J Biosci (2022)47:24
DOI: 10.1007/s12038-022-00253-y

Review

Emerging landscape of molecular interaction networks: Opportunities, challenges and prospects

Gauri Panditrao[1], Rupa Bhowmick[1,2], Chandrakala Meena[1] and Ram Rup Sarkar[1,2]*

[1]*Chemical Engineering and Process Development Division, CSIR-National Chemical Laboratory, Pune 411008, India*

[2]*Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India*

**Corresponding author (Email, rr.sarkar@ncl.res.in)*

MS received 10 September 2021; accepted 3 December 2021

Network biology finds application in interpreting molecular interaction networks and providing insightful inferences using graph theoretical analysis of biological systems. The integration of computational bio-modelling approaches with different hybrid network-based techniques provides additional information about the behaviour of complex systems. With increasing advances in high-throughput technologies in biological research, attempts have been made to incorporate this information into network structures, which has led to a continuous update of network biology approaches over time. The newly minted centrality measures accommodate the details of omics data and regulatory network structure information. The unification of graph network properties with classical mathematical and computational modelling approaches and technologically advanced approaches like machine-learning- and artificial intelligence-based algorithms leverages the potential application of these techniques. These computational advances prove beneficial and serve various applications such as essential gene prediction, identification of drug–disease interaction and gene prioritization. Hence, in this review, we have provided a comprehensive overview of the emerging landscape of molecular interaction networks using graph theoretical approaches. With the aim to provide information on the wide range of applications of network biology approaches in understanding the interaction and regulation of genes, proteins, enzymes and metabolites at different molecular levels, we have reviewed the methods that utilize network topological properties, emerging hybrid network-based approaches and applications that integrate machine learning techniques to analyse molecular interaction networks. Further, we have discussed the applications of these approaches in biomedical research with a note on future prospects.

Keywords. Centrality; disease mechanisms; hybrid network-based models; machine learning; molecular interaction networks; network topology; systems biology

1. Introduction

Molecular entities of biological systems interact with each other at various levels and cooperatively function together to exhibit specific cellular phenotypes. Essentially, every biological entity interacts with other biological entities, forming a network of interactions that maintains the proper functioning of biological systems. The network properties of these biological interactions provide us the opportunity to model biological systems as different types of networks such as protein–protein interactions (Kumar *et al.* 2020a, b; Tomkins and Manzoni 2021), gene regulatory (Grimes *et al.* 2019; Sinha *et al.* 2020a, b) and metabolic networks (Bidkhori *et al.* 2018; Toubiana *et al.* 2019). The availability of a vast expanse of molecular information with the increase in omics analyses has created avenues

This article is part of the Topical Collection: Emergent dynamics of biological networks.

http://www.ias.ac.in/jbiosci
Published online: 20 April 2022

to better understand these molecular interactions (Hawe *et al.* 2019). In this context, systems biology aims to understand biological entities at the system level by studying them not only as discrete components but also as interacting systems with emergent properties. Network biology allows the representation and analysis of biological systems using tools derived from graph theory (Barabási and Oltvai 2004).

Different types of information of biological systems can be represented as networks of mutually interacting entities, where each entity has an effect on the overall function of these networks. The definition of the nodes and edges used in a network representation depends on the type of data used to build the network. In a molecular interaction network, the nodes can be represented by genes, proteins, enzymes, metabolites, transcription factors, etc., and the edges can be represented as the interaction between these nodes. Different types of data produce different network characteristics in terms of structure, connectivity, and complexity, where edges and nodes potentially mediate multiple layers of information (Grennan *et al.* 2014). The integration of constantly evolving complex high-throughput data, such as whole genome sequencing data, single-cell ribonucleic acid (RNA) sequencing, and clustered regularly interspaced short palindromic repeats (CRISPR-cas9) technology, and their ease of availability have further led to improvements and advances in newer techniques and approaches as well as upgradation of the traditional network biology approaches to carry out systems-level studies (Charitou *et al.* 2016; Koh *et al.* 2019; Ma and Zhang 2019).

Due to incomplete information, variability in data resources, and heterogeneity, multiple challenges have been consistently observed while carrying out systems-level studies. Prospects created by molecular interaction networks are instrumental in addressing these challenges (Imam *et al.* 2015). One of the prevalent challenges that limit the applicability of network models is the difficulty in identifying appropriate centrality measures due to variability in the type of molecular network. Universal acceptance of the centrality–lethality hypothesis remains inconsistent owing to the changing network structure and topology of the molecular interaction networks. The centrality measure that defines the central or the most influential nodes in the network changes with changing network structure. The challenge is to identify proper centrality measures that appropriately identify these central nodes (Oldham *et al.* 2019). Furthermore, uncertainty in model structure and parameters that affect the network inferences is an additional challenge in the case of gene regulatory networks (GRNs) (Saint-Antoine *et al.* 2020). One promising avenue is created by hybrid network-based modelling approaches in the analysis of these molecular interaction networks. These approaches are an improvement over the systems modelling methods as they integrate and use network topological properties and also implement advanced computational techniques, such as machine learning-based algorithms, to tackle the aforementioned challenges (Chowdhury *et al.* 2013; Chowdhury and Sarkar 2019; Kang *et al.* 2020; Nandi *et al.* 2020). These, in turn, provide an opportunity to scale up the dynamic genome-scale models for incorporation with network biology and are currently being explored (Stéphanou and Volpert 2016; Bardini *et al.* 2017). The emerging landscape of these molecular interaction networks has enabled better understanding of molecular systems and their prospects (Charitou *et al.* 2016). A comprehensive review that discusses this emerging landscape of molecular interaction networks in the light of the challenges faced and the new approaches and techniques developed in this area is missing.

Hence, in this review, we aim to thoroughly assess available network biology approaches and the progress made in them to decipher the understanding of the molecular interaction networks over the last two decades, with the advancement in biological and computational research. We begin with a brief introduction of the different types of molecular interaction networks, and their structural and topological properties, so as to provide a basic understanding of the molecular interaction networks. In section 3, we discuss several network topology-based methods and their progress in effectively drawing various types of inferences from different types of molecular interaction networks. In each of the following subsections, we briefly highlight how these methodological advances have helped to overcome different limitations and also introduce some recently developed methods that can be useful in future research. Next, we introduce the hybrid network-based approaches that combine traditional systems biology methods with graph theoretical techniques and briefly discuss their applications. We also explain how advanced statistical methods and machine learning (ML)-based computational frameworks help to overcome the limitations of these hybrid approaches. We briefly state a few applications where these recently developed methodologies successfully contribute to advance the molecular network analyses. In section 6, we discuss a few disease-specific studies where these network-based approaches have successfully contributed. We aim to highlight the increasing forte of

network topology-based techniques to analyse molecular-interaction networks with methodological advancements. Finally, we conclude this review by providing suggestions to the readers for possible future prospects in areas that hold scope for advanced molecular network analyses to improve biological and biomedical research. This review will be beneficial to systems biologists who can use emerging graph theoretical approaches, hybrid network-based models, and ML-based applications to study molecular interactions. Network biologists can also gain a holistic view of applications of these emerging approaches, such as deciphering drug–disease interactions, analysing perturbation patterns, characterizing regulatory genes, predicting gene essentiality, etc.

2. Types of molecular interaction networks

Various types of molecular interaction networks emerge from the combination of different interactions among molecular entities that determine the systems-scale behaviour of the cell (Barabási and Oltvai 2004; Han 2008). Some of the most common molecular interaction networks are: (i) protein–protein interaction networks, (ii) metabolic networks, (iii) gene regulatory networks, and (iv) signal transduction networks. In the graph G (V, E) representation of molecular interaction networks, nodes $v \in V$ represents biological entities, i.e., genes, proteins, transcription factors, or miRNAs, and edges $e \in E$ represents interactions among these biological entities. Due to their importance in biological research, these molecular interaction networks are continuously revisited and updated with time. We provide a brief introduction of the different types of molecular interaction networks in the following subsections.

2.1 *Protein–protein interaction networks*

These are mathematical representations of the molecular contacts between the proteins in a cell. These contacts are specific, occur between defined binding regions in the proteins, and have a particular biological meaning (i.e., they serve a specific function) (Schreiber 2021). Protein–protein interactions (PPIs) are essential to almost every process in a cell and play an important role in drug development (Mabonga and Kappo 2019). The interactome is the totality of PPIs that occur in a cell, an organism, or a specific biological context (Safari-Alighiarloo *et al.* 2014).

