction processes, large-scale metabolic networks have become available in recent decades. We provide a list of validated reconstructions that have been converted into predictive gene-scale models in recent years in Table 2.

Biochemically, a metabolic pathway is a linked chain of chemical reactions in the metabolic network that helps to convert molecules and substrates into more usable materials. Enzymes play a crucial role in the reaction to accelerate, slow, or stop the metabolic processes. In general, there are two types of metabolic pathways, catabolic and anabolic. Catabolic pathways break down molecules by releasing energy, such as respiration, while anabolic pathways integrate molecules with the help of energy. There is a large number of metabolic pathways, e.g., glycolysis is one of the important pathways, which is the initial phase in the metabolic system in which sugar is partially oxidized to small molecules, and generates some ATPs for the entire system. Other important metabolic pathways include the pentose phosphate pathway, the Entner–Doudoroff pathway, the tricarboxylic acid cycle, and the glyoxylate cycle [80]. Additionally, all the common

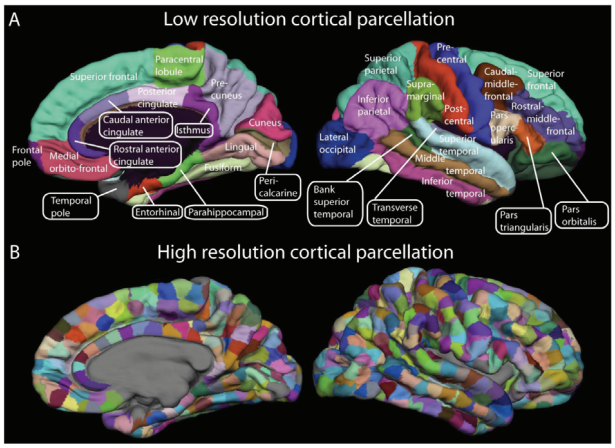

Fig. 5. Low- (A) and the high-resolution (B) cortical parcellations of cerebral cortex.
Source: Figure from Ref. [90].

used metabolic pathways can be collected through the KEGG pathway database,[1] which can be used to construct the metabolic network.

Questions involving metabolic network include the structures, the dynamics, and also the interaction between different nodes in response to diseases such as obesity and diabetes. Jeong et al. [81] analyzed the entire structure of metabolic networks in different organisms. Based on the WIT database [82], they constructed a metabolic network based on the biochemical reactions between educts, and found that, similar to most of the non-biological systems in the world, these metabolic networks showed scaling properties and a power-law degree distribution. Furthermore, the modular structures of the metabolic networks also attracted a series of studies [83,84]. Lee et al. [85] proposed a cell-metabolism-based human disease network by using the metabolic network structure and the mutated enzymes in the metabolic reactions. They found that the disease pairs connected in the constructed disease network had a higher correlated reaction flux rate and higher co-morbidity than the cases of disease pairs without metabolic links between them. This human metabolic network-topology-based method highlighted the possibility using metabolic networks for disease diagnosis and prevention.

2.5. Brain network

The brain network is a skeleton of brain regions with connections. In a brain network, nodes are functional regions in brain, which can be regarded as unitary neural masses [86]. There are multiple techniques to identify the nodes in a brain network, e.g., the anatomical parcellation of the Montreal Neurological Institute (MNI) using structural magnetic resonance imaging (sMRI) data and also functional magnetic resonance imaging (fMRI) [86,87]. Edges are usually from two different processes, which leads to two different configurations of brain networks. One type of link are those links made by physical connections [88]. Using the MRI data, these links are the fiber tracts reconstructed from image processing [89,90], e.g., Fig. 5 shows the network structures based on this imaging method. Another kind of link is made by functional connectivity [91]. These links are based on the functional activities between pairs of nodes, and the link weights are usually calculated from signal series analysis [92]. The connectivity (either structural or functional) can be represented as a connectivity matrix, and the element in the matrix is the weight between two associated brain regions. Based on the processes of identifying brain networks, series of brain networks were identified [93], e.g., the default mode network, dorsal attention network, ventral attention network, salience network, and frontoparietal network.

2.6. Drug-target networks

The drug–target network (as illustrated in Fig. 6) can be described as a bipartite graph $G(D, T, P)$, where the drug set is $D = (d_1, d_2, \ldots, d_n)$, the target (gene) set is $T = (t_1, t_2, \ldots, t_m)$, and the interaction set is $P = (p_{il}, d_i \in D, t_l \in T)$. An interaction is drawn between d_i and t_l when drug d_i binds with target t_l with binding affinity (such as IC_{50}, K_i, or K_d) less than a given threshold value. The types of binding affinities contain the half-maximal inhibitor concentration (IC_{50}), dissociation constant (K_d), and inhibitory constant (K_i). Mathematically, a drug–target bipartite network can be presented

[1] KEGG: Kyoto Encyclopedia of Genes and Genomes, https://www.kegg.jp/.

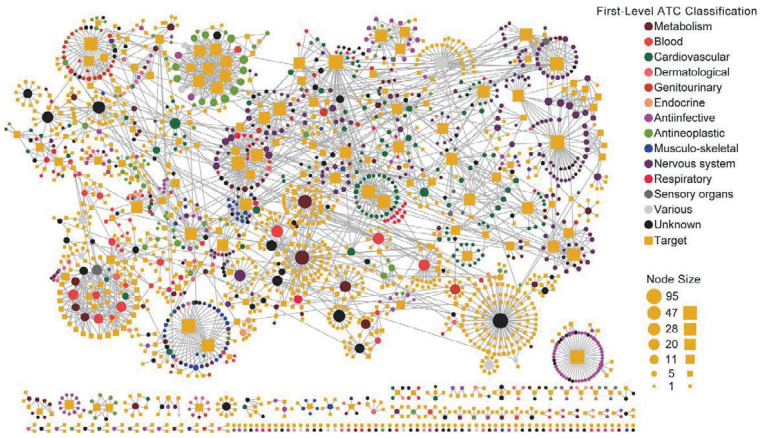

Fig. 6. The known drug–target (DT) bipartite network. In a DT network, a drug node (circle) and a target node (square) are connected to each other by a gray edge if the target is annotated to have known experimental interactions with the drug in DrugBank. The DT network was generated using known FDA-approved small-molecule DT interactions. The size of the drug node is the fraction of the number of targets that the drug linked in DrugBank. The size of the target node is the fraction of the number of drugs that the target linked in DrugBank. Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification.
Source: Figure from Ref. [27].

by an $n \times m$ adjacent matrix p_{ij}, where $p_{ij} = 1$ if the binding affinity between d_i and t_j is less than 10 μM; otherwise, $p_{ij} = 0$, as described in Eq. (1):

$$p_{ij} = \begin{cases} 1 & IC_{50}(K_i) \leq 10\ \mu\text{M} \\ 0 & IC_{50}(K_i) > 10\ \mu\text{M} \end{cases}, \tag{1}$$

There are multiple publicly available databases containing high-quality data for building drug–target interactions, including the DrugBank database (v4.3) [94], Therapeutic Target Database (TTD) [95], and PharmGKB database [96]. Specifically, bioactivity data for drug–target pairs can be collected from ChEMBL (v20) [97], BindingDB [98], and IUPHAR/BPS Guide to PHARMACOLOGY [99]. Detailed drug–target databases are provided in Table 3 (Section 1). To improve the quality of data, we usually focus on physical drug–target interactions based on the following three criteria: (i) the human target is represented by a unique UniProt accession number; (ii) the target is marked as "reviewed" in the UniProt database [100]; and (iii) binding affinities, including K_i, K_d, IC_{50}, or EC_{50} each $\leq 10\ \mu$M. In addition, we can also build functional drug–gene association networks from drug-induced transcriptomics data or proteomics data derived from human cells (Section 2 in Table 3).

Drug targets are nodes within the human interactome, intrinsically coupled between therapeutic indication and adverse effect. The systematic identification of drug targets across many proteins in the human genome is crucial towards the understanding of therapeutic indication versus side effect. A recent study revealed that drug targets can be viewed as special nodes and are not randomly distributed (drug module) in the human interactome (see Fig. 7). Mathematically, we can formulate identifying drug targets from the human interactome as finding the most possible configuration of a network [26]. The network-based location of each drug module characterizes its pharmacological and biological relationships to other drugs. Drugs with overlapping modules show significantly chemical, biological, functional, and therapeutic similarities, whereas drugs residing in the separated network neighborhoods are pharmacologically distinct [26].

3. Biological network structure

Network structure is the fundamental percept of network science, where research on network dynamics, function, and application are all based on the understanding of the network structure. In this section, we briefly review the recent research on the structural properties of biological networks. For convenience, we denote $G = (V, E)$ as a biological network, where $v \in V$ is the set of nodes (genes, proteins, or some other items) and $e \in E$ the set of interactions between these nodes. The number of nodes and edges are N and M, respectively. The network can also be represented as the adjacent matrix $A = a_{i,j}$, where $a_{i,j} = 1$ when there is a link e_{ij} between node v_i and v_j, and $a_{i,j} = 0$ otherwise.

3.1. Degree distribution and hubs

In network science, the degree of a node v_i (denoted k_i, mathematically, $k_i = \sum_j a_{i,j}$) is defined as the number of links to v_i. Many biological networks exhibit the scale-free [111] (or scale-free-like) characteristic, where the node

Table 3
Summary of chemoinformatics and bioinformatics resources for re-constrcuing drug–target networks.

Databases	Description	Number of interactions, drugs, and targets	Website	Refs.
Section 1. Databases for collecting physical drug–target interactions				
DrugBank	Detailed drug data with comprehensive target information.	8250 drug entries, including 2,016 FDA-approved small-molecule drugs and over 6000 experimental drugs.	http://www.drugbank.ca/	[94]
TTD	Information on therapeutic targets.	31,614 drugs and 2589 targets.	http://bidd.nus.edu.sg/group/ttd	[95]
ChEMBL	Chemical properties and biological activities of drug-like molecules.	2,036,512 compounds against 11,224 targets and 14,371,197 bioactivity records.	https://www.ebi.ac.uk/chembldb	[101]
BindingDB	Binding affinities of proteins with small drug-like ligands.	565,136 compounds against 6612 proteins and 1,279,670 binding affinity data.	http://www.bindingdb.org	[102]
PubChem	Repository of small-molecule biological activities.	More than 230 million bioactivities connecting 9.3 million compounds and 9851 targets.	http://puchem.ncbi.nlm.nih.gov	[103]
DGIdb	Drug–gene interaction database.	26,298 unique drug–gene interactions connecting 7569 drugs and 7524 unique genes.	http://dgidb.genome.wustl.edu/	[104]
STITCH	Experimental and predicted compound-protein interactions.	1.6 billion interactions connecting 0.5 million compounds and 9.6 million proteins from 2031 organisms.	http://stitch.embl.de/	[105]
SuperPred	Experimental and predicted compound-protein interactions.	341,000 compounds, 1800 targets, and 665,000 compound-target interactions.	http://prediction.charite.de/index.php	[106]
Section 2. Databases for collecting functional drug–gene interactions				
CMap		Gene-expression signatures to connect 1309 small molecules and 7000 genes in five cancer cell lines.	https://www.broadinstitute.org/cmap	[107]
LINCS		Library of integrated network-based cellular signatures for 1 million gene expression profiles.	http://www.lincscloud.org/	[108]
TG-GATEs		Large-scale toxicogenomics database.	http://toxico.nibio.go.jp/english/index.html	[109]
DrugMatrix		Molecular toxicology reference database and informatics system.	http://ntp.niehs.nih.gov/drugmatrix/index.html	[110]

degree follows a power-law distribution, and the degree distribution $P(k)$ follows $P(k) \sim k^{-\lambda}$, where λ is called the degree exponent. The degree distribution is highly heterogeneous, where many nodes have few links while a few nodes are heavily connected. Jeong et al. [13] found that the degree of protein follows a power-law distribution with an exponential cutoff in the saccharomyces cerevisiae protein–protein interaction network. A similar degree distribution can also be found in the protein–protein interaction network of the human bacterium Helicobacter pylori [112], the gene co-expression network [113,114], BioPlex (biophysical interactions of ORFEOME-derived complexes) [115], the brain functional network [11], and so on. Network theory indicates that scale-free networks are very robust to random attack [116], which is in agreement with systematic mutagenesis results in which the striking capacity of yeast to tolerate the deletion of a substantial number of individual proteins from its proteome [13]. The hubs, the most highly connected proteins in the cell, are always the most important for the cell's survival [13,117]. Evidence from model organisms indicates that hubs are older and have evolved more slowly [8,118,119]. The hubs in the brain networks play important roles in information integration underpinning numerous aspects of complex cognitive function [120], leading to the susceptibility of the hubs to disconnection and dysfunction in brain disorders. In addition, brain hubs are also significantly related to the energy efficiency of the brain [121], where the brain hubs would be more vulnerable to deficits in energy delivery or utilization.

According to the observation of the correlation between a protein's age and its degree, Eisenberg and Levanon [8] proposed the preferential attachment mechanism [111] in the protein network evolution by classifying the proteins into

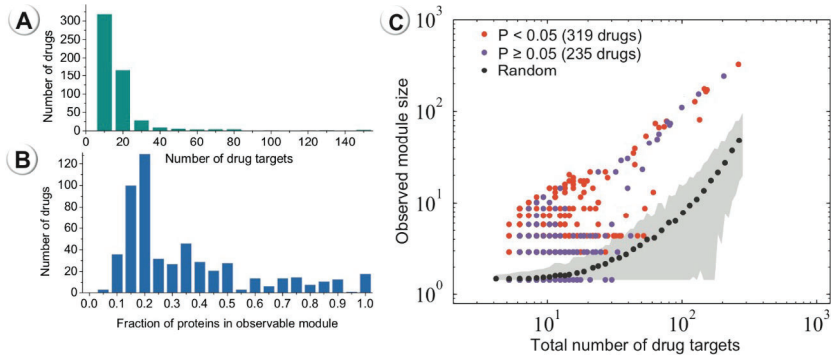

Fig. 7. Proof-of-concept of identifying drug modules in human interactome. (A) Distribution of the number of targets for 554 drugs with at least five targets. (B) Distribution of the fraction of drug targets within the observable drug module. (C) Illustration of the identified largest connected component (LCC, Y axis) for 554 drugs in human interactome compared to random expectation. Each dot denotes a drug. Drugs forming statistically significant modules based on S_i and S_i^r comparison ($P < 0.05$, permutation testing) are shown in red; otherwise, they are shown in purple. Degree-preserving randomizations are shown in black. Gray shading denotes standard derivation during the degree-preserving randomization.

four age groups. Preferential attachment is more likely implemented by interacting with the high degree proteins, which will grow to be the hubs in the network. Gene duplication and divergence would be the origin of the scale-free topology of biological networks [122–124]. In the duplication-divergence processes [125–127], gene duplication is a mechanism that generates genes producing identical proteins that interact with the same protein neighbors, and the degree of the protein connected with a duplicated protein increases. Highly connected proteins have a natural advantage because they are more likely to have a link to a duplicated protein than other proteins. Therefore, the preferential attachment rule is achieved implicitly, with highly connected nodes having more chance to gain new links from the duplicated proteins [4]. In addition, some other evolving models, including fitness-driven preferential attachment [128], the age-dependent stochastic model [129], and preferential depletion [7], are also proposed to generate the heterogeneous degree distribution in biological networks.

Compared to an undirected network, the degree distribution for a directed network would be more complicated, because of the different patterns of in-degree and out-degree. The chemical reactions of metabolites in a cell generate the metabolic networks, in which the direction of the edge represents the chemical reaction direction. Both the in-degree and out-degree distribution are approximately of power-law type (at least broad-tailed), and they also share the similar exponent [12,130]. Cross-regulatory interactions of transcription factors form the gene regulatory network [131], where the edge direction represents the regulatory interactions. The degree distribution of the gene regulatory network for the in-degree and out-degree are very different, where out-degree distribution has power-law characteristics, which is very similar to other biological network [130], while in-degree distribution is an exponential distribution, which is much narrower than that of a power-law distribution [132]. More specifically, for the gene regulatory networks, the broad out-degree distribution would be driven by the mutations of the transcription factors [133]. In addition, such asymmetric degree distribution, where the out-degree distribution is much broader than the in-degree distribution, is also widely obtained in social networks [134], where the edge contains the information propagation direction from peer to peer.

3.2. Network path and distance

3.2.1. Shortest path

The shortest path between two nodes in the network plays a very important role in the network structure and dynamics [135]. For example, in a social network, the shortest paths would be the most effective way to transfer information [136]. As illustrated in Fig. 8, the shortest path is the network path with the shortest length among all the paths between the node pairs (the triangle and circle). For the node pair $(i, j) \in V$, the shortest path length is defined as the distance ($d_{i,j}$) for the node pair. In the context of biological networks, the connection of node pairs via shortest paths is highly motivated. The shortest paths based method are widely used to assign directions to the edges in the protein–protein interaction networks [137,138], to infer the regulatory pathways through the corresponding genes [139] and to identify the cancer-related genes [140] or key components [141].

Based on the network distance, Guney et al. [142] proposed various distance measures, including the closest (d_c), shortest (d_s), kernel (d_k), center (d_{cc}), and separation (d_{ss}) measures (as illustrated in Eq. (2)), to evaluate the proximity between the drug and disease, where the drugs and diseases are represented as set of drug target proteins (e.g., Gliclazide has two target proteins and Daunorubicin has two target proteins as illustrated) and the disease-related genes (e.g., Type 2 diabetes has four related genes and Acute myeloid leukemia has three related genes) in the biological network, respectively. Fig. 8 shows the shortest paths between drug targets and disease proteins for two known drug–disease

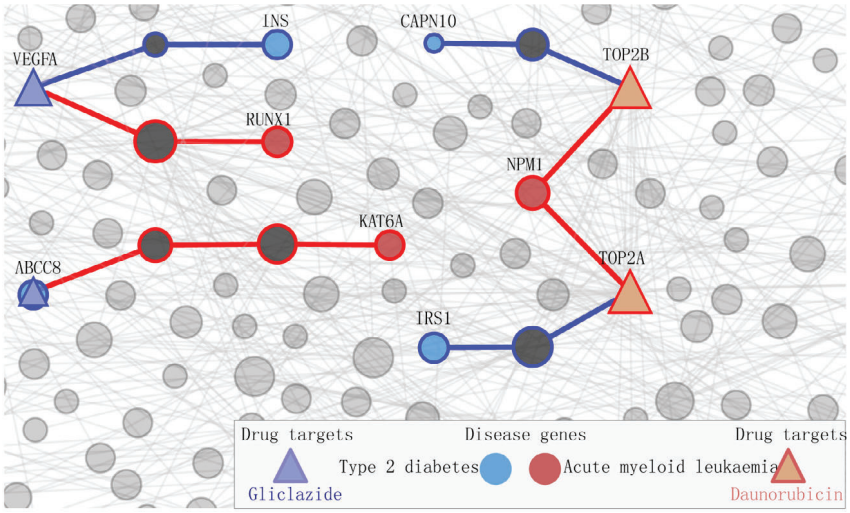

Fig. 8. Shortest paths between drug target proteins and disease proteins in protein–protein network. Triangles and circles represent drug target proteins and disease proteins, respectively. Blue and red links illustrate shortest path from drug target proteins to nearest disease proteins.
Source: Figure from Ref. [142].

associations. The distances can be calculated as follows:

$$\begin{cases} d_c(S,T) = \dfrac{1}{\| T \|} \displaystyle\sum_{t \in T} \min_{s \in S} d_{s,t} \\ d_s(S,T) = \dfrac{1}{\| T \|} \displaystyle\sum_{t \in T} \dfrac{1}{\| S \|} \sum_{s \in S} d_{s,t} \\ d_k(S,T) = \dfrac{-1}{\| T \|} \displaystyle\sum_{t \in T} \ln \sum_{s \in S} \dfrac{e^{-(d_{s,t}+1)}}{\| S \|} \\ d_{cc}(S,T) = \dfrac{-1}{\| T \|} d_{centre_S,t} \\ d_{ss}(S,T) = \dfrac{\| T \| d_c(S,T) + \| S \| d_c(S,T)}{\| T \| + \| S \|} - \dfrac{d'_c(S,S) + d'_c(T,T)}{2}, \end{cases} \tag{2}$$

where $s \in S$ is the set of disease proteins, $t \in T$ the set of drug proteins, $d_{s,t}$ the distance between nodes s and t in the PPI network, $\| * \|$ the number of the proteins in set $*$, $centre_s$ the topological center of S, which is defined as $centre_s = argmin_{u \in S} \sum_{s \in S} d_{s,t}$, and d'_c the modified d_c in which the shortest path length from a node to itself is infinite.

According to the testing of the proximity between drugs and diseases, Guney et al. found that the known drugs (validated drug–disease associations) are much closer to their disease, and for the un-validated drugs of the disease, the closer drugs are more likely to be tested in clinical trails [142]. More specifically, a unique integration of the network proximity and large-scale patient-level longitudinal data can facilitate drug repurposing [143]. In addition, similar distance-based measures are also used to depict the disease–disease relationship through the protein–protein interactome [23].

3.2.2. Network efficiency

Based on the shortest path distance, there is an very important measure called the network efficiency [144], which can be calculated as

$$NE = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{i,j}}. \tag{3}$$

Network efficiency is an indicator of the traffic capacity of a network. Using Eq. (3), Csermely et al. found that the partial inhibition of a small number of the links to target proteins can be more efficient than the complete drug-induced inhibition of a single target protein in the protein–protein interaction network. The multi-target drug-design strategies may be more effective than single drugs [145], leading to the development of drug combinations for complex diseases [26,146]. Moreover, network efficiency is widely used in brain functional networks, which have an efficient structure, supporting parallel information transfer at very low cost [147,148]. The network efficiency would decrease significantly for some

brain-related diseases, including small-vessel disease [149], Type 2 diabetes [150], and Alzheimer's disease [151], and, furthermore, the brain network evolves to more efficient structure during recovery from traumatic brain injury [152]. Human intelligence would be also related to the network efficiency of the brain anatomical network. By dividing 79 healthy young adults into general- and high-intelligence groups according to their intelligence quotient (IQ) test scores, Li et al. found that the network efficiency of the brain network in the high-intelligence group is significantly higher than in the general-intelligence group [153].

3.2.3. Betweenness and bottlenecks

Betweenness is a path-based measure of the node importance in terms of the number of the shortest paths that pass through the node [154], which is defined as

$$B_v = \sum_{i,j,v\in V,i\neq j} \frac{\delta_{i,j}(v)}{\delta_{i,j}}, \tag{4}$$

where $\delta_{i,j}$ is the number of the shortest paths from node i to j and $\delta_{i,j}(v)$ the number of these shortest paths that pass through node v.

The node betweenness also follows the power-law distribution in Cayley trees [155,156] and other scale-free networks [157]. The correlations between the betweenness and the degree are not always positive, and the nodes with small degree and large betweenness are found to be abundant in various networks [158–160]. Fig. 9 gives an example of a node with small degree and large betweenness (the protein CAK1 labeled with a bold circle), which acts as the important connections of different modules (as shown by the red and blue parts in Fig. 9).

The nodes with high betweenness are also referred to as bottlenecks [160]. According to the definition of betweenness, we can find that bottlenecks would control most of the information flow in the network, and deleting the bottlenecks will reduce the network efficiency significantly. Studies on the protein–protein network indicate that proteins with high betweenness are more likely to be highly pleiotropic [161] and tend to be essential [162,163] or disease-related genes [164]. Just as with the hub protein, the evolutionary age of proteins has a positive correlation with betweenness [159], and the hubs vary significantly more slowly than bottlenecks [165]. Based on identifying generalized hierarchies of the transcription factor regulatory network (directed network), betweenness in the middle levels (levels 2 and 3) are significantly larger than in other levels of the hierarchy [166]. Betweenness of the brain functional networks has also attracted much attention from testing the brain activity in different scenes [167]. Makarov et al. [168] applied betweenness to evaluate the transition of the brain functional network during mental task evaluation. Comparing the cumulative number of shortest paths going through each brain region during the transition from a resting state to cognitive task evaluation, they found the gradual transition from a sharp decrease of betweenness across the parietal and frontal lobes in the delta range to a pronounced increase in the gamma range [168]. By constructing whole-brain inter-regional interactions for each participant, Ueda et al. [169] found that the correlation between neuroticism and betweenness are very different for various brain regions, implying that the neuroticism trait is likely formed as a result of interactions among the brain regions related to neuroticism.

Garcia-Vaquero et al. [170] proposed the Double Specific Betweenness (S2B) score to define the node's information traffic between the nodes in different sets. S2B can be obtained as follows:

$$S2B_v(s_1, s_2) = \sum_{i\in s_1, j\in s_2, i\neq j, v\in V} \frac{\delta_{i,j}(v)}{\delta_{i,j}}, \tag{5}$$

where s_1, s_2 are two different sets in node set V. $S2B$ can be used to discover the associated genes with two related diseases, and s_1 and s_2 should be the known disease-associated genes for the two diseases, respectively [170].

The betweenness concept can also be extended to the edges [171], and the edge-based betweenness is defined as

$$B_e = \sum_{i,j\in V, i\neq j, e\in E} \frac{\delta_{i,j}(e)}{\delta_{i,j}}, \tag{6}$$

where $\delta_{i,j}(e)$ is the number of shortest paths from node i to j that pass through edge e. The edge-betweenness is a very powerful clustering tool with which to detect the community structure (Section 3.5) in protein–protein interaction networks [172,173].

3.3. Clustering coefficient

Triangular structures abound in many networks, for example, two individuals with a common friend are more likely to know each other in social networks [174]. This property can be described by the clustering coefficient [175]:

$$C_i = \frac{2l_i}{k_i(k_i-1)}, \tag{7}$$

where l_i is the number of the links between the k_i neighbors of node v_i (number of triangles that go through i) and $\frac{k_i(k_i-1)}{2}$ the number of the possible connections between the k_i neighbors of node i (number of the possible triangles that go

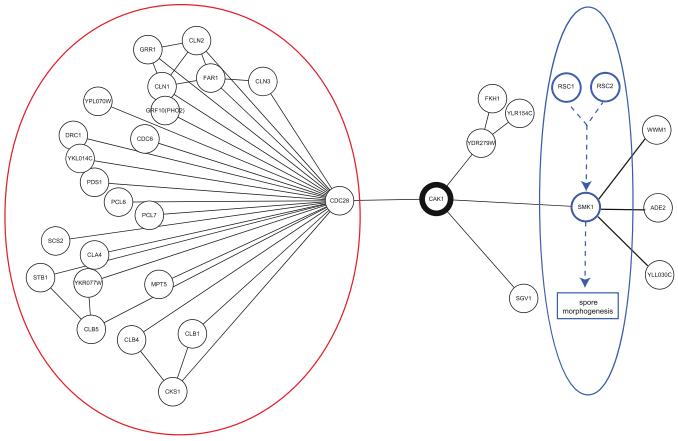

Fig. 9. Biological example of a bottleneck in the protein–protein interaction network. CAK1 (bold circle) is not a hub (total degree is only 4), but a bottleneck (the largest betweenness) in this network. Specifically, CAK1 is a cyclin-dependent kinase-activating kinase involved in two key signaling-transduction pathways: cell cycle (CDC28) and sporulation (SMK1).
Source: Figure from Ref. [160].

through v_i). Taking protein CKS1 (at the bottom of the left-hand red part) in Fig. 9 as an example, there are two pairs (CLB4-CDC28 and CLB1-CDC28) among the three neighbor proteins (CLB4, CLB1, and CDC28) to be linked together, leading to $l_{CKS1} = 2$, and $C_{CKS1} = \frac{2*2}{3*(3-1)} = \frac{2}{3}$. By contrast, none of the four neighbor proteins (CDC28, SGV1, SMK1, and YDR279 W) of protein CAK1 link to each other, leading to $C_{CAK1} = 0$. The clustering coefficient of the entire network $\langle C \rangle$ is defined as the average C_i over all nodes in the network. By definition, $0 \leq C_i \leq 1$ and $0 \leq \langle C \rangle \leq 1$, and large $\langle C \rangle$ indicates that nodes tend to be clustered together to form more triangles.

The average clustering coefficient for most biological networks is significantly larger than that of random networks with similar degree distribution [176,177], and high clustering would be a generic feature of biological networks [4]. The average clustering coefficient follows a scaling law with degree k [for example, $C(k) \sim k^{-1}$] in metabolic networks, which is also an indicator of the hierarchical modular organization (see Section 3.5) [83], and which would be caused by the existence of super-hubs in the network [178]. Comparing the brain functional networks confronted with point-light display stimuli depicting biological motion and scrambled motion, Fraiman et al. found higher clustering coefficients in different brain regions; for example, the right occipital lobe for scrambled motion and central parietal lobe for biological motion [179]. In addition, the edge clustering coefficient [180,181], which is defined as the fraction of the triangles that path through the corresponding edge, is proposed to predict the essential proteins and protein complexes.

3.4. Small world

It is a common feature in nearly all kinds of complex networks that any node pair can be connected within very limited links (small value of the average distance). This feature is known as the small world property, which was first investigated as "six degrees of separation" in a social study [182]. Subsequently, the small world property has been observed in various systems, including biological networks [183], communication networks [184], and co-operation networks [185] [186]. Besides, regarding the small value of average distance, the small world property observed in real networks is often related to a high value of the clustering coefficient, which is not very common in random graph models [111,187]. Watts and Strogatz [175] proposed the random rewiring procedure from a ring lattice to generate the small world network, the average distance of which increases logarithmically with network size N and the clustering coefficient ($\langle C \rangle$) of which is significantly large. Therefore, the small world network would also have very high value of network efficiency (as in Eq. (3)), which brings efficiency to information exchange [144].

With the high efficiency of the small world network, biological networks are more likely to evolve to follow the small world property. A Fibroblastic Reticular Cell (FRC) network exhibits the typical small world properties, and the small world topology can be regenerated within approximately 4 weeks after complete node removal [188]. Small world is an attractive model for the organization of brain anatomical and functional networks [189,190], where small world is economical in the sense of providing high global and local efficiency of parallel information processing for low wiring and energy costs [14,147,191]. Upon observing changes in neuronal connectivity during the cultured neurons, Downes et al. found that a network exhibited random topology in the early cultures stage, and evolved to a small world topology during maturation [192]. The small world network organized by immature cells can promote cell proliferation [193]. The functional brain network is a highly modular structure in which the distance between nodes from different modules is not very short and the precisely organized weak ties provide the optimal global integration of these modules to generate the small world property [194]. The functional brain networks for Alzheimer's disease are abnormally organized for a

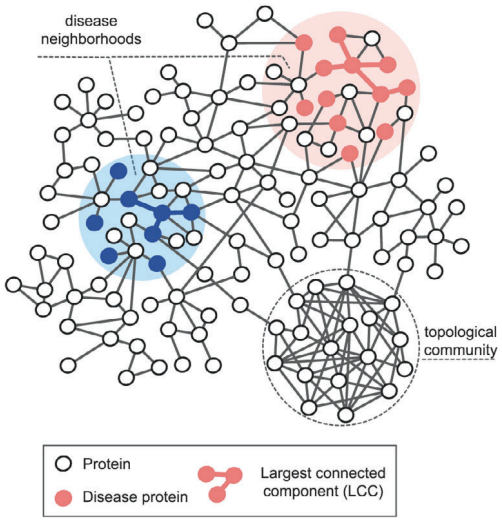

Fig. 10. Illustration of disease module in protein–protein interaction network. Proteins associated with the same disease tend to localize in specific neighborhoods in the network, indicating the approximate location of the corresponding disease modules. Topological network communities are highly interconnected groups of nodes.
Source: Figure from Ref. [201].

loss of small world property, such as the functional disconnection between distant brain areas [195–197]. The small world property is highly correlated with the memory performance, where a better performance in short-term memory as evaluated by digit span tests is detected with the increase of small worldness in a specific frequency band during the resting state from EEG data [198].

3.5. Community structure

3.5.1. Community in network biology

In a social network, many people with a similar interest would interact with each other very frequently to form a friends circle, while the communication of people in different circles would be very little. Such a phenomenon is called a community structure (or module) [199], where the nodes within the communities are highly connected and the links between different communities are sparse (as in the topological module illustrated in Fig. 10, the density of edges is higher within a community than that across communities). As illustrated in Section 3.2.3, the non-hub bottlenecks tend to connect different modules [160]. The community feature is very common in real networks, and it is interesting to find that communities in the network structure are always related to the network function. For example, in the network of the World Wide Web, the communities may be the web pages on similar topics [200].

Community structure is a common feature in biological systems, and the nodes in the systems cluster into several communities to exert their biological functions [15,202]. In a protein–protein interaction network, proteins having the same specific function [203] or forming protein complexes [115,204] are more likely to cluster into a community, which can also be used to predict the protein's function [205]. Huttlin et al. applied the unsupervised Markov clustering method [206] to detect protein communities, and most communities were found to be associated with the disease annotations, in which the candidate disease genes enrich the communities [24]. The modularity structure can promote resilience to the node failure within a network [207]; for example, the chaperome module can maintain the proteostasis in brain aging and neurodegenerative disease [208]. Guimera et al. applied the modularity-based method to analyze the metabolic networks of 12 organisms, finding that 80% of the nodes are only connected to other nodes within their respective modules, and that metabolites whose links are not many but connect different modules (bottlenecks) are more conserved than hubs whose links are mostly within a module [209]. Brain networks also consist of modules, which are associated with specific cognitive functions [210–212]. Modularity plays a key role in the functional organization of brain areas during normal and pathological activity, where healthy brains exhibit a sparse connectivity, whereas the brain networks of patients display a rich connectivity with a clear modular structure [9]. The modularity value calculated from fMRI data increases as children develop into adults [213]. The asymmetric activity propagation and propagation delays in the neural circuits should be supported by the modularity structure in the brain networks [214]. Of note is that the performance of the module structures depends on the detection algorithms, where the modules generated from different algorithms are always very different. Choobdar et al. [215] discovered the "Disease Module Identification DREAM Challenge" to test the predicted network modules from various algorithms for association with complex traits and disease, and provided some useful practical recommendations for module identification, including applying diverse algorithms to

identify complementary modules, using the modules from different methods without forming a consensus, leveraging diverse networks, and applying the algorithm on each network individually first without merging networks.

In addition, the modularity in biological systems always exhibits the property of hierarchical organization [207,216,217], in which many small, highly connected modules combine in a hierarchical manner into larger module [83]. Sales-Pardo et al. [218] proposed an unsupervised method for extracting the hierarchical organization in metabolic networks, and almost all the metabolites in the submodules can be classified as being in the same pathways. Wei [219] constructed a mutilayered hierarchical gene regulatory network that is centered on a given transcription factor, to identify the target genes and collaborative regulators for the observed phenotypic changes in its transgenic lines. Jiao et al. [220] proposed a multi-scale node-similarity-based method to mine the modular structure on protein–protein interaction networks and reveal the hierarchical organization of protein complexes. Xu et al. [221] proposed an adaptive partitioning algorithm to illustrate the hierarchical module structure of the gene expression network, and the risk of death could be well predicted based on the gene expression profiles in the modules. Reyna et al. [222] introduced the hierarchical HotNet method to find the hierarchy of modules, which is effective in predicting cancer genes. Lahav et al. [223] applied the k-shell decomposition method on the human cortical network to reveal the human brain's global functional organization, in which modules were further categorized into three hierarchies reflecting different functional roles. Very recently, Wang et al. [224] applied the eigenmode analysis to reveal the hierarchical modular structural connectome in the brain, which allows a nested functional segregation and integration across multiple spatio-temporal scales. The combination of the hierarchical modular organization and the critical state provides the capacity to maximize the functional diversity. Comparing the brain's structural and functional network architectures, Ashourvan et al. [225] found that the structural networks display a more prominent hierarchical community organization than functional networks. Additionally, parallel processing [226] and information retrieval [227] in biological systems are also related to the hierarchical organization of the modularity.

3.5.2. Module detection

Identifying the topological community structure (also referred to as a dense subgraph; see the dish circle in Fig. 10) of the complex networks is a very active field in complex networks [199,228]. Many approaches for network community detection have been discovered in recent years [215], among which the most popular detection methods aim to find the partition of the network to maximize the modularity [20,229,230], which is defined as follows:

$$Q = \frac{1}{2M}\sum_{i,j\in V}[A_{ij} - P_{ij}]\delta_{C_i,C_j}, \tag{8}$$

where A_{ij} is the adjacency matrix, M the total number of edges of the network, P_{ij} represents the expected number of edges between vertices i and j (e.g., $P_{ij} = \frac{k_i k_j}{2M}$ [231]), and the $\delta_{C_i,C_j} = 1$ if nodes i and j are in the same community ($C_i = C_j$); otherwise, $\delta_{C_i,C_j} = 0$. In addition to the modularity-based method, numerous methods have been proposed to solve community detection problems, including the clique percolation-based methods [232,233] and the link-community-clustering algorithm [234] for the overlapping community structure and random-walk-based methods [235,236] for community detection on weighted and directed networks.

Many results indicate that the disease genes' directed neighbors in a biological network might also be suspected to be the genes associated with the same disease, or, in other words, proteins involved in the same disease are more likely to interact with each other (see the red circle in Fig. 10, which would be very different from the topological module) [1,2,23,201,237]. For example, Menche et al. [23] found that disease genes associated with 226 diseases (out of 299 diseases) show a statistically significant propensity to form modules based on both the largest connected component analysis and the disease diameter.[2] The detection of these modules not only helps people understand various biological processes from a systematic perspective and reveal the mechanism of disease occurrence, but also contributes to the search for new drug targets in medicine, thus promoting drug development [238]. Rather than unfolding the entire community structure of biological networks, we pay more attention to the approaches to identifying the specific modules in biological networks.[3] Although topological communities may often represent meaningful functional modules, they are not able to capture disease modules [201]. The possible reason would be that a disease module is defined in relation to a particular disease, leading to the unique module for each disease, and the disease genes do not constitute particularly dense sub-graphs. Additionally, a gene (protein or metabolite) can be involved in several disease modules, indicating the overlaps of different disease modules, which is also a crucial and difficult problem in topological community detection. In this case, the topological community detection algorithms cannot solve such problems very well.

[2] The disease diameter is defined as the average closest distance $\langle d_s \rangle$, where for each disease gene, d_s is the distance to the next-closest protein associated with the same disease [23]. If $\langle d_s \rangle$ of the disease is significantly smaller than its random expectations, the disease-related genes tend to cluster into a module.

[3] The module detection could be classified into four broad categories according to various data interactions: identification of "active modules" through the integration of networks and molecular profiles, identification of "conserved modules" across multiple species, identification of "differential modules" across different conditions, and identification of "composite modules" through the integration of different interaction types. The detailed approaches on these four categories can be found in Ref. [239].