Knowledge of PPIs can be extended to a wide range of applications, such as understanding complex disease disorders, assigning putative roles to uncharacterized proteins (Lv *et al.* 2015; da Costa *et al.* 2018), and adding fine-grained details about the steps within a signalling pathway (Navlakha *et al.* 2012; Mei and Zhu 2015). Identifying active signalling pathways (Kabir *et al.* 2018), and characterizing the relationships between proteins that form multi-molecular complexes, such as the proteasome (Di Paola *et al.* 2015), are some additional areas where PPI networks find application. Some of the widely popular protein–protein interaction network (PPIN) resources actively used for mining PPI information include the Biological General Repository for Interaction Datasets (BioGRID) (Oughtred *et al.* 2019) and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Szklarczyk *et al.* 2021) Database.

2.2 *Metabolic networks*

These consist of chemical reactions that involve the catalytic conversion of small biomolecules known as metabolites aided by enzymatic reactions. Construction of the network depends on several factors, most importantly the type of analysis to be performed on the network. The most common graph theoretical representation of metabolic networks is considering the metabolites as nodes and the reactions catalysing the conversion of one metabolite to another as edges. Another way is to represent the metabolic network as a reaction adjacency graph, where the nodes are formed by reactions and the connection/edges between the reactions is established if the product of one reaction is the substrate of the second reaction (Kim *et al.* 2019). Metabolite concentration and reaction fluxes are measurable quantities that have been used to infer the properties of the metabolic graph networks at the structural, kinetic, and regulatory levels (Beguerisse-Díaz *et al.* 2018).

2.3 *Gene regulatory networks*

A gene regulatory network (GRN) represents the complex mechanisms that regulate the expression of genes. Regulatory mechanisms occur at different stages of protein production from DNA, such as during the transcription, translation, and splicing phases. Proteins act as both the product and the controller of gene expression in these networks (Junker and Schreiber

2007). In GRNs, each node represents a gene, and a directed link between two genes implies that one gene directly regulates the expression of the other without intervention by any other genes. GRNs are very important in understanding the mechanistic regulation of gene expression and the sequence of events that result in a phenotype. The graphical representation of these networks provide a visualization and intuitive explanation of these complex and interconnected mechanisms.

2.4 *Signal transduction networks*

These are essentially protein interaction networks, but the interaction between the proteins and flow of information is directional. Like the PPI network, the nodes in the signal transduction network are represented by proteins that usually belong to the phosphatase or kinase or similar protein family, e.g., protein tyrosine phosphatases (PTPs), protein serine phosphatases (PSPs), and mitogen-activated protein kinases (MAPKs) (Nguyen *et al.* 2013). The edges are determined by the interaction between two proteins where the first protein interacts to activate the second one, and hence the directionality. Signal transduction networks form the core of information flow for most of the signalling networks. These form the bridge between the receptor-mediated activation of protein complexes whose information is passed down for the activation of transcription factors via these signal transduction networks (Soyer *et al.* 2006).

3. Network topology-based approaches in the study of molecular interaction networks

Representation of molecular interaction data in the form of an interconnected network of biomolecules forms the topology of the information in these networks. This topology is created from the representative graphs, where nodes represent singular entities or the individual biomolecules of the process under study, and the edges represent the relation between them. Elucidation of the topology of these networks is effectively applied to gain a systems-level understanding of the interactive exchange between the entities of the biological system under study (Janjić and Pržulj 2012; Koutrouli *et al.* 2020; Masoomy *et al.* 2021). For example, a recent study used topological analysis on a systems-level curcumin-rewired PPI based on centralities like betweenness and degree, to identify key regulatory proteins that govern the molecular mechanisms, thereby aiding in understanding the anti-cancerous and anti-inflammatory properties of curcumin (Dhasmana *et al.* 2020).

Network topology-based approaches have also helped understand host–pathogen interplay during infection processes (Mulder *et al.* 2014; Saha *et al.* 2018). Recently, Panditrao *et al.* (2021) used betweenness centrality combined with shortest-path analysis to analyse phenotype-specific protein subnetworks of *Leishmania donovani* secretory proteins to delineate infection mechanisms and identify regulatory host proteins that could potentially act as immunomodulatory candidates. Also, the study of molecular networks of SARS-CoV2 during the COVID-10 pandemic has been instrumental in deciphering its viral pathogenesis through virus-host PPI networks (Díaz 2020; Gordon *et al.* 2020; Messina *et al.* 2020). The topological properties of virus–host protein interaction networks have aided in understanding the mechanisms of its pathogenesis. Centrality measures like PageRank, betweenness, eigenvector centrality along with weighted k-shell decomposition analysis have helped identify the most influential nodes of the viral proteins that interfere with the host nucleocytoplasmic trafficking, immune system, and cell cycle which facilitates pathogenesis (Kumar *et al.* 2020a, b). It has also been possible to identify candidate target viral genes for repurposing drugs for treating the COVID-19 infection through the analysis of the fused viral interaction network (VIN) and the drug–target interaction network (DTI) (Zambrana *et al.* 2021). In this analysis, the network structure topological information was utilized for data fusion and the graphlet degree vector (GDV) was used for capturing the local rewiring patterns for functional assessment of gene–drug interactions. Graph theoretical approaches reveal hidden properties and features in molecular interaction networks (Pavlopoulos *et al.* 2011). Thus, such topological network analyses enable several applications such as discovery of drug targets, evaluation of disease genes, and prediction of essential nodes. In this section, we discuss several emerging methodological advances that utilize topology-based approaches to draw various insightful inferences from molecular interaction networks.

3.1 *Centrality measures*

In the past two decades, the use of centrality measures in molecular interaction networks has gained momentum.

Centrality measures depend on the topology-centric parameter of the nodes in the network which would influence the structural properties of the graph. They are instrumental in deducing meaningful interpretations of molecular interaction networks that include PPINs (Ashtiani *et al.* 2018), GRNs (Liseron-Monfils and Ware 2015), signal transduction networks (Alvarez-Ponce *et al.* 2017), and metabolic networks (Resendis-Antonio *et al.* 2012). The commonly used centrality measures include degree, betweenness centrality, closeness centrality, eccentricity, and eigenvector centrality, often referred to as the classical centrality measures (figure 1A). The basic definition and mathematical formulation of the classical centrality measures are provided in table 1.

The past two decades saw a surge in various newly minted concepts of calculating centralities (Jalili *et al.* 2015, 2016). Previous studies show that integration of omics data with topological features can develop improved centrality measures (Li *et al.* 2012, 2010). These centralities are developed through a combination of the classical centralities, utilizing topological features based on connectivity as well as integrating known biological information from experimental outputs. We provide an overview of these newly developed centrality measures in table 2. Figure 1B provides a pictorial representation of how these centralities have evolved from the classical centralities by including additional molecular information. Based on the methods used to derive these new centralities, they can be broadly classified. PageRank, marginal essentiality, subgraph centrality, motif-based centrality, bridging centrality, pairwise disconnectivity index, flux centrality, leverage centrality, perturbation centrality and SSC (source/sink centrality) use solely topological features of the nodes in the network that are based on their connectivity patterns. Annotation transcriptional centrality, neighbourhood functional centrality, game theoretic centrality, DiffSLC and SCNrank (spectral clustering for network-based ranking) additionally integrate biological information in the form of omics data. Perturbation centrality and game centrality are the centrality measures that can be applied to dynamic networks.

Depending on the problem of interest, a certain type of centrality measure may be more important than another. For example, the highest betweenness centrality node in the network has a more pronounced influence on control than the highest closeness centrality node in the network if one wants to control a chaotic metapopulation to the steady states (Meena *et al.* 2017), to protect the resilience of the dynamical networks (Rungta *et al.* 2018), irrespective of the dynamics on the nodes (Meena *et al.* 2020a, b).

Limitations in using these centralities exist in terms of the biological inferences in molecular interaction networks. For example, the universality of the centrality–lethality hypothesis becomes questionable if the measures to identify central nodes in the network are not chosen wisely. In 2005, Hahn and Kern showed that hub proteins in a PPIN were highly essential (Hahn and Kern 2005); however, shortly thereafter Mahadevan and Palsson showed that essentiality was not correlated to the connectivity of the node in GRNs (Mahadevan and Palsson 2005). This idea was further supported by studies which showed that in PPINs, low connectivity could also be considered essential (Tew *et al.* 2007). Hence, the centrality–lethality paradox still exists.