There are roughly two computational scenarios for specific module detection in biological networks. The first type is to identify the module when some disease genes are approved by biological experiments, through which we can obtain the potential disease genes for the hypothesis that genes associated with the same disease are more likely to interact with each other. In this case, the purpose is to design an algorithm to find the gene sets that are most likely to form a module with the known disease genes. Ghiassian et al. [201] proposed the DIAMOnD algorithm to identify proteins that are significantly connected to known disease-associated proteins (the significance of the connection rather than of the density). The probability that a protein with degree k has exactly k_s links to the known disease-associated proteins follows the following hypergeometric distribution:

$$p(k, k_s) = \frac{\binom{s_0}{k_s}\binom{N - s_0}{k - k_s}}{\binom{N}{k}}, \tag{9}$$

where s_0 is the number of the known disease-associated proteins and N the number of total proteins in the network. The significance of a protein connecting to the known disease-associated proteins can be defined as the probability that the protein has more connections to the known disease-associated proteins than expected under the null hypothesis, and it can be calculated as the cumulative probability for observing more connections, as follows:

$$p\text{+}-value(k, k_s) = \sum_{k_i=k_s}^{k} p(k, k_i). \tag{10}$$

The protein with the lowest p value will be added to the set of known disease-associated genes. Iterative application of the algorithm generates a growing disease module by adding the top protein in each step. The DIAMOnD approach can be easily extended to weighted networks by introducing an additional weight in Eq. (9). Testing the performance on some well-controlled synthetic data, including pruned networks and partially rewired networks, the DIAMOnD approach showed very powerful predictive ability to yield robust predictions. Although DIAMOnD can identify a connected disease module efficiently, the coverage of known disease-associated proteins in the module may be very low, with many isolated known disease-associated proteins unexplored. In this case, Wang and Loscalzo [240] proposed the seed connector algorithm (SCA) to identify disease modules by adding as few additional proteins to the known disease-associated protein pool as possible. After adding these proteins, the largest connected component of the new protein pool can increase the coverage of the known disease-associated proteins maximally.

The other type of computational scenario for specific module detection in biological network is identification of the biological module using the scoring and searching methods [239] by integrating the omics data onto the network. On the projection of the omics profiles, each node (gene or protein) in the network can be annotated with a score (e.g., the gene expression levels and mutation frequency). Then, we can define the module score (Z_M) based on the node or edge score (Z_i)[4] using the following equation as an example [241]:

$$Z_M = \frac{1}{\sqrt{m}} \sum_{i \in M} Z_i, \tag{11}$$

where m is the size of the module. Then, the detailed detection method can be carried out as follows: (1) Randomly select a seed module and calculate its module score Z_m. (2) Randomly select a neighbor node of the module (not in the current module), and calculate the new module score Z_{m+1}. (3) Add the node into the module to form the new module with high score (e.g., $Z_{m+1} > Z_m$) using some heuristic solutions, such as the simulated annealing [244], greedy algorithm [242], and random search [25]. (4) Repeat steps (2) and (3) to obtain the raw candidate module until that no neighbor node can be added into the module. (5) Choose the significant module according to the normalization of the Z_m ($Z_m^N = \frac{Z_m - \langle Z_m(r) \rangle}{\sigma_{Z_m(r)}}$, where $\langle Z_m(r) \rangle$ and $\sigma_{Z_m(r)}$ are the average and the standard deviation of score of the raw modules, respectively.) Similar approaches also can be extended to edge score (e.g., differential gene co-expression)- based module detection [245]. The modules identified using these scoring and searching methods are significantly associated with the corresponding diseases, including Type 2 diabetes [246], autism spectrum disorders [243], and cancers [25].

3.6. Motifs

Network motifs are defined as the patterns of sub-networks that recur in the network at frequencies much higher than those found in randomized networks that are believed to be basic building blocks of these networks to perform important functional roles [247,251,252]. Fig. 11 illustrates all 13 possible three-node directed sub-networks [Fig. 11(A)] and 30 possible sub-networks of size two- to five-node undirected sub-networks (Fig. 11(C), also referred to as graphlets [253]).

[4] The node score is defined by integrating the omics data. If p_i is the significance of the expression change for gene i between the disease sample and the controls, Z_i can be transformed from p_i according to $Z_i = \Phi^{-1}(1 - p_i)$, where Φ^{-1} denotes the inverse normal cumulative distribution function [241,242]. Additionally, Z_i is also defined as the mutation frequency in cancer research [25,243].

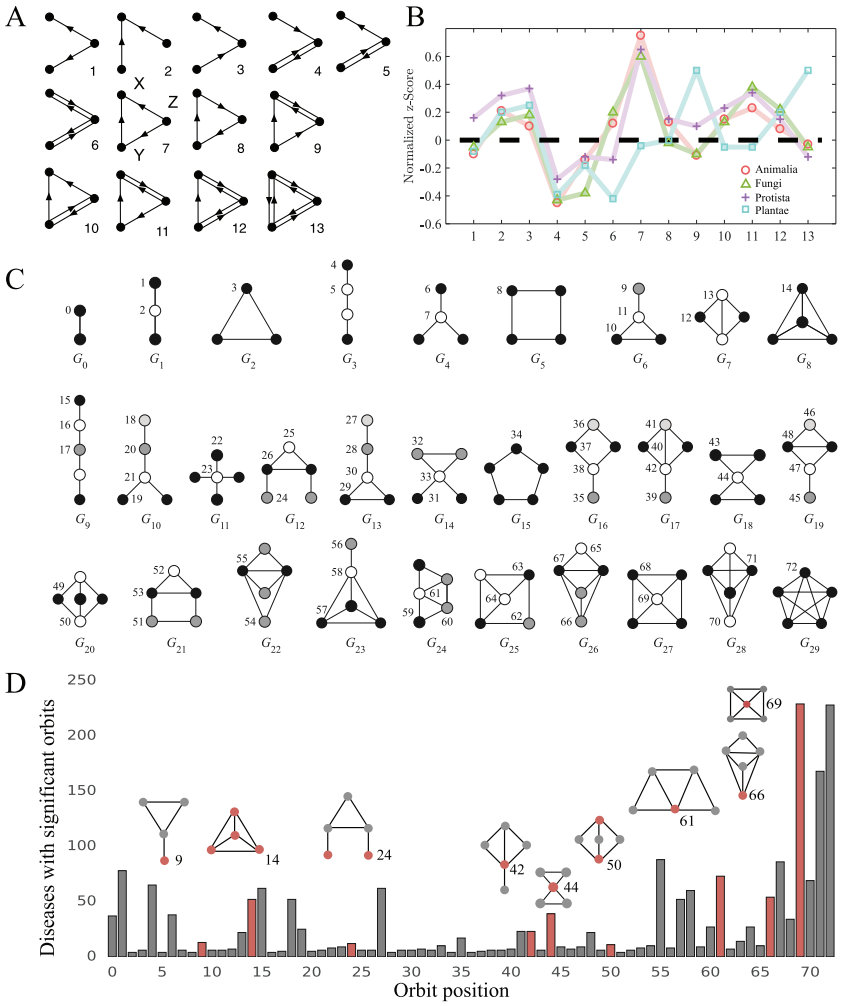

Fig. 11. Illustration of network motif. (A) 13 possible motifs in the three-node directed subgraphs; (B) significance of three-node motifs for the biological networks for different kingdoms of life, and the motifs are consistent with sub-figure (A); (C) 30 possible subnetworks of size 2- to 5-node undirected sub-networks and nodes of the same color belong to the same automorphism orbit within that graphlet; (D) number of the disease (of 519 diseases) -associated protein–protein interactions, in which the corresponding motif is significant at $\alpha = 0.01$ and the structure of the motif is represented as the orbit position corresponding to sub-figure (C).
Source: Sub-figure (A) from Ref. [247], sub-figure (B) from Ref. [248], sub-figure (C) from Ref. [249], and sub-figure (D) from Ref. [250].

Taking motif 7 in Fig. 11 as an example, it consists of three genes (X, Y, and Z), where gene X is influenced by the other two genes, while gene Y is only influenced by gene Z, which is also referred to as a feed-forward loop (FFL). This motif is the most prominent three-node network motif arising in the transcriptional networks in many organisms [as illustrated in Fig. 11(B)]. Specifically, a FFL motif can provide control over the temporal integration of environmental signals, buffering noise or sharpening the response to signals [254]. The statistical significance of a motif is described as the Z score [247],

$$Z_{Motif} = \frac{N_{real} - \langle N_{rand} \rangle}{\sigma_{N_{rand}}}, \tag{12}$$

where N_{real} and N_{rand} are the number of the appearance of the motif in the real network and the randomized networks, respectively, and the symbols $\langle N_{rand} \rangle$ and $\sigma_{N_{rand}}$ represent the mean and standard deviation values of N_{rand}, respectively.

The motif patterns, which can be characterized by the Z score for each sub-network, are very different for various systems [as illustrated in Fig. 11(D)]. For example, Ferreira et al. found that the topological distributions of motifs differ between gene regulatory and signaling networks, where the motifs in signaling networks tend to organize symmetrically while the motifs spread out asymmetrically along the hierarchical layers in gene regulatory networks [10]. The over-represented motifs in biological networks often form essential functional units of biological processes [255], and the network motif analysis has been studied on regulatory networks [256], metabolic networks [257], food webs [258], brain networks [259,260], and so on [261]. Many algorithms are proposed to count the total number of appearance of the

motifs [249,262–265], where the main difficulty in detecting network motifs in large networks lies in the fact that the number of possible sub-networks increases exponentially with network or motif size [263]. Network motifs have been widely applied in biological systems, including identification of the important nodes in biological networks [266], the search for regulatory algorithms of cells [252], druggability modulation of cellular targets [267], and identification of cancer patients [268].

3.7. Network entropy

Network entropy is a measure of complex networks, extended from the key concepts of information theory [269], that reveals the structural complexity and diversity of a network [270]. The most widely used form of network entropy is based on Shannon entropy, and the entropy of node i in the network can be defined as follows: [271–273]:

$$S_i = -\frac{1}{\log k_i} \sum_{j \in \Gamma(i)} p_{ij} \log p_{ij}, \tag{13}$$

where k_i is the degree of node v_i, $\Gamma(i)$ the set of nodes that link to v_i, and $p_{ij} = \frac{w_{ij}}{\sum_{j \in \Gamma(i)} w_{ij}}$ the proportion of the edge weight w_{ij} to the node v_i's total weights. In the protein–protein interaction network, w_{ij} can be the correlation value of the gene co-expression pair between genes v_i and v_j.

The network entropy also can be defined as the Shannon entropy of the distribution of node degrees [274], which can be calculated as follows:

$$S = -\sum_k P(k) \log P(k), \tag{14}$$

where $P(k)$ is the proportion of the nodes with degree k, indicating the degree distribution.

Network entropy can be a useful quantitative measure with which to characterize different disease statuses, like tumor versus normal tissue as well as various stages of progression [275]. Generally, cancer cells often have higher network entropy [272]. By integrating microarray gene expression data into a protein–protein interaction network, Cheng et al. used network entropy to characterize tumor progression and anticancer drug responses [275]. Banerji et al. [276] suggested that signaling entropy provided a potential measure in cancer by investigating microarray gene expression data in 3668 breast cancer samples and 1692 lung adenocarcinoma samples. Juarez-Flores and José [277] also detected the sudden changes in entropy values in four cancer types and identified the most relevant genes involved in carcinogenic processes that would be potential biomarkers or therapeutic targets. Intriguingly, Teschendorff et al. [278] found that increased network entropy of cancer cells would be driven by the scale-free (or near scale-free) property of the interaction network and the positive correlation between differential gene expression and node degree. Further, using over 7000 single-cell RNA-Seq profiles, they revealed that network entropy provides an accurate estimation of the differentiation potency and plasticity of single cells, which would be a way to identify normal and cancer stem-cell phenotypes [279].

In the field of brain networks, the measure of entropy has also proved to be very important in many scenarios. Using the resting-state fMRI data of the brains of 62 healthy subjects and 69 patients with chronic schizophrenia, Jia et al. [280] found that the sample entropy of the amygdala-cortical connectivity decreases with advancing age in healthy subjects, while the age-related loss of entropy could not be detected in patients. Viol et al. [274] compared the entropy of the degree distribution of the brain network on ordinary waking and conscious states by ingestion of the psychedelic Ayahuasca, to detect the increase in network entropy for the networks subsequent to Ayahuasca ingestion, empirically supporting the entropic brain hypothesis [281,282]. The studies on the resting-state fMRI also reveal the high influence of brain entropy on the human intelligence [283] and occupational functional plasticity [284]. Additionally, network entropy is also utilized to rank the brain regions that are associated with different tasks [285], reflecting the event-related dynamic interaction states of the brain [286], and to diagnose neurological disorders (e.g., epilepsy) using EEG data [287].

3.8. Fractal and self-similarity

Fractal properties of the networks are objects that contain self-similarity, in which they exhibit the same pattern on increasingly network-size scales. Fractal and self-similarity properties of complex networks have attracted much attention, since the seminal work of Song et al. [288], with a finding that a variety of real complex networks exist with a self-similarity property. The relation between scale-free networks and self-similar networks has also been discussed [288–290], and fractal networks seem to exhibit scale-free features, while scale-free networks are not always self-similar. The fractal nature of a network can be revealed by the well-known box/ball-covering method [291], by calculating the fractal dimension. Either the node-covering [288] or the edge-covering [292] box-counting method is needed to calculate the minimum number of boxes (N_B) with size l_B that can cover all the nodes or edges of a network. In the node-covering box-counting method, $l_B - 1$ represents the maximum distance of all of the possible pairs of nodes in each box, while the corresponding value is l_B in the edge-covering method. If the network is self-similar, the minimum number N_B of covering boxes scales with respect to box size l_B as

$$N_B \sim l_B^{-D}, \tag{15}$$

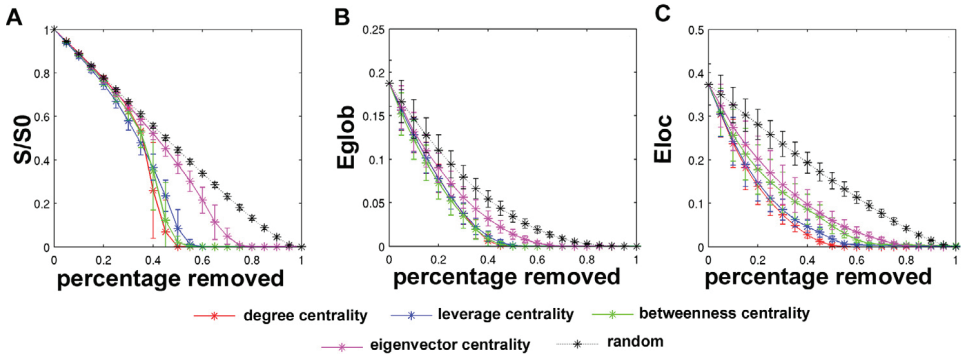

Fig. 12. Topological changes of brain networks upon various network failures. (A) Size of the largest component normalized with the original size; (B) global efficiency; (C) local efficiency [144]. Four centrality metrics were used in the intentional attack: degree centrality (red), leverage centrality (blue), betweenness centrality (green), and eigenvector centrality (pink).
Source: Figure from Ref. [319].

and the exponent D is just the fractal dimension of the network. Based on a similar framework. Wei et al. [293] proposed the modified box-covering algorithm, in which the box size length is obtained by accumulating the edge weight between two nodes connected directly, to calculate the fractal dimension of weighted networks.

The fractal and self-similar properties are empirically observed in many biological networks, such as protein–protein interaction networks [288,294], human cell differentiation networks [295], and brain functional networks [296]. Jin et al. [294] indicated that fractality and multiplicative growth would be the general properties of protein–protein interaction network structure and dynamics with the duplication-divergence model on the ancient and present-day PPI network. Brain functional networks are reported to be the fractal small world [190,297,298], and Gallos et al. [194,299] revealed that the weak ties play a very important role in the optimal global integration of self-similar modules to generate the small world property in the functional brain networks according to renormalization group theory [300]. The fractal properties in the brain network would be an important signature of self-organization [301]. Singh et al. [302] investigated the fractal properties of brain networks in different species, and the fractal dimensions in the high-level species (e.g., human and monkey) are much smaller than that in the low-level species (e.g., C. elegans), indicating a more ordered and systematically self-organized topography in higher-level species [303]. Further, simulation results of activity spreading on the brain network indicate that the fractal dimension would be a very relevant factor in controlling the spreading threshold value [304]. Additionally, the fractal and self-similar properties are widely applied in biological systems, including development of the algorithm to count the motifs in the biological networks [265], prediction of essential genes [305], and measuring the importance of proteins [306]. However, different from the previous view [288] in which metabolic networks exhibit self-similarity (fractality) using a box-counting method, Takemoto [307] found that metabolic networks are almost non-fractal for the increase in network density of the latest metabolic network data, highlighting the needs for a more suitable definition and careful examination of network fractal and self-similarity properties.

3.9. Resilience of network failures

Network resilience (or robustness) is a measure used to depict the ability to adjust the activity to retain the network's basic functionality upon failure (e.g., node deletion or edge breaking) [308,309]. The most simple and frequently used definition of network resilience is to test the network topological changes by removing a proportion of nodes or edges. When the removal proportion (p_r) exceeds a critical value ($p_r > p_{rc}$), the network would break down into smaller and disconnected components [116]. In general, there are two types of network failures: random failures (e.g., random removal of nodes) and intentional attack (e.g., removal of network hubs), and the scale-free (or scale-free-like) networks are very robust against random failures but extremely vulnerable to intentional attacks. For random networks, the influence of the network topology is very similar for both failure types [310,311]. Fig. 12 takes a brain network as an example to illustrate the network topological changes of different network failures, in which the network resilience against intentional attacks is much poorer than against random failures. The network resilience is critically important for reducing the functional losses on the inevitable network failures. To cope with network failures, the correlation between network features (topology and dynamics) and network resilience is deeply studied from ecological [312–314], biological to social [315], and economic systems [316], to find and design principles to enhance network resilience and prevent system collapse [317,318].

Focusing on biological systems, the highly resilient network would help the life system retain biological functionality against physical damage or environmental variation. Network failures in biological systems include edgetic mutations in protein–protein interactions [320], removal of enzymes in metabolic networks [321], a nonsense/missense gene mutation or gene dysfunction in the genetic network [320], and traumatic brain injury on the brain network. Taking the metabolic

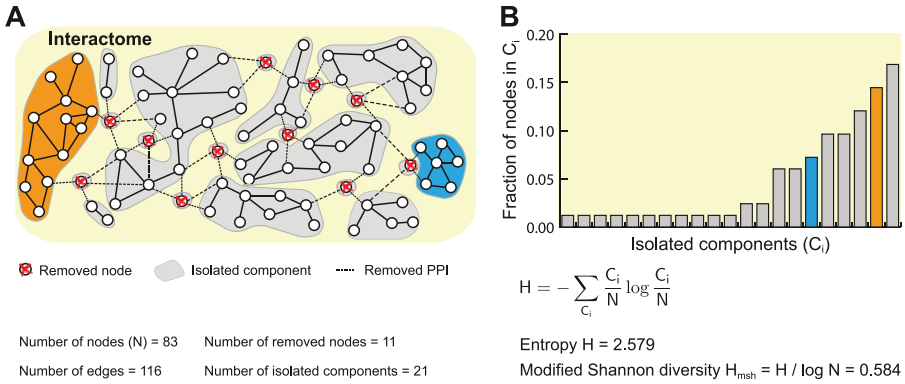

Fig. 13. Disintegration of network into isolated components with node removal. (A) Hypothetical interactome, the nodes of which represent proteins and edges indicating physical protein–protein interactions (PPIs). (B) Fragmentation of interactome characterized by isolated components and quantified by modified Shannon diversity H_{msh}.
Source: Figure from Ref. [325].

network as an example, the removal of enzymes in the metabolic network would influence the corresponding reactions, in which the damage would extend to additional reactions [321] to generate a cascading failure. Using the cascading failure model, Smart et al. [322] found that the metabolic networks are exceptionally robust against node failures compared with null models, and the organization of the branched metabolites [323] would be the origin of the enhanced robustness. Comparing both interaction and regulatory networks with null model networks, which are generated with a rewiring algorithm that preserves the degree of distribution, Maslov and Sneppen [324] found that increased overall network resilience in real networks was achieved by localizing node failure effects for the topological properties, in that the highly connected protein tends to be linked with low-connected proteins rather than the other highly connected proteins. The dynamical robustness to a perturbation of the gene expression would generate the asymmetry degree distributions in gene regulatory networks [132], where the out-degree distribution is broader than the in-degree distribution.

The good resilience of protein–protein interactions show phenotype robustness to environmental fluctuations, which would be highly related to life evolution. Zitnik et al. [325] collected interactomes data of 1840 species across the Tree of Life involving 8,762,166 protein–protein interactions, and defined the resilience of interactome G as follows (as shown in Fig. 13):

$$\text{Resilience}(G) = 1 - \int_0^1 H_{\text{msh}}(G_f)df, \tag{16}$$

where $H_{\text{msh}}(G_f) = -\frac{1}{\log N}\sum_{i=1}^{k} p_i \log p_i$ is the modified Shannon diversity of G_f, N the number of proteins in the interactome, $p_i = |C_i|/N$ the proportion of proteins in component C_i, and f the network failure rate. A higher value of *Resilience*(G) indicates the more resilient interactome. Using this definition, Zitnik et al. found that interactomes become more resilient during evolution, wherein a more resilient interactome is associated with the greater ability of the organism to survive in a more complex, variable, and competitive environment. Considering the protein–protein interaction networks for four species, i.e., yeast Saccharomyces cerevisiae, worm Caenorhabditis elegans, fly Drosophila melanogaster, and Homo sapiens, in order from simple to complex, Rodrigues et al. [326] revealed that the more complex species tended to exhibit more robust networks, especially when subjected to intentional attacks.

In brain research, it is critically important to quantify to what extent of failure the brain can withstand, which also can be described as resilience from the network perspective [327]. Many results indicate that human brain networks are remarkably resilient to different types of damage, even damage to some highly connected regions [319], compared to other types of complex networks, such as random or scale-free networks. Generally, the damage on the network hub regions would cause the larger disturbances in network organization, where the damages are significantly more likely to be located in hub regions in many brain-related diseases, such as Alzheimer's disease [328] and schizophrenia [329]. Studying the functional brain network from the resting-state fMRI data, acquired from 25 patients with schizophrenia, 25 first-degree relatives of patients and 29 healthy volunteers, Lo et al. [330] found that all of the networks were highly resilient to random failure, and that the global network efficiency remained high even after removing more than 50% of the nodes. In the case of an intentional attack, however, the resilience of the network of schizophrenia patients and their relatives was much greater than that of the healthy volunteers, suggesting that this is a system-level marker of familial risk for schizophrenia. The brain hubs are more biologically costly for energy delivery and utilization [121] and therefore they are more vulnerable to a diverse range of pathogenic processes in their susceptibility to disconnection and dysfunction in brain disorders [327,331]. However, the real damage may be very different from random attacks and intentional attacks. Crossley et al. modeled a pathological attack on a brain network, and global efficiency test results showed that the network was more resilient to pathological attack than to intentional attack on hubs [329]. Encouragingly, both empirical

and modeling studies have indicated that, after focal damage, the connectome carries the potential to recover at least to some extent, with normalization of graph metrics being related to improved behavioral and cognitive functioning. Abeysuriya et al. [332] simulated a cortical brain network and used inhibitory synaptic plasticity to dynamically achieve a spatially local balance between excitation and inhibition, to prove the potential reason for the robust dynamics of neuronal networks against perturbations.

4. Network-based methods

This section is intended to introduce the network-based modeling approaches that are widely used in network biology. We start from the network-structure-oriented methods, including network centrality, network propagation, structural-similarity-based link prediction and structural control. Going beyond the specific description of the computational methods, we focus on how to extract biological information using these approaches. Nowadays, the combination of complex networks and data mining [333] has attracted much attention for its predictive power. Therefore, we also review machine learning, especially the deep-learning-based approaches in biological networks, which would be a very attractive field in computational network biology in the future.

4.1. Network centrality

Biological networks exhibit a heterogeneous nature, with nodes playing very different roles in structure and function. In this subsection, we depict the node importance by network centrality, which is a measure that assigns a value to each node in the network, so we can rank the node importance using the value. It is very meaningful to identify the important nodes in the network biology for their crucial roles in biological function; for example, the hub proteins in the cell are very important for its survival, which is referred to as the centrality–lethality rule [13]. Since the meanings of importance are very different for various network dynamics, many network-centrality approaches have been proposed. We briefly review several centralities, including degree-, path-, and eigenvector-based centralities, and introduce how to use them in biological networks to identify the important nodes associated with certain biological functions. Detailed information on vital node-identification algorithms can be found in a more specific review [17].

4.1.1. Degree-based centrality

(i) Degree centrality:

In a network, a node with large degree would be very influential because it can impact numerous nodes with the immediate connections. Degree centrality is the simplest but the most widely used measure to identify a node's importance. It can be simply calculated as $DC(i) = k_i$, and assumes that a node with high-degree centrality would have larger influence. Although it exhibits simplicity and low computational complexity, degree centrality performs very well in some aspects. As illustrated in Section 3.1, the biological networks always exhibit power-law (or power-law-like) degree distribution, where several nodes have very high degree centrality (also referred to as hubs). Empirical and modeling results reveal the importance of hubs in the biological systems; for example, the hubs in the protein–protein network evolve more slowly [8,118,119] and are more likely to be essential for cell survival [13,117,334]. Moreover, the removal of hubs can significantly affect network structure (see Section 3.9).

(ii) Coreness centrality:

Degree centrality only considers the number of a node's immediate neighbors while ignoring its location in the network. However, the location of a node is a significant effective index with which to depict the node's importance. For example, the influence of a node located in the central part of the network would be higher than that in the periphery. Kitsak et al. [335] used the k-core decomposition [336] approach to obtain the coreness of each node, where a node with a larger coreness means that the node is located in a more central place and it is much more influential in network propagation than the high-degree nodes with smaller coreness. The k-core decomposition approach decomposes the network iteratively according to the nodes' remaining degree, and the method can be described as follows.

As illustrated in Fig. 14(A), at the initial step, all the nodes with degree $k = 1$ will be removed at first. The removal will lead to the reduction of the degree of nodes that are connected with the removed nodes in the original network (as shown by the yellow node in Fig. 14(A)). Continue to remove the nodes with remaining degree $k \leq 1$ iteratively until there is no node left with degree $k \leq 1$ in the system. All of the removed nodes in this step form a 1-shell with coreness value $k_s = 1$. At the second step, all of the nodes with degree $k = 2$ will be removed at first. Continue to remove the nodes with remaining degree $k \leq 2$ iteratively until there is no node left with degree $k \leq 2$ in the system. All of the removed nodes in this step form a 2-shell with coreness value $k_s = 2$. The decomposition process will continue until all of the nodes are removed, and the latest removed shell is the most central part of the network. In addition to the original decomposition, many modified methods [337,338] have been proposed to obtain more accurate coreness values, while the original coreness is highly coarse grained with many indistinguishable nodes sharing the same coreness.

Because of its nice performance and low computational complexity, coreness has been widely applied to identify the influential nodes in many real systems [340]. Lahav et al. [223] applied k-shell decomposition to reveal the hierarchical cortical organization of the human brain, and found that the nodes with the largest coreness value (the nucleus) serve as a

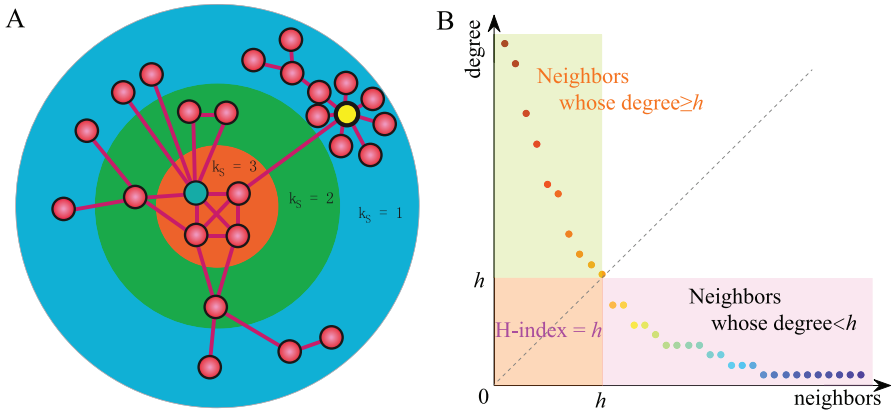

Fig. 14. Illustration of calculation of two types of degree-based centralities: (A) k-core decomposition; (B) H-index of network nodes. *Source:* (A) from Ref. [335] and (B) from [339].

global interconnected collective and demonstrate high correlation with consciousness-related regions. Many empirical and modeling results indicate that the nodes in the inner layer in k-core decomposition are highly related to some essential genes, disease progression, etc., which would even be helpful in drug development. Narang et al. [341] found that the most influential genes in a gene regulatory network are found within the innermost layer or core through k-shell decomposition. Ashraf et al. [342] integrated the coreness, functional connectivity, and some other centralities on the cancer interactome, which is comprised of 129,276 interactions arising from 8177 proteins, to extract the proteins involved in the pathways leading to cancer, and found 10 effective candidate proteins for drug targets. Ahmed et al. [162] applied weighted k-shell decomposition on the Arabidopsis interactome to find that coreness surpasses other centrality measures for effector target discovery. They also integrated weighted k-shell decomposition and some other centrality measures to analyze the cell surface interactome, and predicted the 35 most influential proteins that physically interact with pathogen effectors and modulate plant immunity. In weighted k-shell decomposition, the weighted degree of node i is defined as

$$k_i^w = \alpha k_i + (1-\alpha)\sum_{j\in\Gamma i} w_{i,j}, \tag{17}$$

where w_{ij} is the weight of edge $e_{i,j}$ and $\alpha \in [0\ 1]$ is a tunable parameter.

(iii) H-index centrality:

H index is a widely accepted measure to appraise the academic impacts of researchers or journals based on their publications and citations. Very recently, the H index was applied to quantify the network node influence (sometimes named the Lobby index [343,344]), where the H index of node v_i is defined as the largest h that v_i has at least h neighbors whose degrees are no less than h (as illustrated in Fig. 14(B)). The high H index indicates that the node has many high-quality neighbors.

Lü et al. [339] constructed an operator, H, to calculate the H index of a list of reals $(x_1, x_2, \ldots, x_n)$, which can be denoted as $h = H(x_1, x_2, \ldots, x_n)$; h is the maximum integer that satisfies the condition that there are at least h elements in $(x_1, x_2, \ldots, x_n)$ with the value $x_i \geq h$. Denoting $(j_1, j_2, \ldots, j_{k_i})$ as the sequence of the node v_i's neighbors in the network, and the m-order $(m > 0)$ H index of node i can be iteratively defined as

$$h_i^{(m)} = H(h_{j_1}^{(m-1)}, h_{j_2}^{(m-1)}, \ldots, h_{j_{k_i}}^{(m-1)}). \tag{18}$$

The zero-order H index of v_i is defined as its degree, $h_i^{(0)} = k_i$, and the first-order H index is just the original H index, $h_i = h_i^{(1)}$. Interestingly, Lü et al. [339] mathematically proved that the H index of node v_i ($h_i^{(\infty)}$) would converge to its coreness (k_{si}) within finite iterative steps. Thus, the node's degree, H-index, and coreness can be unified by operator H acting on its neighbors' degree sequence, and the H index in other orders also can be used to measure the node's importance.

4.1.2. Path-based centrality

Betweenness centrality, which is the most important path-based centrality, measures a node's influence on information flow in the network according to its betweenness value. As illustrated in Section 3.2.3, the betweenness value is defined as the ratio of the shortest paths passing through the target node among all of the shortest paths in the network. It can be very effective to interpret the node importance, especially the bridge nodes that connect different communities (Eqs. (4)–(6) in Section 3.2.3). However, for calculating all of the shortest paths crossing the entire network, the time

complexity of the betweenness centrality is very high, and it is not easy to obtain the betweenness centrality for the large real networks. Here, we introduce another path-based centrality, named closeness centrality.

The closeness centrality of node v_i is characterized by the distances between node v_i and all of the other nodes. Generally, the node is more central when it is closer to the others. Furthermore, closeness centrality can be defined as the mean value of the inverse of the distance from v_i to other nodes:

$$CC(i) = \frac{1}{N-1} \sum_{i,j \in V, j \neq i} \frac{1}{d_{ij}}, \tag{19}$$

where d_{ij} is the distance between node v_i and v_j in the network, and if nodes v_i and v_j do not connect with each other, $\frac{1}{d_{ij}} = 0$. The node with larger closeness value would be more central in the network, and has the shorter average propagation length of information to the others. Therefore, the node's closeness centrality reveals how efficiently it exchanges information with other nodes, and the network efficiency, illustrated as Eq. (3) in Section 3.2.2, is just the average closeness value over the entire network.

For directed network, the closeness centrality can be divided into output closeness centrality when d_{ij} is the distance from v_i to v_j and input closeness centrality when d_{ij} is the distance from v_j to v_i [345]. Ma and Zeng identified the top 10 output and input closeness central metabolites in the metabolic network of E. coli (with five overlapping metabolites), in which eight metabolites are in the central metabolism, namely the glycolysis and citrate acid cycle pathway.

4.1.3. Eigenvector-based centrality

(i) Eigenvector centrality

In contrast to the degree-based centralities, which focus on the number of a node's neighbors, the eigenvector-based centralities also take into account the influence of a node's neighbors. In this consideration, the nodes connecting to nodes that are themselves influential within the network would be more important, and the entire pattern of the network is considered. The eigenvector centrality of node v_i can be obtained as the sum of its neighbors' centralities [346], which can be defined as

$$EC(i) = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} EC(j), \tag{20}$$

and the matrix form is

$$\overrightarrow{EC} = \frac{1}{\lambda} A \overrightarrow{EC}, \tag{21}$$

where λ is the largest eigenvalue of the adjacent matrix A. Mathematically, the eigenvector centrality $EC(i)$ is just the ith element in the normalized eigenvector belonging to the largest eigenvalue of A.

Eigenvector centrality can be obtained by calculating the eigenvector of matrix A directly or via the power iteration method [347]. Negre et al. [348] applied the eigenvector centrality to identify the critical amino acid residues on the protein residue network, where each node represents the α-carbon of an amino acid in the protein, and each edge represents the mutual information between the two residues. The most important amino acid residues identified by the eigenvector centrality that are involved in the allosteric mechanism triggered upon effector binding are highly consistent with the solution of nuclear-magnetic-resonance (NMR) relaxation experiment results. In neuroscience studies, eigenvector centrality has been employed as a mapping tool for the brain network, which can be subjected to statistical tests to detect differences in centrality in various states [347]; for example, gender and age differences [349], the network alterations in Alzheimer's disease [350], and Type 1 diabetes mellitus [351].

(ii) PageRank and LeaderRank

The PageRank algorithm is proposed to rank websites in the Google search engine [352]. As a variant of eigenvector centrality, PageRank supposes that the importance of a web page is determined by both the quantity and quality of the pages linked to it. In the PageRank algorithm, each node in the network should be assigned a PageRank value (PR), and the node with the larger PR value would be more important. Generally, PageRank is executed by the iterative method. Initially, the PR value for each node is set to be 1 [$PR_i(0) = 1$]. At each time step, every node will evenly distribute the PR value to its neighbors along its outgoing links. The PR value of node v_i at step t is

$$PR_i(t) = \sum_{j=1}^{N} a_{ji} \frac{PR_j(t-1)}{k_j^{out}}, \tag{22}$$

where $a_{ji} = 1$ if node v_j links to v_i, and $a_{ji} = 0$ otherwise, and k_j^{out} is the out-degree of node v_j. The iteration will continue until the PR values of all of the nodes reach the steady states, and we can rank the nodes with the converged PR values in descending order. It should be noted that PageRank is actually a random-walk process on a directed network, and if there exist some dangling nodes ($k^{out} = 0$, without outgoing links), the walker will be trapped at these nodes and Eq. (22) will not converge. To guarantee convergence, PageRank introduces a random jumping factor α by assuming that the web

surfer will browse the web pages along the links with probability α, and leave the current page and open a random page with probability $1-\alpha$. The iteration steps can be modified as

$$PR_i(t) = \alpha \sum_{j=1}^{N} a_{ji} \frac{PR_j(t-1)}{k_j^{out}} + (1-\alpha)\frac{1}{N}, \tag{23}$$

where $\alpha \in [0, 1]$ is a tunable parameter that illustrates the random jumping probability. The parameter is usually set to approximately 0.85; however, that is just the experienced value, and how to determine the parameter α to achieve the best ranking on various scenarios is a significant challenge.

To solve this problem, Lü et al. [353] proposed a parameter-free algorithm, the LeaderRank algorithm, by adding a ground node that connects to all other nodes through bidirectional links. The network is strongly connected where the largest distance between two nodes is 2. The initial LR values are arranged by $LR_i(0) = 1$ for all nodes v_i (except for the ground node) and $LR_g(0) = 0$ for the ground node. Analogously, at each time step, every node will evenly distribute the LR value to its neighbors along its outgoing links. The LR value of node v_i at step t is

$$LR_i(t) = \sum_{j=1}^{N+1} a_{ji} \frac{LR_j(t-1)}{k_j^{out}}. \tag{24}$$

At the steady state, the value of the ground node will be evenly distributed to all of the other nodes, generating the final LR value of node v_i:

$$LR_i = LR_i(t_c) + \frac{LR_g(t_c)}{N}, \tag{25}$$

where $LR_i(t_c)$ and $LR_g(t_c)$ are the LR values of node v_i and the ground node at steady state, respectively.

Both PageRank and LeaderRank can be applied to real networks easily owing to their rapid convergence and good ranking performance, and the node-ranking methods are widely used to identify the essential proteins, disease-related genes, brain regions, neurons, etc. in biological systems. Winter et al. [354,355] employed the important genes identified by PageRank as the significant features with which to train the classification model to improve cancer outcome prediction. The modified PageRank algorithm on the weighted protein–protein network was applied to predict the essential proteins [356] as well as the tumor biomarkers [357], where the modified PageRank algorithm used the following formula:

$$PR_i(t) = \alpha \sum_{j \in \Gamma_i^{in}} w_{ji} \frac{PR_j(t-1)}{\sum_{q \in \Gamma_j^{out}} w_{jq}} + (1-\alpha) PR_i(0), \tag{26}$$

where Γ_i^{in} is the set of nodes pointed to v_i, Γ_j^{out} the set of nodes that node v_j pointed to, w_{ji} the weight of link ji, and $PR_i(0) = \frac{\sum_{j \in \Gamma_i^{in}} w_{ji}}{\sum_{i=1}^{N} \sum_{j \in \Gamma_i^{in}} w_{ji}}$.

In addition, the PageRank algorithm is also used to identify the influential protein in the bipartite graph. Jiang et al. [358] proposed the AptRank method, which utilizes an adaptive diffusion mechanism to adjust the fixed jumping probability in PageRank, to predict the protein function on a bi-relational graph comprised of both protein–protein association and function–function hierarchical networks.

4.1.4. Combined methods

Prediction of the essential nodes in a biological network is a challenging task, where a single centrality measure would be not enough to predict the essential nodes well in biological networks [359]. de Rio et al. [360] analyzed the 16 different centralities on various metabolic networks, where no single centrality measure can identify essential genes in a statistically significant way. There are two aspects to consider: (i) each centrality may just be advantageous in describing a unique structural feature [360,361], and (ii) the performance of a centrality on various networks would be even very different; for example, PageRank is sensitive to perturbations in random networks, while it is stable in scale-free networks [362].

Therefore, integrating various centrality measures would be a more reliable way to predict the essential nodes in the biological systems [359]. Further research by de Rio et al. [360] revealed that the combination of at least two centrality measures (e.g., closeness and clustering coefficient) achieved a reliable prediction of most essential genes, while no improvement was achieved when three or four centrality measures were combined. Ran et al. [363] found that blood-pressure variation is orchestrated by an integrated network centered on NOS3, which is identified by comprehensive consideration of the betweenness and closeness centrality on the protein–protein interaction network related to essential hypertension. Wang et al. [364] proposed a structurally dominant proteins index, which is the combination of eight centralities using principal component analysis (PCA), to identify the more important nodes in protein–protein interactions. Mistry et al. [365] proposed the DiffSLC centrality measure, which is a weighted combination of eigenvector centrality and co-expression-biased degree centrality, to identify more essential proteins. Apart from the combination of various centralities, integrating other topological and biological information into the centrality measure is also employed to improve the accuracy of influential node identification. Li et al. [366] proposed a method of combining

neighborhood closeness centrality and orthology information based on the effectiveness of triangular structure in a PPI network and the orthologous information in detecting essential proteins [367]. Co-expression [368,369] and protein complex [370,371] also comprise very important information related to the essential proteins; thereby, we can improve the efficiency of identification by integrating these sources of information with network topological centrality.

4.2. Network propagation

4.2.1. Techniques of network propagation

We assume that there is some information located on several nodes prior, and these nodes can transmit the information to their neighbors through links, and then extend to their neighbors' neighbors, and so on. This propagation would execute in an iterative way for a fixed number of steps or until convergence. This process is called *network propagation*, and is also named *network spreading* or *network diffusion*, which is a fundamental network dynamics in various systems. Based on nonlinear dynamic models, Hens et al. [372] predicted that the propagation rules condense around three highly distinctive dynamic regimes, characterized by the interplay between network paths, degree distribution, and interaction dynamics. The most researched field of network propagation is epidemic spreading [373] or news diffusion [136] on social networks. Imagine that several people acquired infectious diseases, and the diseases would propagate among the people through the contact between the infected and the susceptible people. The studies of the network propagation on these fields focus on the estimation of the popularity of the disease, the influence of the network structure on the spreading process, how to inhibit the disease spreading, and so on. The network resilience (Section 3.9) can also be considered as the propagation of network failures, where the network exhibits high resilience when the failure propagation is very limited.

In biological systems, a single mutation in a gene (e.g., driver mutation), which can influence genes in its neighborhood and amplify the mutation signal through network propagation, is sometimes enough to perturb an entire pathway and drive disease for an individual patient [18,374]. Many experimental results indicate that the disease genes are always biologically relevant, including the physical interaction, belonging to the same protein complexes, and involving in the same metabolic pathway. As for the biological network, the disease-associated genes are more likely to have biological interactions with each other than with randomly chosen genes. This may encourage us to predict that the neighbors of disease-associated genes in the network are also potential disease-associated genes [375], and the performance is better than that with randomly chosen genes. Network propagation would be more powerful in such prediction tasks by simultaneously considering all of the possible paths between genes. Specifically, network propagation is schematically illustrated in Fig. 15. The red nodes in the initial step ($t = 0$) in Fig. 15(A) and (B) have prior information (e.g., disease-associated gene), which will diffuse to the other nodes through the edges of the network. At step t, the amount of information at a node, which is described as the node color in Fig. 15, depends on the information at its neighbors at the previous step. The node receiving the most information (labeled D in Fig. 15(A) and (B)) in the final step (steady-state $t = \infty$) would be predicted to be a potential disease-associated gene.

We denote a node vector $F = [f_1, f_2, \ldots, f_i, \ldots, f_N]^T$, the elements of which represent the amount of information located in the corresponding nodes. The network propagation can be mathematically described as

$$F = \kappa F_0, \tag{27}$$

where κ is a $N \times N$ matrix and F_0 is the initial node state. The value of F_0 would represent our prior knowledge or experimental measurements of the corresponding genes [18]. For example, in disease genetics, we can set that $F_0(i) = 1$ if v_i is the known disease-associated gene; in cancer research, we can set F_0 as the somatic mutation for each patient, where $F_0(i) = 1$ indicates the observation of mutation on gene v_i for this patient [376], or the frequency of the somatic mutations in the cancer cohort, where $F_0(i)$ indicates the number of patients with the mutated gene v_i [69]. All of the network-propagation techniques aim to detect the expression of the global transmission matrix κ, the element $\kappa(i, j)$ of which represents the propagation-based similarity between nodes v_i and v_j in the network. Many mathematically techniques, including diffusion processes, random walk, and diffusion kernel, have been discovered to interpret network propagation.

Diffusion process:

Network propagation assumes that the node pairs with shorter distance are more likely to interplay with each other. For example, the influence of the directed neighbors should be more important than that of the second-order neighbors. Supposing that the information diffusing from node v_i to node v_j and the length of a path between node v_i to node v_j is l, then the amount of information that node v_j receives from this path is set to be α^l, where $\alpha \in (0, 1)$ is a tunable parameter, indicating that the information decays with propagation path length. The total amount of information that v_j received from v_i through the paths with length l is $\alpha^l n_{pl}$, where n_{pl} is the number of paths between node v_i to v_j with length l, which can be represented as the power of the adjacent matrix A^l. Therefore, the information propagation between two nodes with path length l is $\alpha^l A^l$, and the total propagation process can be described as the global transmit matrix:

$$\kappa = \sum_{l=1}^{n} \alpha^l A^l. \tag{28}$$

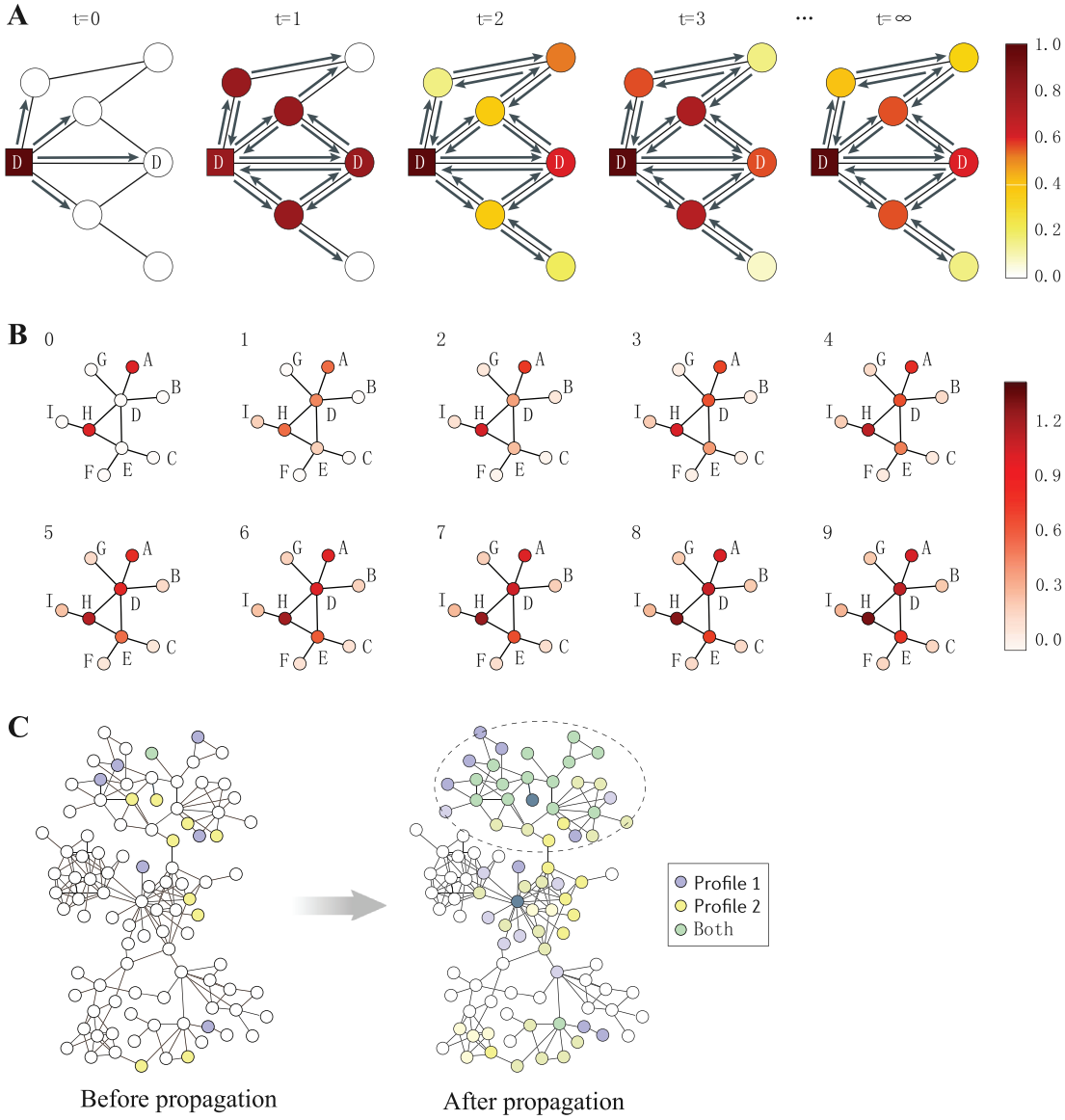

Fig. 15. Schematic of network propagation. (A) Step-by-step demonstration of network propagation. The propagation process is depicted at different time points until convergence ($t = \infty$). Arrows depict the direction of the flow or walk. Nodes are color-coded according to the amount of flow that they receive. D indicates nodes that are known (square node) or that are predicted (circular node) to be associated with a disease phenotype. (B) Example network with initial high scores for two of nine nodes (step 0, nodes A and H; score shown by color bar). These scores are allowed to propagate over stepwise iterations 0–9; note that convergence is reached by approximately step 5 and thus the colors do not change markedly in subsequent steps. (C) Illustration of biological network with gene scores before and after propagation, performed independently for two datasets (profiles 1 and 2). Propagation results in greater concordance between the datasets, as is evident from the greater number of green nodes (dashed oval).
Source: Figure from Ref. [18].

When α is less than the reciprocal of the largest eigenvalue of A, we can obtain the mathematical solution of the limitation of κ with

$$\kappa = \lim_{n\to\infty} \sum_{l=1}^{n} \alpha^l A^l = (I - \alpha A)^{-1} - I, \tag{29}$$

where I is the identity matrix. It should be noted that Eq. (28) is also known as the famous Katz centrality, where the Katz centrality of node v_i can be described as $KC(i) = \sum_{j\in V, j\neq i} \kappa_{ij}$ [17,377].

Random walk:

In the context of networks, a random walk describes a process in which a "walker" moves from one node to another with a probability that is proportional to the weight of their edge. The random walk can be considered as a Markov process, where the result in the current step depends on the last step's state:

$$F_t = WF_{t-1}, \tag{30}$$

where $W = AD^{-1}$ and D is the diagonal degree matrix of the adjacent matrix A, the diagonal entries of which hold the node degree ($D(i,i) = k_i$) and all other entries are 0. If we run this random-walk process for t steps, iterating Eq. (30) generates $F_t = W^t F_0$. When t is small, the information distribution is similar to that of the initial state F_0, but when t is large the information diffuses away from the initial state and reflects the network topology. Actually, a random-walk process is the origin of PageRank, where Eqs. (22) and (30) are, in fact, two formulas for the same process.

Random walk with restart (RWR) is very widely used random-walk-based propagation process. It states that a random walk starts from a node v_i, and at each time step t moves to one randomly selected neighbor of the current node v_j with probability α, while the walk restarts from v_i with probability $1-\alpha$. This propagation process can be described as follows:

$$F_t = (1-\alpha)F_0 + \alpha WF_{t-1}, \tag{31}$$

where the parameter $\alpha \in [0, 1]$ describes the trade-off between initial information and network propagation. Running this process until equilibrium is reached, the amount of information on each node at steady state depends on the initial state (F_0), the network topology, and the α value. The steady state can be described as

$$F = (1-\alpha)(I - \alpha W)^{-1}F_0. \tag{32}$$

Although the above propagation system can be solved exactly using Eq. (32), the inverse of the large matrix would be very time-consuming to calculate. Therefore, an iterative propagation-based algorithm works faster for large networks and is guaranteed to converge to the system's solution. For example, the propagation function (Eq. (31)) is run iteratively with $t = [0, 1, 2, \ldots]$ until F_t converges ($|F_t - F_{t-1}| < \varepsilon$, where ε is close to 0, e.g., $\varepsilon = 10^{-6}$). Additionally, the definition of the matrix W is not unique, and using different ways to define W can generate different propagation processes. In addition to $W = AD^{-1}$ for the random-walk process, some other approaches, such as $W = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ [378,379], are also used to define the propagation process.

Diffusion kernel:

Lazy random walk [380,381] is another version of the random-walk-based propagation process. Assuming that the random walker is at node v_i at time step t, and the next step, it will walk to node v_i's neighbor with probability β and remain at node v_i with probability $1-k_i\beta$. The transmit matrix of this process is $I - \beta L$, where $L = D - A$ is the network's Laplacian matrix. In lazy random walk, the probability β decays with the walk steps, denoted $\beta = \alpha/t$. We can obtain the continuous time limit of lazy random walk as

$$\lim_{t\to\infty}(1 + \frac{\alpha}{t}L)^t = e^{-\alpha L}. \tag{33}$$

Here, the $\kappa = e^{-\alpha L}$ is called the diffusion kernel [380,382], which can be interpreted as a similarity matrix, in which each element κ_{ij} gives the amount of information propagated to node v_i, given that the initial information F_0 with only $F_0(j) = 1$.

4.2.2. Propagation in network biology

Many studies have shown that network propagation is very powerful in predicting potential disease genes. For a given biological network structure (A) and known disease genes (F_0), we can obtain the score of each gene at the steady state (F) through Eq. (27), where the genes with high score would be predicted as the potential disease genes (see Fig. 15(A) and (B)). The transmit matrix κ, interpreted as the Katz process [383], random walk [378], and diffusion kernel [382] are all very effective in potential disease gene prediction. It should be stressed that the propagation process would be very sensitive to noise in the network [18], and choosing the suitable network would be very important in the prediction. Huang et al. [384] evaluated 21 human genome-wide interaction networks for their ability to recover disease genes through network propagation under the RWR model. They found that all networks have some ability to recover disease genes, where the STRING (Search Tool for Recurring Instances of Neighboring Genes), ConsensusPathDB, and GIANT (Genome-scale Integrated Analysis of gene Networks in Tissues) networks have the best performance, and the composite networks outperform any single network.

Network-propagation technology can be extended to the networks that are integrated with some other biological information, such as gene expression and phenotype information, to obtain more accurate disease gene prediction. Li and Patra [385] proposed a RWR on heterogeneous networks (RWR-H), which was composed of a protein–protein network ($A_P[N1 \times N1]$), disease–disease similarity network ($A_D[N2 \times N2]$), and protein–disease association network ($A_{PD}[N1 \times N2]$).

Therefore, the adjacency matrix of the heterogeneous network can be represented as $A = \begin{bmatrix} A_P & A_{PD} \\ A_{PD}^T & A_D \end{bmatrix}$, and the propagation process on the heterogeneous network can be modified as follows:

$$F_t = (1-\alpha)F_0 + \alpha W^H F_{t-1}. \tag{34}$$

Let $W^H = \begin{bmatrix} W_{PP} & W_{PD} \\ W_{DP} & W_{DD} \end{bmatrix}$ denote the transmission matrix on the heterogeneous networks, where W_{PP} and W_{DD} describe the transmission matrix within a network (protein–protein and disease–disease), and W_{PD} and W_{DP} describe the jumps between networks (protein–disease and disease–protein). The definition of the initial state F_0 in the heterogeneous networks is $F_0 = \begin{bmatrix} (1-\eta)u_0 \\ \eta v_0 \end{bmatrix}$, where u_0 and v_0 represent the initial information on the protein–protein interactions and disease–disease similarity networks, respectively, and $\eta \in (0, 1)$ is a tunable parameter used to describe the trade-off between genotype and phenotype. Some modification of the RWR, such as the DRWR (Discriminative Random Walk with Restarts, a two-stage RWR) [386] and RWHN (Randomly Walking in the Heterogeneous Network, considering the weighted protein–protein interaction and disease–disease association networks) [387], have also been developed to predict the genes most related to a given gene set on a heterogeneous biological network. More specifically, Valdeolivas et al. [388] extended the RWR algorithm to multiplex-heterogeneous networks to predict the disease-associated genes, and it outperformed the random-walk-based methods on monoplex or heterogeneous networks by leveraging the different interaction sources.

Integrating the somatic mutation data on a large set of patients to identify frequently mutated subnetworks is very important in cancer research, because somatically mutated contribute to the growth and development of cancer. The network-propagation method, in which the "influence" of each mutation diffuses through the network, is very effective to compensate the sparseness of the mutation profiles. The very famous methods, named HotNet and HotNet2 (the improved variant) [222,389,390], applied network propagation to integrate the somatic mutation data and protein–protein interaction network to identify significantly mutated pathways in cancer. HotNet and HotNet2 used diffusion kernel and RWR, respectively, to model the propagation process, and the initial value (F_0) can be set to be the observed mutation frequency of each gene. Hofree et al. [376] employed the network-propagation method to identify the cancer subtypes [391] in the TCGA datasets. They obtained the smoothed mutation profiles of each patient by the propagation of the observed mutated genes in the RWR way (see Fig. 15(C)), and then applied the non-negative matrix factorization method to cluster cancer patients. Cheng et al. [69] proposed the gene gravity model based on network propagation to quantify the cancer genome evolution, and one further application of the gene gravity model was the identification of the cancer driver genes [392]. In the gene gravity model, the smoothed mutation profiles of each gene are considered the "gene mass" and the path length of gene pairs in the network is considered the "biological distance". Additionally, network-propagation methods can also be applied to facilitate the identification of functional changes [393], prediction of drug targets [25,394], and drug synergy [395] in cancers.

Additionally, a network-propagation method is also an efficient way to identify the network module (see 3.5), where the connected nodes with large information flow would be considered as the network module. For example, HotNet2 [390] applied a directed heat diffusion model to find the significantly mutated modules in cancers. Through network propagation (RWR process), HotNet2 can generate a non-symmetric weighted diffusion matrix, which represents the transmission similarity between the gene pairs. The gene pairs with similarity larger than the threshold value are selected, and the strongly connected components can be considered as the potential modules. Using HotNet2, Leiserson et al. [390] identified 16 significantly mutated modules in which some well-known cancer signaling pathways were enriched. Interestingly, the identified modules also contained dozens of genes with rare somatic mutations, and their role in cancer is supported by additional evidence. Very recently, Cáceres and Paccanaro [396] proposed the Cardigan approach, which utilized network propagation of the query weight set[5] to retrieve the disease module.

4.3. Link prediction

Link prediction aims to infer missing links or predict new interactions between non-connected nodes in networks [397,398]. It can help complement missing data in real networks, as well as facilitate better understanding of the evolution processes of networks. Numerous algorithms for link prediction have been set up according to network topology. Some of these algorithms have been successfully applied in biological networks to solve a wide range of biological and medical problems, such as the prediction of gene regulatory relations [399], protein–protein interactions [400], lncRNA–protein interactions [401], drug–target interactions [27], drug–drug interactions [402], gene–disease associations [403], miRNA–disease associations [404], drug sensitivity [405], and brain network connections [406]. Most of the predictions are performed in unipartite networks, which consist of one set of nodes, while some predictions are conducted in bipartite networks, i.e., networks consisting of two non-overlapping sets of nodes with only links joining the nodes in different sets.

[5] For a given (query) disease, Cardigan calculates its phenotypic similarity to every other disease first, and the query weight set represents the weight of each gene, which is the similarity between the query disease and gene-associated disease.

Table 4
Illustration of commonly used local indices for node similarity.

Name of similarity index	Equation (S_{ij} =)	Illustration
Common neighbor index (CN)	$\mid \Gamma(i) \cap \Gamma(j) \mid (= [A^2]_{ij})$	$\Gamma(i)$ is the set of neighbors of node v_i
Jaccard index (JC)	$\frac{\mid \Gamma(i) \cap \Gamma(j) \mid}{\mid \Gamma(i) \cup \Gamma(j) \mid}$	
Sørensen index (SEN)	$\frac{2 \mid \Gamma(i) \cap \Gamma(j) \mid}{k_i + k_j}$	k_i is degree of node v_i
Salton index (SAL)	$\frac{\mid \Gamma(i) \cap \Gamma(j) \mid}{\sqrt{k_i k_j}}$	
Leicht–Holme–Newman (LHN)	$\frac{\mid \Gamma(i) \cap \Gamma(j) \mid}{k_i k_j}$	
Hub depressed index (HDI)	$\frac{\mid \Gamma(i) \cap \Gamma(j) \mid}{\min(k_i k_j)}$	
Hub promoted index (HPI)	$\frac{\mid \Gamma(i) \cap \Gamma(j) \mid}{\max(k_i k_j)}$	
Common neighbors & distance (CND)	$\frac{\mid \Gamma(i) \cap \Gamma(j) \mid}{d_{ij}}$	d_{ij} is distance between nodes v_i and v_j
Adamic–Adar index (AA)	$\sum_{v_z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log k_z}$	
Resource allocation index (RA)	$\sum_{v_z \in \Gamma(i) \cap \Gamma(j)} \frac{1}{k_z}$	
Local community links (LCL)	$\sum_{v_z \in \Gamma(i) \cap \Gamma(j)} \frac{\gamma(v_z)}{2}$	$\gamma(v_z)$ is subset of neighbors of node v_z that are also common neighbors of v_i and v_j. $\mid \gamma(v_z) \mid$ is the number of edges which link with node v_z and other common neighbors of v_i and v_j.
Cannistraci–Alanis–Ravasi (CAR)	$\mid \Gamma(i) \cap \Gamma(j) \mid \sum_{v_z \in \Gamma(i) \cap \Gamma(j)} \frac{\gamma(v_z)}{2}$	
Cannistraci-Resource-Allocation (CRA)	$\sum_{v_z \in \Gamma(i) \cap \Gamma(j)} \frac{\gamma(v_z)}{k_z}$	

Here, we survey link prediction algorithms frequently applied in biological networks, including unipartite and bipartite networks.

The existing link-prediction algorithms can be roughly categorized as belonging to three classes [397,398]: similarity-based algorithms, probabilistic and statistical models, and machine-learning-based methods. The first class comprises the mainstreaming class of methods and has been the most widely used in researching biological networks. Thus, we focus on this class of algorithms in this subsection, while the machine-learning-based methods are illustrated in Section 4.5.

4.3.1. Similarity-based methods

Similarity-based methods assume that a pair of non-connected nodes tends to establish a link if they are similar or close to each other in the network. Different similarity indices S_{ij} were proposed to measure the similarity or proximity of two nodes v_i and v_j in the network. A larger value of the score suggests a higher probability of a link between the nodes. Similarity indices are further classified into three classes, local, global, and quasi-local indices, according to the topological information used in the indices.

(i) Local indices:

Local indices are designed from triadic closure principle and use different normalization or adjusting methods. Triadic closure suggests that, if two links exist among three nodes, the two non-connected nodes tend to build a link. The commonly used local indices for the similarity are illustrated in Table 4.

(ii) Global indices:

Global indices utilize the entire topological information to measure the possibility that two non-connected nodes establish a link. For simplicity, some indices below are represented by matrix form S, which is the similarity matrix, and each element S_{ij} represents the similarity between nodes v_i and v_j.

Cosine based on L^+ (Cos^+) [407]: Cosine based on L^+ index is defined as the cosine of the node vectors, as follows:

$$S_{ij} = cos(V_i, V_j) = \frac{V_i^T V_j}{|V_i \parallel V_j|} = \frac{L_{ij}^+}{\sqrt{L_{ii}^+ L_{jj}^+}}, \tag{35}$$

where $V_i = \Lambda^{\frac{1}{2}} U^T \overrightarrow{e_i}$, U is the orthonormal matrix made of the eigenvectors L^+ (pseudo inverse of Laplacian matrix) ordered in decreasing order of corresponding eigenvalue λ_i, Λ is $diag(\lambda_i)$, and $\overrightarrow{e_i}$ is an $N \times 1$ vector with the ith element equal to 1, while the other elements are 0. $|V_i|$ is the length of V_i, L_{ij}^+ is the entry of L^+, and $L_{ij}^+ = V_i^T V_j$.

Structural perturbation method (SPM) [408]: The structural perturbation index is derived from structural consistency, which is defined as the consistency of structural features before and after the removal of some randomly selected links. It is calculated as follows:

- Randomly select a fraction of links from the network (with edge set E) to constitute a perturbation set ΔE; then, the adjacency matrix of the remaining network (with link set $E_R = E - \Delta E$) is as follows:

$$A_R = \sum_{k=1}^{N} \lambda_k \overrightarrow{x} \overrightarrow{x}^T. \tag{36}$$

- Compute $\Delta\lambda_k$ as follows:

$$\Delta\lambda_k \approx \frac{\overrightarrow{x}^T \Delta A \overrightarrow{x}}{\overrightarrow{x}^T \overrightarrow{x}}, \tag{37}$$

  where ΔA is the adjacency matrix corresponding to ΔE.
- Compute structural perturbation matrix:

$$\widetilde{A} = \sum_{k=1}^{N} (\lambda_k + \Delta\lambda_k) \overrightarrow{x} \overrightarrow{x}^T. \tag{38}$$

- Repeat above steps, and then the similarity matrix S is the average of all of the structural perturbation matrices $\widetilde{A}$.

In terms of structural perturbation, a perturbation-based framework based on non-negative matrix factorization (NMF) is proposed for link prediction by considering random noises and irregular links [409].

In addition, the Katz index and RWR index, which can be obtained according to Eqs. (29) and (32) in Section 4.2, respectively, are also effective in link prediction.

(iii) Quasi-local indices:

Quasi-local indices use topological information more than local indices and less than global indices. Their computational complexity is below the complexity of global methods and can reach the prediction precision as well as global methods do.

Degree-normalized L3 index(L3) [400]:

$$S_{ij} = \sum_{u,v} \frac{a_{iu} a_{uv} a_{vj}}{\sqrt{k_u k_v}}, \tag{39}$$

where k_u is the degree of node v_u and a_{uv} the (u, v) element of the adjacent matrix. This index counts the number of length-3 paths existing between node v_i and v_j, and meanwhile normalizes it with degrees of intermediate nodes.

Local path index (LPI) [410]:

$$S = \sum_{k=2}^{l} a^{k-2} A^k, \tag{40}$$

where l is the length of the longest path under consideration and $\alpha \in (0, 1)$ is a tunable parameter. $(A^k)_{ij}$ equals to the number of paths of length k between v_i and v_j. This index shows that the shorter the path between nodes v_i and v_j, the more similar the two nodes are to each other. When l is taken as 2, this index is just the common neighbor index.

Node graphlet degree vector similarity (Node-GDV-similarity) [411]:

This index measures the similarity of two- to five-node graphlet degree vectors (GDVs) of two nodes. A graphlet is a small induced subgraph of the network (as illustrated in Fig. 11(C) in Section 3.6). Since a node v_i can be included in at most 73 different graphlets with two to five nodes (considering one predicted edge), the GDV of node v_i has 73 dimensions, in which each element is the number of corresponding types of graphlets that the node v_i touches. Supposing that i_k is the kth element of the GDV of node v_i, the distance between the kth elements of nodes v_i and v_j is defined as

$$D_k(i, j) = w_k \frac{|\log(i_k + 1) - \log(j_k + 1)|}{\log(\max(i_k, j_k) + 2)}, \tag{41}$$

and the node-GDV similarity between v_i and v_j is defined as

$$S_{i,j} = 1 - D(i,j) = 1 - \frac{\sum_{k=1}^{73} D_k(i,j)}{\sum_{k=1}^{73} w_k}, \tag{42}$$

where w_k is the weight of the kth graphlet type that accounts for the dependencies.

4.3.2. Link prediction for bipartite networks

Link-prediction algorithms in bipartite networks have been used in biological networks for the prediction of drug–target [27], miRNA–disease [412], and miRNA-lncRNA interactions [413]. Unlike unipartite networks, in which any pair of nodes can make links, bipartite networks only allow links between nodes of two disjoint sets. Thus, the topological methods based on the triadic closure model cannot directly work in the bipartite case. There are two main directions for link prediction in bipartite networks: the first is to extend the similarity indices of unipartite networks to the bipartite case (called extension methods), and the second is to project a bipartite network into two unipartite networks and use one or both of them for link prediction (called projection methods). Here, we focus on these two directions of link-prediction methods.

(i) Extension methods:

Instead of triadic closure in unipartite networks, this class of methods extends the definition of common neighbors to quadratic closures in bipartite networks. Supposing a bipartite network has two disjoint node sets U and V, for a pair of non-connected nodes ($i \in U, j \in V$) the common neighbor index can be defined as follows [412]:

$$S_{ij} = |\Gamma(i) \bigcap \hat{\Gamma}(j)|, \tag{43}$$

where $\Gamma(i)$ is the set of neighbors of node i and $\hat{\Gamma}(j)$ is the set of neighbors of j's neighbors. According to the similar extension of $\hat{\Gamma}(j)$, all of the local indices listed in Table 4 can be generalized to bipartite networks easily.

Similarly, global similarity indices can also be generalized to bipartite cases. Taking the Katz index (KI) as an example, the KI in the bipartite network can be obtained as follows:

$$S_{ij} = \sum_{l=1}^{\infty} a^l |ps_{ij}^l|, \tag{44}$$

where ps_{ij}^l is the set of all of the length-l paths from node i to j.

(ii) Projection methods:

For bipartite network that has two disjoint node sets U and V, V projection is used to construct a network containing only V nodes, where two V nodes are linked if they have common U neighbors. Since the projection loses much information in the original bipartite networks, many algorithms have been proposed to construct weighted projection networks to better reflect the topology of bipartite networks, some of which still use the similarity between two nodes as the weight of their link.

Once the weight matrix $W = (w_{ij})$ for the V-projection network is produced, the likelihood matrix for link prediction can be calculated as

$$S = WA, \tag{45}$$

where $A = (a_{lm})$ is a $|V| \times |U|$-dimensional adjacent matrix of the bipartite network. Therefore, for any V node v_i, its likelihood index with non-collected U node u_j is

$$S_{ij} = \sum_{l=1}^{|V|} w_{jl} a_{li}, \tag{46}$$

and the node pairs with highest similarity scores are predicted to be linked. Therefore, the key problem for the projection method lies in how to quantify the weight for the V-projection network.

The first class of weights is defined based on node similarity. For a pair of V nodes v_i and v_j, similarity-based weights in the projection network are defined as follows (taking the common neighbor as an example):

$$w_{ij} = |\Gamma(v_i) \bigcap \Gamma(v_j)|, \tag{47}$$

and the local similarities listed in Table 4 can all be defined as the similarity-based weights. In addition to the local similarities, there are some other similarity definitions that can be used to define the link weight in the V-projection network.

Euclidean [414]:

$$w_{ij} = \sqrt{\sum_{l=1}^{|U|} (v_i(U) - v_j(U))^2}, \tag{48}$$

where $v_i(U)$ is a vector denoting node v_i's neighbors in U, in which an element is 1 if the corresponding U node links with v_i; otherwise, it is equal to 0.

Pearson [414]:

$$w_{ij} = corr(v_i(U), v_j(U)) = \frac{cov(v_i(U), v_j(U))}{\sigma(v_i(U))\sigma(v_j(U))}. \tag{49}$$

The second class of weights is defined based on information propagation in bipartite networks, and the typical algorithms are as follows:

Probabilistic spreading (ProbS) [415]:

$$w_{ij} = \frac{1}{k(v_j)} \sum_{l=1}^{|U|} \frac{a_{il}a_{jl}}{k(u_l)}, \tag{50}$$

where (a_{lm}) is a $|V| \times |U|$-dimensional adjacent matrix of the bipartite network and $k(v_j)$ the degree of node v_j. The probabilistic spreading (ProbS) considers resource allocation in bipartite networks. It assumes that the relatedness between a pair of V nodes v_i and v_j depends on the resource flow from v_i and v_j to the U nodes and back.

Heat spreading (HeatS) [416]:

$$w_{ij} = \frac{1}{k(v_i)} \sum_{l=1}^{|U|} \frac{a_{il}a_{jl}}{k(u_l)}, \tag{51}$$

and HeatS represents a discrete analogy of a heat diffusion process. It is similar to ProbS, while the only difference is that HeatS is a row-normalized matrix while ProbS is a column-normalized one.

Hybrid method [416]:

$$w_{ij} = \frac{1}{k(v_i)^{1-\lambda}k(v_j)^{\lambda}} \sum_{l=1}^{|U|} \frac{a_{il}a_{jl}}{k(u_l)}, \tag{52}$$

where $\lambda \in [0, 1]$ is a tunable parameter and Hybrid a combination of ProbS and HeatS.

4.3.3. Evaluation metrics

Usually, area under the receiver operating characteristic curve (*AUC*), Precision, and Recall are used to measure the performance of link-prediction algorithms, defined as follows [397]:

AUC: In link prediction, a simplified method is used to compute the AUC value. A missing link and a non-existent link are randomly picked and their scores are compared. If among n independent comparisons, there are n' times that the missing link has a higher score and n'' times that they have the same score, then $AUC = \frac{n'+0.5n''}{n}$.

Precision: Precision is the ratio of real predicted missing links to all of the predicted links. If the top L links are considered predicted links, while L_r of which appear in the test set, then $Precision = \frac{L_r}{L}$.

Recall: Recall is the ratio of real predicted missing links to all of the missing links. If the top L links are considered predicted links, while L_r of which appear in the test set, and the total number of missing links in the test set is L_m, then $Precision = \frac{L_r}{L_m}$.

4.4. Network control

4.4.1. Structural controllability

In recent years, the control of complex systems has become an important research topic in statistical physics [19,417]. According to control theory, a dynamic system is controllable if suitable external inputs can drive the system from any initial state to any desired final state in a finite time interval [19,418]. The ability to control the complex systems would be the ultimate aim in our search for understanding of the systems. However, controlling complex systems is still an outstanding challenge. In this subsection, we focus only on the structural controllability of complex networks and identification of the minimum number of driver nodes [418].

Although the dynamics of biological systems are nonlinear in practical applications, linear systems have been successfully used as tools in modeling, analysis, and control technology systems in control theory. To study the controllability of biological networks, researchers have used linear dynamic models to represent the dynamics of biological networks. Kalman [420] proposed that the structural controllability of linear systems can provide a sufficient condition for the controllability of most nonlinear systems. For linear time-invariant networks with N nodes, the dynamic representation is as follows [418]:

$$\frac{dx(t)}{dt} = Ax(t) + Bu(t), \tag{53}$$

where $x(t) = (x_1(t), x_2(t), \ldots, x_N(t))^T$ is an N-dimensional vector describing the state of all the nodes (e.g., the amount of transcription factor concentration in a gene regulatory network). A is an $N \times N$ state transition matrix, which represents the

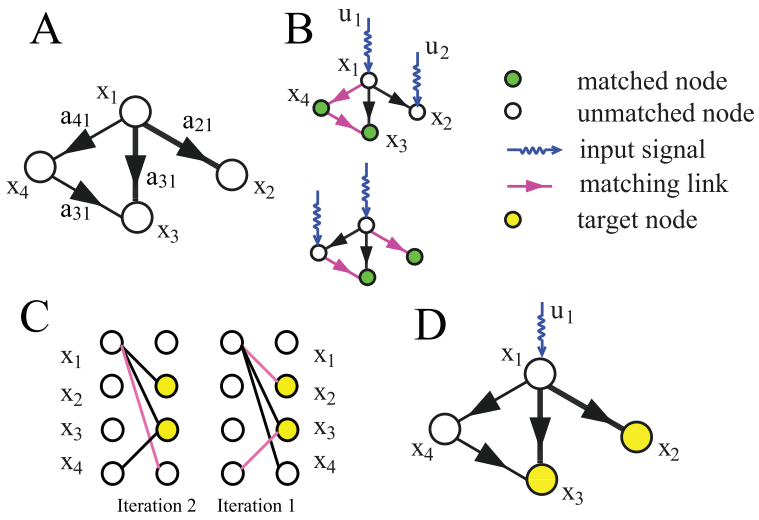

Fig. 16. Schematic of controlling a simple network. (A) Sample network. (B) Full controlled network with maximum matching approach. Two links can be part of a maximum matching for the network, yielding two unmatched nodes (driver nodes). There are two different maximum matchings for this network. (C) Greedy algorithm for target control with the target set $\{x_2, x_3\}$ (in yellow). All target nodes are matched by an induced bipartite graph in the first iteration. Target nodes x_2 and x_3 are matched by node x_1 and x_4, which are the new target nodes considered in the next iteration. We find that node x_1 is the driver node for the target set $\{x_2, x_3\}$ after two iterations. (D) Target controlled network. By controlling the unmatched node x_1, the target nodes x_2 and x_3 are controllable, while we must control two nodes to obtain full control.
Source: Adapted from Refs. [418] and [419].

regulatory relationship between nodes in the network. The input A_{ij} represents the impact strength of node v_j on node v_i. $u(t) = (u_1(t), u_2(t), \ldots, u_M(t))^T$ is the M-dimensional vector of M independent input control signals. B is the input matrix of the node driven directly by the input control signal. The network system described by Eq. (53) is represented as system (A, B).

The system is controllable if it can be transferred from any initial state to any final state in a limited time with a suitable choice of inputs. Kalman's controllability theorem [420] proposed that the system $(A\text{-}B)$ is controllable, if and only if the $N \times NM$ controllability matrix,

$$C = [B\ AB\ A^2B\ \cdots\ A^{N-1}B], \tag{54}$$

has full row rank of N. In a complex network, the structural controllability problem can be transformed to identify the minimum number of driver nodes (inputs control signals) to satisfy that the rank of the controllability matrix C is N. It has been proved that the minimum number of driver nodes needed to maintain full control of the network is determined by the "maximum matching" in the network [418]. In the maximum-matching approach, the maximum matching is the maximum set of links that do not share start or end nodes. A node is said to be matched if a link in the maximum matching points at it; otherwise, it is unmatched. The unmatched nodes in the maximum-matching approach would be the driver nodes, where the network can be fully controlled if each unmatched node is directly controlled (the input control signals). Generally, there are many kinds of maximum matching for a given network, leading to various driver-node combinations. Results show that the number of driver nodes is mainly determined by the networks' degree distribution [418].