These above-mentioned new centrality measures are derived and continuously added to address the limitations previously faced by the classical centrality measures and to improve our ability to extract more useful information from complex molecular networks (Roy 2012). For example, leverage centrality is observed to be highly useful in analysing hierarchical networks, such as brain networks, which harbour assortative behaviour, as it helps to identify nodes which are critical for the function of the global network as well as local communities (modules) in that network (Joyce *et al.* 2010). This helps capture the assortative or disassortative behaviour of the network since it captures the nodes which control the quality of information received by its neighbours. Classical centralities like degree, betweenness, and eigenvector centrality fail to analyse this assortative behaviour. This is possible as leverage centrality relies on the principle that a node in the network is central if its immediate neighbours rely on that node for information. Game centrality is another example that effectively contributes to the identification of functionally and dynamically important network nodes, which has previously been a difficult task (Simko and Csermely 2013). It significantly outperforms the classical centrality measures in predicting genetic buffering of evolutionary changes, i.e., the contribution of a protein to the overall robustness of the cell. This was possible due to the ability of game centrality to precisely discriminate hubs with different dynamic parameters. Degree centrality, although able to capture essential proteins, is highly dependent on the degree-sorted nodes list and thus misses out on a few other known essential proteins that have fewer interactions. DiffSLC promotes potentially essential proteins with low node degree by elevating eigenvector centrality values with additional weights from co-expression data (Mistry *et al.* 2017). The source/sink centrality is a useful measure if one wishes to clearly distinguish and identify *a priori* important genes from pathways, as it accounts for the importance of

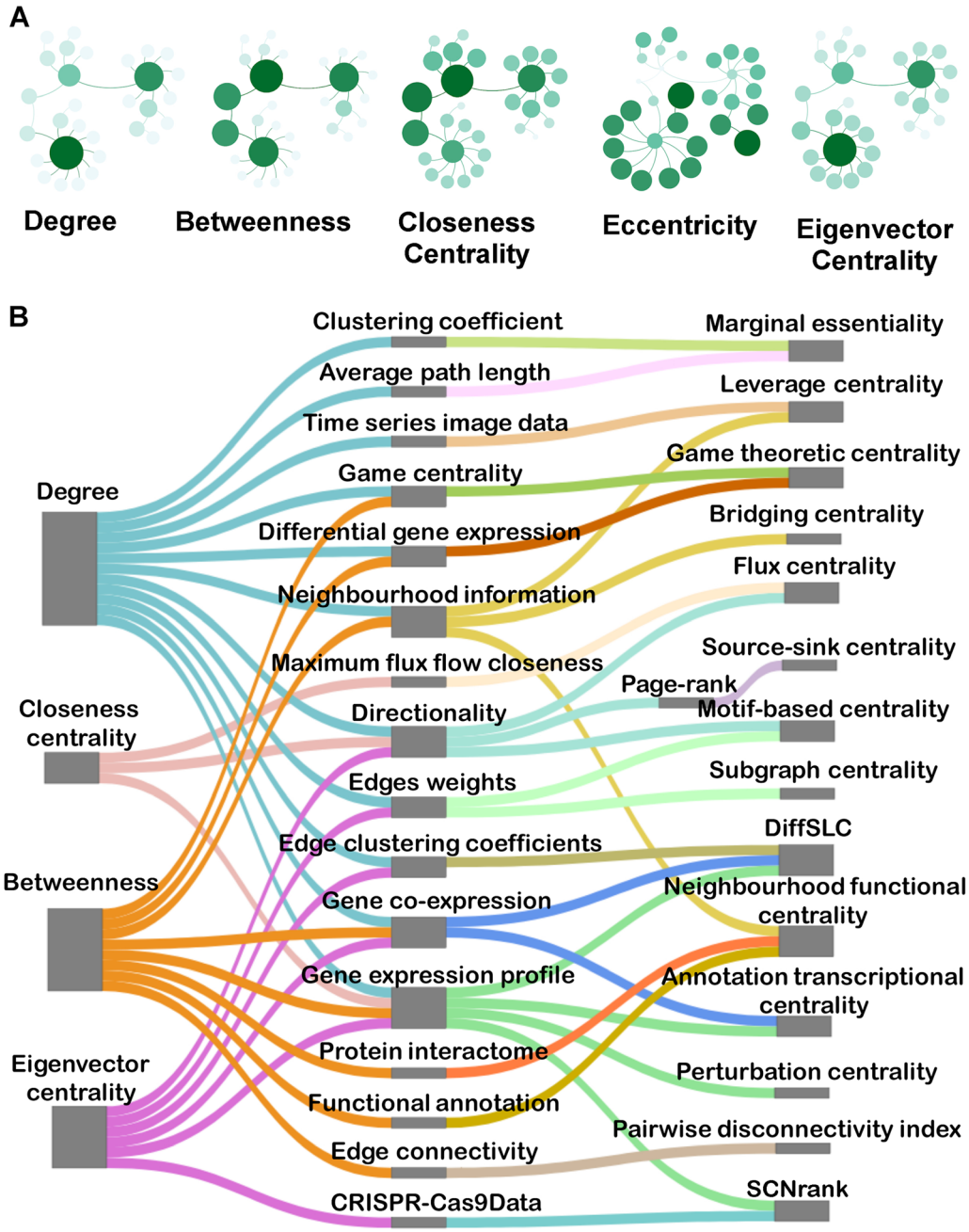

Figure 1. Centrality measures in molecular interaction networks. (**A**) A schematic representation of the commonly used centrality measures, where the colour gradient and the size of the nodes correspond to the respective centrality value for that node in the representative network. Highest centrality in the network is the largest size and darkest colour of the node. (**B**) A Sankey plot representation of some of the new centrality measures (shown in the right side of the plot) developed in the recent past which are derived from the traditionally used basic centrality measures (shown in the left side of the plot) by incorporating various types of OMICs and network data (shown in the centre of the plot).

pathway elements with respect to the upstream and downstream positions (Naderi Yeganeh *et al.* 2020).

Thus, advancements in these centrality measures continue to prove to be valuable tools that can tackle the complexities and limitations in the emerging molecular networks and help to gain meaningful inferences.

3.2 *Methods for integrating large-scale genomic screen data in PPINs*

Large-scale genomic screens involve functional genomic approaches employing diverse experimental techniques such as transcriptome profiling, loss-of-function screening, RNA interference screens, and CRISPR

Table 1. Definition and mathematical representation of classical network centrality measures

Centrality	Definition	Mathematical representation	References
Betweenness Centrality (BC)	Ratio of the number of shortest paths passing through a node v out of all the shortest paths existing between all node pairs in a network	$BC(v) = \sum_{i \neq j \neq v \in V} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$ σ_{ij}= number of shortest paths between node i and j $\sigma_{ij}(v)$ = number of shortest paths in σ_{ij} that pass through the node v	Freeman (1977)
Closeness Centrality (CC)	A measure of the average farness of a node v from all other nodes belonging to V	$CC(v) = \frac{1}{\sum_{t \in V \vee \{v\}} d(v,t)}$ $d(v,t)$ = distance between the node v from t and $i \in V$	Freeman (1978)
Degree Centrality (DC)	The most classical centrality measure that represents the number of connected neighbours of a node v	$DC(v) = N(v)$ $N(v)$ = number of nodes connected to the node $v \in V$	Proctor and Loomis (1951)
Eccentricity (Ecc)	A measure of proximity of node v to all other nodes belonging to V in G. A node v with high eccentricity means all other nodes are in proximity to node v	$Ecc(v) = \frac{1}{max\{dist(v,i) \vee v,i \in V\}}$ $dist(v,i)$= distance between the node v and node i, where v, $i \in V$	Hage and Harary (1995)
Eigenvector Centrality (EC)	A measure of the influence of a node. EC of a node v is defined as the weighted sum of the centralities of all the nodes that are connected to v by an edge in an adjacency matrix $A_{v,i}$	$EC(v) = \frac{1}{\lambda} \sum_{i \in V} A_{v,i}.EC(i)$ $A_{v,i} = (a_{v,i})$= adjacency matrix; if node v and i are connected, then $a_{v,i} = 1$, otherwise 0 $EC(v)$ = eigenvector value of node v associated with the eigenvalue λ of A	Ruhnau (2000)

libraries to investigate gene functions. Molecular interaction networks potentiate the identification of functionally related biological components from large-scale genomic screen datasets. Genomic information is incorporated into network topology using the principle of guilt-by-association to develop network-based scoring methods that allow detection of false-positive and false-negative screens (Wang *et al.* 2009). These genome-scale interaction networks that harbour co-localized genes or proteins having similar topological roles are more likely to be functionally correlated. Thus, the 'guilt-by-association' process allows inferring properties of unknown proteins or genes by transferring knowledge from these similar co-localized genes or proteins. This principle can be effectively used for identification of drug–target interaction (Li *et al.* 2016), integrating gene regulatory pathways (Grimes *et al.* 2019). Furthermore, incorporation of network neighbour information improves the quality of functional genomic screens (Jiang *et al.* 2015), and the selection of an optimally functionally enriched network allows easier identification and interpretation of diagnostic or predictive gene signatures for diseases (Kairov *et al.* 2012). These strategies enable better utilization of genomic screening data in conjunction with more topological properties to associate the information with phenotypes. For example, recently, Rubanova and co-workers developed a new method called MasterPATH, which utilizes the results from functional screening data such as loss-of-function data and uses shortest path-based subnetwork extraction to elucidate members of molecular pathways that influence the studied phenotypes (Rubanova *et al.* 2020).