Fig. 16 provides an example of a controlled network with the maximum-matching approach. Fig. 16(A) shows a sample directed network in which the weight of the edge for node v_j point to v_i is represented as a_{ij}. As illustrated in Fig. 16(B), there are two different maximum matchings, where the matching edges are shown in purple, matched nodes in green, and unmatched nodes in white. Taking the maximum matching of the upper panel in Fig. 16(B) as an example (the driver nodes are x_1 and x_2), we can calculate the controllability matrix C as follows: $C = \begin{bmatrix} 0 & 0 & 0 & 0 & & & & \\ b_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_2 & a_{21}b_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{31}b_1 & 0 & a_{34}a_{41}b_1 & 0 & 0 & 0 \\ 0 & 0 & a_{41}b_1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, where $A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ a_{31} & 0 & 0 & a_{41} \\ a_{41} & 0 & 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$. The rank of C is 4, which is just the number of the total nodes in the networks, suggesting that the mathematical condition for controllability is satisfied. To obtain the maximum matching of the directed network efficiently, we can convert a directed network into a bipartite graph with two disjoint sets of out- and in-nodes (as illustrated in Fig. 16(C)). A directed link from node v_i to v_j corresponds to a connection between node v_i in the out-set and node v_j in the in-set. By finding the maximum matching of the bipartite graph, the minimum driver nodes are unmatched nodes in the in-set. Maximum matching has the power to solve the structural controllability of a directed network, while it cannot be extended into the

undirected network arbitrarily. Nacher and Akutsu [421] addressed the controllability problem on the undirected network from the perspective of minimum dominating set (MDS), from which all the remaining nodes can be reached by one link. The results show that it is easier to control a network with a more heterogeneous degree distribution. Yuan et al. [422] proposed the exact-controllability paradigm based on maximum multiplicity to identify the minimum set of driver nodes, which can be applied to analyze the controllability of undirected networks.

However, sometimes it is not necessary to be completely controllable, which would waste many resources in practical problems. Gao et al. [419] proposed target control, which is an efficient control of a pre-selected subset of nodes (the yellow nodes in Fig. 16(D)). In target control, the linear time-invariant dynamics of the system can be represented as

$$\begin{cases} \dfrac{dx(t)}{dt} = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases}, \tag{55}$$

where $x(t) = (x_1(t), x_2(t), \ldots, x_N(t))^T$, $u(t) = (u_1(t), u_2(t), \ldots, u_M(t))^T$, and $y(t) = (y_1(t), y_2(t), \ldots, y_S(t))^T$ represent the system's state, input, and output vectors, respectively. A is an $N \times N$ state transition matrix. B is an $N \times M$ input matrix, which identifies the M driver nodes controlled by the external controller. C is an $M \times S$ output matrix, which identifies the S target nodes to be controlled. The network system described by Eq. (55) is represented as system (A, B, C). The system (A, B, C) is said to be target controllable if the target nodes can be transferred from the initial state to any desired final state in finite time with the suitable input vector $u(t)$. Target controllability can be viewed as a special type of output controllability, and the system (A, B, C) is target controllable if and only if the rank of the target controllability matrix satisfies

$$rank([CB\ CAB\ CA^2B\ \cdots\ CA^{N-1}B]) = S, \tag{56}$$

which is the mathematical condition for target controllability. To identify the minimum set of driver nodes sufficient for target control, Gao et al. [419] developed a "k-walk" theory for a directed tree network, in which one node can control a target set of nodes if the length of the path from the control node to each target node is unique. The greedy algorithm (as illustrated in Fig. 16(C)) is proposed to approximate the minimum set of driver nodes sufficient for target control for more general cases [419].

4.4.2. Controllability of biological networks

Structural controllability has been widely applied in biological networks. First, network controllability facilitates the identification of the essential or disease genes, whereas the driver nodes in biological networks tend to be significantly related to some specific biological functions. Vinayagam et al. [423] applied structural controllability theory to identify a minimum set of driver proteins on protein–protein interaction networks. Based on the influence of the driver-node number, the proteins could be classified into three parts: "indispensable", "neutral", and "dispensable", and removing the corresponding protein leads to increasing, not changing, or decreasing the number of driver nodes, respectively. Results indicate that indispensable proteins are the primary targets of disease-causing mutations, human viruses, and drugs, suggesting that altering a network's control property is critical for the transition between healthy and disease states. Wuchty [424] used the MSD model [421] to identify the driver proteins, and enrichment analysis showed that the predicted proteins not only carry important functional characteristics (e.g., essential proteins, cancer-related proteins, and virus-targeted proteins), but they also play a key role in controlling the entire network (e.g., transcription factors and protein kinases). Bidkhori et al. [425] applied the MSDs to identify the hepatocellular carcinoma (HCC) subtype-specific genes with pivotal roles in controlling the metabolic network, and in silico knockout of these genes led to lethality in their respective subtype. Zhang et al. [426] developed a centrality-corrected minimum dominating set (CC-MDS) model that can capture more driver proteins than the MDS model. Schwartz et al. [427] proposed the probabilistic minimum dominating set model (PMDS) to identify a minimum set of nodes that act as driver nodes to control the entire network. Comparing the metabolic networks corresponding to healthy and cancer tissues, PMDS analysis shows that cancer states require fewer controllers than their corresponding healthy states. Nowadays, some direct experimental proofs are also provided to validate the efficiency of the control principles on biological systems [428].

Many results indicate that drug target identification can be formulated as a problem of network control by finding driver nodes in biological networks [429,430]. For example, the critical network control proteins, which belong to every MDS configuration, may be of interest as drug targets and possibly useful for drug design and development [431]. Based on the concept of target control, Guo et al. proposed constrained target controllability (CTC) [432] and the target control problem with objectives-guided optimization (TCO) [433] on biological networks to facilitate the identification of drug targets, where just a subset of nodes must be controlled (e.g., disease-associated genes). The constrained target controllability model required that all of the selected driver nodes must be in the constrained-nodes set [432], while the target control problem with the objectives-guided optimization model [433] considered both the number of driver nodes (minimum) and the correlation between the identified driver nodes and the prior known drug–target genes (maximum). In addition to great advances in linear control of single networks, structural controllability on some more complicated systems also can generate similar observations. Zheng et al. [434] studied the control of multilayer biological networks by identifying the minimal driver nodes that can steer a multilayered nonlinear dynamical system toward any desired dynamical attractor. Results on disease-related multilayer networks show that the known drug targets are enriched in the

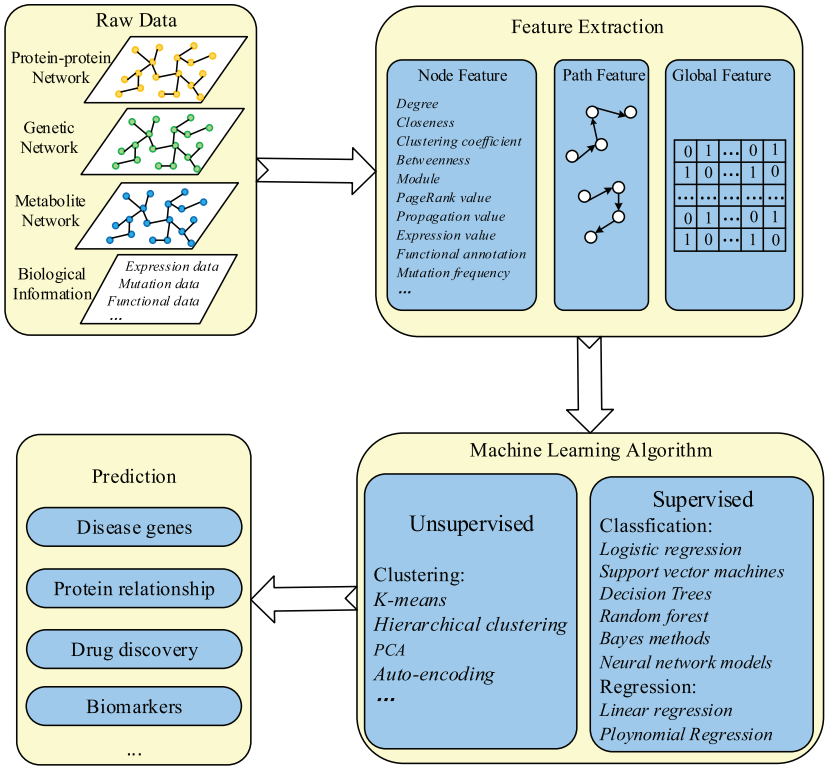

Fig. 17. Simple schematic of machine-learning framework of biological network analysis.
Source: Adapted from Ref. [21].

identified minimal set of the driver nodes. Sun [435] studied the co-controllability of a drug–disease–gene network, which consists of a drug–drug network, a disease–disease network, and a gene–gene network. The results show that driver nodes tend to be highly correlated, e.g., diseases highly associated with driver nodes of a drug–drug network tend to be driver nodes in a disease–disease network and drugs tend to dominate diseases by controlling a protein–protein interaction network. Network control would be an efficient way to identify drug targets for leading the phenotype transitions of underlying biological networks [433].

4.5. Machine learning in network biology

4.5.1. Basics of machine learning

Machine learning, which is at the core of artificial intelligence and data science, has progressed dramatically over the past two decades [436]. Machine-learning methods are the practice of generating a predictive model based on learning feature patterns from the data and predicting the final label of the new data through the model. Fig. 17 is a simple schematic of a machine-learning framework of biological network analysis. The general machine-learning workflow consists of four parts: raw data processing, feature extraction, model training or learning process, and predictions of new data based on the trained model.

To make accurate predictions, the machine-learning model must use many different types of data. The quality of the data is key to the entire machine-learning process, where the performance of any given machine-learning algorithm is dependent on the data used to train the model. With the development of high-throughput technology, we can conveniently obtain a huge amount of experimental data, but how to use the incredibly complex biological dataset effectively is still a significant challenge. Features hidden in the data are the directed inputs of the machine-learning methods. In machine learning in network biology, features include two types of information: the network-based features (such as node centrality, interaction, local structure, subgraph, network propagation results, and network-based similarities) and the biological information (such as the gene expression profile, gene mutation frequency, and gene functional annotation). It is very important to identify the informative features, where the performance is highly sensitive to the quality of the features, and including irrelevant features would lower the performance of the machine-learning model. An example for informative feature selection is to correlate all of the input features with the labels and retain only those significant features.

The learning process aims to find the optimal model that can map the features in the input data into accurate predictions of the labels. Machine-learning methods can be roughly divided into two categories: supervised and unsupervised-learning (see Fig. 17). The supervised-learning methods are used when labels of the input data are available, which are applied to train the machine-learning model and then predict the labels of the new data. As labels can be continuous or discrete, the corresponding supervised methods can be considered regression or classification tasks, respectively. Regression tasks aim to predict a continuous label value of the new data and classification tasks aim to predict the label category. For example, in disease gene prediction, a set of genes are the known disease genes. A machine-learning model based on the prior information must be obtained first to predict whether the others are disease genes, which falls under a classification task. Commonly used supervised approaches include linear regression, support vector machines (SVM), logistic regression, k-nearest neighbors, decision tree, and random forest. Generally, in supervised-learning, $x = (x_0, x_1, \ldots, x_m) \in X$ is set as the input (feature) space and $y \in Y$ as the output space (label). The goal of supervised-learning is to choose a function $f_\omega(x)$ with parameter vector ω that can predict the supervisor's label in the best possible way. The goodness of a function f with parameter ω can be obtained by minimizing the follow loss function:

$$L(\omega) = \sum_{x \in X, y \in Y} \Theta(y, f_\omega(x)), \tag{57}$$

where $\Theta(*, *)$ is the metric used to evaluate the gap between the label and the prediction. Taking the widely used logistic regression model as an example, the target of logistic regression is to find a classification model that can map a label to a piece of sample data. $f_\omega(x)$ can be defined as a sigmoid function:

$$f_\omega(x) = \frac{1}{1 + e^{-\omega^T x}} = \frac{1}{1 + e^{-(\omega_0 x_0 + \omega_1 x_1 + \cdots + \omega_m x_m)}}, \tag{58}$$

which can be considered as the probability that maps label 1 to the corresponding sample, while $1 - f_\omega(x)$ is the probability that maps label 0 to this sample. The probability that the prediction is true can be calculated as follows:

$$P(true) = f_\omega(x)^y * (1 - f_\omega(x))^{1-y}, \tag{59}$$

where $y = 1$ or 0 is the label of the corresponding sample x. To obtain the function f_ω, we must select the vector ω when the $P(true)$ for all the samples ($x \in X$) is maximized. According to maximum likelihood, we can obtain the loss function as follows:

$$L(\omega) = \sum_{x \in X, y \in Y} -y log(f_\omega(x)) - (1 - y) log(1 - f_\omega(x)), \tag{60}$$

where $\Theta(*, *)$ in Eq. (57) is defined as the cross-entropy loss, evaluating the gap between the label y and the prediction $f_\omega(x)$. The regression model f_ω can be obtained by minimizing the loss function $L(\omega)$. In general, all of the supervised machine learning methods can be studied within the framework described in Eq. (57).

Unlike the supervised-learning methods, the unsupervised-learning methods are used when the labels for the input data are unknown. Without the labels to supervise the machine learning, the unsupervised methods can only be trained from the patterns hidden in the input data. The widely used unsupervised models applied to biological data include clustering, PCA, and outlier detection. For instance, in the task of community detection in biological networks (which can also be considered a clustering task), the label of each node in the network is not known, and the sub-graphs are sought using the network structure features. In unsupervised approaches, an algorithm that can automatically divide the similar data into closely related subsets or clusters is desired. Different unsupervised methods are proposed based on the different definitions of "similar". For example, the assumption in the k-means method is that nodes close to each other are more likely to be clustered into the same group, which generates the following loss function (here, the data is set that can be clustered in to K groups):

$$L = \sum_{x \in X} \min_k \parallel x - \mu_k \parallel^2, \tag{61}$$

where $x \in X$ is the input space and μ_k the center vector of the kth cluster. Upon minimizing L, the best clustering of the input data can be obtained. Sometimes, the results of unsupervised-learning, e.g., the clusters of the input data, can be used as the input features of supervised-learning [437].

When the model is well trained, it can be used to predict the results of the new data. Many problems in network biology can be transformed into a classification or regression task, which can be solved in the machine-learning framework very well [21,438]. For instance, Moore et al. [437] proposed a prediction model based on 10,243 features using random-forest and SVM algorithms to predict specialized metabolism genes. Results indicate that the random-forest method generated the most accurate prediction with a true positive rate of 87%. Tokheim et al. [439] applied the random-forest approach to identify the cancer-driver genes based on their patterns of mutation in large patient cohorts. Additionally, machine-learning methods have also demonstrated their utility in drug discovery and development [440], cellular decision-making [441], community prioritization of biological networks [442], and so on. However, using the machine-learning approach in network biology effectively is not straightforward [21]. Generally, the performance of machine-learning methods can be significantly affected by multiple factors, such as features, user-defined parameters,

and the methods in which any incorrect operation would generate pool outcomes. Fortunately, Saez-Rodriguez et al. [443] extracted several rules, including that a simple model is often better, that prior knowledge boosts performance, and that multitasking models produce robust results, for applying machine-learning approaches in network biology.

4.5.2. Deep learning in network biology

Deep learning, as one of the most active fields in machine learning, is making major advances in artificial intelligence and is transforming many fields, including computer vision and natural language processing [436,444]. Deep-learning methods are mainly composed of the artificial neural network model, which was initially inspired by the human brain's neural networks (a very important class of biological networks). An artificial neural network consists of three layers of interconnected computing units (neurons): the input layer, hidden layer, and output layer (as illustrated in Fig. 18). The raw data are processed and fed into the input layer, transformed in a nonlinear way through the hidden layer, and generated as the final prediction in the output layer. As in machine learning, the purpose of the artificial neural network is to obtain a trained neural network model to accurately predict output values for new sets of input data. The depth of a neural network is the number of hidden layers, and when neural networks with large depth are trained, the artificial neural networks are called "deep neural networks", which is referred to as "deep learning". Just like the representation-learning methods with multiple levels of representation, deep-learning methods have turned out to be very good at discovering intricate structures in high-dimensional data [444].

In deep learning, each unit in the hidden layers computes a weighted sum of its inputs from the previous layer and passes the outputs using a nonlinear function to the unit in the next layer. For instance, the outputs of the unit i in the lth layer can be calculated as follows[6]:

$$h_i^{(l)} = \sigma\Big(\sum_{j\in\Gamma_{i^{(l)}}} h_j^{(l-1)} w_{j^{(l-1)}i^{(l)}}\Big), \tag{62}$$

where $\Gamma_{i^{(l)}}$ is the set of the units in the layer $(l-1)$ that link to the unit i^l, $w_{j^{(l-1)}i^{(l)}}$ the weighted value from unit $j^{(l-1)}$ to $i^{(l)}$, and $\sigma(*)$ the nonlinear active function, such as the rectified linear unit (ReLU, $\sigma(z) = max(0, z)$) and the sigmoid function ($\sigma(z) = \frac{1}{1+e^{-z}}$). The parameter w can be estimated using back-propagation methods [444]. Next, the outputs of the new input data can be predicted using the well-learned neural network model.

Typically, the network architecture of the hidden layers can be designed according to the purpose of the learning model. Several major classes of neural networks have been intensively researched in recent years, as follows.

- Fully connected neural networks (FCNNs) (Fig. 18(A)): Fully connected is the most commonly used network architecture, in which each computed unit receives inputs from all of the units of the previous layer. The problem in a fully connected network is that the number of parameters increases dramatically with increasing network depth, yielding new problems, such as overfitting or trapping in the local optimum.
- Convolutional neural networks (CNNs) (Fig. 18(B)): CNNs are designed to process data in the form of multidimensional arrays, e.g., two-dimensional images with three color channels. The hidden layer in a CNN typically consists of multiple convolutional and pooling layers and fully connected layers. A convolutional layer consists of multiple feature maps, in which each unit is connected to local patches in the feature maps of the previous layer through a convolution kernel (a set of parameters that must be learned). At each unit, the dot-product operation is conducted between its inputs and the convolution kernel, and then a nonlinear activation function (typically ReLU) is applied to obtain the outputs. Local connectivity and parameter sharing, where all of the units in a feature map share the same convolution kernel, result in the significant reduction of the number of parameters compared with fully connected networks. A pooling operation is applied after the convolutional layer, which summarizes adjacent units of the feature map by typically computing the maximum or average over their outputs. The pooling layer aims to smooth the feature map by merging semantically similar features into one. The outputs of the pooling operation are fed into the next stage of the convolutional and pooling layer. Finally, the outputs of the last convolutional and pooling layer can be used as inputs to a fully connected layer to generate the final prediction.
- Recurrent neural networks (RNNs) (Fig. 18(C)): RNNs are designed to process one-dimensional sequential data. The network architecture of a RNN is not very complicated, in which the hidden layer can be designed as a fully connected network. The difference is that the units in a RNN get input from other units at previous time steps rather than from the previous layer as in a CNN. At each time step, the outputs of the previous time steps and the new raw input make up the inputs for the current time step. Of note is that the outputs of each time step can be considered the predictions, or used as inputs for the next time step. When a RNN is unfolded in time, it can be considered a very deep feedforward network (the depth is the number of recurrent time steps) in which all of the layers share the same parameters.

For leveraging large datasets very effectively, deep-learning methods have generated many dramatic achievements in computational biology [22,445,446], such as somatic mutation detection in cancer analysis [447], cancer screening

[6] $l = 0$ represents the input layer and $h_i^{(0)}$ is just the raw input data to unit i.

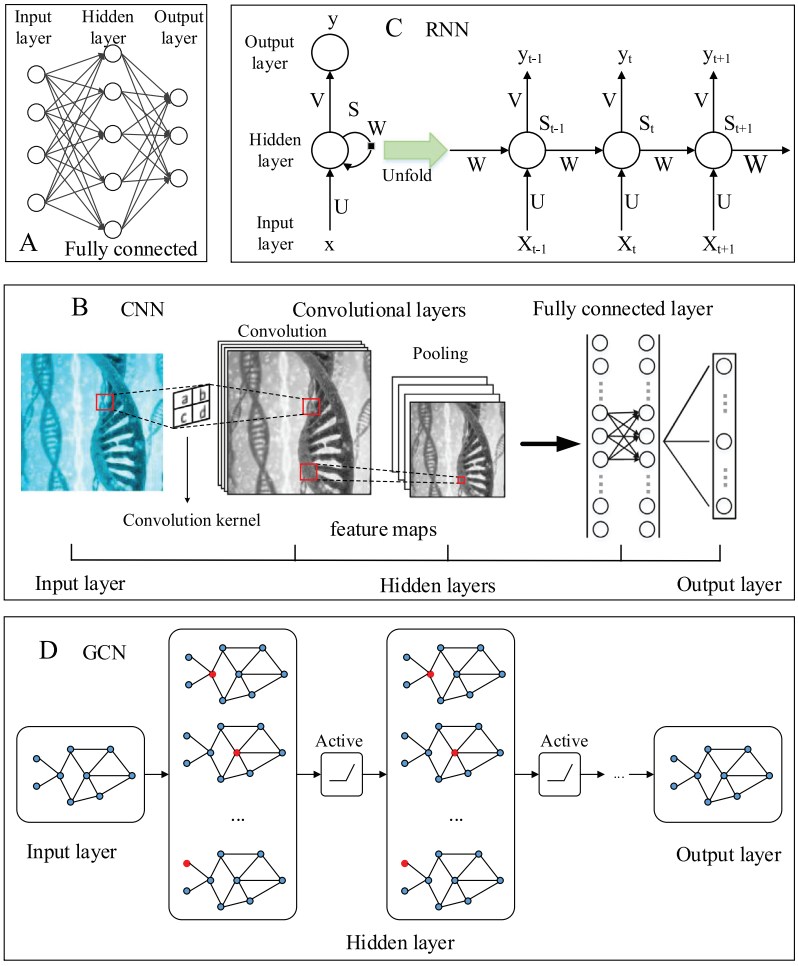

Fig. 18. Simple schematic of several commonly used deep-learning models: (A) fully connected neural network; (B) convolutional neural network; (C) recurrent neural network; (D) graph convolutional neural network.
Source: Adapted from Ref [21].

using computed tomography [448], survival prediction of patients diagnosed with cancers using microscopic images of tissue biopsies [449], drug–target interaction prediction using protein sequences [450], and identification of the role of non-coding mutation in autism spectrum disorders using genomic sequence data [451]. With the development of deep-learning frameworks, such as PyTorch and TensorFlow, users can build neural networks conveniently by designing the network architecture. Unfortunately, modern deep-learning toolboxes are designed for grid or sequence data. However, network data are far more complicated than sequence or grid data; for example, there is no spatial locality like grids or images, and the performance of a CNN and RNN would be limited for a network problem.

Graph convolutional neural networks (GCNs) have been proposed to process network data and they can be considered extensions of CNNs [452]. GCNs use the individual features of nodes and the node connectivity in a graph to solve machine-learning tasks. GCNs sequentially apply multiple graph layers (Fig. 18(D)), whereby each graph layer aggregates features from the neighboring nodes or edges in a nonlinear manner and represents nodes or edges with a new set of features. The state in the l-layer can be obtained with the following layer-wise propagation rule:

$$H^{l} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{\frac{1}{2}}H^{(l-1)}W^{(l-1)}), \tag{63}$$

where $\hat{A} = A + I$ is the adjacency matrix (A) with added self-connections (the identity matrix I), $\hat{D}$ is the diagonal degree matrix of the adjacent matrix $\hat{A}$, $W^{(l-1)}$ is a layer-specific trainable parameter matrix, and $\sigma(*)$ is the activation function.

Transformation has been shown to be very effective at accumulating and encoding features from network neighborhoods and has led to significant improvements in various prediction tasks in network biology. For instance, Luo et al. [453] predicted cancer-driver genes by applying a CNN to learn information within mutation data and similarity networks simultaneously, and the AUC scores were superior to those of competing algorithms. Zitnik et al. [454] proposed the

Decagon method, which developed the GCNN for multi-relational link prediction in multi-modal networks (e.g., protein–protein interactions and drug–protein target interactions), to model polypharmacy side effects of drug combinations. The approach generated very accurate predictions while outperforming the baselines by up to 69%; it can thus open up new opportunities for development of combinatorial drug therapies. Rhee et al. [455] proposed an approach of merging the GCNN and relation network to investigate human disease subtypes classification using the gene expression profile and protein–protein interaction data. Currently, increasingly more researchers are focused on how to apply deep learning to network biology, and this could be a very active field in computational network biology in the near future.

4.5.3. Network embedding

Deriving most informative features is essential for performance, but the process can be labor-intensive and requires domain knowledge. Network embedding approaches, which extract the feature automatically and map network nodes to d-dimensional embeddings, such that similar nodes in the graph are embedded close together, have been widely used in biological network analysis recently. The network embedding approaches aim to find an encoder function that maps node v_i in the original network to the embedding Z_i. The encoder function satisfies the following expression[7]:

$$Sim(v_i, v_j) \approx Z_i^T Z_j, \tag{64}$$

indicating that similarity in the embedding space (e.g., dot product; right-hand side of Eq. (64)) approximates similarity in the original network (left-hand side of Eq. (64)). The embedding Z can be obtained by minimizing the loss function:

$$L = \sum_{v_i, v_j \in V} \| Z_i^T Z_j - Sim(v_i, v_j) \|^2 . \tag{65}$$

Defining the similarity as the connection (or the edge weight) between two nodes, the graph factorization (GF) method [456] can obtain the network embedding by minimizing the following loss function:

$$L = \frac{1}{2} \sum_{v_i, v_j \in V} \| Z_i^T Z_j - A_{ij} \|^2 + \frac{\lambda}{2} \| Z_i \|^2, \tag{66}$$

where $\frac{\lambda}{2} \| Z_i \|^2$ is the regularization item used to avoid over-fitting and λ the regularization coefficient. In addition to the directed connection, all the network structure-based similarities (as illustrated in Section 4.3) can be used. In High-Order Proximity-preserved Embedding (HOPE) [457], the authors experimented with different similarity measures, including KI, PageRank, Common Neighbors, and Adamic–Adar score. The node similarity was represented as $S = M_g^{-1} M_l$; for example, $\begin{cases} M_g = I - \alpha W \\ M_l = (1-\alpha) I \end{cases}$ for the PageRank index (the matrix formula of Eq. (23), where W is the probability transition matrix). Therefore, generalized singular value decomposition (SVD) can be used to obtain the embedding upon minimizing Eq. (65) efficiently.

The second category of network embedding is the random-walk-based approaches and the representative methods include DeepWalk [458] and Node2Vec [459]. They use random walks to generate sequences of nodes from the network, and nodes close in the network will tend to have similar contexts that would be near one another in the embeddings. Technically, for each node v_i, the network neighborhood $\Gamma_R(i)$ can be obtained, which is a sequence of nodes that appear in short random walks starting at node v_i. The algorithms aim to find the embedding Z_i by maximizing the conditional probability of observing $\Gamma_R(i)$, which can be described as follows:

$$L = \sum_{v_i \in V} \sum_{v_j \in \Gamma_R(i)} \log(P(j|Z_i)). \tag{67}$$

The crucial difference between DeepWalk and Node2vec is that DeepWalk uses truncated random walks [458], while Node2vec employs biased random walks that provide a trade-off between breadth-first (BFS) and depth-first (DFS) graph searches [459]. The biased random walks in Node2vec can provide more flexibility when generating the node context, leading to higher quality and more informative embeddings than DeepWalk. Apart from the assumption that nearby nodes in the graph have similar embeddings, struc2vec [460] and GraphWave [461] were developed with specific designs to capture structural roles. Additionally, network embedding is also an important application of deep learning, in which many deep-neural-network-based methods were also proposed to embed networks. For additional insight into the network embedding algorithms, we refer the reader to several excellent review articles (Refs. [462–464]).

Once embeddings are learned, each node is embedded as a low-dimensional vector that can be intuitively used for any downstream prediction task, including node classification, link prediction, and clustering, through the application of machine learning. For example, node classification can be conducted by applying a classifier, such as logistic regression, SVM, and random forest classification, on the set of labeled node embedding for training. For an unlabeled node, the trained classifier can predict its label with its embedding. For high performance over the traditional methods, the network

[7] It should be noted that other similarity measures other than dot products in the embedding space could be used (e.g., Euclidean distance), but the dot product is the standard measure of similarity used.

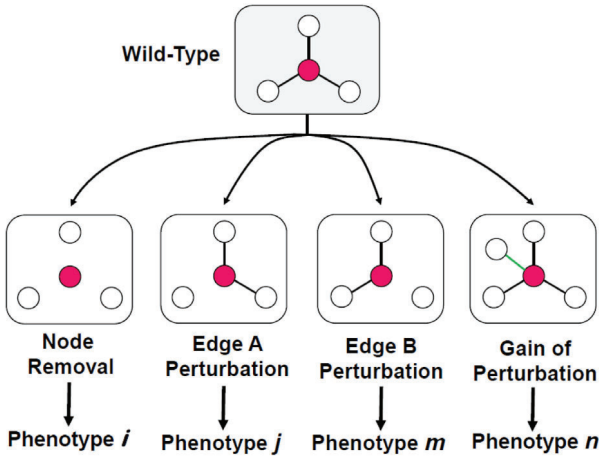

Fig. 19. Diagram illustrating different types of network perturbation altered by disease-causing mutations leading to distinct phenotypes (e.g. diseases). There are two types of networm perturbations: (a) nodetic effects that proteins completely lose expression by disease mutations (e.g., nonsense mutations) or alter protein folding or protein stability by point mutations (e.g., missense mutations); and (b) edgetic mutations or protein–protein interaction (PPIs) disruptive mutations that alter PPIs (including lose of PPIs [edges or links] and gain of PPIs. The details are given in previous studies [119,320].

embedding approaches are widely used in computational biology [464,465], including for the prediction of the specific pathways [466], associations between microRNA and human disease [467], disease-associated genes [468], drug target associations [469,470], relations between brain structure and function [471], protein function [472], and protein functional similarity and synthetic lethality across species [473]. Of note is that the performances of different embedding methods are always different in various case; for example, Node2Vec performs better on node classification while HOPE performs better on link prediction [462]. Therefore, the suitable embedding approaches should be selected for the specific prediction task in computational network biology.

Although machine-learning approaches perform very well in many cases, their "black box" nature, which is the mechanistic uninterpretability of most machine-learning based models, would be the critical challenge for biological applications [21]. For machine-learning (especially deep-learning) methods, the training procedure is so complicated that it can be very difficult to determine the relative importance of the input features or whether a feature is positively or negatively correlated with the outcome. However, it is very important to provide the interpretation of how particular models relate to input features [474]. Very recently, Yang et al. [475] proposed a "white-box" machine-learning approach to reveal the antibiotic mechanisms of action. The proposed approach integrated a genome scale computational model of metabolism with metabolic network simulations to provide mechanistically linked training data for machine learning. In this case, the "white-box" machine-learning analysis, underpinned with mechanistic metabolic information [476], can identify the complex causal mechanisms of drug efficacy rather than only the correlative relationships. Therefore, to extend the application of machine-learning approaches to network biology, more attention should be paid to the development of such "white box" machine-learning approaches, which can be interpreted from the biological perspective [21,477].

5. Applications of computational network biology

5.1. Human disease

5.1.1. Network perturbation by disease mutations

Biological entities are involved in intricate and complex interactions, thereby forming highly complex dynamic systems. Understanding human diseases from the point-of-view of how sub-cellular systems and molecular "interactome" network perturbations underlie disease initiation and progression is the essence of the fields of systems biology and network medicine. The main hypothesis of network medicine is that sub-cellular networks gradually rewire throughout disease initiation, progression, and maintenance, leading to progressive shifts of local and global network properties and systems states, all of which, in turn, underlie disease-causing factors [1]. For example, although being often described as a disease of the genome, it is perhaps more appropriate to describe complex disease as a "disease of the interactome". Obviously, genome alterations, such as amplification, deletion, translocation, and mutations, are the primary events of disease initiation and progression, but such events can only be selected in cells if they encode the appropriate changes or perturbations in the human interactome and systems properties of the affected cells.

Recent remarkable advances in genome sequencing technologies have enabled detailed maps of identified and interpreted genomic mutations. The availability of 10,000 whole-exome/genome sequencing data in patient samples has

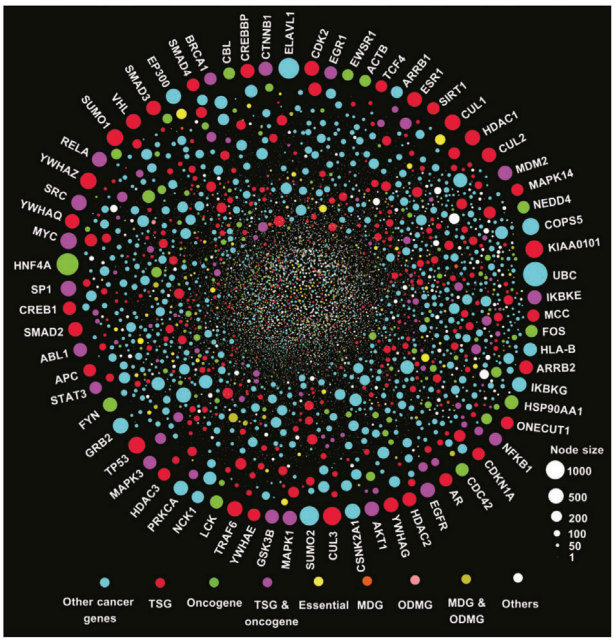

Fig. 20. Network representation of global cancer protein interactome. Node size indicates node connectivity. Edges are hidden. TSG, tumor suppressor genes; MDG, Mendelian disease genes; ODMG, orphan disease mutated genes.
Source: Figure from Ref. [119].

prompted the emergence of multiple challenges for personalized diagnosis and treatment in the clinical setting. To date, the consequences of mutations in enzyme active sites or in ligand–protein binding sites are more straightforward to interpret, named "nodetic" effects on the human interactome [478,479]. However, recent functional studies show that disease-associated alleles commonly alter distinct PPIs or gene regulation rather than grossly affect folding and stability of proteins [320,480]. A new conceptual paradigm (see Fig. 19) that incorporates both innovative experimental and computational strategies, such as "edgetic" mutations [320] or "network-attacking" mutations [119,481,482], enables better assessment of the intrinsic complexities of human diseases from the sub-cellular-molecular level of the human protein–protein interactome. A recent study revealed the unique network centrality of cancer proteins, which is largely independent of gene essentiality, and the network representation of the global cancer protein interactome is illustrated in Fig. 20. By inspecting the network-attacking perturbations by somatic mutations derived from 3268 tumors across 12 cancer types, Cheng et al. found a positive correlation between protein connectivity (degree) and the number of non-synonymous somatic mutations, whereas there is a weaker or insignificant correlation between protein connectivity and the number of synonymous somatic mutations. These comprehensive network analyses support the proof-of-concept of somatic mutational network-attacking perturbations to hub proteins playing a crucial role in tumorigenesis and tumor genome evolution [119].

5.1.2. Disease module

There is convincing evidence that genes associated with complex diseases tend to interact with each other, participating in the same biological pathways [1,2]. Analysis of the disease modules (the proteins that associate with and functionally govern a disease phenotype are localized in the corresponding disease module or subnetwork within the comprehensive protein–protein interaction network) shows that genes that contribute to common disorders (i) show an increased tendency for their products to interact with each other via PPIs; (ii) tend to be expressed together in specific tissues; (iii) tend to display high co-expression; and (iv) tend to share Gene Ontology (GO) terms [23]. These findings support a network-based model for the disease module, whereby cellular networks are modular, consisting of groups of highly interconnected proteins responsible for specific cellular functions. A disease then represents a variation-induced perturbation of a specific disease module, producing developmental and pathophysiological abnormalities. For convenience, we provide several representative bioinformatics resources for building disease modules from the human protein–protein interactome in Table 5.

Most diseases cannot be fully described by a single gene in the gene network; instead, most diseases are associated with multiple genes or a set of genes. Recent studies have found that most of the disease-related proteins show a modular structure in the interactome space, and the module depends on the completeness of the interactome space. From a disease-disorder bipartite network, Goh et al. [2] projected the human disease network and disease gene network. The disease

Table 5
Summary of representative bioinformatics resources for building disease modules from human protein–protein interactome.

Databases	Description	Website	Refs.
OMIM	Comprehensive collection covering literature-curated human disease genes with experimental evidence.	http://www.omim.org/	[483]
CTD	Database containing literature-curated interactions connecting chemical, genes, and diseases.	http://ctdbase.org/	[484]
ClinVar	Public archive of relationships among sequence variation and various human phenotypes.	https://www.ncbi.nlm.nih.gov/clinvar/	[485]
GWAS Catalog	Database containing unbiased SNP-trait associations with genome-wide significance.	https://www.ebi.ac.uk/gwas/	[486]
GWASdb	Data curation and knowledge database for SNP-trait associations from GWAS for PubMed.	http://jjwanglab.org/gwasdb	[487]
PheWAS Catalog	Catalog containing SNP-trait associations identified by the phenome-wide association study (PheWAS).	phewas.mc.vanderbilt.edu	[488]
HuGE Navigator	Integrated disease candidate gene database based on the core data from PubMed abstracts using text-mining algorithms.	https://phgkb.cdc.gov/PHGKB/	[489]
DisGeNET	Disease-gene database created by assembling expert-curated databases and text-mined data.	http://www.disgenet.org/	[490]

network showed that diseases associated with disorders in the same class have a modular structure. By using percolation theory, Menche et al. [23] showed that, despite the incompleteness of the entire interactome, the fragmentation of the disease modules is still enough to investigate the mechanisms and the pathological complexity between diseases.

Disease proteins are not scattered randomly in the human protein–protein interactome, but form one or several connected sub-graphs, defining the disease module [23]. Yet, whether the mutant proteins directly derived from individual patient sequencing data (e.g., whole-exome sequencing) form a statistically significant module in the human protein–protein interactome (or sub-network in a disease module) remains unknown. A recent study demonstrated that the significantly mutated genes (SMGs), identified from large-scale genome sequencing projects across 15 cancer types [25], are more likely to form modules [Fig. 21(C)]. For instance, in lung adenocarcinoma, 83.1% of genes [172/207, $P = 1.6 \times 10^{-62}$ permutation test, as illustrated in Fig. 21(B)] form significantly the largest connected component compared to the same number of randomly selected genes with similar connectivity (degree) as the original SMGs in the human interactome. Fig. 21(A) illustrates the connectivity of products of several highly mutated genes (such as EGFR, TP53, and NF1) compared to genes with low somatic mutation frequency in lung adenocarcinoma. The strong modularity of gene products with high mutation frequency have motivated the development of new network-based methodologies to identify patient-specific disease modules by mapping individual patient DNA sequencing data to the human protein–protein interactome network model. Therefore, Cheng et al. [25] proposed a Genome-wide Positioning Systems network (GPSnet) algorithm for drug target identification by specifically targeting disease modules derived from individual patients' DNA and RNA sequencing profiles mapped to the human protein–protein interactome disease network model.

Numerous bioinformatics and network-based approaches have been proposed for disease module identifications, including Kernel clustering, modularity optimization, random-walk-based, local methods, and ensemble methods. For example, a recent study comprehensively evaluated 75 module identification methods across diverse biological networks, such as PPIs, signaling, gene co-expression, and homology and cancer-gene networks [215]. Through systematic metric evaluation, five practical recommendations for disease module identification were made [215]: (i) Methods from diverse categories should be applied to identify complementary module; (ii) the resulting modules from different methods should be used as is, without forming a consensus network; (iii) diverse networks with complementary information should be leveraged; (iv) module identification methods should be applied to each network individually; and (v) multi-network methods should be applied to illustrate modules in layered networks.

5.2. Network theory in neuroscience

5.2.1. Controlled brain and neuro network

Network control theory was well studied with a systematic mathematical framework in the last decade, both for static networks and temporal networks [19,418,419,491]. As described in Section 4.4, the minimum number of driver nodes is identified in network control, and the network can be fully controlled by the driver nodes so that any signals passed through will lead to a desired network state. However, few experiments were conducted to validate the control theory in real biological systems due to the difficulties of, and accuracy in, mapping the complete structural connectome in the neural level. Yan et al. [428] conducted an experiment to validate the control theory by predicting the neuron function in

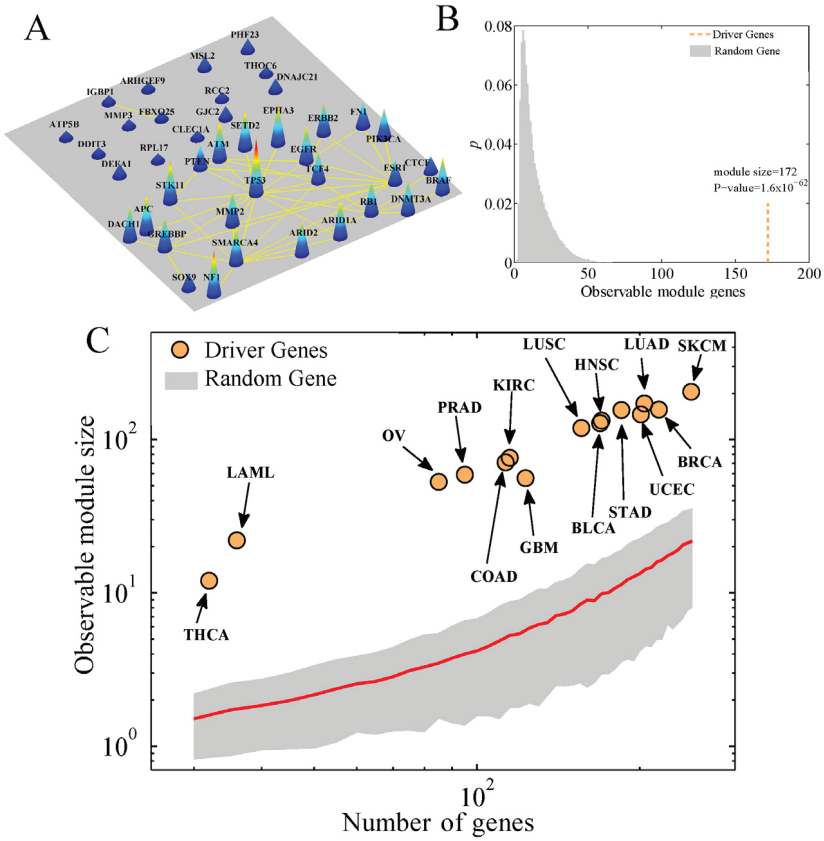

Fig. 21. Proof-of-concept of disease module for mutant genes derived from patient-specific DNA sequencing data. (A) A subgraph illustrating a subnetwork of highly mutated genes versus genes with low mutation frequency in lung adenocarcinoma. (B) The sizes of the largest connected component of significantly mutated genes and of highly mutated proteins are shown for lung adenocarcinoma. The observed module sizes, 172 (SMGs, orange line) and 76 (highly mutated gene, cyan line), are significantly larger than the random expectation. (C) Both significantly mutated genes identified by statistical approaches and highly mutated genes ranked by mutation frequency form a significant largest connected component (LCC) compared to random genes with matching connectivity (degree) distributed in the human protein–protein interactome, across 15 cancer types. *Source:* Figure from Ref. [25].

the Caenorhabditis elegans (C. elegans) connectome, which is the only one with relatively complete mapping. The entire nervous system can be modeled as a directed network, in which part of the nodes are neurons that accept the input signals, part of the nodes are muscles that capture the locomotion status, and the electrical and neuromuscular junctions connect these parts together. It was found that by a total of 12 neuronal classes could the C. elegans entire motor neuron system be fully controlled. Yan et al. also found one neuron, PDB, that was not fully understood before, and that has a special role in locomotion. This work provided strong experiment evidence for controlling real-world biological complex systems.

Control of the brain network offers great opportunities for fully understanding the brain circuit and the cognitive roles in human brain regions [492,493]. Gu et al. [492] applied control theories to the controllability of the brain network. Based on the different control strategies, they reported the distribution of hub driver nodes in different cognitive systems (see Fig. 22). In detail, they determined the following. (i) The average control strategy that identifies the set of nodes that can drive the network into various states with little input energy, which is related to the trace of the controllability matrix. 30% of the hub controllers are located in the default mode system, followed by the visual system (19%), somatosensory system (18%), and frontoparietal system (11%). (ii) The modal control strategy that identifies the set of nodes that can drive the network dynamics to states that are difficult to control and usually related to the states that need substantial input energy. The hub controllers are mainly located in the somatosensory systems (18%), cingulo-opercular system (17%), auditory system (16%), frontoparietal system (15%), and dorsal attention system (12%). (iii) The boundary control strategy, which identifies the set of nodes that are located at the boundaries of network modules and that leverage the segregation of network modules. They found that the hub controllers are relatively uniformly located in each cognitive system, with dorsal attention system possessing 19%, which is the largest one, and most of the systems studied have approximately 10% hub controllers. Therefore, network control theory can be used to understand how the brain moves between cognitive states drawn from the network organization of white-matter microstructure. The global controllability analysis indicated

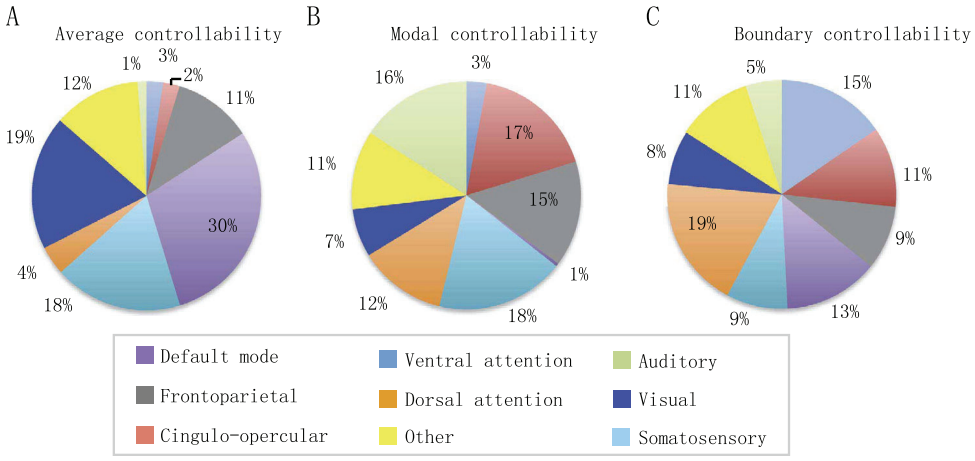

Fig. 22. Distribution of hub drivers in different cognitive systems in the brain using three control strategies: (A) average controllability, (B) modal controllability, and (C) boundary controllability.
Source: Figure from Ref. [492].

that, while the brain is consistently theoretically controllable through a single region (node), it is extremely difficult to control the system through a single region in practice. Regional controllability analysis revealed that densely connected areas, preferentially located in the default mode system, are postulated to facilitate the movement of the brain to many easily reachable states; weakly connected areas, predominantly located in cognitive control systems, are postulated to facilitate the movement of the brain to difficult-to-reach states; and areas at the boundary between network communities, predominantly located in attentional control systems, are postulated to facilitate the integration or segregation of cognitive systems. These results suggest that structural network differences between the default mode, cognitive control, and attentional control systems dictate their distinct roles in brain network function. Network control theory can also be used to predict if the effects of stimulation to a single region on the brain remained focal or were spread globally, and structural connectivity differentially constrains the effects of regional stimulation [494]. Moreover, preferentially optimized dynamic network control would be a possible mechanism of human brain development [495].

5.2.2. Identifying influential nodes in brain network

A complex network is an interactive system in which the action on one node will influence the entire system via the interaction links. Studies have found that a much smaller subset of structural nodes play an important role in information spreading and epidemic propagations [335]. A series of works have been done to measure and identify the influential nodes in the network. For example, recently Morone et al. [496] proposed a method that maps the problem to the optimal percolation, which offers strong insights into the weakly connected nodes in the network.

Identifying the essential nodes is of importance for brain networks. On the one hand, these nodes are associated with many brain-related disorders, which can help understand the disorder-related information transmission in the brain network. On the other hand, these nodes are functionally connected to the entire brain network, which can be used to efficiently modulate the entire brain network via the minimal set of nodes. Ferraro et al. [497] applied the optimal percolation theory to identify the influential nodes in the brain network and found that a small set of nodes with lower degree in the nucleus cumulatively influences the entire brain memory network, and the results were confirmed by targeted inactivation regions in the brain network.

5.2.3. Brain network dynamics

Along with the change of our live environment and social communications during human living, the functional configurations of our brain also change and develop accordingly, based on our adaptive systems [93]. Studying brain network dynamics can help us understand the interdependence of brain regions in cognitive function evolution, the dynamic control of the brain network, and the early warning signals of diseases, all of which requires accurate detection of functional changes in brain regions and sufficient real data samples across time and space. Here, we list some of the advances in cognitive system dynamics in brain networks. Wang et al. [498] found that during the stroke recovery process the motor execution network in the brain was evolving to random mode by using the linear mixed regression model, which integrated the network characteristics, including clustering coefficient, shortest path length, and betweenness centrality. Grefkes and Fink [499] reviewed the motor system reorganization after a stroke, discussed the advantages and disadvantages of the connectivity approaches, and highlighted the network analysis in understanding the motor system. Khambhati et al. [500,501] studied brain network dynamics during neocortical epilepsy episodes using connectivity matrix analysis constructed from brain function networks. Bassett et al. [502–504] studied the dynamics of functional regions in the brain during learning and lexical processing.

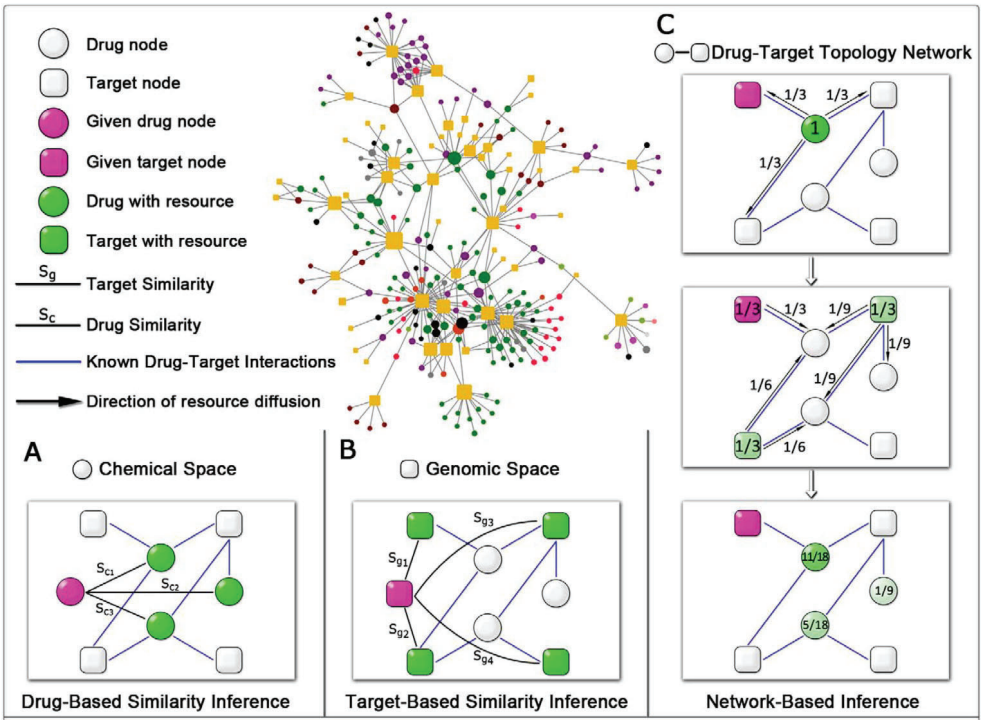

Fig. 23. Schematic of network-based approach in prediction of drug–target interactions: (A) drug-based similarity inference (DBSI), (B) target-based similarity inference (TBSI), and (C) network-based inference (NBI) methods. In ACC, given drug node (pink circle) denotes the drug that we want to predict a new target for, given target node (pink square) denotes the target that we want to predict a new drug for, drug with resource (green circle) denotes that this drug has a resource, and target with resource (green square) denotes that this target has a resource. The more resources a node possesses, the darker the color is. Blue edges denote the drug–target interactions with known experimental evidence and black arrows denote the resource diffusion direction.
Source: Figure from Ref. [27].

5.3. Drug development

Although investment in biomedical and pharmaceutical research and development has increased significantly over the past two decades, the annual number of new treatments approved by the U.S. Food and Drug Administration has remained relatively constant and limited [505]. While there are many factors that contribute to this limited approval rate, one important and often overlooked determinant is the continued adherence to the classical "one gene, one drug, one disease" hypothesis of drug development dating to the work of Ehrlich. As drug targets do not operate in isolation from the complex system of proteins that comprise the molecular machinery of the cell with which they associate, we believe that each drug–target interaction must be examined in appropriate integrative context.

5.3.1. Prediction of drug–target interactions

There are several different categories of computational drug repositioning approaches, including ligand-based [506], target-based [507], chemogenomics-based [506], and network-aided approaches [27,508,509]. Here, we focus on three types of network-based methodologies for prediction of DTIs.

Drug similarity-based network inference:

The underlying hypothesis of drug similarity-based network inference posits that if a drug interacts with a target, then other drugs similar to this drug will be inferred to the given target with a high probability [27] (see Fig. 23(A)). An interaction between d_i and t_l can be determined by the following equation:

$$v_{il}^{d} = \frac{\sum_{j=1,j\neq i}^{N} S_x(d_i, d_j)a_{jl}}{\sum_{j=1,j\neq i}^{N} S_x(d_i, d_j)}, \tag{68}$$

where N is the number of drugs in the system, if there is an interaction between drug d_j, $a_{jl} = 1$; otherwise, $a_{jl} = 1$, and target t_l, $S_x(d_i, d_j)$ denotes the similarity between drugs d_i and d_j, which can be measured by different approaches as below.

Chemical similarity. Chemical similarity $S_c(d_i, d_j)$ between drugs d_i and d_j is often measured by various chemo-physical descriptors or fingerprints [510]. For example, the Tanimoto similarity for drugs d_i and d_j can be calculated using molecular fingerprints calculated by OpenBabel [511] or other tools [512]:

$$S_c(d_i, d_j) = \frac{c_{ij}}{a_i + b_j - c_{i,j}}, \tag{69}$$

where a_i and b_j denote the bit set in d_i and d_j fingerprint bit-strings, respectively, c_{ij} represents these bits being set in both d_i and d_j. In addition to Tanimoto similarity, several other metrics, including Cosine, Hamming, Russell-Rao, and Forbes, are often used as well [513].

Drug side-effect similarity. Mathematically, each drug is coded by a given number of side-effect bit vectors. Each bit denotes a side effect for a specific drug annotated in the publicly available clinical databases [514]. If a side-effect event is reported to be associated with a given drug in clinical settings, the corresponding bit is set to *1*; otherwise, it is set to *0*. Then, side-effect similarity $S_{SE}(d_i, d_j)$ between drugs d_i and d_j can be measured by Tanimoto, Cosine, Hamming, Russell-Rao, or Forbes.

Drug therapeutic similarity. Drug therapeutic profiles, such as the Anatomical Therapeutic Chemical Classification System (ATC) codes, offer a new dimension with which to quantify drug similarity [513]. The drug ATC codes can be downloaded from several public databases, such as DrugBank (Table 3). Then, the kth level drug therapeutic similarity (S_k) between d_i and d_j is calculated by the ATC codes as below:

$$S_k(d_i, d_j) = \frac{ATC_k(d_i) \cap ATC_k(d_j)}{ATC_k(d_i) \cup ATC_k(d_j)}, \tag{70}$$

where $ATC_k(d)$ denotes all of the ATC codes at the kth level of drug d. Finally, a score $S_{ATC}(d_i, d_j)$ between d_i and d_j at all of the ATC codes is used to define the therapeutic similarity between d_i and d_j as

$$S_{ATC}(d_i, d_j) = \frac{\sum_{k=1}^{n} S_k(d_i, d_j)}{n}, \tag{71}$$

where n denotes all five-level ATC codes (1 to 5). However, for some drugs with multiple ATC codes, $S_{ATC}(d_i, d_j)$ is often calculated for each code, and the average therapeutic similarity is used for network building as described in Eq. (70).

Target similarity-based network inference

The biological hypothesis for target similarity-based network inference posits that, if a drug interacts with a target protein, then the drug will be inferred to other targets that share a similarly biological distance (e.g. protein sequence identity) or a closely functional similarity (e.g. network topological similarity in the human protein–protein interaction network or the GO similarity [515]) to a specific protein (see Fig. 23(B)). An interaction between d_i and t_l is calculated by the following predicted score:

$$v_{il}^{t} = \frac{\sum_{j=1, j\neq l}^{M} S_x(t_l, t_j) a_{ij}}{\sum_{j=1, j\neq l}^{M} S_x(t_l, t_j)}, \tag{72}$$

where M is the number of targets in the systems and $S_x(t_l, t_j)$ denotes the protein evolutionary distance or functional similarity between targets t_l and t_j as described below.

Protein sequence identity. Protein biological distances are often measured using protein sequence alignment. For example, a normalized version of Smith–Waterman scores [516] is commonly used to quantify protein sequence identity between corresponding proteins/targets t_l and t_j, as described in our previous studies [27,506].

Network topological similarity in protein–protein interaction network. The network distance $S_N(t_l, t_j)$ between corresponding proteins/targets t_l and t_j is calculated using all-pair network distances on the human protein–protein interaction network via the shortest path distance measure. Then, network distance can be transformed to similarity measure as below [517]:

$$S_N(t_l, t_j) = Ae^{-S(t_l, t_j)}, \tag{73}$$

where A is often assigned an empirical value 0.9 [517] and self-similarity is assigned a value of 1.

Functional similarity by sharing the terms of the Gene Ontology. The functional similarity of drug corresponding proteins/targets is often measured by their shared GO terms [515]. A functional similarity asserts that drug target-coding genes sharing very specific functions are more similar to each other than those only sharing generic GO terms. The biological or functional similarity $S_{GO}(t_l, t_j)$ between corresponding targets t_l and t_j is measured by the most specific GO terms they share:

$$S_{GO}(t_l, t_j) = \frac{2}{min(n_y)}, \tag{74}$$

where n_y denotes the total number of genes (proteins) annotated in the entire GO corpus. The value of $S_{GO}(t_l, t_j)$ ranges from 0 (no shared GO terms) to 1 (drug corresponding proteins/targets t_l and t_j are the only two according target-coding genes (proteins) annotated to a specific GO term).

Network-based inference:

Unweighted network-based inference. Considering the bipartite graph $G(D, T, E)$, a mass diffusion-based method can be used to generate the predicted list. For a given drug d_i, supposing that a kind of resource is initially located in the targets interacting with d_i, the resource will diffuse to all of the targets in the DTI network after the network-based resource allocation process [27]. Each target node will averagely distribute its resource to all of the neighboring drugs and then each drug redistributes the received resource to all of the neighboring targets (see Fig. 23(C), and this process can be represented as Eq. (50) in Section 4.3)). The final resource on the targets that are not connected with the drug d_i in $G(D, T, E)$ could be considered the predicted score for each target, and the targets with high scores are more likely to interact with d_i. This shows the initial resource of a_{ij} between d_i (cycle) and t_j (square) defined in Eq. (1).

Denoting $F_{0N\times M}$ as the initial resource and $F_{0(i,j)} = a_{ij}$, $R_{N\times N}$ as the total resource (e.g., degree) for each drug,

$$R = diag(\sum_{j=1}^{M} a_{1j}, \sum_{j=1}^{M} a_{2j}, \ldots, \sum_{j=1}^{M} a_{Nj}), \tag{75}$$

and $H_{M\times M}$ as the total resource (degree) for each target,

$$R = diag(\sum_{i=1}^{N} a_{i1}, \sum_{i=1}^{N} a_{i2}, \ldots, \sum_{i=1}^{N} a_{iM}). \tag{76}$$

Finally, the resource matrix ($F_{1N\times M}$) is obtained as

$$F_1 = F_0 W_{M\times M} \quad or \quad F_1^T = F_0^T W_{N\times N}, \tag{77}$$

where the transfer matrix is $\begin{cases} W_{M\times M} = (F_0 H^{-1})^T (R^{-1} F_0) \\ W_{N\times N} = (R^{-1} F_0)(F_0 H^{-1})^T \end{cases}$. The large value in the matrix F_1 would be the predicted drug–target interactions, and the in vitro assays show the predictive power [27].

Node-weighted network-based inference. In general, hub nodes with more receiving resources often tend to generate the higher predicted scores, which often leads to a potential risk of false positive rate. Compared to the unweighted NBI method, the node-weighted network-based inference (NWNBI) utilizes a new expression of initial resource distribution by taking into account the influence of resources associated with the receiver nodes in the DTI network [509]. For the initial resource matrix, the resource for each drug and target node is the same as the unweighted NBI. The final resource matrices can be calculated as $F_1^d = F_0 W_{M\times M}^d$ for drugs and $F_1^t = F_0^T W_{N\times N}^t$ for targets. The transfer matrix is $\begin{cases} W_{M\times M}^d = (F_0 H^{-1})^T (R^{-1} F_0 H'^{-1}), \ where H'_{ij} = H_{ij}^{\beta} \\ W_{N\times N}^t = (R^{-1} F_0)(R'^{-1} F_0 H^{-1})^T, \ where R'_{ij} = R_{ij}^{\beta} \end{cases}$. β is a tunable parameter used to balance the influence. Compared with a uniform case ($\beta = 0$), a positive β value strengthens the influence of hub nodes, while a negative β value weakens it [509].

Edge-weighted network-based inference. Mathematically, an edge between a drug and target can be weighted using binding affinity data, such as IC_{50} or K_i value. For edge-weighted network-based inference (EWNBI), each edge of the DTI network will be weighted by the drug-protein binding affinity, where a_{ij} is the specific value rather than the binary (0 or 1). We can obtain the prediction of the edge-weighted DTI interaction by substituting the binary value of a_{ij} in Eqs. (75), (76), and (77) for the weighted value.

5.3.2. Drug repurposing

Novel approaches, such as network-based drug–disease proximity, that shed light on the relationship between drugs (i.e., drug targets) and diseases (i.e., molecular disease determinants in disease modules within the PPI network) [23,518,519] can serve as useful tools for efficient screening of potentially new indications for approved drugs, or for previously unidentified adverse events [142,520,521]. The basis for network-based drug repurposing rests on the notions that (i) the proteins that associate with and functionally govern a disease phenotype are localized in the corresponding disease module or sub-network within the comprehensive PPI network [23], and (ii) proteins that serve as drug targets for a specific disease may also be suitable drug targets for another disease owing to common PPIs and functional pathways elucidated by the PPI (see Fig. 24).

Given the set of drug targets (T) and the set of disease proteins (S), we can quantify the network topological distance $d(s, t)$ between nodes $s \in S$ and $t \in T$ in the human interactome. As illustrated in Section 3.2.1, there are four different distance measures that take into account the path lengths between drug targets and the set of disease proteins [142]: (a) the closest measure (d_c), representing the average shortest path length between targets of T and the nearest proteins of S; (b) the shortest measure (d_s), representing the average shortest path length among all targets of drugs; (c) the kernel measure (d_k), down-weighting longer paths via an exponential penalty; and (d) the center measure (d_{cc}), representing the shortest path length among all targets of drugs with the greatest closeness centrality among proteins in S and T. These four measures can be calculated according to Eq. (2) in Section 3.2.1.

Finally, the significance of the measure can be evaluated by comparison to the reference distance distribution corresponding to the expected network topological distance between two randomly selected groups of proteins matched to size and degree (connectivity) distribution as the original disease proteins and drug targets in the human interactome.

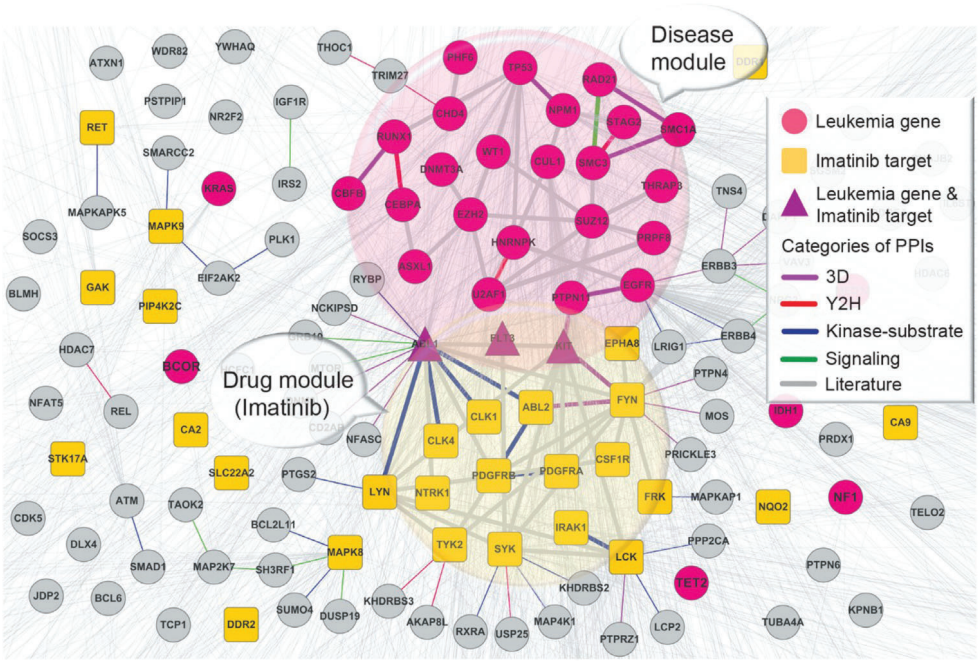

Fig. 24. Schematic of network-based relationship of drug–disease proximity. A small neighborhood of the interactome revealing the pharmacological nature of each interaction of the origin of drug–target interactions and disease-gene associations. Drug targets associated with Imatinib are shown in yellow squares. Leukemia-associated gene-coding proteins are shown in red circles. Shared proteins are shown as drug or disease modules by a connected sub-graph.

This procedure is often repeated 1000 or 10,000 times in order to perform permutation test. For closest distance measure [d_c in Eq. (2)], the mean distance ($\bar{d}_c$) and standard deviation (σ_{d_c}) of the reference distribution are used to calculate a z score (z) by converting an observed closest distance to a normalized distance using

$$z = \frac{d_c - \bar{d}_c}{\sigma_{d_c}}. \tag{78}$$

If the distance is significantly smaller than the randomly selected groups (e.g., $z < -2$), there would be a therapeutic effect of the drug on the corresponding disease. Cheng et al. [143] selected four network-predicted drug–disease associations and tested their causal relationship using large health-care databases with over 220 million patients. Two of the four predictions were validated with patient-level data: carbamazepine is associated with an increased risk of coronary artery disease (CAD) and hydroxychloroquine is associated with a decreased risk of CAD.

5.3.3. Network based drug combinations

The clinical safety and effectiveness of monotherapeutic agents are intrinsically characterized by their biological activity spectrum on many proteins in the human proteome, which is named network pharmacology [518,522]. A "magic bullet" drug in monotherapy is commonly said to have effects far broader than its molecular targets according to the interdependencies of cellular and molecular effector components in biological systems [523]. The promiscuous off-target profiles of drugs are commonly associated with various side effects of monotherapeutic agents and adverse drug–drug interactions (DDIs) of polypharmacy [524,525]. Combination therapies (see Fig. 25) that use pairwise or multiple drugs in combination to target the complementary cellular pathways have suggested higher effectiveness compared to monotherapy [526,527]. In addition, combination therapy can also reduce side effects of monotherapies, as the dosage of each active ingredient in combination is lower than the dosage of the respective drug during monotherapy [528]. Hence, drug combinations have boosted clinical outcomes for many notoriously complex diseases, such as hypertension [529], cancer [530,531], and viral infection [532], via synergistically targeting multiple disease proteins or pathways. However, the effectiveness of beneficial drug combinations and adverse DDIs are intimately coupled by the shared biological pathways among different diseases and cellular heterogeneity [533]. Identifying drug combinations with high efficacy and low toxicity – to achieve to synergistically target a specific disease cellular pathway over thousands of types of diseases towards minimizing the toxic profiles (e.g., adverse DDIs) – is an exceedingly difficult task in drug discovery and patient care.

The traditionally experimental assays carried out to map and understand the systems-level view of disease and drug pathways is expensive, time-consuming, and not necessarily feasible. Despite recent rapid developments in chemistry,

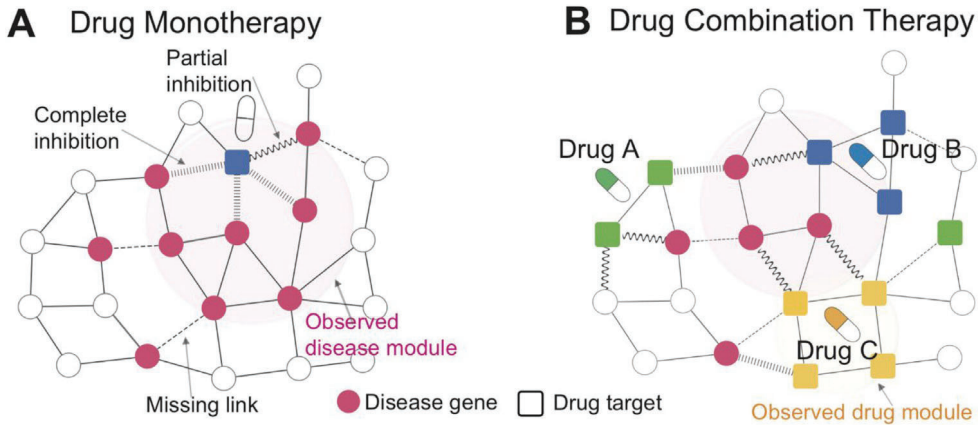

Fig. 25. Schematic of network-based relationship of drug–drug interactions in human interactome network. (A) Traditional "magic bullet" drugs that focus on designing maximally selective ligands to target a single disease protein [134] have contributed to a significant decrease in pharmaceutical research and development productivity due to lack of efficacy or toxic effect. (B) Drug combinations that show partial inhibition of a small set of disease proteins (e.g., disease modules) have been recognized to be more efficient than "magic bullet"-drug-based monotherapy.
Source: Figure from Ref. [26].

biology, and medicine that have been remarkably productive over the past century, experimental assays for pairwise combinations of thousands of FDA-approved drugs are infeasible without an efficient method to test vast amounts of pairs [534–536]. For example, for 100 drugs, there are 4950 possibly pairwise combinations on one disease; for 1000 drugs, 499,500; and for 10,000 drugs, 49,995,000. We are still far from even a cursory searching of the vast number of possible pairwise combinations that could exist for hundreds of different complex diseases and for different doses of individual active ingredients. In the past decade, two major types of computational methods for prediction of drug combinations have emerged: (i) data-driven approaches [523,534,537–539], and (ii) systems-biology approaches [526]. Data-driven approaches mainly include machine-learning-based approaches [537,538] and mathematical modelings [539]. However, most of the currently reported approaches are machine-learning-based "black box" models. Highly predictive, mechanism-based models have been described very rarely. Developing new systems-biology-based, mechanism models are urgently needed for the emerging development of combination therapies.

Drug targets are not scattered randomly in the human interactome, but they tend to interact with each other, forming one largest connected sub-graph that we call a drug module (see Fig. 26). The agglomeration of drug targets has been used for the development of various network approaches to assess adverse DDIs and drug combinations [537,540,541]. A recent study described an interactome-based solution to uncover drug–drug relationships for the emerging development of drug combination therapies. Specifically, they measured the network proximity of drug–target modules A and B as reflected in their target localizations using the recently introduced separation measure [23],

$$S_{AB} = \langle d_{AB} \rangle - \frac{\langle d_{AA} \rangle + \langle d_{BB} \rangle}{2}, \tag{79}$$

which compares the mean shortest distance within the interactome between the targets of each drug, $\langle d_{AA} \rangle$ and $\langle d_{BB} \rangle$, to the mean shortest distance $\langle d_{AB} \rangle$ between A–B target pairs [Fig. 26(A)]. In $\langle d_{AB} \rangle$, targets associated with both drugs A and B have a zero distance by definition. For $S_{AB} < 0$, the targets of the two drugs are located in the same network neighborhood [Fig. 26(B)], while for $S_{AB} \geq 0$ the two drug targets are topologically separated [Fig. 26(C)]. For example, imatinib (I) is a FDA-approved agent for the treatment of chronic myeloid leukemia [542]. Fig. 26(A) and (B) show that imatinib's targets are in the same network neighborhood as the targets of tandutinib (T), a FMS-like tyrosine kinase 3 (FLT3) inhibitor under phase III trial for treating acute myeloid leukemia; consequently, the separation score between their targets is negative, $S_{IT} = -0.35$. However, the targets of natalizumab (N), an FDA-approved drug for treating multiple sclerosis, are in a topologically distinct neighborhood from the targets of imatinib and tandutinib, having a positive $S_{IN} = 0.59$ [Fig. 26(C)]. These network analyses offer proof-of-concept of drug–drug–target relationships in the human interactome network.

In this work, Cheng et al. [26] further explored the network-based relationship between two drug–target modules and a disease module (drug–drug–disease combinations). Mathematically, all of the possible drug–drug–disease combinations can be classified into four topologically distinct classes: (a) overlapping exposure, in which two overlapping drug–target modules also overlap with the disease module of interest [Fig. 27(A)]; (b) complementary exposure, in which two separated drug–target modules overlap individually with the disease module [Fig. 27(B)]; (c) single (indirect) exposure [Fig. 27(C)]; and (d) non-exposure, in which two overlapping drug–target modules are topologically separated from the disease module [Fig. 27(D)]. The question is do these four classes manifest in detectable differences in clinical efficacy for drug combinations?

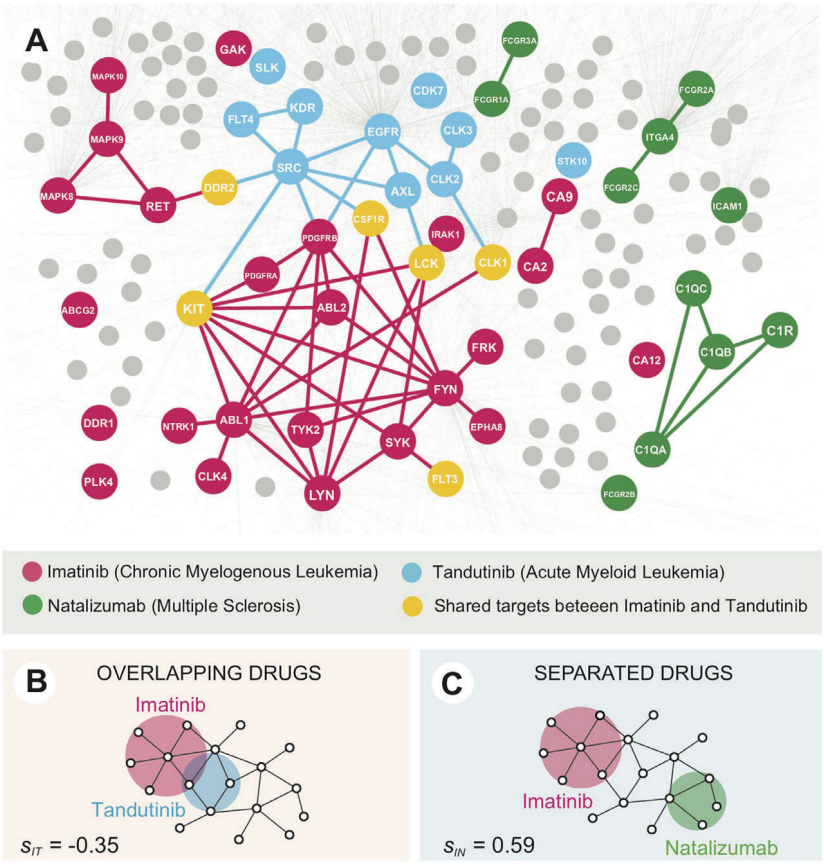

Fig. 26. Network separation and drug–drug relationships. (A) Subnetwork of global interactome illustrating the network-based relationship between drug targets associated with three drugs (imatinib, lapatinib, and natalizumab) shown in the legend. (B) and (C) Cartoon representing the definition of drug pairs that are topologically overlapped [$S_{AB} < 0$, (B)] and separated [$S_{AB} > 0$, (C)], respectively.
Source: Figure from Ref. [26].

To test this hypothesis, Cheng et al. assembled two benchmark datasets: (i) clinically reported adverse DDIs and (ii) FDA-approved or experimentally validated pairwise drug combinations [26]. One key finding is that a drug combination is therapeutically effective only if it follows a specific relationship to the disease module, as captured by the complementary exposure pattern in the targets' modules of both drugs (Fig. 27). Importantly, if we are looking for therapeutically synergistic combinations, the two drugs must not only overlap with the disease module, but the two drug–target modules must also be separated in the human interactome without overlapping toxic mechanisms (Fig. 27). This study offers the first proof-of-concept of rational design of drug combination from the human interactome network perspective.