In an attempt to address the challenges of accurately integrating cell-line and CRISPR-Cas9 data within the network structure, a method called SCNrank has been developed. This method prioritizes potential drug targets in tumour cell-line screens by combining expression profiles from tumour tissues, normal tissues and cell-lines, PPI network and CRISPR-cas9 data to construct tissue-specific networks that are aligned based on graph structure similarity (Liu *et al.* 2020). The Network-augmented Gene Set Enrichment Analysis (NGSEA) method has been developed to utilize the information from Gene Set Enrichment Analysis (GSEA) of functional networks by calculating the enrichment score for gene sets using expression difference not only for individual genes but also from

Table 2. Description of the new developed centrality measures

New centrality measures	Description	Additional features	Year of development	References
Page Rank Centrality	A variant of eigenvector centrality which scores a vertex as a fraction of time spent visiting that vertex measured over time in a random walk over all vertices in the network with an additional probability for jumping to any vertex	Uses directionality in the network	1998	Brin and Page (1998)
Marginal Essentiality	A quantitative measure that calculates the importance of no-essential gene to a cell using protein networks by using the topological information in the form of local interconnectivity measures	Characterized by two classical centralities, namely, degree and clustering co-efficient	2004	Yu *et al.* (2019)
Subgraph Centrality	Quantifies the participation of a vertex in all subgraphs of a network. Subgraph centrality (SC) of a node is a weighted sum of the numbers of all closed walks of different lengths in the network starting and ending at the node	Built over eigenvector centrality and uses edge weights	2005	Estrada and Rodríguez-Velázquez (2005)
Neighbourhood Functional Centrality	Quantifies the extent to which a protein is surrounded by functionally consistent neighbouring proteins in a PPI network and thus help in mining lethal proteins	Uses protein interactome, neighbourhood information and functional annotation	2007	Tew *et al.* (2007)
Motif-based Centrality	Calculates the occurrence of motifs (subnetwork patterns in local interconnections) by calculating motif match sets to analyse gene regulatory and protein interaction patterns	Uses directionality and edge weights	2007	Koschützki *et al.* (2007)
Bridging Centrality	Identification of a node or edge that is located between and connects modules (modular subregions) in a network. Based on the principle that the number of edges entering or leaving the direct neighbour subgraph of a vertex is high at the bridge	Uses neighbour subgraph information and betweenness centrality	2008	Hwang *et al.* (2008)
Pairwise Disconnectivity Index	Calculated for a vertex as a fraction of those initially connected pair of vertices in a directed network which becomes disconnected after removal of the vertex from the network. It evaluates the importance of an individual element in a network for sustaining the communication ability between connected pairs of a directed network	Built upon betweenness centrality and uses edge connectivity	2008	Potapov *et al.* (2008)
Annotation transcriptional centrality	Delimits representative functional domains in co-expression networks and uses this information to identify key nodes (proteins/genes) that modulate the propagation of functional influences within the network	Integrates gene expression profiles with co-expression network information for centrality calculation	2010	Prifti *et al.* (2010)
Flux centrality	Developed on the concept of maximum flow which is defined by the largest flow that is observed for all possible paths between the two vertices in a network using shortest path closeness centrality	Built upon the closeness centrality by integrating directionality and maximum flux flow closeness	2010	Koschützki *et al.* (2010)
Leverage Centrality	For a vertex in a network, it determines the extent to which its immediate neighbouring vertices reply on the vertex for information	Built upon eigenvector and degree centrality by integrating neighbourhood information	2010	Joyce *et al.* (2010)

Table 2. (Continued)

New centrality measures	Description	Additional features	Year of development	References
Perturbation Centrality	A measure of dynamic network centrality which uses weighted degree and is defined as the reciprocal of silencing time retrieved by using a Dirac delta type starting perturbation of 10n units, where n is the number of nodes in the network, using a dissipation value of 1	Identifies intermodular hubs from dynamic networks	2013	Szalay and Csermely (2013)
Game Centrality	A dynamic centrality measure that measures the individually defecting nodes in the network to convert other nodes in the network to its own strategy	Applies strategy update rule on the network dynamics data	2013	Simko and Csermely (2013)
DiffSLC	Combines multiple centrality measures and exploits the advantage of eigenvector centrality and edge clustering co-efficient to identify essential genes/proteins. This centrality is a weighted combination of eigenvector centrality and co-expression biased degree centrality	Gene co-expression values are used in conjunction with eigenvector and edge clustering co-efficient	2017	Mistry *et al.* (2017)
Game Theoretic Centrality	Determines a gene's contribution to the overall connectivity of its corresponding node in the network by calculating the gen's synergistic influence in a gene-to-gene interaction network	Uses differential gene expression profile (GWAS data) and is built upon game centrality	2020	Sun *et al.* (2020)
Source-Sink centrality (SSC)	This centrality measures the importance of the node separately in the upstream and downstream of a pathway, as sender and receiver of biological signals	Built upon the PageRank centrality by incorporating directionality	2020	Naderi Yeganeh *et al.* (2020)
Spectral clustering for Network-based Ranking (SCNrank)	Centrality developed for network-based target ranking based on spectral clustering. It calculates a target influential score by integrating PPI, CRISPR-cas9 and gene expression data to calculate the influence of target towards a cluster it belongs to for ranking and scoring each drug target	Integrates data from three sources: gene expression, protein–protein interaction network and CRISPR-cas9 data	2020	Liu *et al.* (2020)

their neighbours. This method facilitates the repurposing of approved drugs with pathway interpretation of gene expression phenotypes (Han *et al.* 2019).

3.3 *Methods for analysing metabolic networks*

The inferences of graph theoretical analyses of metabolic systems are highly dependent on graph construction, which includes multiple options to define the nodes and edges of the network. Generally, in metabolic networks, nodes are represented as metabolites and edges as reactions, or vice versa, in reaction adjacency graphs. In bipartite networks, nodes can be metabolites or can also represent reactions. Metabolic graph networks are essentially directional. However, it was observed that directionality information, which was considered the sole defining factor of metabolic graph networks, might not be the only defining factor for the metabolic function (Wagner and Fell 2001). Previously, metabolic reaction graphs had limitations in analysing context-specific metabolic events under different growth conditions of the cell (Sauer *et al.* 1999).

In conjunction with this limitation is the challenge that arises from the reversibility of these metabolic graphs. The prior approaches do not ensure justice by generalizing the direction of reaction based on one condition and also do not take into consideration varying physiological conditions (Wagner and Fell 2001). Beguerisse-Díaz's group addressed this complexity by emphasizing the inclusion of directionality information, as well as capturing environment-specific metabolic connectivity (Beguerisse-Díaz *et al.* 2018). This approach accounts for the utilization of metabolic directionality for representing the natural flow of chemical mass from reactants to products. It provides a flux-based strategy using Flux Balance Analysis (FBA)-based solutions to build different metabolic graphs under different growth conditions, creating opportunities to convert genome-scale metabolic models into directed graphs. Further, to improve upon the above-discussed challenges, several tools that employ network science to improve FBA pipelines have been consistently explored (Lewis *et al.* 2012). The integration of graph theory with FBA by constructing flux-weighted graphs is recently being proposed as a promising solution to overcome these previous shortcomings (Dusad *et al.* 2021) and holds application in several areas of industrial biotechnology such as maximizing production from metabolic cell factories and dynamic control of gene expression (Brockman and Prather 2015; de Lorenzo *et al.* 2018; Liu *et al.* 2018).

Several tools have been developed in the recent past in an effort to improve and provide a seamless experience in developing and executing pipelines for simulation of metabolic networks and also to integrate several other optimizing functions which include FBA-based solutions (Ebert *et al.* 2012; Rowe *et al.* 2018). A useful tool which can assist in carrying out FBA and several other network analysis for metabolic networks is MetaNET (Narang *et al.* 2014). Along with simulation studies, it also provides an option to conduct topological analysis.