5.3.4. Personalized treatment

At the root of so-called "precision medicine" is the hypothesis that disease treatment would be considerably better if therapies were guided by a patient's genomic mutations/alterations. This hypothesis has sparked major initiatives focusing on whole-genome/-exome sequencing, creation of large databases, and developing tools for their statistical analyses — all aspiring to identify actionable mutations, and, molecular targets, in a patient. At the center of the massive amount of collected sequence data is their interpretation, which largely rests on statistical analysis and phenotypic observations. From the network biology and protein conformational standpoints, statistics is vital, since it guides identification of disease-associated mutations. However, statistics of mutations does not identify a change in protein conformation and network perturbations in cells; thus, it may not define sufficiently accurately actionable mutations, neglecting those which are rare. Understanding biological networks, pathways, and broadly sub-cellular systems will require dynamic information at multi-dimensional levels, including cells, tissues, organs, and organisms, which are missing in current experimental and computational approaches. Currently, disease therapy is moving from a drug-centered to a patient-centered approach with different levels of personalization. This requires paradigm shifts along the entire drug development process and multi-omics data integration using network-based approaches. For example, Cheng proposed a network-based approach for accelerated development of personalized medicine, termed genome-wide positioning systems network (GPSnet)

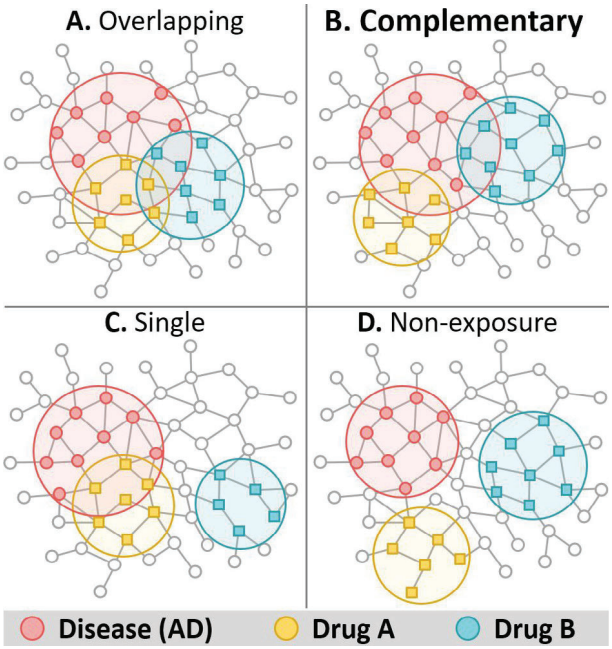

Fig. 27. Network hypothesis of efficacious drug combinations. (A–D) Diagrams of distinct exposure classes capturing the network-based relationship between two drug–target modules and one disease module.
Source: Adapted from Ref. [26].

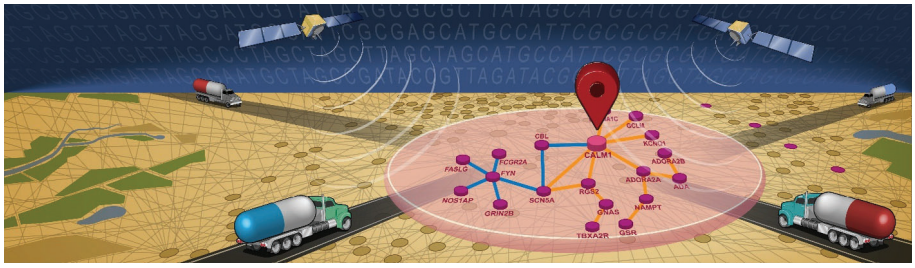

Fig. 28. A diagram illustrating the genome-wide positioning systems network (GPSnet) algorithm for personalized medicine. Individualized disease network modules built from multi-omics data and the human interactome could guide therapeutic evaluation during clinical trials and optimize new treatment for current patients. Specifically, GPSnet can built disease module (subnetworks) by a unique integration of multi-omics data, including genomics, proteomics, metabolomics, and other types of omics data. And GPSnet then can utilize disease modules for patient clustering (subtype identification) and for guiding personalized treatment.

algorithm (Fig. 28) [25]. Specifically, GPSnet integrates individual patient's DNA and RNA sequencing profiles into the human protein–protein interactome networks. They demonstrated that GPSnet-predicted disease modules were highly correlated with drug responses and prioritize new therapeutic discovery by the "precise" targeting of individualized disease modules. More details about network-based approaches for accelerating personalized medicine are given in several reviews [543,544].

6. Outlook

The application of network perspective to the study of biological systems is generating great interest and attention from scientific researchers, and describing a complex disease as a "disease of the interactome" has become a widely accepted concept [545]. Ranging from biological molecular–molecular interaction to cell–cell interaction, and from unipartite networks (e.g., PPI interactions) to heterogeneous networks (e.g., drug–target networks), we have reviewed how to understand the biological insights from various kinds of biological network data and help disease diagnosis and drug discovery via computational network-based approaches. Despite all of the achievements outlined in this review, many open questions and unexplored directions still must be reckoned with in future studies. In this section, we list some of the critical challenges in computational network biology.

First, more high-quality biological data are needed to extract more reliable biological insights. In the so-called Big Data era, the increasing amount of available biological data encourages us to study biological systems from the point of view of data analysis, where all of the achievements outlined in this review will all benefit from the advances of network data science. However, data quality is the critical limitation in the application of computational network biology. The biological network data quality issues arise from two aspects.

(i) *Data incompleteness*. For instance, despite impressive advances in high-throughput methods, we have obtained only less than 20% of all of the potential pairwise protein interactions in the human cell [23,32]. That is to say, many network biological studies are fulfilled relying on interactome maps that are more than 80% incomplete. In this case, we must ask that to what extent the current achievements are reliable and the current data are sufficient for the network-based studies (e.g., mapping out the disease module). The data incompleteness would hinder the extraction of novel biological insights.

(ii): *Data bias*. The current biological network data are biased toward much-studied genes [546]. These genes, to which much attention has been paid in many previous studies, could be identified in more interactions. In this way, the important genes will more likely be the hubs because of the data bias. An open question is how to eliminate the data bias influence in order to mine more reliable biological insights in computational network biology. Therefore, high-quality datasets and effective pre-treatments are needed to improve the data quality. Additionally, biological network datasets are collected from various experiments (Section 2), and the extra biological data, such as somatic mutation information and differential expression information, can be downloaded from different databases, such as Online Mendelian Inheritance in Man (OMIM),[8] The Cancer Genome Atlas (TCGA),[9] Gene Expression Omnibus (GEO),[10] and International Cancer Genome Consortium (ICGC).[11] Integrating the biological data from various datasets well is of great importance to future studies in computational network biology.

Second, in-depth understanding of the biological network structure should be presented. Although some new insights have been obtained from computational network biology, there is still no clear mapping between cell function and its topological features within the biological network. As illustrated in this review, the current studies are focused on the identification of topological structural properties in biological networks, such as the scale-free (-like) degree distribution, betweenness centrality, module structure, and good resiliency. Some studies should test the correlation between the structural properties and some biological functions, e.g., the hub nodes tend to be the essential proteins (centrality–lethality rule) and the disease genes tend to be linked together. However, the questions of how the corresponding network structure forms through biological activities and how the structure properties influence cell function are still unclear. This blind spot limits the identification of biological insights from network biology and the application of the network-based approaches. The interpretation of biological network structure calls for deep cooperation between communities, including physicists, biologists, and computer scientists.

The third aspect is that the approaches in computational network biology need further development. We have reviewed many kinds of network-based approaches that are widely applied in biological systems, including network centrality, network propagation, link prediction, structural controllability, and machine-learning-based approaches. To deal with the specific problems, we must select suitable methods. For instance, many kinds of approaches can be used to identify disease genes, but different methods may have very significant impacts on the identification, and generate diverse results. It would be very challenging to choose the best method for a specific disease-gene prediction problem. Generally, most current network-based approaches are developed in the social networks, and they are directly transferred to the biological systems to solve the network biology problems. With the different hypotheses on the patterns of the social networks, the performance of the corresponding approaches on biological systems would be limited. In this way, we hope for some biological-network-oriented approaches in computational network biology studies. Because of the power of predictive and descriptive tasks, designing machine-learning-based, especially deep-learning-based, approaches by merging different types of bio-information would be of great potential for future studies. Additionally, the application of the computational biology network must be extended, for instance, network-based approaches would shed some light on traditional Chinese medicine [547–549].

The above-mentioned issues and open directions mainly concern theoretical and methodological developments. In biology, many experiments are proposed to validate network biology predictions. However, this is far from sufficient, and we are also eager to develop evaluations based on the real biological experiments or even clinical trials of computational biology network approaches. As we are convinced that the computational network biology framework is really an effective tool for mining new biological insight, we believe this review will offer many valuable enlightenments for related researchers, and we hope to see an increasing range of communities – including physicists, biologists, computer scientists, and interdisciplinary researchers – contributing to this field.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8 https://omim.org/.

9 http://cancergenome.nih.gov/.

10 https://www.ncbi.nlm.nih.gov/geo/.

11 https://icgc.org/.

Acknowledgments

This work was partially supported by Natural Science Foundation of China (Grant Nos. 61873080 and 61673151), Natural Science Foundation of Zhejiang Province (Grant Nos. LY18A050004 and LR18A050001), the Major Project of The National Social Science Fund of China (Grant No. 19ZDA324), the Natural Science Foundation of Chongqing (Grant No. cstc2018jcyjAX0090), and the Swiss National Science Foundation (Grant No. 200020_182498). This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH) under Award Number K99 HL138272 and R00 HL138272 to F.C. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E to R.N. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. This research was supported [in part] by the Intramural Research Program of NIH, National Cancer Institute, Center for Cancer Research.

References

[1] A.-L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genet. 12 (1) (2011) 56–68.
[2] K.-I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.-L. Barabási, The human disease network, Proc. Natl. Acad. Sci. USA 104 (21) (2007) 8685–8690.
[3] J.X. Hu, C.E. Thomas, S. Brunak, Network biology concepts in complex disease comorbidities, Nat. Rev. Genet. 17 (10) (2016) 615–629.
[4] A.-L. Barabási, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev. Genet. 5 (2) (2004) 101–113.
[5] M. Vidal, M.E. Cusick, A.-L. Barabási, Interactome networks and human disease, Cell 144 (6) (2011) 986–998.
[6] T. Ideker, R. Nussinov, Network approaches and applications in biology, PLoS Comput. Biol. 13 (10) (2017) e1005771.
[7] C.M. Schneider, L. de Arcangelis, H.J. Herrmann, Modeling the topology of protein interaction networks, Phys. Rev. E 84 (1) (2011) 016112.
[8] E. Eisenberg, E. Levanon, Preferential attachment in the protein network evolution, Phys. Rev. Lett. 91 (13) (2003) 138701.
[9] M. Chavez, M. Valencia, V. Navarro, V. Latora, J. Martinerie, Functional modularity of background activities in normal and epileptic brain networks, Phys. Rev. Lett. 104 (11) (2010) 118701.
[10] G.R. Ferreira, H.I. Nakaya, L. Costa, Gene regulatory and signaling networks exhibit distinct topological distributions of motifs, Phys. Rev. E 97 (4) (2018) 042417.
[11] V.M. Eguiluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, A.V. Apkarian, Scale-free brain functional networks, Phys. Rev. Lett. 94 (1) (2005) 018102.
[12] R. Tanaka, Scale-rich metabolic networks, Phys. Rev. Lett. 94 (16) (2005) 168101.
[13] H. Jeong, S. Mason, A.-L. Barabási, Z. Oltvai, Lethality and centrality in protein networks, Nature 411 (6833) (2001) 41–42.
[14] X. Liao, A.V. Vasilakos, Y. He, Small-world human brain networks: perspectives and challenges, Neurosci. Biobehav. Rev. 77 (2017) 286–300.
[15] D.M. Lorenz, A. Jeng, M.W. Deem, The emergence of modularity in biological systems, Phys. Life Rev. 8 (2) (2011) 129–160.
[16] M. Zanin, J.M. Tuñas, E. Menasalvas, Understanding diseases as increased heterogeneity: a complex network computational framework, J. R. Soc. Interface 15 (2018) 20180405.
[17] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, T. Zhou, Vital nodes identification in complex networks, Phys. Rep. 650 (2016) 1–63.
[18] L. Cowen, T. Ideker, B.J. Raphael, R. Sharan, Network propagation: a universal amplifier of genetic associations, Nat. Rev. Genet. 18 (9) (2017) 551–562.
[19] Y.-Y. Liu, A.-L. Barabási, Control principles of complex systems, Rev. Modern Phys. 88 (3) (2016) 035006.
[20] M.E.J. Newman, Communities, modules and large-scale structure in networks, Nat. Phys. 8 (1) (2012) 25–31.
[21] D.M. Camacho, K.M. Collins, R.K. Powers, J.C. Costello, J.J. Collins, Next-generation machine learning for biological networks, Cell 173 (7) (2018) 1581–1592.
[22] G. Eraslan, Z. Avsec, J. Gagneur, F.J. Theis, Deep learning: new computational modelling techniques for genomics, Nat. Rev. Genet. 20 (7) (2019) 389–403.
[23] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabási, Uncovering disease-disease relationships through the incomplete interactome, Science 347 (6224) (2015) 1257601.
[24] E.L. Huttlin, R.J. Bruckner, J.A. Paulo, J.R. Cannon, L. Ting, K. Baltier, G. Colby, F. Gebreab, M.P. Gygi, H. Parzen, J. Szpyt, S. Tam, G. Zarraga, L. Pontano-Vaites, S. Swarup, A.E. White, D.K. Schweppe, R. Rad, B.K. Erickson, R.A. Obar, K.G. Guruharsha, K. Li, S.A. Rtavanis-Tsakonas, S.P. Gygi, J.W. Harper, Architecture of the human interactome defines protein communities and disease networks, Nature 545 (7655) (2017) 505–509.
[25] F. Cheng, W. Lu, C. Liu, J. Fang, Y. Hou, D.E. Handy, R. Wang, Y. Zhao, Y. Yang, J. Huang, D.E. Hill, M. Vidal, C. Eng, J. Loscalzo, A genome-wide positioning systems network algorithm for in silico drug repurposing, Nature Commun. 10 (2019) 3476.
[26] F. Cheng, I.A. Kovacs, A.-L. Barabási, Network-based prediction of drug combinations, Nature Commun. 10 (2019) 1197.
[27] F. Cheng, C. Liu, J. Jiang, W. Lu, W. Li, G. Liu, W.-X. Zhou, J. Huang, Y. Tang, Prediction of drug-target interactions and drug repositioning via network-based inference, PLoS Comput. Biol. 8 (5) (2012) e1002503.
[28] R. Nussinov, H. Jang, C.-J. Tsai, F. Cheng, Precision medicine review: rare driver mutations and their biophysical classification, Biophys. Rev. 11 (1) (2019) 5–19.
[29] F. Cheng, H. Liang, A.J. Butte, C. Eng, R. Nussinov, Personal mutanomes meet modern oncology drug discovery and precision health, Pharmacol Rev. 71 (1) (2019) 1–19.
[30] L.Y. Lee, J. Loscalzo, Network medicine in pathobiology, Am. J. Pathol. 189 (7) (2019) 1311–1326.
[31] Z.-M. Ren, A. Zeng, Y.-C. Zhang, Structure-oriented prediction in complex networks, Phys. Rep. 750 (2018) 1–51.
[32] T. Rolland, et al., A proteome-scale map of the human interactome network, Cell 159 (5) (2014) 1212–1226.
[33] J.F. Rual, et al., Towards a proteome-scale map of human protein-protein interaction network, Nature 437 (7062) (2005) 1173–1178.
[34] S.K. Burley, et al., RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy, Nucleic Acids Res. 47 (2019) D464–D474.
[35] R. Mosca, A. Ceol, P. Aloy, Interactome3D: adding structural details to protein networks, Nature Methods 10 (1) (2013) 47–53.
[36] M.J. Meyer, J. Das, X. Wang, H. Xu, INstruct: a database of high-quality 3D structurally resolved protein interactome networks, Bioinformatics 29 (12) (2013) 1577–1579.
[37] M.J. Meyer, J.F. Beltran, S. Liang, R. Fragoza, A. Rumack, J. Liang, X. Wei, H. Yu, Interactome INSIDER: a structural interactome browser for genomic studies, Nature Methods 15 (2018) 107–144.
[38] F. Cheng, P. Jia, Q. Wang, Z. Zhao, Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy, Oncotarget 5 (11) (2014) 3697–3710.
[39] R.H. Newman, et al., Construction of human activity-based phosphorylation networks, Mol. Syst. Biol. 9 (2013) 655.

[40] P.V. Hornbeck, B. Zhang, B. Murray, J.M. Kornhauser, V. Latham, E. Skrzypek, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations, Nucleic Acids Res. 43 (2015) D512–D520.
[41] C.-T. Lu, K.-Y. Huang, M.-G. Su, T.-Y. Lee, N.A. Bretana, W.-C. Chang, Y.-J. Chen, Y.-J. Chen, H.-D. Huang, dbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications, Nucleic Acids Res. 41 (2013) D295–D305.
[42] H. Dinkel, C. Chica, A. Via, C.M. Gould, L.J. Jensen, T.J. Gibson, F. Diella, Phospho.ELM: a database of phosphorylation sites-update 2011, Nucleic Acids Res. 39 (2011) D261–D267.
[43] D. Fezekas, M. Koltai, D. Turei, D. Modos, M. Palfy, Z. Dul, L. Zsakai, M. Szalay-Beko, K. Lenti, I.J. Farkas, T. Vellai, P. Csermely, T. Korcsmaros, SignaLink2 - a signaling pathway resource with multi-layered regulatory networks, BMC Syst. Biol. 7 (2013) 7.
[44] S. Peri, et al., Human protein reference database as a discovery resource for proteomics, Nucleic Acids Res. 32 (2004) D497–D501.
[45] A. Chatr-Aryamontri, B.J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, C. Stark, A. Breitkreutz, N. Kolas, L. O'Donnell, T. Reguly, J. Nixon, L. Ramage, A. Winter, A. Sellam, C. Chang, J. Hirschman, C. Theesfeld, J. Rust, M.S. Livstone, K. Dolinski, M. Tyers, The BioGRID interaction database: 2015 update, Nucleic Acids Res. 43 (2015) D470–D478.
[46] M.J. Cowley, M. Pinese, K.S. Kassahn, N. Waddell, J.V. Pearson, S.M. Grimmond, A.V. Biankin, S. Hautaniemi, J. Wu, PINA v2.0: mining interactome modules, Nucleic Acids Res. 40 (2012) D862–D865.
[47] L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Lannuccelli, E. Galeota, F. Sacco, A. Palma, A.P. Nardozza, E. Santonico, L. Castagnoli, G. Cesareni, MINT, the molecular interaction database: 2012 update, Nucleic Acids Res. 40 (2012) D857–D861.
[48] S. Orchard, et al., The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2014) D358–D363.
[49] K. Breuer, A.K. Foroushani, M.R. Laird, C. Chen, A. Sribnaia, R. Lo, G.L. Winsor, R.E. Hancock, F.S. Brinkman, D.J. Lynn, InnateDB: systems biology of innate immunity and beyond - recent updates and continuing curation, Nucleic Acids Res. 41 (2013) D1228–D1233.
[50] I.N. Smith, S. Thacker, M. Seyfi, F. Cheng, C. Eng, Conformational dynamics and allosteric regulation landscapes of germline PTEN mutations associated with autism compared to those associated with cancer, Am. J. Hum. Genet. 104 (5) (2019) 861–878.
[51] T.S.K. Prasad, et al., Human protein reference database–2009 update, Nucleic Acids Res. 37 (2009) D767–D772.
[52] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic Acids Res. 45 (2017) D362–D368.
[53] J. Hu, H.-S. Rho, R.H. Newman, J. Zhang, H. Zhu, J. Qian, PhosphoNetworks: a database for human phosphorylation networks, Bioinformatics 30 (1) (2014) 141–142.
[54] Q. Pan, O. Shai, L.J. Lee, J. Frey, B.J. Blencowe, Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing, Nat. Genet. 40 (12) (2008) 1413–1415.
[55] X. Yang, et al., Widespread expansion of protein interaction capabilities by alternative splicing, Cell 164 (4) (2016) 805–817.
[56] L.M. Smith, N.L. Kelleher, The Consortium for Top Down Proteomics Proteoform: a single term describing protein complexity, Nature Methods 10 (3) (2013) 186–187.
[57] J. Gao, S.V. Buldyrev, S. Havlin, H.E. Stanley, Robustness of a network formed by n interdependent networks with a one-to-one correspondence of dependent nodes, Phys. Rev. E 85 (6) (2012) 066134.
[58] N.J. O'Neil, M.L. Bailey, P. Hieter, Synthetic lethality and cancer, Nat. Rev. Genet. 18 (10) (2017) 613–623.
[59] D.P. McLornan, A. List, G.J. Mufti, Applying synthetic lethality for the selective targeting cancer, New Engl. J. Med. 371 (18) (2014) 1725–1735.
[60] B.M. Emerling, et al., Depletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-null tumors, Cell 155 (4) (2013) 844–857.
[61] T. Wang, H. Yu, N.W. Hughes, B. Liu, A. Kendirli, K. Klein, W.W. Chen, E.S. Lander, D.M. Sabatini, Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras, Cell 168 (5) (2017) 890–903.
[62] J.P. Shen, et al., Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions, Nature Methods 14 (6) (2017) 573–576.
[63] D. Du, A. Roguev, D.E. Gordon, M. Chen, S.-H. Chen, M. Shales, J.-P. Shen, T. Ideker, P. Mali, L.S. Qi, N.J. Krogan, Genetic interaction mapping in mammalian cells using CRISPR interference, Nature Methods 14 (6) (2017) 577–580.
[64] D.M. Munoz, et al., CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions, Cancer Discov. 6 (8) (2016) 900–913.
[65] R.M. Meyers, et al., Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells, Nat. Genet. 49 (12) (2017) 1779–1784.
[66] M. Costanzo, et al., The genetic landscape of a cell, Science 327 (5694) (2010) 425–431.
[67] L. Jerby-Arnon, N. Pfetzer, Y.Y. Waldman, L. McGarry, D. James, E. Shanks, B. Seashore-Ludlow, A. Weinstock, P.A. Clemons, E. Gottlieb, E. Ruppin, Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality, Cell 158 (5) (2014) 1199–1209.
[68] S. Sinha, D. Thomas, S. Chan, Y. Gao, D. Brunen, D. Torabi, A. Reinisch, D. Hernandez, A. Chan, E.B. Rankin, R. Bernards, R. Majeti, D.L. Dill, Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data, Nature Commun. 8 (2017) 15580.
[69] F. Cheng, C. Liu, C.-C. Lin, J. Zhao, P. Jia, W.-H. Li, Z. Zhao, A gene gravity model for the evolution of cancer genomes: a study of 3000 cancer genomes across 9 cancer types, PLoS Comput. Biol. 11 (9) (2015) e1004497.
[70] D.E. Dykhuizen, A.M. Dean, D.L. Hartl, Metabolic flux and fitness, Genetics 115 (1) (1987) 25–31.
[71] P.D. Keightley, H. Kacser, Dominance, pleiotropy and metabolic structure, Genetics 117 (2) (1987) 319–329.
[72] A.M. Feist, M.J. Herrgård, I. Thiele, J.L. Reed, B.Ø. Palsson, Reconstruction of biochemical networks in microorganisms, Nat. Rev. Microbiol. 7 (2) (2009) 129–143.
[73] N. Ottman, M. Davids, M. Suarez-Diez, S. Boeren, P.J. Schaap, V.A.P.M. Martins Dos Santos, H. Smidt, C. Belzer, W.M. de Vos, Genome-scale model and omics analysis of metabolic capacities of akkermansia muciniphila reveal a preferential mucin-degrading lifestyle, Appl. Environ. Microbiol. 83 (18) (2017) e01014–17.
[74] F. Branco Dos Santos, B.G. Olivier, J. Boele, V. Smessaert, P. De Rop, P. Krumpochova, G.W. Klau, M. Giera, P. Dehottay, B. Teusink, P. Goffin, Probing the genome-scale metabolic landscape of bordetella pertussis, the causative agent of whooping cough, Appl. Environ. Microbiol. 83 (21) (2017) e01528–17.
[75] Y. Zhang, J. Cai, X. Shang, B. Wang, S. Liu, X. Chai, T. Tan, Y. Zhang, T. Wen, A new genome-scale metabolic model of corynebacterium glutamicum and its application, Biotechnol. Biofuels 10 (2017) 169.
[76] A. Ahmad, H.B. Hartman, S. Krishnakumar, D.A. Fell, M.G. Poolman, S. Srivastava, A genome scale model of geobacillus thermoglucosidasius (C56-YS93) reveals its biotechnological potential on rice straw hydrolysate, J. Biotechnol. 251 (2017) 30–37.
[77] S.N. Mendoza, P.M. Canon, A. Contreras, M. Ribbeck, E. Agosin, Genome-scale reconstruction of the metabolic network in oenococcus oeni to assess wine malolactic fermentation, Front. Microbiol. 8 (2017) 534.
[78] M.P. Cortes, S.N. Mendoza, D. Travisany, A. Gaete, A. Siegel, V. Cambiazo, A. Maass, Analysis of piscirickettsia salmonis metabolism using genome-scale reconstruction, modeling, and testing, Front. Microbiol. 8 (2017) 2462.

[79] L. Toro, L. Pinilla, C. Avignone-Rossa, R. Rios-Estepa, An enhanced genome-scale metabolic reconstruction of streptomyces clavuligerus identifies novel strain improvement strategies, Bioprocess Biosyst. Eng. 41 (5) (2018) 657–669.
[80] D. Voet, J.G. Voet, C.W. Pratt, Fundamentals of Biochemistry: Life at the Molecular Level, John Wiley & Sons, 2013.
[81] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks, Nature 407 (6804) (2000) 651–654.
[82] R. Overbeek, N. Larsen, G.D. Pusch, M. D'Souza, J.E. Selkov, N. Kyrpides, M. Fonstein, N. Maltsev, E. Selkov, WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction, Nucleic Acids Res. 28 (1) (2000) 123–125.
[83] E. Ravasz, A. Somera, D.A. Mongru, Z.N. Oltvai, A.-L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (5586) (2002) 1551–1555.
[84] P. Holme, Metabolic robustness and network modularity: a model study, PLoS One 6 (2) (2011) e16605.
[85] D.S. Lee, J. Park, K.A. Kay, N.A. Christakis, Z.N. Oltvai, A.-L. Barabási, The implications of human metabolic network topology for disease comorbidity, Proc. Natl. Acad. Sci. USA 105 (29) (2008) 9880–9885.
[86] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain, Neuroimage 15 (1) (2002) 273–289.
[87] A. Zalesky, A. Fornito, I.H. Harding, L. Cocchi, M. Yucel, C. Pantelis, E.T. Bullmore, Whole-brain anatomical networks: does the choice of nodes matter? Neuroimage 50 (3) (2010) 970–983.
[88] V.J. Wedeen, D.L. Rosene, R.P. Wang, G. Dai, F. Mortazavi, P. Hagmann, J.H. Kaas, W.Y.I. Tseng, The geometric structure of the brain fiber pathways, Science 335 (6076) (2012) 1628–1634.
[89] J.D. Schmahmann, D.N. Pandya, R. Wang, G. Dai, H.E. D'Arceuil, A.J. de Crespigny, V.J. Wedeen, Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography, Brain 130 (2007) 630–653.
[90] C.J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.P. Thiran, R. Meuli, P. Hagmann, Predicting human resting-state functional connectivity from structural connectivity, Proc. Natl. Acad. Sci. USA 106 (6) (2009) 2035–2040.
[91] L. Furlong, V. Cropley, S. Rossell, T. Van Rheenen, The structural, functional, and effective connectivity of the facial emotion processing neural circuitry in bipolar disorder: A review, Bipolar Disord. 21 (2019) 82–83.
[92] A. Zalesky, A. Fornito, E.T. Bullmore, On the use of correlation as a measure of network connectivity, Science 60 (4) (2012) 2096–2106.
[93] S.L. Bressler, V. Menon, Large-scale brain networks in cognition: emerging methods and principles, Trends Cogn. Sci. 14 (6) (2010) 277–290.
[94] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A.-C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z.T. Dame, B. Han, Y. Zhou, D.S. Wishart, DrugBank 4.0: shedding new light on drug metabolism, Nucleic Acids Res. 42 (2014) D1091–D1097.
[95] H. Yang, C. Qin, Y.-H. Li, L. Tao, J. Zhou, C.-Y. Yu, F. Xu, Z. Chen, F. Zhu, Y.-Z. Chen, Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information, Nucleic Acids Res. 44 (2016) D1069–D1074.
[96] T. Hernandez-Boussard, M. Whirl-Carrillo, J.M. Hebert, L. Gong, R. Owen, M. Gong, W. Gor, F. Liu, C. Truong, R. Whaley, M. Woon, T. Zhou, R.B. Altman, T.E. Klein, The pharmacogenetics and pharmacogenomics knowledge base: accentuating the knowledge, Nucleic Acids Res. 36 (2008) D913–D918.
[97] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, Nucleic Acids Res. 40 (2012) D1100–D1107.
[98] T. Liu, Y. Lin, X. Wen, R.N. Jorissen, M.K. Gilson, BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities, Nucleic Acids Res. 35 (2007) D198–D201.
[99] A.J. Pawson, J.L. Sharman, H.E. Benson, E. Faccenda, S.P.H. Alexander, O.P. Buneman, A.P. Davenport, J.C. McGrath, J.A. Peters, C. Southan, M. Spedding, W.-Y. Yu, A.J. Harmar, The IUPHAR/BPS guide to pharmacology: an expert-driven knowledgebase of drug targets and their ligands, Nucleic Acids Res. 42 (2014) D1098–D1106.
[100] R. Apweiler, A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S.L. Yeh, UniProt: the universal protein knowledgebase, Nucleic Acids Res. 32 (2004) D115–D119.
[101] A.P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Kruger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, J.P. Overington, The ChEMBL bioactivity database: an update, Nucleic Acids Res. 42 (2014) D1083–D1090.
[102] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res. 44 (2016) D1045–D1053.
[103] Y. Wang, E. Bolton, S. Dracheva, K. Karapetyan, B.A. Shoemaker, T.O. Suzek, J. Wang, J. Xiao, J. Zhang, S.H. Bryant, An overview of the PubChem BioAssay resource, Nucleic Acids Res. 38 (2010) D255–D266.
[104] A.H. Wagner, A.C. Coffman, B.J. Ainscough, N.C. Spies, Z.L. Skidmore, K.M. Campbell, K. Krysiak, D. Pan, J.F. McMichael, J.M. Eldred, J.R. Walker, R.K. Wilson, E.R. Mardis, M. Griffith, O.L. Griffith, DGIdb 2.0: mining clinically relevant drug-gene interactions, Nucleic Acids Res. 44 (2016) D1036–D1044.
[105] M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild, T.H. Blicher, C. von Mering, L.J. Jensen, P. Bork, STITCH 4: integration of protein-chemical interactions with user data, Nucleic Acids Res. 42 (2014) D401–D407.
[106] J. Nickel, B.-O. Gohlke, J. Erehman, P. Banerjee, W.W. Rong, A. Goede, M. Dunkel, R. Preissner, SuperPred: update on drug classification and target prediction, Nucleic Acids Res. 42 (2014) W26–W31.
[107] J. Lamb, The Connectivity Map: a new tool for biomedical research, Nat. Rev. Cancer 7 (1) (2007) 54–60.
[108] Q. Duan, C. Flynn, M. Niepel, M. Hafner, J.L. Muhlich, N.F. Fernandez, A.D. Rouillard, C.M. Tan, E.Y. Chen, T.R. Golub, P.K. Sorger, A. Subramanian, A. Ma'ayan, LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures, Nucleic Acids Res. 42 (2014) W449–W460.
[109] Y. Igarashi, N. Nakatsu, T. Yamashita, A. Ono, Y. Ohno, T. Urushidani, H. Yamada, Open TG-GATEs: a large-scale toxicogenomics database, Nucleic Acids Res. 43 (2015) D921–D927.
[110] B. Ganter, R.D. Snyder, D.N. Halbert, M.D. Lee, Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database, Pharmacogenomics 7 (7) (2006) 1025–1044.
[111] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[112] J.C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, P. Legrain, The protein-protein interaction map of helicobacter pylori, Nature 409 (6817) (2001) 211–215.
[113] J. Stuart, E. Segal, D. Koller, S. Kim, A gene-coexpression network for global discovery of conserved genetic modules, Science 302 (5643) (2003) 249–255.
[114] E. Gross, Statistical mechanics of scale-free gene expression networks, Europhys. Lett. 100 (5) (2012) 58004.
[115] E.L. Huttlin, et al., The BioPlex network: a systematic exploration of the human interactome, Cell 162 (2) (2015) 425–440.
[116] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Resilience of the internet to random breakdowns, Phys. Rev. Lett. 85 (2000) 4626–4628.
[117] X. He, J. Zhang, Why do hubs tend to be essential in protein networks? PLoS Genet. 2 (6) (2006) e88.
[118] M.W. Hahn, A.D. Kern, Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks, Mol. Biol. Evol. 22 (4) (2014) 803–806.

[119] F. Cheng, P. Jia, Q. Wang, L. Chen-Ching, W.-H. Li, Z. Zhao, Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome, Mol. Biol. Evol. 31 (8) (2014) 2156–2169.
[120] M.P. van den Heuvel, O. Sporns, Network hubs in the human brain, Trends Cogn. Sci. 17 (12) (2013) 683–696.
[121] D. Tomasi, G.-J. Wang, N.D. Volkow, Energetic cost of brain functional connectivity, Proc. Natl. Acad. Sci. USA 110 (33) (2013) 13642–13647.
[122] J. Kim, P.L. Krapivsky, B. Kahng, S. Redner, Infinite-order percolation and giant fluctuations in a protein interaction network, Phys. Rev. E 66 (2) (2002) 055101.
[123] R. Pastor-Satorras, E. Smith, R. Sole, Evolving protein interaction networks through gene duplication, J. Theoret. Biol. 222 (2003) 199–201.
[124] S. Konini, E.J. Janse van Rensburg, Mean field analysis of algorithms for scale-free networks in molecular biology, PLoS One 12 (12) (2017) e0189866.
[125] I. Ispolatov, P.L. Krapivsky, A. Yuryev, Duplication-divergence model of protein interaction network, Phys. Rev. E 71 (6) (2005) 061911.
[126] K. Evlampiev, H. Isambert, Conservation and topology of protein interaction networks under duplication-divergence evolution, Proc. Natl. Acad. Sci. USA 105 (29) (2008) 9863–9868.
[127] S. Cai, Z. Liu, H.C. Lee, Mean field theory for biology inspired duplication-divergence network model, Chaos 25 (8) (2015) 083106.
[128] K. Takemoto, C. Oosawa, Modeling for evolving biological networks with scale-free connectivity, hierarchical modularity, and disassortativity, Math. Biosci. 208 (2) (2007) 454–468.
[129] W.K. Kim, E.M. Marcotte, Age-dependent evolution of the yeast protein interaction network suggests a limited role of gene duplication and divergence, PLoS Comput. Biol. 4 (11) (2008) e1000232.
[130] R. Albert, Scale-free networks in cell biology, J. Cell Sci. 118 (21) (2005) 4947–4957.
[131] S. Neph, A. Stergachis, A. Reynolds, R. Sandstrom, E. Borenstein, J. Stamatoyannopoulos, Circuitry and dynamics of human transcription factor regulatory networks, Cell 150 (6) (2012) 1274–1286.
[132] N. Ichinose, T. Yada, H. Wada, Asymmetry in indegree and outdegree distributions of gene regulatory networks arising from dynamical robustness, Phys. Rev. E 97 (6) (2018) 062315.
[133] Z. Bruda, A. Krzywicki, O.C. Martin, M. Zagorski, Distribution of essential interactions in model gene regulatory networks under mutation-selection balance, Phys. Rev. E 82 (1) (2010) 011908.
[134] C. Liu, X.-X. Zhan, Z.-K. Zhang, G.-Q. Sun, P.M. Hui, How events determine spreading patterns: information transmission via internal and external influences on social networks, New J. Phys. 7 (2015) 113045.
[135] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics, Phys. Rep. 424 (2006) 175–308.
[136] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang, Y.-C. Zhang, Dynamics of information diffusion and its applications on complex networks, Phys. Rep. 651 (2016) 1–34.
[137] D. Blokh, D. Segev, R. Sharan, The approximability of shortest path-based graph orientations of protein-protein interaction networks, J. Comput. Biol. 20 (12) (2013) 945–957.
[138] D. Silverbush, R. Sharan, Network orientation via shortest paths, Bioinformatics 30 (10) (2014) 1449–1455.
[139] Y.-K. Shih, S. Parthasarathy, A single source k-shortest paths algorithm to infer regulatory pathways in a gene network, Bioinformatics 28 (12) (2012) 49–58.
[140] B.-Q. Li, T. Huang, L. Liu, Y.-D. Cai, K.-C. Chou, Identification of colorectal cancer related genes with mRMR and shortest path in protein-protein interaction network, PLoS One 7 (4) (2012) e33393.
[141] Y. Ren, A. Ay, T. Kahveci, Shortest path counting in probabilistic biological networks, BMC Bioinformatics 19 (2018) 465.
[142] E. Guney, J. Menche, M. Vidal, A.-L. Barabási, Network-based in silico drug efficacy screening, Nature Commun. 7 (2016) 10331.
[143] F. Cheng, R.J. Desai, D.E. Handy, R. Wang, S. Schneeweiss, A.-L. Barabási, J. Loscalzo, Network-based approach to prediction and population-based validation of in silico drug repurposing, Nature Commun. 9 (2018) 2691.
[144] V. Latora, M. Marchiori, Efficient behavior of small-world networks, Phys. Rev. Lett. 87 (19) (2001) 198701.
[145] P. Csermely, V. Ágoston, S. Pongor, The efficiency of multi-target drugs: the network approach might help drug design, Trends Pharmacol. Sci. 26 (4) (2005) 178–182.
[146] A. Vazquez, Optimal drug combinations and minimal hitting sets, BMC Syst. Biol. 3 (2009) 81.
[147] S. Achard, E.T. Bullmore, Efficiency and cost of economical brain functional networks, PLoS Comput. Biol. 3 (2) (2007) e17.
[148] E.T. Bullmore, O. Sporns, The economy of brain network organization, Nat. Rev. Neurosci. 13 (5) (2012) 336–349.
[149] A.J. Lawrence, A.W. Chung, R.G. Morris, H.S. Markus, T.R. Barrick, Structural network efficiency is associated with cognitive impairment in small-vessel disease, Neurology 83 (4) (2014) 304–311.
[150] D.-J. Kim, J.H. Yu, M.-S. Shin, Y.-W. Shin, M.-S. Kim, Hyperglycemia reduces efficiency of brain networks in subjects with Type 2 Diabetes, PLoS One 11 (6) (2016) e0157268.
[151] J.D. Medaglia, W. Huang, S. Segarra, C. Olm, J. Gee, M. Grossman, A. Ribeiro, C.T. McMillan, D.S. Bassett, Brain network efficiency is influenced by the pathologic source of corticobasal syndrome, Neurology 89 (13) (2017) 1373–1381.
[152] A. Roy, R.A. Bernier, J. Wang, M. Benson, J.J. French, D.C. Good, F.G. Hillary, The evolution of cost-efficiency in neural networks during recovery from traumatic brain injury, PLoS One 12 (4) (2017) e0170541.
[153] Y. Li, Y. Liu, J. Li, W. Qin, K. Li, C. Yu, T. Jiang, Brain anatomical network and intelligence, PLoS Comput. Biol. 5 (5) (2009) e1000395.
[154] L.C.A. Freeman, A set of measures of centrality based on betweenness, Sociometry 40 (1977) 35–41.
[155] G. Szabó, M. Alava, J. Kertész, Shortest paths and load scaling in scale-free trees, Phys. Rev. E 66 (2) (2002) 026101.
[156] A. Kirkley, H. Barbosa, M. Barthelemy, G. Ghoshal, From the betweenness centrality in street networks to structural invariants in random planar graphs, Nature Commun. 9 (2018) 2501.
[157] K.I. Goh, B. Kahng, D. Kim, Universal behavior of load distribution in scale-free networks, Phys. Rev. Lett. 87 (27) (2001) 278701.
[158] R. Guimera, S. Mossa, A. Turtschi, L.A.N. Amaral, The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles, Proc. Natl. Acad. Sci. USA 102 (22) (2005) 7794–7799.
[159] M.P. Joy, A. Brock, D.E. Ingber, S. Huang, High-betweenness proteins in the yeast protein interaction network, J. Biomed. Biotechnol. 2005 (2) (2005) 96–103.
[160] H. Yu, P.M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics, PLoS Comput. Biol. 3 (4) (2007) e59.
[161] L. Zou, S. Sriswasdi, B. Ross, P.V. Missiuro, J. Liu, H. Ge, Systematic analysis of pleiotropy in C.elegans early embryogenesis, PLoS Comput. Biol. 4 (2) (2008) e1000003.
[162] H. Ahmed, T.C. Howton, Y. Sun, N. Weinberger, Y. Belkhadir, M.S. Mukhtar, Network biology discovers pathogen contact points in host protein-protein interactomes, Nature Commun. 9 (2018) 2312.
[163] E. Estrada, G.J. Ross, Centralities in simplicial complexes. Applications to protein interaction networks, J. Theor. Biol. 438 (2018) 46–60.
[164] A. Al-Aamri, K. Taha, Y. Al-Hammadi, M. Maalouf, D. Homouz, Analyzing a co-occurrence gene-interaction network to identify disease-gene association, BMC Bioinformatics 20 (2019) 70.
[165] E. Pang, Y. Hao, Y. Sun, K. Lin, Differential variation patterns between hubs and bottlenecks in human protein-protein interaction networks, BMC Evol. Biol. 16 (2016) 260.

[166] H. Yu, M. Gerstein, Genomic analysis of the hierarchical structure of regulatory networks, Proc. Natl. Acad. Sci. USA 103 (40) (2006) 14724–14731.
[167] K.W. Thee, H. Nisar, C.S. Soh, Graph theoretical analysis of functional brain networks in healthy subjects: visual oddball paradigm, IEEE Access 6 (2018) 64708–64727.
[168] V.V. Makarov, M.O. Zhuravlev, A.E. Runnova, P. Protasov, V.A. Maksimenko, N.S. Frolov, A.N. Pisarchik, A.E. Hramov, Betweenness centrality in multiplex brain network during mental task evaluation, Phys. Rev. E 98 (6) (2018) 062413.
[169] I. Ueda, S. Kakeda, K. Watanabe, K. Sugimoto, N. Igata, J. Moriya, K. Takemoto, A. Katsuki, R. Yoshimura, O. Abe, Y. Korogi, Brain structural connectivity and neuroticism in healthy adults, Sci. Rep. 8 (2018) 16491.
[170] M.L. Garcia-Vaquero, M. Gama-Carvalho, J. De Las Rivas, F.R. Pinto, Searching the overlap between network modules with specific betweeness (S2B) and its application to cross-disease analysis, Sci. Rep. 8 (2018) 11555.
[171] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99 (12) (2002) 7821–7826.
[172] J. Yoon, A. Blumer, K. Lee, An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality, Bioinformatics 22 (24) (2006) 3106–3108.
[173] R. Dunn, F. Dudbridge, C.M. Sanderson, The use of edge-betweenness clustering to investigate biological function in protein interaction networks, BMC Bioinformatics 6 (2005) 39.
[174] M. Boguna, R. Pastor-Satorras, A. Diaz-Guilera, A. Arenas, Models of social networks based on social distance attachment, Phys. Rev. E 70 (5) (2004) 056122.
[175] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[176] S. Wuchty, Scale-free behavior in protein domain networks, Mol. Biol. Evol. 18 (9) (2001) 1694–1702.
[177] N. Zaki, D. Efimov, J. Berengueres, Protein complex detection using interaction reliability assessment and weighted clustering coefficient, BMC Bioinformatics 14 (2013) 163.
[178] D. Hao, C. Ren, C. Li, Revisiting the variation of clustering coefficient of biological networks suggests new modular structure, BMC Syst. Biol. 6 (2012) 34.
[179] D. Fraiman, G. Saunier, E.F. Martins, C.D. Vargas, Biological motion coding in the brain: analysis of visually driven EEG functional networks, PLoS One 9 (1) (2014) e84612.
[180] P. Du, L. Wang, Predicting human protein subcellular locations by the ensemble of multiple predictors via protein-protein interaction network with edge clustering coefficients, PLoS One 9 (1) (2014) e86879.
[181] C.-Y. Ma, Y.-P.P. Chen, B. Berger, C.-S. Liao, Identification of protein complexes by integrating multiple alignment of protein interaction networks, Bioinformatics 33 (11) (2017) 1681–1688.
[182] S. Milgram, The small world problem, Psychol. Today 2 (1967) 60–67.
[183] D. Fell, A. Wagner, The small world of of metabolism, Nat. Biotechnol. 18 (11) (2000) 1121–1122.
[184] M. Karsai, M. Kivelä, R.K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, J. Saramäki, Small but slow world: how network topology and burstiness slow down spreading, Phys. Rev. E 83 (2) (2011) 025102(R).
[185] L. Cohen, A. Frazzini, C. Molloy, The small world of investing: board connections and mutual fund returns, J. Polit. Econ. 116 (5) (2008) 951–979.
[186] L.A. Amaral, A. Scala, M. Barthelemy, H.E. Stanley, Classes of small-world networks, Proc. Natl. Acad. Sci. USA 97 (21) (2000) 11149–11152.
[187] P. Erdös, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17–61.
[188] M. Novkovic, L. Onder, J. Cupovic, J. Abe, D. Bomze, V. Cremasco, E. Scandella, J.V. Stein, G. Bocharov, S.J. Turley, B. Ludewig, Topological small-world organization of the Fibroblastic Reticular Cell network determines lymph node functionality, PLoS Biol. 14 (7) (2016) e1002515.
[189] D.S. Bassett, E.T. Bullmore, Small-world brain networks, Neuroscientist 12 (6) (2006) 512–523.
[190] D.S. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, E.T. Bullmore, Adaptive reconfiguration of fractal small-world human brain functional networks, Proc. Natl. Acad. Sci. USA 103 (51) (2006) 19518–19523.
[191] M. Rubinov, R.J. Ypma, C. Watson, E.T. Bullmore, Wiring cost and topological participation of the mouse brain connectome, Proc. Natl. Acad. Sci. USA 112 (32) (2015) 10032–10037.
[192] J.H. Downes, M.W. Hammond, D. Xydas, M.C. Spencer, V.M. Becerra, K. Warwick, B.J. Whalley, S.J. Nasuto, Emergence of a small-world functional network in cultured neurons, PLoS Comput. Biol. 8 (5) (2012) e1002522.
[193] S. Malmersjo, et al., Neural progenitors organize in small-world networks to promote cell proliferation, Proc. Natl. Acad. Sci. USA 110 (16) (2013) 1524–1532.
[194] L.K. Gallos, H.A. Makse, M. Sigman, A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks, Proc. Natl. Acad. Sci. USA 109 (8) (2012) 2825–2830.
[195] X. Delbeuck, M. Van der Linden, F. Collette, Alzheimer' disease as a disconnection syndrome, Neuropsychol. Rev. 13 (2) (2003) 79–92.
[196] C.J. Stam, B.F. Jones, G. Nolte, M. Breakspear, P. Scheltens, Small-world networks and functional connectivity in Alzheimer's disease, Cerebral Cortex 17 (1) (2007) 92–99.
[197] E.J. Sanz-Arigita, M.M. Schoonheim, J.S. Damoiseaux, S.A.R.B. Rombouts, E. Maris, F. Barkhof, P. Scheltens, C.J. Stam, Loss of 'small-world' networks in Alzheimer's disease: graph analysis of fMRI resting-state functional connectivity, PLoS One 5 (11) (2010) e13788.
[198] F. Vecchio, F. Miraglia, D. Quaranta, G. Granata, R. Romanello, C. Marra, P. Bramanti, P.M. Rossini, Cortical connectivity and memory performance in cognitive decline: a study via graph theory from EEG data, Neuroscience 316 (2016) 143–150.
[199] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (2009) 75–174.
[200] Y. Dourisboure, F. Geraci, M. Pellegrini, Extration and classification of dense communities in the web, in: Proceedings of the 16th ACM International Conference on World Wide Web, ACM Press, New York, 2007, pp. 461–470.
[201] S.D. Ghiassian, J. Menche, A.-L. Barabási, A disease module detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome, PLoS Comput. Biol. 11 (4) (2015) e1004120.
[202] A. Rives, T. Galitski, Modular organization of cellular networks, Proc. Natl. Acad. Sci. USA 100 (3) (2003) 1128–1133.
[203] J. Chen, B. Yuan, Detecting functional modules in the yeast protein-protein interaction network, Bioinformatics 22 (18) (2006) 2283–2290.
[204] T. Nepusz, H. Yu, A. Paccanaro, Detecting overlapping protein complexes in protein-protein interaction networks, Nature Methods 9 (5) (2012) 471–472.
[205] J. Lee, S.P. Gross, J. Lee, Improved network community structure improves function prediction, Sci. Rep. 3 (2013) 2197.
[206] A.J. Enright, S. Van Dongen, C.A. Ouzounis, An efficient algorithm for large-scale detection of protein families, Nucleic Acids Res. 30 (7) (2002) 1575–1584.
[207] H.K. Norton, D.J. Emerson, H. Huang, J. Kim, K.R. Titus, S. Gu, D.S. Bassett, J.E. Phillips-Cremins, Detecting hierarchical genome folding with network modularity, Nature Methods 15 (2) (2018) 119–122.
[208] M. Brehme, C. Voisine, T. Rolland, S. Wachi, J.H. Soper, Y. Zhu, K. Orton, A. Villella, D. Garza, M. Vidal, H. Ge, R.I. Morimoto, A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease, Cell Rep. 9 (3) (2014) 1135–1150.
[209] R. Guimera, L.A.N. Amaral, Functional cartography of complex metabolic networks, Nature 433 (7028) (2005) 895–900.

[210] N.A. Crossley, A. Mechelli, P.E. Vertes, T.T. Winton-Brown, A.X. Patel, C.E. Ginestet, P. McGuire, E.T. Bullmore, Cognitive relevance of the community structure of the human brain functional coactivation network, Proc. Natl. Acad. Sci. USA 110 (28) (2013) 11583–11588.
[211] C.J. Stam, Modern network science of neurological disorders, Nat. Rev. Neurosci. 15 (10) (2014) 683–695.
[212] M.A. Bertolero, B.T. Yeo, M. D'Esposito, The modular and integrative functional architecture of the human brain, Proc. Natl. Acad. Sci. USA 112 (49) (2015) 6798–6807.
[213] M. Chen, M.W. Deem, Development of modularity in the neural activity of children's brains, Phys. Biol. 12 (1) (2015) 016009.
[214] M. Shein-Idelson, G. Cohen, E. Ben-Jacob, Y. Hanein, Modularity induced gating and delays in neuronal networks, PLoS Comput. Biol. 12 (4) (2016) e1004883.
[215] S. Choobdar, et al., Assessment of network module identification across complex diseases, Nature Methods 16 (9) (2019) 843–852.
[216] C. Zhou, L. Zemanova, G. Zamora, C.C. Hilgetag, J. Kurths, Hierarchical organization unveiled by functional connectivity in complex brain networks, Phys. Rev. Lett. 97 (23) (2006) 238103.
[217] J. Dutkowski, M. Kramer, M.A. Surma, R. Balakrishnan, J.M. Cherry, N.J. Krogan, T. Ideker, A gene ontology inferred from molecular networks, Nat. Biotechnol. 31 (1) (2013) 38–45.
[218] M. Sales-Pardo, R. Guimerà, A.A. Moreira, L.A.N. Amaral, Extracting the hierarchical organization of complex systems, Proc. Natl. Acad. Sci. USA 104 (39) (2007) 15224–15229.
[219] H. Wei, Construction of a hierarchical gene regulatory network centered around a transcription factor, Brief. Bioinform. 20 (3) (2019) 1021–1031.
[220] Q.-J. Jiao, Y. Huang, H.-B. Shen, A new multi-scale method to reveal hierarchical modular structures in biological networks, Mol. Biosyst. 12 (12) (2016) 3724–3733.
[221] J. Xu, R. Jing, Y. Liu, Y. Dong, Z. Wen, M. Li, A new strategy for exploring the hierarchical structure of cancers by adaptively partitioning functional modules from gene expression network, Sci. Rep. 6 (2016) 28720.
[222] M.A. Reyna, M.D.M. Leiserson, B.J. Raphael, Hierarchical HotNet: identifying hierarchies of altered subnetworks, Bioinformatics 34 (17) (2018) 972–980.
[223] N. Lahav, B. ksherim, E. Ben-Simon, A. Maron-Katz, R. Cohen, S. Havlin, K-shell decomposition reveals hierarchical cortical organization of the human brain, New J. Phys. 18 (2016) 083013.
[224] R. Wang, P. Lin, M. Liu, Y. Wu, T. Zhou, C. Zhou, Hierarchical connectome modes and critical state jointly maximize human brain functional diversity, Phys. Rev. Lett. 123 (3) (2019) 038301.
[225] A. Ashourvan, Q.K. Telesford, T. Verstynen, J.M. Vettel, D.S. Bassett, Multi-scale detection of hierarchical community architecture in structural and functional brain networks, PLoS One 14 (5) (2019) e0215520.
[226] P. Sollich, D. Tantari, A. Annibale, A. Barra, Extensive parallel processing on scale-free networks, Phys. Rev. Lett. 113 (23) (2014) 238106.
[227] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, Retrieval capabilities of hierarchical neworks: from dyson to hopfield, Phys. Rev. Lett. 114 (2) (2015) 028103.
[228] S. Fortunato, D. Hric, Community detection in networks: A user guide, Phys. Rep. 659 (2016) 1–44.
[229] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2) (2004) 026113.
[230] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (3) (2006) 036104.
[231] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (6) (2004) 066111.
[232] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[233] G. Palla, A.-L. Barabási, T. Vicsek, Quantifying social group evolution, Nature 446 (7136) (2007) 664–667.
[234] Y.-Y. Ahn, J.P. Bagrwo, S. Lehmann, Link communities reveal multiscale complexity in networks, Nature 466 (7307) (2010) 761–764.
[235] M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. USA 105 (4) (2008) 1118–1123.
[236] M. Rosvall, C.T. Bergstrom, An information-theoretic framework for resolving community structure in complex networks, Proc. Natl. Acad. Sci. USA 104 (18) (2007) 7327–7331.
[237] L.H. Hartwell, J.J. Hopfield, S. Leibler, A.W. Murray, From molecular to modular cell biology, Nature 402 (1999) 47–52.
[238] J.M.J. Derry, L.M. Mangravite, C. Suver, M.D. Furia, D. Henderson, X. Schildwachter, B. Bot, J. Izant, S.K. Sieberts, M.R. Kellen, S.H. Friend, Developing predictive molecular maps of human disease through community-based modeling, Nat. Genet. 44 (2) (2012) 127–130.
[239] K. Mitra, A.-R. Carvunis, S.K. Ramesh, T. Ideker, Integrative approaches for finding modular structure in biological networks, Nat. Rev. Genet. 14 (10) (2013) 719–732.
[240] R.-S. Wang, J. Loscalzo, Network-based disease module discovery by a novel seed connector algorithm with pathobiological implications, J. Mol. Biol. 430 (18) (2018) 2939–2950.
[241] T. Ideker, O. Ozier, B. Schwikowski, A.F. Siegel, Discovering regulatory and signalling circuits in molecular interaction networks, Bioinformatics 18 (s1) (2002) 233–240.
[242] P. Jia, S. Zheng, J. Long, W. Zheng, Z. Zhao, dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks, Bioinformatics 27 (1) (2011) 95–102.
[243] F. Hormozdiari, O. Penn, E. Borenstein, E.E. Eichler, The discovery of integrated gene networks for autism and related disorders, Genome Res. 25 (1) (2015) 142–154.
[244] A. Gouy, J.T. Daub, L. Excoffier, Detecting gene subnetworks under selection in biological pathways, Nucleic Acids Res. 45 (16) (2017) e149.
[245] Q. Wang, H. Yu, Z. Zhao, P. Jia, EW_dmGWAS: edge-weighted dense module search for genome-wide association studies and gene expression profiles, Bioinformatics 31 (15) (2015) 2591–2594.
[246] C. Fuchsberger, et al., The genetic architecture of type 2 diabetes, Nature 536 (7614) (2016) 41–47.
[247] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.
[248] E.R. Shellman, C.F. Burant, S. Schnell, Network motifs provide signatures that characterize metabolism, Mol. Biosyst. 9 (3) (2013) 352–360.
[249] T. Hocevar, J. Demsar, A combinatorial approach to graphlet counting, Bioinformatics 30 (4) (2014) 559–565.
[250] M. Agrawal, M. Zitnik, J. Leskovec, Large-scale analysis of disease pathways in the human interactome, Pac. Symp. Biocomput. 23 (2018) 111–122.
[251] S.S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of Escherichia coli., Nat. Genet. 31 (1) (2002) 64–68.
[252] U. Alon, Network motifs: theory and experimental approaches, Nat. Rev. Genet. 8 (6) (2007) 450–461.
[253] N. Przulj, Biological network comparison using graphlet degree distribution, Bioinformatics 23 (2) (2007) E177–E183.
[254] O.C. Martin, A. Krzywicki, M. Zagorski, Drivers of structural features in gene regulatory networks: From biophysical constraints to biological function, Phys. Life Rev. 17 (2016) 124–158.
[255] R.J. Prill, P.A. Iglesias, A. Levchenko, Dynamic properties of network motifs contribute to biological network organization, PLoS Biol. 3 (11) (2005) e343.

[256] T. Ruths, L. Nakhleh, Neutral forces acting on intragenomic variability shape the Escherichia coli regulatory network topology, Proc. Natl. Acad. Sci. USA 110 (19) (2013) 7754–7759.
[257] N. Pearcy, J.J. Crofts, N. Chuzhanova, Network motif frequency vectors reveal evolving metabolic network organisation, Mol. BioSyst. 11 (1) (2015) 77–85.
[258] D.P. Giling, A. Ebeling, N. Eisenhauer, S.T. Meyer, C. Roscher, M. Rzanny, W. Voigt, W.W. Weisser, J. Hines, Plant diversity alters the representation of motifs in food webs, Nature Commun. 10 (2019) 1226.
[259] O. Sporns, R. Kötter, Motifs in brain networks, PLoS Biol. 2 (11) (2004) e369.
[260] F. Battiston, V. Nicosia, M. Chavez, V. Latora, Multilayer motif analysis of brain networks, Chaos 27 (4) (2017) 047404.
[261] L. Stone, D. Simberloff, Y. Artzy-Randrup, Network motifs and their origins, PLoS Comput. Biol. 15 (4) (2019) e1006749.
[262] M. Ansariola, M. Megraw, D. Koslicki, IndeCut evaluates performance of network motif discovery algorithms, Bioinformatics 34 (9) (2018) 1514–1521.
[263] E. Wong, B. Baur, S. Quader, C.-H. Huang, Biological network motif detection: principles and practice, Brief. Bioinform. 13 (2) (2011) 202–215.
[264] N.H. Tran, K.P. Choi, L. Zhang, Counting motifs in the human interactome, Nature Commun. 4 (2013) 2241.
[265] M.B.Z. Joveini, J. Sadri, Application of fractal theory on motifs counting in biological networks, IEEE ACM Trans. Comput. Biol. Bioinform. 15 (2) (2018) 613–623.
[266] P. Wang, J. Lü, X. Yu, Identification of important nodes in directed biological networks: a network motif approach, PLoS One 9 (8) (2014) e106132.
[267] F. Wu, C. Ma, C. Tan, Network motifs modulate druggability of cellular targets, Sci. Rep. 6 (2016) 36626.
[268] L. Chen, X. Qu, M. Cao, Y. Zhou, W. Li, B. Liang, W. Li, W. He, C. Feng, X. Jia, Y. He, Identification of breast cancer patients based on human signaling network motifs, Sci. Rep. 3 (2013) 3368.
[269] K. Anand, G. Bianconi, Entropy measures for networks: toward an information theory of complex topologies, Phys. Rev. E 80 (4) (2009) 045102.
[270] N. Eagle, M. Macy, R. Claxton, Network diversity and economic development, Science 328 (5981) (2010) 1029–1031.
[271] A.E. Teschendorff, S. Severini, Increased entropy of signal transduction in the cancer metastasis phenotype, BMC Syst. Biol. 4 (2010) 104.
[272] J. West, G. Bianconi, S. Severini, A.E. Teschendorff, Differential network entropy reveals cancer system hallmarks, Sci. Rep. 2 (2012) 802.
[273] C.R. Banerji, D. Miranda-Saavedra, S. Severini, M. Widschwendter, T. Enver, J.X. Zhou, A.E. Teschendorff, Cellular network entropy as the energy potential in Waddington's differentiation landscape, Sci. Rep. 3 (2013) 3039.
[274] A. Viol, F. Palhano-Fontes, H. Onias, D.B. de Araujo, G.M. Viswanathan, Shannon entropy of brain functional complex networks under the influence of the psychedelic Ayahuasca, Sci. Rep. 7 (2017) 7388.
[275] F. Cheng, C. Liu, B. Shen, Z. Zhao, Investigating cellular network heterogeneity and modularity in cancer: a network entropy and unbalanced motif approach, BMC Syst. Biol. 10 (s3) (2016) 65.
[276] C.R. Banerji, S. Severini, C. Caldas, A.E. Teschendorff, Intra-tumour signalling entropy determines clinical outcome in breast and lung cancer, PLoS Comput. Biol. 11 (3) (2015) e1004115.
[277] A. Juarez-Flores, M.V. José, Multivariate entropy characterizes the gene expression and protein-protein networks in four types of cancers, Entropy 8 (3) (2018) 154.
[278] A.E. Teschendorff, C.R.S. Banerji, S. Severini, R. Kuehn, P. Sollich, Increased signaling entropy in cancer requires the scale-free property of protein interaction networks, Sci. Rep. 5 (2015) 9646.
[279] A.E. Teschendorff, T. Enver, Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome, Nature Commun. 8 (2017) 15599.
[280] Y. Jia, H. Gu, Q. Luo, Sample entropy reveals an age-related reduction in complexity of dynamic brain, Sci. Rep. 7 (2017) 7990.
[281] R.L. Carhart-Harris, R. Leech, P.J. Hellyer, M. Shanahan, A. Feilding, E. Tagliazucchi, D.R. Chialvo, D. Nutt, The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs, Front. Hum. Neurosci. 8 (2014) 20.
[282] A.V. Lebedev, M. Kaelen, M. Lovden, J. Nilsson, A. Feilding, D.J. Nutt, R.L. Carhart-Harris, LSD-induced entropic brain activity predicts subsequent personality change, Hum. Brain Mapp. 37 (9) (2016) 3203–3213.
[283] G.N. Saxe, D. Galderone, L.J. Morales, Brain entropy and human intelligence: a resting-state fMRI study, PLoS One 13 (2) (2018) e0191582.
[284] N. Wang, H. Wu, M. Xu, Y. Yang, C. Chang, W. Zeng, H. Yan, Occupational functional plasticity revealed by brain entropy: a resting-state fMRI study of seafarers, Hum. Brain Mapp. 39 (7) (2018) 2997–3004.
[285] B. Sen, S.-H. Chu, K.K. Parhi, Ranking regions, edges and classifying tasks in functional brain graphs by sub-graph entropy, Sci. Rep. 9 (2019) 7628.
[286] C. Zhang, F. Cong, T. Kujala, W. Liu, J. Liu, T. Parviainen, T. Ristaniemi, Network entropy for the sequence analysis of functional connectivity graphs of the brain, Entropy 20 (5) (2018) 311.
[287] U.R. Acharya, H. Fujita, V.K. Sudarshan, S. Bhat, J.E.W. Koh, Application of entropies for automated diagnosis of epilepsy using EEG signals: A review, Knowl-Based Syst. 88 (2015) 85–96.
[288] C. Song, S. Havlin, H.A. Makse, Self-similarity of complex networks, Nature 433 (7024) (2005) 392–395.
[289] C. Song, S. Havlin, H.A. Makse, Origins of fractality in the growth of complex networks, Nat. Phys. 2 (4) (2006) 275–281.
[290] K.-I. Goh, G. Salvi, B. Kahng, D. Kim, Skeleton and fractal scaling in complex networks, Phys. Rev. Lett. 96 (1) (2006) 018701.
[291] L. Wang, Q. Wang, L. Xi, J. Chen, S. Wang, L. Bao, Z. Yu, L. Zao, On the fractality of complex network: covering problem, algorithms and Ahlfors regularity, Sci. Rep. 7 (2017) 41385.
[292] W.-X. Zhou, Z.-Q. Jiang, D. Sornette, Exploring self-similarity of complex cellular networks: the edge-covering method with simulated annealing and log-periodic sampling, Physica A 375 (2) (2007) 741–752.
[293] D.-J. Wei, Q. Liu, H.-X. Zhang, Y. Hu, Y. Deng, S. Mahadevan, Box-covering algorithm for fractal dimension of weighted networks, Sci. Rep. 3 (2013) 3049.
[294] Y. Jin, D. Turaev, T. Weinmaier, T. Rattei, H.A. Makse, The evolutionary dynamics of protein-protein interaction networks inferred from the reconstruction of ancient network, PLoS One 8 (3) (2013) e58134.
[295] V. Galvão, J.G.V. Miranda, R.F.S. Andrade, J.S. Andrade, L.K. Gallos, H.A. Makse, Modularity map of the network of human cell differentiation, Proc. Natl. Acad. Sci. USA 107 (13) (2010) 5750–5755.
[296] F. Klimm, D.S. Bassett, J.M. Carlson, P.J. Mucha, Resolving structural variability in network models and the brain, PLoS Comput. Biol. 10 (3) (2014) e1003491.
[297] T.M. Reese, A. Brzoska, D.T. Yott, D.J. Kelleher, Analyzing self-similar and fractal properties of the C.elegans neural network, PLoS One 7 (10) (2012) e40483.
[298] P. Moretti, M.A. Muñoz, Griffiths phases and the stretching of criticality in brain networks, Nature Commun. 4 (2013) 2521.
[299] L.K. Gallos, M. Sigman, H.A. Makse, The conundrum of functional brain networks: small-world efficiency or fractal modularity, Front. Physiol. 3 (2012) 123.
[300] H.D. Rozenfeld, C. Song, H.A. Makse, Small-world to fractal transition in complex networks: a renormalization group approach, Phys. Rev. Lett. 104 (2) (2010) 025701.

[301] S.S. Singh, D. Haobijam, M.Z. Malik, R. Ishrat, R.K.B. Singh, Fractal rules in brain networks: signatures of self-organization, J. Theoret. Biol. 437 (2018) 58–66.
[302] S.S. Singh, B. Khundrakpam, A.T. Reid, J.D. Lewis, A.C. Evans, R. Ishrat, B.I. Sharma, R.K.B. Singh, Scaling in topological properties of brain networks, Sci. Rep. 6 (2016) 24926.
[303] A.J.E. Seely, K.D. Newman, C.L. Herry, Fractal structure and entropy production within the central nervous system, Entropy 16 (8) (2014) 4497–4520.
[304] A. Safari, P. Moretti, M.A. Muñoz, Topological dimension tunes activity patterns in hierarchical modular networks, New J. Phys. 19 (2017) 113011.
[305] Y. Yu, L. Yang, Z. Liu, C. Zhu, Gene essentiality prediction based on fractal features and machine learning, Mol. BioSyst. 13 (3) (2017) 577–584.
[306] E. Fadhal, J. Gamieldien, E.C. Mwambene, Self-similarity of human protein interaction networks: a novel strategy of distinguishing proteins, Sci. Rep. 5 (2015) 7628.
[307] K. Takemoto, Metabolic networks are almost nonfractal: a comprehensive evaluation, Phys. Rev. E 90 (2) (2014) 022802.
[308] J. Gao, B. Barzel, A.-L. Barabási, Universal resilience patterns in complex networks, Nature 530 (7590) (2016) 307–312.
[309] G. Dong, J. Fan, L.M. Shekhtman, S. Shai, R. Du, L. Tian, X. Chen, H.E. Stanley, S. Havlin, Resilience of networks with community structure behaves as if under an external field, Proc. Natl. Acad. Sci. USA 115 (27) (2018) 6911–6915.
[310] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, Nature 406 (6794) (2000) 378–382.
[311] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Breakdown of the internet under intentional attack, Phys. Rev. Lett. 86 (16) (2001) 3682–3685.
[312] L.J. Gilarranz, B. Rayfield, G. Linan-Cembrano, J. Bascompte, A. Gonzalez, Effects of network modularity on the spread of perturbation impact in experimental metapopulations, Science 357 (6347) (2017) 199–201.
[313] C.N. Kaiser-Bunbury, J. Mougal, A.E. Whittington, T. Valentin, R. Gabriel, J.M. Olesen, N. Bluthgen, Ecosystem restoration strengthens pollination networks resilience and function, Nature 542 (7640) (2017) 223–227.
[314] R.D. Batt, S.R. Carpenter, J.J. Cole, M.L. Pace, R.A. Johnson, Changes in ecosystem resilience detected in automated measures of ecosystem metabolism during a whole-lake manipulation, Proc. Natl. Acad. Sci. USA 110 (43) (2013) 17398–17403.
[315] L. Zhang, G. Zeng, D. Li, H.-J. Huang, H.E. Stanley, S. Havlin, Scale-free resilience of real traffic jams, Proc. Natl. Acad. Sci. USA 116 (18) (2019) 8673–8678.
[316] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, D. Garlaschelli, Reconstruction methods for networks: the case of economic and financial systems, Phys. Rep. 757 (2018) 1–47.
[317] S.D.S. Reis, Y. Hu, A. Babino, J.S. Andrade, S. Canals, M. Sigman, H.A. Makse, Avoiding catastrophic failure in correlated networks of networks, Nat. Phys. 10 (10) (2014) 762–767.
[318] W. Li, Y. Li, Y. Tan, Y. Cao, C. Chen, Y. Cai, K.Y. Lee, M. Pecht, Maximizing network resilience against malicious attacks, Sci. Rep. 9 (2019) 2261.
[319] K.E. Joyce, S. Hayasaka, P.J. Laurienti, The human functional brain network demonstrates structural and dynamical resilience to targeted attack, PLoS Comput. Biol. 9 (1) (2013) e1002885.
[320] N. Sahni, et al., Widespread macromolecular interaction perturbations in human genetic disorders, Cell 161 (3) (2015) 647–660.
[321] N. Lemke, F. Heredia, C.K. Barcellos, A.N. Dos Reis, J.C.M. Mombach, Essentiality and damage in metabolic networks, Bioinformatics 20 (1) (2004) 115–119.
[322] A.G. Smart, L.A.N. Amaral, J.M. Ottino, Cascading failure and robustness in metabolic networks, Proc. Natl. Acad. Sci. USA 105 (36) (2008) 13223–13228.
[323] K. Takemoto, T. Tamura, T. Akutsu, Theoretical estimation of metabolic network robustness against multiple reaction knockouts using branching process approximation, Physica A 392 (21) (2013) 5525–5535.
[324] S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks, Science 296 (5569) (2002) 910–913.
[325] M. Zitnik, R. Sosic, M.W. Feldman, J. Leskovec, Evolution of resilience in protein interactomes across the tree of life, Proc. Natl. Acad. Sci. USA 116 (10) (2019) 4426–4433.
[326] F.A. Rodrigues, L.F. Costa, A.L. Barbieri, Resilience of protein-protein interaction networks as determined by their large-scale topological features, Mol. BioSyst. 7 (4) (2011) 1263–1269.
[327] H. Aerts, W. Fias, K. Caeyenberghs, D. Marinazzo, Brain networks under attack: robustness properties and the impact of lesions, Brain 139 (2016) 3063–3083.
[328] W. de Haan, K. Mott, E.C.W. van Straaten, P. Scheltens, C.J. Stam, Activity dependent degeneration explains hub vulnerability in Alzheimer's disease, PLoS Comput. Biol. 8 (8) (2012) e1002582.
[329] N.A. Crossley, A. Mechelli, J. Scott, F. Carletti, P.T. Fox, P. McGuire, E.T. Bullmore, The hubs of the human connectome are generally implicated in the anatomy of brain disorders, Brain 137 (2014) 2382–2395.
[330] C.Y.Z. Lo, T.W. Su, C.C. Huang, C.C. Hung, W.-L. Chen, T.-H. Lan, C.-P. Lin, E.T. Bullmore, Randomization and resilience of brain functional networks as systems-level endophenotypes of schizophrenia, Proc. Natl. Acad. Sci. USA 112 (29) (2015) 9123–9128.
[331] M. van den Heuvel, O. Sporns, Network hubs in human brain, Trends Cogn. Sci. 17 (12) (2013) 683–696.
[332] R.G. Abeysuriya, J. Hadida, S.N. Sotiropoulos, S. Jbabdi, R. Becker, B.A.E. Hunt, M.J. Brookes, M.W. Woolrich, A biophysical model of dynamic balancing of excitation and inhibition in fast oscillatory large-scale networks, PLoS Comput. Biol. 14 (2) (2018) e1006007.
[333] M. Zanin, D. Papo, P.A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik, S. Boccaletti, Combining complex networks and data mining: why and how, Phys. Rep. 635 (2016) 1–44.
[334] S. Khuri, S. Wuchty, Essentiality and centrality in protein interaction networks revisited, BMC Bioinformatics 16 (2015) 109.
[335] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6 (11) (2010) 888–893.
[336] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, K-core organization of complex networks, Phys. Rev. Lett. 96 (4) (2006) 040601.
[337] A. Zeng, C.-J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A 377 (14) (2013) 1031–1035.
[338] J.-G. Liu, Z.-M. Ren, Q. Guo, Ranking the spreading influence in complex networks, Physica A 392 (18) (2013) 4154–4159.
[339] L. Lü, T. Zhou, Q.-M. Zhang, H.E. Stanley, The H-index of a network node and its relation to degree and coreness, Nature Commun. 7 (2016) 10168.
[340] F. Morone, G. Del Ferraro, H.A. Makse, The k-core as a predictor of structural collapse in mutualistic ecosystems, Nat. Phys. 15 (1) (2019) 95–102.
[341] V. Narang, M.A. Ramli, A. Singhal, P. Kumar, G. de Libero, M. Poidinger, C. Monterola, Automated identification of core regulatory genes in human gene regulatory networks, PLoS Comput. Biol. 11 (9) (2015) e1004504.
[342] M.I. Ashraf, S.-K. Ong, S. Mujawar, S. Pawar, P. More, S. Paul, C. Lahiri, A side-effect free method for identifying cancer drug targets, Sci. Rep. 8 (2018) 6669.
[343] A. Korn, A. Schubert, A. Telcs, Lobby index in networks, Physica A 388 (11) (2009) 2221–2226.
[344] M.G. Campiteli, A.J. Holanda, L.D.H. Soares, P.R.C. Soles, O. Kinouchi, Lobby index as a network centrality measure, Physica A 392 (21) (2013) 5511–5515.