Recent studies implement global centrality measures such as in-degree, out-degree, closeness centrality, and modularity for directed networks rather than a centralized focus on only the high-degree nodes, to identify targets in metabolic pathways (Newman 2006; Kim *et al.* 2019). For topological analyses, reaction-centric bipartite graphs that use centrality metrics independent of a node's degree are being explored (Kim *et al.* 2019). These studies focus on calculating the influence of a node on the downstream flow of information in the network by calculating the bridging centrality and cascade number (Kim *et al.* 2019) (figure 2A). These newly developed approaches help in the analyses of directed reaction graphs to prioritize the nodes and their associated genes, which are essential for global and local connectivity and can help identify crucial targets in metabolic engineering. Development of network-based approaches for understanding the evolution of metabolic genes at different evolutionary timescales that tackle the challenge of procuring the gene's likelihood to be under adaptive selection is another emerging application of network biology in understanding metabolic networks (Dobon *et al.* 2019).

3.4 *Methods for integration of gene co-expression networks*

The incorporation of omics datasets in computational networks using correlational analysis is gaining momentum. The development of the Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath 2008) to identify highly correlated genes, or eigengene-based highly correlating gene clusters, has paved the way for transcriptomic data integration and analysis for relating modules to one another and for measuring module membership (Niu *et al.* 2019). The co-expression network analyses have garnered applications in uncovering various disease

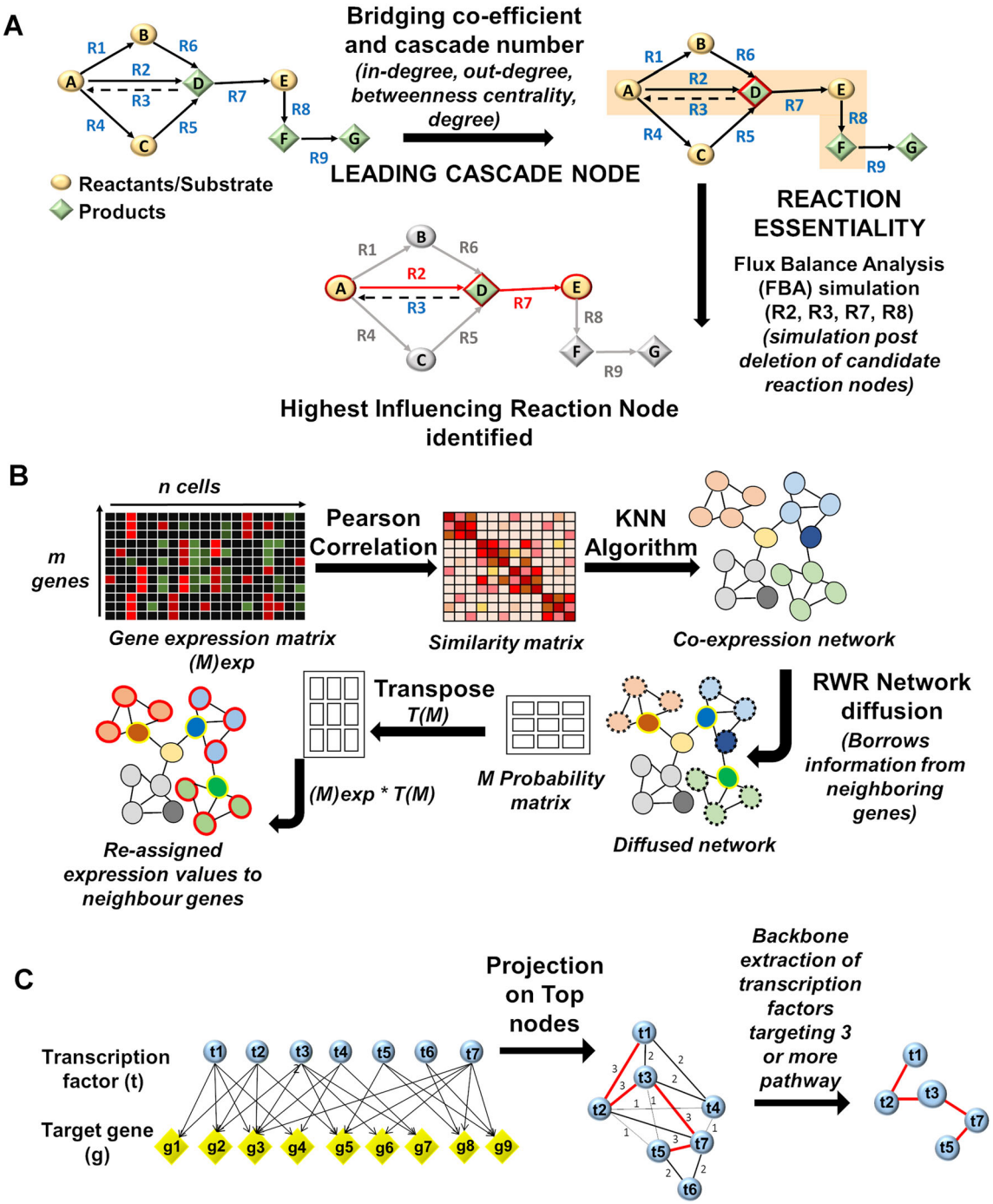

Figure 2. Schematic representations of recent methodological developments in molecular network analyses techniques. (**A**) Cascade number representing local controllability of the node demonstrated on a schematic metabolic network model which measures the influence of a node on its downstream flow of information and identifies the highest influencing reaction node. (**B**) Graphical workflow representation of the netImpute algorithm which employs diffusion of co-expression network to improvise the dropout issue in single-cell data. (**C**) Backbone extraction: A schematic representation through a toy transcription factor regulatory bipartite graph demonstrating the working of the backbone extraction technique to extract a subset of important regulatory transcription factors.

mechanisms, from identification of rheumatoid arthritis-related diagnostic genes, potential disease genes, and vital microRNAs (Ren *et al.* 2021) to biomarkers and prognostic signatures in multiple cancers (Kadkhoda *et al.* 2020; Terkelsen *et al.* 2020). The previously used co-expression network analyses methods could not distinguish between the regulatory and regulated genes or provide information on

causality. Improvements have been made in these network studies by including differential co-expression analysis performed under different regulatory conditions (Chowdhury *et al.* 2020). Gene Sets Net Correlation Analysis (GSNCA) is another method that has enabled analysis of differentially co-expressed pathways by inferring differences in co-expression networks (Rahmatallah *et al.* 2014). Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) (Margolin *et al.* 2006) and Gene Network Inference with Ensemble of trees (GENIE3) (Huynh-Thu *et al.* 2010) are popular tools used to construct regulatory networks from co-expression data. The Generalised Single Value Decomposition (GSVD) is another approach that relies on spectral decomposition to identify modules of co-regulated genes (van Dam *et al.* 2018). Furthermore, Higher-Order GSVD (HO-GSVD) (Ponnapalli *et al.* 2011) helps in multi-tissue analysis (Xiao *et al.* 2014).

Single-cell RNA sequencing technology plays an influential role in obtaining transcriptomes at single-cell resolution. One drawback of this technique is that only a small fraction of the transcript gets sequenced, resulting in dropout events (excess zero counts). Zand and Ruan (2020) have recently developed a gene co-expression network-based method called 'netImpute' to alleviate this dropout issue (figure 2B). A similar effort was made using a Bayesian factor model (Sekula *et al.* 2020). In the case of single-cell RNA-seq experiments where the dataset consists of several biologically distinct unknown sample groups, identifying differentially expressed clusters with similar expression patterns is challenging. An alternative method is biclustering, which can identify such patterns without prior sample classification (Cheng and Church 2000). Several new tools have also been recently developed for improving the analyses of co-expression networks, which include CoExp (García-Ruiz *et al.* 2021), Gene Whole Co-Expression Network Analysis (GWENA) (Lemoine *et al.* 2021), Translational Bioinformatics Tool Suite for Network Analysis and Mining (TSUNAMI) (Huang *et al.* 2021) and Conserved and Comparative Co-expression Network (CococoNet) (Lee *et al.* 2020a, b).

3.5 *Prediction of perturbation patterns*

The unavailability or lack of information about kinetic parameters while studying interactions between biochemical entities leads to loss of information. The consequences often reflect in the perturbation patterns. Collective perturbations affect disease states. The patterns of these perturbations help to understand differential expression patterns. The study of perturbation patterns in biochemical networks provides the opportunity for drug development, better understanding of drug combinations, and improved therapies (Santolini and Barabási 2018). Increasingly accurate topological models have provided us with improved confidence to approximate the impact of perturbation patterns (Santolini and Barabási 2018). Santolini and Barabási have proved that the topology-based linear response matrix or correlation matrix alone provides more than 65% accuracy in predicting these perturbation and biochemical influence patterns (Santolini and Barabási 2018). Their proposed method predicts perturbation patterns with higher accuracy for the networks that, upon link removal, can be decoupled into sparse networks (i.e., nodes with a low degree and link density). Additionally, the topology-based method, along with the integration and inference from experimental perturbation data, plays a key role in predicting physiological and phenotypic perturbations. However, the interplay between network topology and inherent dynamics predicts the emergent patterns in biochemical networks (Meena *et al.* 2020a, b).

The impact of the perturbations can be effectively studied by characterizing transient cell states that reflect in the cellular responses. Dynamic gene interactions and pathway behaviour are of central importance to characterize these transient populations. Dynamic modelling techniques are being developed that utilize the RNA velocity from single-cell RNA sequencing data that help observe dynamics of single genes and thus assist in interpreting the impact of perturbations (La Manno *et al.* 2018; Bergen *et al.* 2020). Network representations of this dynamic gene interaction data hold potential to improve the prediction of perturbation impact through dynamic network analysis.

3.6 *Identification of clustering patterns in dynamic networks*

Clustering patterns in molecular interaction networks find applications in several analyses, such as identifying similar gene expression patterns (Oyelade *et al.* 2016) and classifying cellular subtypes. Several algorithms have been developed that identify such clustering patterns in the form of community detection methods to analyse large biological datasets (Oyelade *et al.* 2016; Sharma and Ali 2017; Kanter *et al.* 2021). The concept behind network clustering is to partition

networks into clusters or groups of 'topologically related' nodes expected to 'correlate' well in terms of their function or phenotype. Clustering in dynamic networks is considered to be challenging and complex due to the additional factor of timescales. In the case of dynamic networks, where the data essentially consist of a series of snapshots of networks through time, two popular approaches, snapshot clustering (Chi *et al.* 2007) and consensus clustering (Lancichinetti and Fortunato 2012), are used. Dynamic network clustering (DNC) approaches like Louvain and Infomap (Held *et al.* 2016) assume that topological similarity naturally implies dense interconnectedness. This disparity of assumption is being addressed by a newly developed approach called 'ClueNet' (Crawford and Milenković 2018), which evaluates the need for some dynamic networks to be partitioned based on topological similarity. Another method uses a combined approach of partitioning based on topological similarity combined with denseness (Crawford and Milenković 2018). Other clustering methods have also been developed where the metadata of the nodes in network are considered prior to forming clusters and not just used in post-clustering module segregation steps (Newman and Clauset 2016; Peel *et al.* 2017). Timely improvements in these heuristic methods would further remove current limitations such as the incorporation of overlapping clustering that are evident in real-world networks where one node can belong to multiple functional modules.

3.7 *Gene prioritization methods*

Identifying causal genes and refining candidate genes for experimental verification is an important step in high-throughput analyses, specifically in studying diseases. One of the most widely used prediction servers is GeneMANIA (Warde-Farley *et al.* 2010). The prioritization of genes in this server is highly dependent on network topological properties and the likeliness of shared phenotypes. New approaches like Hybrid-Ranker exploit topological properties to prioritize genes based on their proximity to the causal genes of a particular disease of interest and information on its corresponding co-morbid disease (Razaghi-Moghadam and Nikoloski 2020). Arete is a similar tool incorporated as an app in the Cytoscape graph analysis suite (Lysenko *et al.* 2017). Novel gene prioritization tools like GenePANDA (Yin *et al.* 2017) and TopControl (Nazarieh and Helms 2019) use additional features like the relative distance of the candidate disease gene to the known disease genes and dominating sets on co-regulatory networks instead of high-degree nodes. Target gene prioritization has also been demonstrated through the identification of the important regulatory modules within large GRNs, such as transcription factors regulating the downstream target genes in regulatory pathways usually represented as bipartite graphs. A recent methodological development to identify these regulatory target modules in such bipartite graphs is being explored through the technique of backbone extraction (Pavlopoulos *et al.* 2018). Backbone extraction delivers a subgraph composed of the most significant nodes and edges in a network. It has been used in projected graph networks obtained from the bipartite projection of regulatory and target nodes, which helps to identify important regulatory modules within the large regulatory network (figure 2C).

3.8 *Methods for analysing amino acid networks*

Amino acid networks (AANs) or protein topological networks (PTNs) are used for the graphical representation of functional domains of proteins. The edges represent interactions based on the amino acid distance cut-off set at their primary, secondary, or tertiary structural arrangement levels. Topological parameters effectively represent the structural and functional properties of protein networks (Bagler and Sinha 2007). The structural organization of these networks shows small-world network properties (Bagler and Sinha 2005). These networks are assortative and have a hierarchy and are limited to the subnetwork of hydrophobic amino acids (Yan *et al.* 2014). These networks are helpful in distinguishing the folding states of the protein structures from the decoys (Zhou *et al.* 2014), predicting protein fold (Bhavani *et al.* 2011) and understanding disease mutational landscapes such as identifying the epitopes of topological importance for rational immunogen design (Yan *et al.* 2014). The Protein Topological Graph Library (PTGL) was developed to provide a fast search for secondary structure classification and characterization of proteins by abstracting the structure in the form of undirected labelled graphs (May *et al.* 2004). Network Analysis of Protein Structures (NAPS) (Chakrabarty and Parekh 2016; Chakrabarty *et al.* 2019) and Amino acid Network Construction and Analysis (ANCA) (Yan *et al.* 2020) are web servers developed that facilitate the qualitative and quantitative topological analyses and visualization to study residue–residue relationships and

help gain insights into structure–function relationships. A protocol developed by Sinha *et al.* effectively determined the allosteric residues regulating drug binding activities by constructing a protein-contact network (PCN) and subsequently employed the network propagation theory based on a heat diffusion model (Sinha *et al.* 2020a, b) (figure 3). Additional applications of AANs include studying mutation patterns to effectively design HIV vaccine target T-cell epitopes (Gaiha *et al.* 2019) and to study allosteric changes leading to protein stability (Srivastava and Sinha 2014), designing of thermostable mutants (Kandhari and Sinha 2017), and understanding the molecular basis for resistivity and specificity of proteins in drug resistance (Sinha *et al.* 2020a, b).

4. Emerging hybrid network-based approaches

Mathematical models based on interaction graphs allow the investigation of complex biological systems (Sinha 1997). However, with increasing size of these systems, their dynamics and complexity grows exponentially, consequently making the screening of possible interventions infeasible (Cohen and Harel 2007). Static network topology-based analysis of large biological systems allows the identification of dynamically relevant components of the whole network. Systems biology approaches unravel different intracellular and intercellular signalling mechanisms and metabolisms to study emerging molecular systems. Hybrid network-based models are designed by combining graph theoretical analysis with two or more systems biology modelling methods such as Boolean modelling (Chowdhury and Sarkar 2019), FBA-based metabolic modelling (Dusad *et al.* 2021), ODE (ordinary differential equation) (Kang *et al.* 2020) and PDE (partial differential equation)-based models (Bardini *et al.* 2017). Integration of high-throughput data into these hybrid models has helped to overcome previous limitations and challenges, such as identification of system-level continuous and discrete dynamic functional modules, timescale integration, handling data heterogeneity, and elucidation of complete pathway topology. In this section, we briefly explain some recent hybrid network-based modelling approaches developed to understand the temporal dynamics of regulatory molecular mechanisms, identification of tumour heterogeneity, and elucidation of pathway topological modules.

(i) *Regulatory mechanisms:* Integrative measures to identify dynamically relevant modules from large-scale systems-level molecular interaction networks have been employed. A sequential evaluation of the hedgehog signalling pathway in different types of cancer using network topology-based approach followed by Boolean analysis of the important regulatory modules provided a promising measure to predict the dynamic behaviour of biological networks (Chowdhury *et al.* 2013). Boolean models account for genes as either specific activators or repressors of the target genes, due to which the analysis of gene regulatory behaviour at the subfunction level is compromised. This limitation has been recently addressed by the data-driven Fundamental Boolean Model (FBM), which facilitates subfunction-level analysis over a period of time by generating dynamic trajectories. This model is implemented in R for use as a package, 'FBNNet', and can be effectively used to study dynamic gene regulatory behaviour (Chen *et al.* 2018).