[345] H.-W. Ma, A.-P. Zeng, The connectivity structure, giant strong component and centrality of metabolic networks, Bioinformatics 19 (11) (2003) 1423–1430.
[346] P. Bonacich, Some unique properties of eigenvector centrality, Social Networks 29 (4) (2007) 555–564.
[347] G. Lohmann, D.S. Margulies, A. Horstmann, B. Pleger, J. Lepsien, D. Goldhahn, H. Schloegl, M. Stumvoll, A. Villringer, R. Turner, Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain, PLoS One 5 (4) (2010) e10232.
[348] C.F.A. Negre, U.N. Morzan, H.P. Hendrickson, R. Pal, G.P. Lisi, J.P. Loria, I. Rivalta, J. Ho, V.S. Batista, Eigenvector centrality for characterization of protein allosteric pathways, Proc. Natl. Acad. Sci. USA 115 (52) (2018) 12201–12208.
[349] A.M. Wink, Eigenvector centrality dynamics from resting-state fMRI: gender and age difference in healthy subjects, Front. Neurosci. 13 (2019) 648.
[350] M.A.A. Binnewijzend, S.M. Adriaanse, W.M. Van der Flier, C.E. Teunissen, J.C. de Munck, C.J. Stam, P. Scheltens, B.N.M. van Berckel, F. Barkhof, A.M. Wink, Brain network alterations in Alzheimer's disease measured by eigenvector centrality in fMRI are related to cognition and CSF biomarkers, Hum. Brain Mapp. 35 (5) (2014) 2383–2393.
[351] E. van Duinkerken, M.M. Schoonheim, R.G. IJzerman, A.C. Moll, J. Landeira-Fernandez, M. Klein, M. Diamant, F.J. Snoek, F. Barkhof, A.M. Wink, Altered eigenvector centrality is related to local resting-state network functional connectivity in patients with longstanding type 1 diabetes mellitus, Hum. Brain Mapp. 38 (7) (2017) 3623–3636.
[352] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1) (1998) 107–117.
[353] L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou, Leaders in social networks, the delicious case, PLoS One 6 (6) (2011) e21202.
[354] C. Winter, et al., Google goes cancer: improving outcome prediction for cancer patients by network-based ranking of marker genes, PLoS Comput. Biol. 8 (5) (2012) e1002511.
[355] J. Roy, C. Winter, Z. Isik, M. Schroeder, Network information improves cancer outcome prediction, Brief. Bioinform. 15 (4) (2014) 612–625.
[356] Y. Fan, X. Tang, X. Hu, W. Wu, Q. Ping, Prediction of essential proteins based on subcellular localization and gene expression correlation, BMC Bioinformatics 18 (s13) (2017) 470.
[357] J. Choi, S. Park, Y. Yoon, J. Ahn, Improved prediction of breast cancer outcome by identifying heterogeneous biomarkers, Bioinformatics 33 (22) (2017) 3619–3626.
[358] B. Jiang, K. Kloster, D.F. Gleich, M. Gribskov, AptRank: an adaptive PageRank model for protein function prediction on bi-relational graphs, Bioinformatics 33 (12) (2017) 1829–1836.
[359] M. Jalili, A. Salehzadeh-Yazdi, S. Gupta, O. Wolkenhauer, M. Yaghmaie, O. Resendis-Antonio, K. Alimoghaddam, Evolution of centrality measurements for the detection of essential proteins in biological networks, Front. Physiol. 7 (2016) 375.
[360] G. del Rio, D. Koschützki, G. Coello, How to identify essential genes from molecular networks, BMC Syst. Biol. 3 (2009) 102.
[361] M.T. Kuhnert, C. Geier, C.E. Elger, K. Lehertz, Identifying important nodes in weighted functional brain networks: a comparison of different centrality approaches, Chaos 22 (2) (2012) 023142.
[362] G. Ghoshal, A.-L. Barabási, Ranking stability and super-stable nodes in complex networks, Nature Commun. 2 (2011) 394.
[363] J. Ran, H. Li, J. Fu, L. Liu, Y. Xing, X. Li, H. Shen, Y. Chen, X. Jiang, Y. Li, H. Li, Construction and analysis of the protein-protein interaction network related to essential hypertension, BMC Syst. Biol. 7 (2013) 32.
[364] P. Wang, X. Yu, J. Lü, Identification and evolution of structurally dominant nodes in protein-protein interaction networks, IEEE Trans. Biomed. Circ. Syst. 8 (1) (2014) 87–97.
[365] D. Mistry, R.P. Wise, J.A. Dickerson, DiffSLC: A graph centrality method to detect essential proteins of a protein-protein interaction network, PLoS One 12 (11) (2017) e0187091.
[366] G. Li, M. Li, J. Wang, Y. Li, Y. Pan, United neighborhood closeness centrality and orthology for predicting essential proteins, IEEE ACM Trans. Comput. Biol. Bioinform. (2018) http://dx.doi.org/10.1109/TCBB.2018.2889978.
[367] W. Peng, J. Wang, W. Wang, Q. Liu, F.-X. Wu, Y. Pan, Iteration method for predicting essential proteins based on orthology and protein-protein interaction networks, BMC Syst. Biol. 6 (2012) 87.
[368] M. Li, H. Zhang, J.-X. Wang, Y. Pan, A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data, BMC Syst. Biol. 6 (2012) 15.
[369] X. Zhang, J. Xu, W.-X. Xiao, A new method for the discovery of essential proteins, PLoS One 8 (3) (2013) e58763.
[370] J. Luo, Y. Qi, Identification of essential proteins based on a new combination of local interaction density and protein complexes, PLoS One 10 (6) (2015) e0131418.
[371] M. Li, Y. Lu, Z. Niu, F.-X. Wu, United complex centrality for identification of essential proteins from PPI networks, IEEE ACM Trans. Comput. Biol. Bioinform. 14 (2) (2017) 370–380.
[372] C. Hens, U. Harush, S. Haber, R. Cohen, B. Barzel, Spatiotemporal signal propagation in complex networks, Nat. Phys. 15 (4) (2019) 403–412.
[373] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Modern Phys. 87 (3) (2015) 925–979.
[374] Y.-A. Kim, D.-Y. Cho, T.M. Przytycka, Understanding genotype-phenotype effects in cancer via network approaches, PLoS Comput. Biol. 12 (3) (2016) e1004747.
[375] M. Oti, B. Snel, M.A. Huynen, H.G. Brunner, Predicting disease genes using protein-protein interactions, J. Med. Genet. 43 (8) (2006) 691–698.
[376] M. Hofree, J.P. Shen, H. Carter, A. Cross, T. Ideker, Network-based stratification of tumor mutations, Nature Methods 10 (11) (2013) 1108–1115.
[377] L. Katz, A new status index derived from sociometric analysis, Psychometrika 18 (1) (1953) 39–43.
[378] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, R. Sharan, Associating genes and protein complexes with disease via network propagation, PLoS Comput. Biol. 6 (1) (2010) e1000641.
[379] S. Santolini, A.-L. Barabási, Predicting perturbation patterns from the topology of biological networks, Proc. Natl. Acad. Sci. USA 115 (27) (2018) 6375–6383.
[380] R.I. Kondor, J. Lafferty, Diffusion kernels on graphs and other discrete input spaces, in: Proceedings of the 9th International Conference on Machine Learning, ACM New York, 2002, pp. 315–322.
[381] S. Zhang, X.-M. Ning, X.-S. Zhang, Graph kernels, hierarchical clustering, and network community structure: experiments and comparative analysis, Eur. Phys. J. B 57 (1) (2007) 67–74.
[382] S. Köhler, S. Bauer, D. Horn, P.N. Robinson, Walking the interactome for prioritization of candidate disease genes, Am. J. Hum. Genet. 82 (4) (2008) 949–958.
[383] J. Zhao, T.-H. Yang, Y. Huang, P. Holme, Ranking candidate disease genes from gene expression and protein interaction: a katz-centrality based approach, PLoS One 6 (9) (2011) e24306.
[384] J.K. Huang, D.E. Carlin, M.K. Yu, W. Zhang, J.F. Kreisberg, P. Tamayo, T. Ideker, Systematic evaluation of molecular networks for discovery of disease genes, Cell Syst. 6 (4) (2018) 484–495.
[385] Y. Li, J.C. Patra, Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network, Bioinformatics 26 (9) (2010) 1219–1224.
[386] C. Blatti, S. Sinha, Characterizing gene sets using discriminative random walks with restart on heterogeneous biological networks, Bioinformatics 32 (14) (2016) 2167–2175.

[387] B. Zhao, Y. Zhao, X. Zhang, Z. Zhang, F. Zhang, L. Wang, An iteration method for identifying yeast essential proteins from heterogeneous network, BMC Bioinformatics 20 (2019) 355.
[388] A. Valdeolivas, L. Tichit, C. Navarro, S. Perrin, G. Odelin, N. Levy, P. Cau, E. Remy, A. Baudot, Random walk with restart on multiplex and heterogeneous biological networks, Bioinformatics 35 (3) (2019) 497–505.
[389] F. Vandin, E. Upfal, B. Raphael, Algorithms for detecting significantly mutated pathways in cancer, J. Comput. Biol. 18 (3) (2011) 507–522.
[390] M.D. Leiserson, et al., Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes, Nat. Genet. 47 (2) (2015) 106–114.
[391] W. Zhang, J. Ma, T. Ideker, Classifying tumors by supervised network propagation, Bioinformatics 34 (13) (2018) 484–493.
[392] Y. Hou, B. Gao, G. Li, Z. Su, MaxMIF: a new method for identifying cancer driver genes through effective data integration, Adv. Sci. 5 (9) (2018) 1800640.
[393] S. Patkar, A. Magen, R. Sharan, S. Hannenhalli, A network diffusion approach to inferring sample-specific function reveals functional changes associated with breast cancer, PLoS Comput. Biol. 13 (11) (2017) e1005793.
[394] X. Chen, M.-X. Liu, G.-Y. Yan, Drug-target interaction prediction by random walk on the heterogeneous network, Mol. Biosyst. 8 (7) (2012) 1970–1978.
[395] H.-Y. Li, T.-Y. Li, D. Quang, Y.-F. Guan, Network propagation predicts drug synergy in cancers, Cancer Res. 78 (18) (2018) 5446–5457.
[396] J.J. Cáceres, A. Paccanaro, Disease gene prediction for molecularly uncharacterized diseases, PLoS Comput. Biol. 15 (7) (2019) e1007078.
[397] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (6) (2011) 1150–1170.
[398] V. Martinez, F. Berzal, J.-C. Cubero, A survey of link prediction in complex networks, ACM Comput. Surv. 49 (4) (2016) 69.
[399] B. Barzel, A.-L. Barabási, Network link prediction by global silencing of indirect correlations, Nat. Biotechnol. 31 (8) (2013) 720–725.
[400] I.A. Kovács, K. Luck, K. Spirohn, Y. Wang, C. Pollis, S. Schlabach, W. Bian, D.-K. Kim, N. Kishore, T. Hao, M.A. Calderwood, M. Vidal, A.-L. Barabási, Network-based prediction of protein interactions, Nature Commun. 10 (2019) 1240.
[401] H. Hu, C. Zhu, H. Ai, L. Zhang, J. Zhao, Q. Zhao, H. Liu, LPI-ETSLP: LncRNA-protein interaction prediction using eigenvalue transformation-based semi-supervised link prediction, Mol. BioSyst. 13 (9) (2017) 1781–1787.
[402] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, X. Li, Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data, BMC Bioinformatics 18 (2017) 18.
[403] L. Lin, T. Yang, L. Fang, J. Yang, F. Yang, J. Zhao, Gene gravity-like algorithm for disease gene prediction based on phenotype-specific network, BMC Syst. Biol. 11 (2017) 121.
[404] X. Chen, Z. Zhou, Y. Zhao, ELLPMDA: Ensemble learning and link prediction for miRNA-disease association prediction, RNA Biol. 15 (6) (2018) 807–818.
[405] T. Turki, Z. Wei, A link prediction approach to cancer drug sensitivity prediction, BMC Syst. Biol. 11 (S5) (2017) 94.
[406] S. Sulaimany, M. Khansari, P. Zarrineh, M. Daianu, N. Jahanshad, P.M. Thompson, A. Masoudi-Nejad, Predicting brain network changes in Alzheimer's disease with link prediction algorithms, Mol. Biosyst. 13 (4) (2017) 725–735.
[407] F. Fouss, A. Pirotte, J. Renders, M. Saerens, Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation, IEEE Trans. Knowl. Data Eng. 19 (3) (2007) 355–369.
[408] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, H.E. Stanley, Toward link predictabilityof complex networks, Proc. Natl. Acad. Sci. USA 112 (8) (2015) 2325–2330.
[409] W. Wang, F. Cai, P. Jiao, L. Pan, A perturbation-based framework for link prediction via non-negative matrix factorization, Sci. Rep. 6 (2016) 38938.
[410] T. Zhou, L. Lü, Y.-C. Zhang, Predicting missing links via local information, Eur. Phys. J. B 71 (4) (2009) 623–630.
[411] Y. Hulovatyy, R.W. Solvava, T. Milenković, Revealing missing parts of the interactome via link prediction, PLoS One 9 (3) (2014) e90073.
[412] M. Chen, Y. Zhang, A. Li, Z. Li, W. Liu, Z. Chen, Bipartite heterogeneous network method based on co-neighbor for MiRNA-disease association prediction, Front. Genet. 10 (2019) 385.
[413] Z.-A. Huang, Y.-A. Huang, Z.-H. You, Z. Zhu, Y. Sun, Novel link prediction for large-scale miRNA-lncRNA interaction network in a bipartite graph, BMC Med. Genomics 11 (S6) (2018) 113.
[414] M.A. Yildirim, M. Coscia, Using random walks to generate associations between objects, PLoS One 9 (8) (2014) e104813.
[415] T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, Bipartite network projection and personal recommendation, Phys. Rev. E 76 (4) (2007) 046115.
[416] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J.R. Wakeling, Y.-C. Zhang, Solving the apparent diversity accuracy dilemma of recommender systems, Proc. Natl. Acad. Sci. USA 107 (10) (2010) 4511–4515.
[417] J. Ruths, D. Ruths, Control profiles of complex networks, Science 343 (6177) (2014) 1373–1376.
[418] Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks, Nature 473 (7346) (2019) 167–173.
[419] J. Gao, Y.-Y. Liu, R.M. D'Souza, A.-L. Barabási, Target control of complex networks, Nature Commun. 5 (2014) 5415.
[420] R.E. Kalman, Mathematical description of linear dynamical systems, J. Soc. Ind. Appl. Math. Ser. A 1 (1963) 152–192.
[421] J.C. Nacher, T. Akutsu, Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control, New. J. Phys. 14 (2012) 073005.
[422] Z. Yuan, C. Zhao, Z.-R. Di, W.-X. Wang, Y.-C. Lai, Exact controllability of complex networks, Nature Commun. 4 (2013) 2447.
[423] A. Vinayagam, T.E. Gibson, H.-J. Lee, B. Yilmazel, C. Roesel, Y. Hu, Y. Kwon, A. Sharma, Y.-Y. Liu, N. Perrimon, A.-L. Barabási, Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets, Proc. Natl. Acad. Sci. USA 113 (18) (2016) 4976–4981.
[424] S. Wuchty, Controllability in protein interaction networks, Proc. Natl. Acad. Sci. USA 111 (19) (2014) 7156–7160.
[425] G. Bidkhori, R. Benfeitas, M. Klevstig, C. Zhang, J. Nielsen, M. Uhlen, J. Boren, A. Mardinoglu, Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes, Proc. Natl. Acad. Sci. USA 115 (50) (2018) 11874–11883.
[426] X.-F. Zhang, L. Ou-Yang, Y. Zhu, M.-Y. Wu, D.-Q. Dai, Determining minimum set of driver nodes in protein-protein interaction networks, BMC Bioinformatics 16 (2015) 146.
[427] J.-M. Schwartz, H. Otokuni, T. Akutsu, J.C. Nacher, Probabilistic controllability approach to metabolic fluxes in normal and cancer tissues, Nature Commun. 10 (2019) 2725.
[428] G. Yan, P.E. Vértes, E.K. Towlson, Y.L. Chew, D.S. Walker, A.-L. Barabási, Network control principles predict neuron function in the Caenorhabditis elegans connectome, Nature 550 (7677) (2017) 519–523.
[429] Y. Asgari, A. Salehzadeh-Yazdi, F. Schreiber, A. Masoudi-Nejad, Controllability in cancer metabolic networks according to drug targets as driver nodes, PLoS One 8 (11) (2013) e79397.
[430] L. Wu, Y. Shen, M. Li, F.-X. Wu, Network output controllability-based method for drug target identification, IEEE Trans. Nanobiosci. 14 (2) (2015) 184–191.
[431] M. Ishitsuka, T. Akutsu, J.C. Nacher, Critical controllability in proteome-wide protein interaction network integrating transcriptome, Sci. Rep. 6 (2016) 23541.
[432] W.-F. Guo, S.-W. Zhang, Z.-G. Wei, T. Zeng, F. Liu, J. Zhang, F.-X. Wu, L. Chen, Constrained target controllability of complex networks, J. Stat. Mech. Theory Exp. 2017 (2017) 063402.

[433] W.-F. Guo, S.-W. Zhang, Q.-Q. Shi, C.-M. Zhang, T. Zeng, L. Chen, A novel algorithm for finding optimal driver nodes to target control complex networks and its application for drug targets identification, BMC Genomics 19 (S1) (2018) 924.
[434] W. Zheng, D. Wang, X. Zou, Control of multilayer biological networks and applied to target identification of complex diseases, BMC Bioinformatics 20 (2019) 271.
[435] P.-G. Sun, Co-controllability of drug-disease-gene network, New J. Phys. 17 (2015) 085009.
[436] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (6245) (2015) 255–260.
[437] B.M. Moore, P. Wang, P. Fan, B. Leong, C.A. Schenck, J.P. Lloyd, M.D. Lehti-Shiu, R.L. Last, E. Pichersky, S.-H. Shiu, Robust predictions of specialized metabolism genes through machine learning, Proc. Natl. Acad. Sci. USA 116 (6) (2019) 2344–2353.
[438] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, M.M. Hoffman, Machine learning for integrating data in biology and medicine: principles, practice, and opportunities, Inform. Fusion 50 (2019) 71–91.
[439] C.J. Tokheim, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, R. Karchin, Evaluating the evaluation of cancer driver genes, Proc. Natl. Acad. Sci. USA 113 (50) (2016) 14330–14335.
[440] S. Ekins, A.C. Puhl, K.M. Zorn, T.R. Lane, D.P. Russo, J.J. Klein, A.J. Hickey, A.M. Clark, Exploiting machine learning for end-to-end drug discovery and development, Nature Mater. 18 (5) (2019) 435–441.
[441] T.J. Rademaker, E. Bengio, P. Francois, Attack and defense in cellular decision-making: lessons from machine learning, Phys. Rev. X 9 (3) (2019) 031012.
[442] M. Zitnik, R. Sosič, J. Leskovec, Prioritizing network communities, Nature Commun. 9 (2018) 2544.
[443] J. Saez-Rodriguez, J.C. Costello, S.H. Friend, M.R. Kellen, L. Mangravite, P. Meyer, T. Norman, G. Stolovitzky, Crowdsourcing biomedical research: leveraging communities as innovation engines, Nat. Rev. Genet. 17 (8) (2016) 470–486.
[444] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[445] S. Webb, Deep learning for biology, Nature 554 (7693) (2018) 555–557.
[446] C. Angermueller, T. Pärnamaa, L. Parts, O. Stegle, Deep learning for computational biology, Mol. Syst. Biol. 12 (7) (2016) 878.
[447] S.M.E. Sahraeian, R. Liu, B. Lau, K. Podesta, M. Mohiyuddin, H.Y.K. Lam, Deep convolutional neural networks for accurate somatic mutation detection, Nature Commun. 10 (2019) 1041.
[448] D. Ardila, A.P. Kiraly, S. Bharadwaj, B. Choi, J.J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D.P. Naidich, S. Shetty, End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nat. Med. 25 (6) (2019) 954–961.
[449] P. Mobadersany, S. Yousefi, M. Amgad, D.A. Gutman, J.S. Barnholtz-Sloan, J.E.V. Vega, D. Brat, L.A.D. Cooper, Predicting cancer outcomes from histology and genomics using convolutional networks, Proc. Natl. Acad. Sci. USA 115 (13) (2018) 2970–2979.
[450] I. Lee, J. Keum, H. Nam, DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences, PLoS Comput. Biol. 15 (6) (2019) e1007129.
[451] J. Zhou, C.Y. Park, C.L. Theesfeld, A.K. Wong, Y. Yuan, C. Scheckel, J.J. Fak, J. Funk, Y. Yao, A. Packer, R.B. Darnell, O.G. Troyanskaya, Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk, Nat. Genet. 51 (6) (2019) 973–980.
[452] T. Xie, J.C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. 120 (14) (2018) 145301.
[453] P. Luo, Y. Ding, X. Lei, F.-X. Wu, DeepDriver: predicting cancer driver genes based on somatic mutations using deep convolutional neural networks, Front. Genet. 10 (2019) 13.
[454] M. Zitnik, M. Agrawal, J. Leskovec, Modeling polypharmacy side effects with graph convolutional networks, Bioinformatics 34 (13) (2018) 457–466.
[455] S. Rhee, S. Seo, S. Kim, Hybrid approach of relation networks and localized graph convolutional filtering for breast cancer subtype classification, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, AAAI Press, 2018, pp. 3527–3534.
[456] A. Ahmed, N. Shervashidze, S. Narayanamurthy, V. Josifovski, A.J. Smola, Distributed large-scale natural graph factorization, in: Proceedings of the 22nd International Conference on World Wide Web, ACM press, New York, 2013, pp. 37–48.
[457] M. Ou, P. Cui, J. Pei, Z. Zhang, W. Zhu, Asymmetric transitivity preserving graph embedding, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2016, pp. 1105–1114.
[458] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: online learning of social representation, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2014, pp. 701–710.
[459] A. Grover, J. Leskovec, Node2vec: scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2016, pp. 855–864.
[460] L.F.R. Ribeiro, P.H.P. Saverese, D.R. Figueiredo, Struc2vec: Learning node representations from structural identity, in: Proceedings of the 23th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2017, pp. 385–394.
[461] C. Donnat, M. Zitnik, D. Hallac, J. Leskovec, Learning structural node embedding via diffusion wavelets, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2018, pp. 1320–1329.
[462] P. Goyal, E. Ferrara, Graph embedding techniques, applications and performance: A survey, Konwl-Based Syst. 151 (2018) 78–94.
[463] H. Cai, V.W. Zheng, K.C.-C. Chang, A comprehensive survey of graph embedding: problems, techniques and applications, IEEE Trans. Knowl. Data Eng. 30 (9) (2018) 1616–1637.
[464] W. Nelson, M. Zitnik, B. Wang, J. Leskovec, A. Goldenberg, R. Sharan, To embed or not: network embedding as a paradigm in computational biology, Front. Genet. 10 (2019) 381.
[465] C. Su, J. Tong, Y. Zhu, P. Cui, F. Wang, Network embedding in biomedical data science, Brief. Bioinform. (2018) http://dx.doi.org/10.1093/bib/bby117.
[466] S. Wang, E. Huang, J. Cairns, J. Peng, L. Wang, S. Sinha, Identification of pathways associated with chemosensitivity through network embedding, PLoS Comput. Biol. 15 (3) (2019) e1006864.
[467] X. Wu, W. Zeng, Y. Xu, B. Wang, X. Liu, F. Lin, G. Alterovitz, Predicting of associations between microRNA and human disease based on multiple similarities and arbitrarily-order proximity network embedding, IEEE Access 7 (2019) 86625–86634.
[468] J.-J. Peng, J.-J. Guan, X.-Q. Shang, Predicting parkinson's disease genes based on Node2vec and autoencoder, Front. Genet. 10 (2019) 226.
[469] N. Zong, H. Kim, V. Ngo, O. Harismendy, Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations, Bioinformatics 33 (15) (2017) 2337–2344.
[470] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, J. Zeng, A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information, Nature Commun. 8 (2017) 573.
[471] G. Rosenthal, F. Váša, A. Griffa, P. Hagmann, E. Amico, J. Goñi, G. Avidan, O. Sporns, Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes, Nature Commun. 9 (2018) 2178.
[472] M. Kulmanov, M.A. Khan, R. Hoehndorf, DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier, Bioinformatics 34 (4) (2018) 660–668.
[473] J. Fan, A. Cannistra, I. Fried, T. Lim, T. Schaffner, M. Crovella, B. Hescott, M.D.M. Leiserson, Functional protein representations from biological networks enable diverse cross-species inference, Nucleic Acids Res. 47 (9) (2019) e51.

[474] H. Lakkaraju, S.H. Bach, J. Leskovec, Interpretable decision sets: a joint framework for description and prediction, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2016, pp. 1675–1684.
[475] J.H. Yang, S.N. Wright, M. Hamblin, D. McCloskey, M.A. Alcantar, L. Schrübbers, A.J. Lopatkin, S. Satish, A. Nili, B.O. Palsson, G.C. Walker, J.J. Collins, A white-box machine learning approach for revealing antibiotic mechanisms of action, Cell 177 (6) (2019) 1649–1661.
[476] D.J. Burgess, Illuminating the dark side of machine learning, Nat. Rev. Genet. 20 (7) (2019) 374–375.
[477] M.K. Yu, J. Ma, J. Fisher, J.F. Kreisberg, B.J. Raphael, T. Ideker, Visible machine learning for biomedicine, Cell 173 (7) (2018) 1562–1565.
[478] H. Vuong, F. Cheng, C.-C. Lin, Z. Zhao, Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach, Genome Med. 6 (2014) 81.
[479] J. Zhao, F. Cheng, Y. Wang, C.L. Arteaga, Z. Zhao, Systematic prioritization of druggable mutations in similar to 5000 genomes across 16 cancer types using a structural genomics-based approach, Mol. Cell. Proteomics 15 (2) (2016) 642–656.
[480] P.K. Ng, et al., Systematic functional annotation of somatic mutations in cancer, Cancer Cell 33 (3) (2018) 450–462.
[481] P. Creixell, et al., Kinome-wide decoding of network-attacking mutations rewiring cancer signaling, Cell 163 (1) (2015) 202–217.
[482] J. Zhao, F. Cheng, Z. Zhao, Tissue-specific signaling networks rewired by major somatic mutations in human cancer revealed by proteome-wide discovery, Cancer Res. 77 (11) (2017) 2810–2821.
[483] J.S. Amberger, C.A. Bocchini, F. Schiettecatte, F. Scott, A. Hamosh, OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders, Nucleic Acids Res. 43 (2015) D789–D798.
[484] A.P. Davis, C.J. Grondin, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B.L. King, T.C. Wiegers, C.J. Mattingly, The comparative toxicogenomics database's 10th year anniversary: update 2015, Nucleic Acids Res. 43 (2015) D914–D920.
[485] M.J. Landrum, J.M. Lee, G.R. Riley, W. Jang, W.S. Rubinstein, D.M. Church, D.R. Maglott, ClinVar: public archive of relationships among sequence variation and human phenotype, Nucleic Acids Res. 42 (2014) D980–D985.
[486] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, H. Parkinson, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations, Nucleic Acids Res. 42 (2014) D1001–D1006.
[487] M.J. Li, Z. Liu, P. Wang, M.P. Wong, M.R. Nelson, J.P. Kocher, M. Yeager, P.C. Sham, S.J. Chanock, Z. Xia, J. Wang, GWASdb v2: an update database for human genetic variants identified by genome-wide association studies, Nucleic Acids Res. 44 (2016) D869–D876.
[488] J.C. Denny, et al., Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data, Nat. Biotechnol. 31 (12) (2013) 1102–1110.
[489] W. Yu, W. Gwinn, M. Clyne, A. Yesupriya, M.J. Khoury, A navigator for human genome epidemiology, Nat. Genet. 40 (2) (2008) 124–125.
[490] J. Pinero, N. Queralt-Rosinach, A. Bravo, J. Deu-Pons, A. Bauer-Mehren, M. Baron, F. Sanz, L.I. Furlong, DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes, Database 2015 (2015) bav028.
[491] A. Li, S.P. Cornelius, Y.-Y. Liu, L. Wang, A.-L. Barabási, The fundamental advantages of temporal networks, Science 358 (6366) (2017) 1042–1046.
[492] S. Gu, F. Pasqualetti, M. Cieslak, Q.K. Telesford, A.B. Yu, A.E. Kahn, J.D. Medaglia, J.M. Vettel, M.B. Miller, S.T. Grafton, D.S. Bassett, Controllability of structural brain networks, Nature Commun. 6 (2015) 8414.
[493] E. Tang, D.S. Bassett, Colloquium: Control of dynamics in brain networks, Rev. Modern Phys. 90 (3) (2018) 031003.
[494] S.F. Muldoon, F. Pasqualetti, S. Gu, M. Cieslak, S.T. Grafton, J.M. Vettel, D.S. Bassett, Stimulation-based control of dynamic brain networks, PLoS Comput. Biol. 12 (9) (2016) e1005076.
[495] E. Tang, C. Giusti, G.L. Baum, S. Gu, E. Pollock, A.E. Kahn, D.R. Roalf, T.M. Moore, K. Ruparel, R.C. Gur, R.E. Gur, T.D. Satterthwaite, D.S. Bassett, Developmental increases in white matter network controllability support a growing diversity of brain dynamics, Nature Commun. 8 (2017) 1252.
[496] F. Morone, H.A. Makse, Influence maximization in complex networks through optimal percolation, Nature 524 (7563) (2015) 65–68.
[497] G. Del Ferraro, A. Moreno, B. Min, F. Morone, U. Perez-Ramirez, L. Perez-Cervera, L.C. Parra, A. Holodny, S. Canals, H.A. Makse, Finding influential nodes for integration in brain networks using optimal percolation theory, Nature Commun. 9 (2018) 2274.
[498] L. Wang, C.-S. Yu, H. Chen, W. Qin, Y. He, F. Fan, Y.-J. Zhang, M.-L. Wang, K.-C. Li, Y.-F. Zang, T.S. Woodward, C.-Z. Zhu, Dynamic functional reorganization of the motor execution network after stroke, Brain 133 (2010) 1224–1238.
[499] C. Grefkes, G.R. Fink, Reorganization of cerebral networks after stroke: new insights from neuroimaging with connectivity approaches, Brain 134 (2011) 1264–1276.
[500] A.N. Khambhati, K.A. Davis, B.S. Oommen, S.H. Chen, T.H. Lucas, B. Litt, D.S. Bassett, Dynamic network drivers of seizure generation, propagation and termination in human neocortical epilepsy, PLoS Comput. Biol. 11 (12) (2015) e1004608.
[501] A.N. Khambhati, K.A. Davis, T.H. Lucas, B. Litt, D.S. Bassett, Virtual cortical resection reveals push-pull network control preceding seizure evolution, Neuron 91 (5) (2016) 1170–1182.
[502] D.S. Bassett, N.F. Wymbs, M.A. Porter, P.J. Mucha, J.M. Carlson, S.T. Grafton, Dynamic reconfiguration of human brain networks during learning, Proc. Natl. Acad. Sci. USA 108 (18) (2011) 7641–7646.
[503] D.S. Bassett, M. Yang, N.F. Wymbs, S.T. Grafton, Learning-induced autonomy of sensorimotor systems, Nat. Neurosci. 18 (5) (2015) 744–751.
[504] K.W. Doron, D.S. Bassett, M.S. Gazzaniga, Dynamic network structure of interhemispheric coordination, Proc. Natl. Acad. Sci. USA 109 (46) (2012) 18661–18668.
[505] A. Mullard, 2016 FDA drug approvals, Nat. Rev. Drug Discov. 16 (2) (2017) 73–76.
[506] F. Cheng, Y. Zhou, J. Li, W. Li, G. Liu, Y. Tang, Prediction of chemical-protein interactions: multitarget-QSAR versus computational chemogenomic methods, Mol. Biosyst. 8 (9) (2012) 2373–2384.
[507] W. Lu, F. Cheng, J. Jiang, C. Zhang, X. Deng, Z. Xu, S. Zou, X. Shen, Y. Tang, J. Huang, FXR antagonism of NSAIDs contributes to drug-induced liver injury identified by systems pharmacology approach, Sci. Rep. 5 (2015) 8114.
[508] J. Fang, C. Liu, Q. Wang, P. Lin, F. Cheng, In silico polypharmacology of natural products, Brief. Bioinform. 19 (6) (2018) 1153–1171.
[509] F. Cheng, Y. Zhou, W. Li, G. Liu, Y. Tang, Prediction of chemical-protein interactions network with weighted network-based inference method, PLoS One 7 (7) (2012) e41064.
[510] F. Cheng, W. Li, Z. Wu, X. Wang, C. Zhang, J. Li, G. Liu, Y. Tang, Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space, J. Chem. Inf. Model. 53 (4) (2013) 753–762.
[511] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: An open chemical toolbox, J. Cheminform. 3 (2011) 33.
[512] J. Shen, F. Cheng, Y. Xu, W. Li, Y. Tang, Estimation of ADME properties with substructure pattern recognition, J. Chem. Inf. Model. 50 (6) (2010) 1034–1041.
[513] P. Willett, Similarity-based virtual screening using 2D fingerprints, Drug Discov. Today 11 (23–24) (2006) 1046–1053.
[514] F. Cheng, W. Li, X. Wang, Y. Zhou, Z. Wu, J. Shen, Y. Tang, Adverse drug events: database construction and in silico prediction, J. Chem. Inf. Model. 53 (4) (2013) 744–752.
[515] J.A. Blake, et al., Gene ontology annotations and resources, Nucleic Acids Res. 41 (2013) D530–D535.
[516] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug-target interaction networks from the integration of chemical and genomic spaces, Bioinformatics 24 (13) (2008) I232–I240.

[517] L. Perlman, A. Gottlieb, N. Atias, E. Ruppin, R. Sharan, Combining drug and gene similarity measures for drug-target elucidation, J. Comput. Biol. 18 (2) (2011) 133–145.
[518] M.A. Yildirim, K.I. Goh, M.E. Cusick, A.-L. Barabási, M. Vidal, Drug-target network, Nat. Biotechnol. 25 (10) (2007) 1119–1126.
[519] R.-S. Wang, J. Loscalzo, Illuminating drug action by network integration of disease genes: a case study of myocardial infarction, Mol. Biosyst. 12 (5) (2016) 1653–1666.
[520] S. Zhao, T. Nishimura, Y. Chen, E.U. Azeloglu, O. Gottesman, C. Giannarelli, M.U. Zafar, L. Benard, J.J. Badimon, R.J. Hajjar, J. Goldfarb, R. Iyengar, Systems pharmacology of adverse event mitigation by drug combinations, Sci. Transl. Med. 5 (206) (2013) 206ra140.
[521] D.S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S.-L. Chen, D. Hadley, A. Green, P. Khankhanian, S.E. Baranzini, Systematic integration of biomedical knowledge prioritizes drugs for repurposing, Elife 6 (2017) e26726.
[522] A.L. Hopkins, Network pharmacology: the next paradigm in drug discovery, Nat. Chem. Biol. 4 (11) (2008) 682–690.
[523] D. Moodley, H. Yoshida, S. Mostafavi, N. Asinovski, A. Ortiz-Lopez, P. Symanowicz, J.B. Telliez, M. Hegen, J.D. Clark, D. Mathis, C. Benoist, Network pharmacology of JAK inhibitors, Proc. Natl. Acad. Sci. USA 113 (35) (2016) 9852–9857.
[524] J.J. Moslehi, Cardiovascular toxic effects of targeted cancer therapies, N. Engl. J. Med. 375 (15) (2016) 1457–1467.
[525] E. Lounkine, M.J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J.L. Jenkins, P. Lavan, E. Weber, A.K. Doak, S. Cote, B.K. Shoichet, L. Urban, Large-scale prediction and testing of drug activity on side-effect targets, Nature 486 (7403) (2012) 361–367.
[526] K.A. Ryall, A.C. Tan, Systems biology approaches for advancing the discovery of effective drug combinations, J. Cheminform. 7 (2015) 7.
[527] X. Sun, S. Vilar, N.P. Tatonetti, High-throughput methods for combinatorial drug discovery, Sci. Transl. Med. 5 (205) (2012) 205rv1.
[528] M.A. Ali, S. Rizvi, B.A. Syed, Trends in the market for antihypertensive drugs, Nat. Rev. Drug Discov. 16 (2017) 309–310.
[529] T.D. Giles, M.A. Weber, J. Basile, A.H. Gradman, D.B. Bharucha, W. Chen, M. Pattathil, Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study, Lacnet 383 (9932) (2014) 1889–1898.
[530] K.M. Mahoney, P.D. Rennert, G.J. Freeman, Combination cancer immunotherapy and new immunomodulatory targets, Nat. Rev. Drug Discov. 14 (8) (2015) 561–584.
[531] B. Al-Lazikani, U. Banerji, P. Workman, Combinatorial drug therapy for cancer in the post-genomic era, Nat. Biotechnol. 30 (7) (2012) 679–692.
[532] X. Tan, L. Hu, L.J. Luquette, G. Gao, Y. Liu, H. Qu, R. Xi, Z.J. Lu, P.J. Park, S.J. Elledge, Systematic identification of synergistic drug pairs targeting HIV, Nat. Biotechnol. 30 (11) (2012) 1125–1130.
[533] A. Ballesta, P.F. Innominato, R. Dallmann, D.A. Rand, F.A. Levi, Systems chronotherapeutics, Pharmacol. Rev. 69 (2) (2017) 161–199.
[534] K.C. Bulusu, R. Guha, D.J. Mason, R.P. Lewis, E. Muratov, Y.K. Motamedi, M. Cokol, A. Bender, Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives, Drug Discov. Today 21 (2) (2016) 225–238.
[535] C. Lipinski, A. Hopkins, Navigating chemical space for biology and medicine, Nature 432 (7019) (2004) 855–861.
[536] K. Han, E.E. Jeng, G.T. Hess, D.W. Morgens, A. Li, M.C. Bassik, Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions, Nat. Biotechnol. 35 (5) (2017) 463–474.
[537] Y. Sun, Z. Sheng, C. Ma, K. Tang, R. Zhu, Z. Wu, R. Shen, J. Feng, D. Wu, D. Huang, D. Huang, J. Fei, Q. Liu, Z. Cao, Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer, Nature Commun. 6 (2017) 8481.
[538] M. Bansal, et al., A community computational challenge to predict the activity of pairs of compounds, Nat. Biotechnol. 32 (12) (2014) 1213–1222.
[539] A. Zimmer, I. Katzir, E. Dekel, A.E. Mayo, U. Alon, Prediction of multidimensional drug dose responses based on measurements of drug pairs, Proc. Natl. Acad. Sci. USA 113 (37) (2016) 10442–10447.
[540] B.A. Kidd, A. Wroblewska, M.R. Boland, J. Agudo, M. Merad, N.P. Tatonetti, B.D. Brown, J.T. Dudley, Mapping the effects of drugs on the immune system, Nat. Biotechnol. 34 (1) (2016) 47–54.
[541] J.H. Woo, Y. Shimoni, W.S. Yang, P. Subramaniam, A. Iyer, P. Nicoletti, M.R. Martínez, G. Lopez, M. Mattioli, R. Realubit, C. Karan, B.R. Stockwell, M. Bansal, A. Califano, Elucidating compound mechanism of action by network perturbation analysis, Cell 162 (2) (2015) 441–451.
[542] L. Kalmanti, et al., Safety and efficacy of imatinib in CML over a period of 10 years: data from the randomized CML-study IV, Leukemia 29 (5) (2015) 1123–1132.
[543] J. Loscalzo, Systems biology and personalized medicine: a network approach to human disease, Proc. Am. Thorac. Soc. 8 (2) (2011) 196–198.
[544] C. Savoia, M. Volpe, G. Grassi, C. Borghi, E.A. Rosei, R.M. Touyz, Personalized medicine - a modern approach for the diagnosis and management of hypertension, Clin. Sci. 131 (22) (2017) 2671–2685.
[545] L. Chen, J. Wu, Bio-network medicine, J. Mol. Cell Biol. 7 (3) (2015) 185–186.
[546] M.E. Cusick, et al., Literature-curated protein interaction datasets, Nature Methods 6 (1) (2009) 39–46.
[547] S. Li, Mapping ancient remedies: applying a network approach to traditional Chinese medicine, Science 350 (6262) (2015) S72–S74.
[548] J. Zhao, C. Lv, Q. Wu, H. Zeng, X. Guo, J. Yang, S. Tian, W. Zhang, Computational systems pharmacology reveals an antiplatelet and neuroprotective mechanism of Deng-Zhan-Xi-Xin injection in the treatment of ischemic stroke, Pharmacol. Res. 147 (2019) 104365.
[549] J. Zhao, J. Yang, S. Tian, W. Zhang, A survey of web resources and tools for the study of TCM network pharmacology, Quant. Biol. 7 (1) (2019) 17–29.