(ii) *Tumour heterogeneity:* Mechanistic models with prior knowledge derived from topological measures have contributed to improved understanding of intra-tumour heterogeneity and dynamic regulations behind the emergence of tumorigenic phenotype lineages and regulation of plasticity in cancer (Chowdhury and Sarkar 2019). The protocol provides a platform for personalized and target-based glioblastoma tumour therapy (figure 3). Dynamic graphical model frameworks are developed for comprehensive analyses of tumour heterogeneity by integrating different genome-level datasets. These frameworks are useful to understand the role of mutations in conferring heterogeneity at different stages of cancer progression and additional complexities in tumour evolution (Lysenko *et al.* 2017).

(iii) *Pathway topology:* In the current era of precision medicine, a system-wide pathway-level understanding plays a crucial role. Elucidation of the entire pathway topology has been a challenge in systems biology for quite some time. In this context, a recent approach developed by Liang and co-workers handles the property of mutual exclusivity in the pathway perturbation of tumours by applying an OR-gated network that infers modules of patient-specific dysregulated pathways (Liang *et al.* 2021). The Boolean variables for generating OR-gate functions in this model are obtained from mutation and gene expression data, which are then converted to a OR-gated network and thus it effectively handles

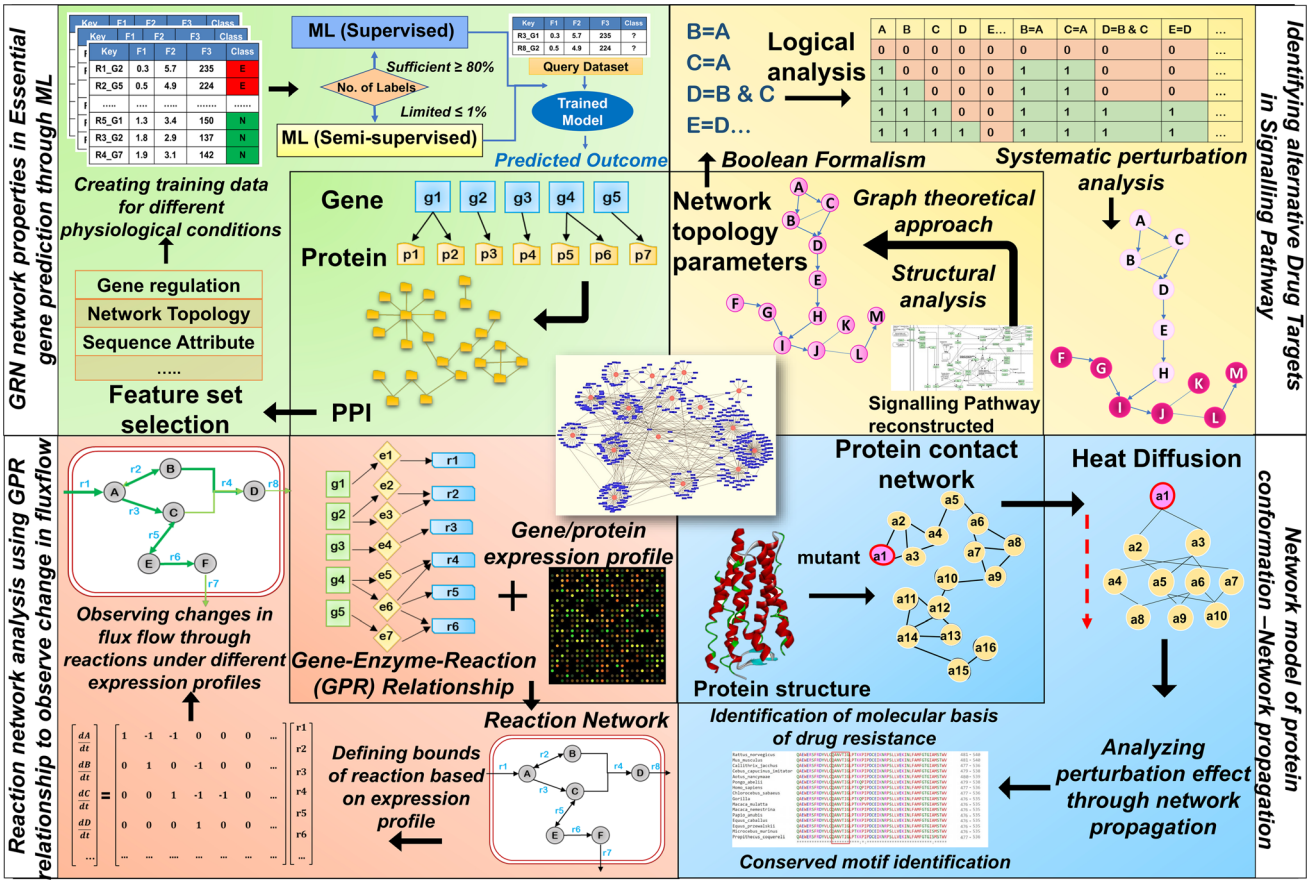

Figure 3. Schematic workflows of emerging hybrid network models. Overview of emerging integrative approaches where network biology is combined with different molecular interaction data resources such as gene regulatory interactions, signalling and metabolic networks, and amino acid interaction in proteins. The implementation of different systems biology techniques such as Boolean formalism, flux balance analysis, heat diffusion, and integration of ML is demonstrated as various applications in deciphering molecular mechanisms.

co-occurrence of genes, mutual exclusivity, and other properties that effectively contribute to elucidate patient-specific pathway modules.

Hence, we observe that the hybrid network-based approaches contribute to the understanding of complex system behaviour and generate testable hypotheses for experimental validation. However, these hybrid network-based approaches face issues in obtaining phenotype-specific data as well as incorporating disparate variable types in the networks (Walker *et al.* 2014). Although the hybrid modelling techniques ameliorate the classical methods, their extensive application is limited by complexities during *in silico* implementation. The scope and scalability of each modelling approach, like the Boolean, constraint-based, ordinary and partial differential equation-based, or network-based approaches, are different. Assimilating the information generated by one into another approach is often challenged by loss of information. Furthermore, the availability of appropriate data and parameterization of the model are factors adding to the challenges. Since the hybrid models concatenate the information generated by tools and techniques with different scalability, the integration, calibration, and evaluation of the model outcomes using appropriate experimental evidence is more challenging than in the classical techniques. For example, paramterizing all variables obtained from a subnetwork of hub protein interactions for an ordinary differential model is challenged by the availability of adequate information. Although the parameter estimation techniques can be helpful to deduce unknown parameters, their reliability depends on the preciseness of the experimental data used for calibrating the biological context to be studied. In the case of a hybrid network-based approach for metabolic modelling, the connectivity between the metabolic components in the network method is defined solely by the metabolic stoichiometry, whereas in the FBA

model, it depends on the choice of objective functions (Dusad *et al.* 2021). As a result, while translating the FBA model into a network, it can be constructed as various types of graphs, such as bipartite graphs, graphs where nodes represent metabolites, or graphs where nodes represent a reaction. This creates a lack of consensus on the type of graph that can be built for that metabolic model and strongly influences the conclusions drawn from their network analyses (Beguerisse-Díaz *et al.* 2018). Another challenge arises while constructing networks that consider the direction of reaction fluxes in the metabolic models. However, these limitations are increasingly being addressed by recent developments of flux-weighted graphs and mass-flow graphs (Beguerisse-Díaz *et al.* 2018). In the case of a hybrid network-based approach that integrates Boolean modelling, the Boolean equations can often incorporate timescale-dependent behaviour, but this information might not be translated to the model's graph representation, as, at one time point, the network can only represent a static interaction for one specific condition. These limitations and challenges further press the need for advanced statistical tools to augment non-continuous data and variables. With the surge in the approaches that automatically learn to encode network structure into low-dimensional representations, the use of transformation techniques with ML-based approaches and their hybridization with other first-principle modelling techniques have gained momentum (Lee *et al.* 2020a, b). In the next section, we will discuss some of the areas where network topology-integrated ML-based approaches find extensive application in analysing molecular interactions.

5. Applications of emergent machine learning-based approaches in molecular networks analysis

The above discussions provide a brief overview of the hybrid network-based approaches that are being constantly updated to incorporate network topological information with different systems biology approaches. With avenues of dimensionality reduction of the vast genome-wide association data for automated extraction of information, the ML-based techniques have revolutionized the prospects of network biology in understanding molecular interaction networks (Mochida *et al.* 2018). ML techniques, such as semi-supervised algorithms, effectively contribute to the classification problems with limited availability of data and parameters, which often remains as a limitation of the hybrid network-based approaches, as discussed in the previous section. They also enable easy automation of integration of multi-level heterogenous data with network models, thus enabling the use of a vast spectrum of information in an automated fashion. The use of network topological features in semi-supervised ML algorithms has enabled automation even with limited availability of relevant information (Nandi *et al.* 2020). In this section, we briefly state areas where these emerging ML techniques that amalgamate molecular network information find applications.

5.1 *Gene essentiality prediction*

Predicting essential genes through the analysis of molecular networks has significantly contributed to drug development and understanding of synthetic biology (Hwang *et al.* 2009). Single-gene knockout studies and genome-wide RNAi screens have unveiled the multifaceted nature of gene essentiality that is context-dependent and evolvable rather than just binary and static (Rancati *et al.* 2018). This provides scope for developing predictive models for identifying essential genes using genome-wide data by applying advanced computational techniques such as machine learning on genome-wide association studies (GWAS), e.g., EssRank (Xu *et al.* 2019). Integration of gene expression data and network topological features has been employed for predicting gene essentiality (Zhong *et al.* 2021).

ML algorithms use gene expression, functional annotation, sequence, and network topology as features to identify gene essentiality (Zhang *et al.* 2016). Along with PPINs, transcriptional and metabolic network features have also been increasingly incorporated into the ML models (da Silva *et al.* 2008; Plaimas *et al.* 2010). Nandi *et al.* (2020) have recently addressed the shortcomings of limited availability of experimental data and, thus, the lack of labelled data by developing a semi-supervised ML strategy. This Laplacian support vector machine (SVM)-based strategy revealed topological measures of reaction networks as one of the important determining features for classifying essential and non-essential genes in prokaryotes and eukaryotes (figure 3). These ML strategies contribute to identifying the deterministic features that help distinguish class labels (e.g., essential and non-essential gene), which was previously challenging due to limited availability of data. DEEPLYESSENTIAL is another method that uses deep neural network architecture to predict essential microbial genes using sequence information (Hasan

and Lonardi 2020). Campos and co-workers' ML-based workflow to predict essential genes in *Caenorhabditis elegans* has shown that essential genes are positively correlated with low single nuclear polymorphism (SNP) densities and epigenetic markers in promoter regions (Campos *et al.* 2020).

5.2 *Prediction of drug–disease interactions*

Drug repurposing is known to accelerate drug discovery research and development processes (Novac 2013). Identification of drug–disease interactions plays an important role in drug repurposing and thus accelerates *de novo* drug discovery. The advantage of using network-based models for identifying drug–disease interactions is that it utilizes complete large-scale high-throughout data to build complex biological interaction networks. Several network-based models to identify these drug–disease interactions have been developed in the recent past (Wang *et al.* 2014; Martínez *et al.* 2015; Luo *et al.* 2016). The prediction of potential drug–disease interactions by integrating multiple layers of network data has been helpful in assessing molecular actions and studying disease implications (Oh *et al.* 2014). Recent developments using ML-based prediction models designed for drug–disease association studies employ a number of methods ranging from logistic regression-based methods (Gottlieb *et al.* 2011) to Laplacian regularized sparse subspace learning (LRSSL)-based methods (Liang *et al.* 2021). Wu *et al.* (2017) proposed a semi-supervised graph cut (SSGC) algorithm to predict drug–disease pairs by integrating the information on drug substructures, disease phenotypes, and gene–gene interactions with known drug–disease interaction treatment relationships in a hierarchical framework. This proposed algorithm enabled the integration of three different layers of disease phenotype, treatment and gene mechanism data, and optimally identified drug–disease similarity associations. Network similarities have also shown to contribute to drug–disease associations and can be effectively used in ML-based training algorithms to improve predictions. A novel method was recently developed that combines network similarities of drugs and diseases with their chemical and semantic similarities to predict novel drug–disease interactions and effectively handles the unwanted disease interaction pairs, which have been a challenge in some previously developed methods (Cui *et al.* 2019). Integration of similarity measures in heterogenous networks and deep learning models to predict drug–disease interactions can further significantly benefit drug repurposing. A novel framework was recently proposed by Jarada and co-workers that uses similarity selection and similarity network fusion combined with neural network deep learning model to efficiently predict drug–disease interactions (Jarada *et al.* 2021). This method resolves the challenge of limited availability of known interactions by integrating similarity information along with tackling data noise and redundancy issues which were previously faced by other methods, and thus has improved prediction accuracy. Another recently developed methodology of ensemble-based strategy uses weighted K-nearest known neighbours to construct drug and disease similarity networks (Wang *et al.* 2021). Such statistically improved methods are increasingly being proposed to improve accurate drug–disease interaction predictions by developing novel strategies using molecular network information, which will advance the development towards precision medicine (Zhu *et al.* 2018; Yu *et al.* 2019).

5.3 *Characterization of regulatory genes*

ML methods contribute significantly in predicting and inferring GRNs using transcriptomic data (Mochida *et al.* 2018). GRN inferences face limitations due to noise, low sample size and incomplete characterization of regulatory dynamics, leading to networks with missing and anomalous links (Banf and Rhee 2017). A semi-supervised network reconstruction algorithm has been developed that enables the synthesis of information from partially known GRNs with time course gene expression data (Nguyen and Braun 2018). This method successfully identifies novel and anomalous connections. A recent advancement addresses the problem of two potential regulators in GRNs having high correlation or matching expression patterns, making it challenging to differentiate between them. A novel method called linear profile likelihood (LiPLike) predicts gene-to-gene regulation with high accuracy by selecting interactions that are uniquely inferred by measured data (Magnusson and Gustafsson 2020). A recent supervised-learning-based method, GRADIS, incorporates graph distance profiles from transcriptomic data to reconstruct GRNs (Razaghi-Moghadam and Nikoloski 2020). This approach offers the possibility to use network representations of large-scale data that help characterize cellular networks and analyse GRNs effectively.

5.4 *Prediction of protein abundance and protein complexes from PPINs*

ML models are being implemented to predict protein abundance from single-cell RNA-Seq data using PPI and prior knowledge embedded into neural graph networks (Niu *et al.* 2020; Dai *et al.* 2021). The PIKE-R2P (PPIN-based knowledge embedding with graph neural network for single-cell RNA to protein prediction) model proposed by Dai and co-workers uses graph neural networks (GNN) which enable multi-label modelling. The information from PPIs in these GNNs thus help in the cross-modality prediction of protein abundances at the single-cell level. DeepHE, a network embedding method, automatically learns features from PPINs and additionally uses sequence features. These two types of feature data are used to train multi-layer neural networks and address the imbalanced learning problem using a cost-sensitive technique (Zhang *et al.* 2020). New algorithms have been proposed for mining the best topological features to predict protein complexes from PPINs. The Sequential Forward Feature Selection (SFFS) algorithm, recently proposed by Younis and co-workers, uses random forest-based Boruta feature selection to integrate a wide variety of topological and biological features as well as protein interaction information (Younis *et al.* 2021).

6. Understanding disease mechanisms and identification of potential therapeutic targets

Network topology-based approaches have contributed to the development of advanced therapeutic applications to curb metastasis-driven cancer progression. Topology-based approaches have been applied to classifying breast cancer subtypes by searching for significant sub-networks (Chuang *et al.* 2007). Their application has expanded to a broader pan-analysis perspective with experimental and theoretical advancements in cancer diagnosis. Multi-layer frameworks combining network topology and spectral graph theory have enabled the study of the cancer complexome to identify important proteins across multiple cancers (Ramadan *et al.* 2016; Rai *et al.* 2017; Hari *et al.* 2020; Buffard *et al.* 2021). WGCNA on the PPI networks has been used to identify gene co-expression modules between the differentially expressed genes (DEGs) through hierarchical clustering to identify gene expression signatures associated with acquired gefitinib resistance (Lee *et al.* 2015). The potential network analysis techniques applied explicitly for developing precision cancer medicine have been thoroughly reviewed by Ozturk *et al.* (2018).

Evaluation of the inter-species heterogeneity in molecular interaction networks has contributed significantly towards delineating infection mechanisms by analysing cause–effect relationships in treatment strategies. For example, Singh *et al.* (2020) applied exhaustive topological analysis using the parameters in-degree, out-degree, and directed and undirected average path lengths to study the comprehensive transcriptional regulatory network of *Mycobacterium tuberculosis* (MTB) H37rv. Furthermore, global proteomic datasets analysis of virus-infected patients with human immunodeficiency virus (HIV) and hepatitis C virus (HCV) demonstrated that using degree and betweenness for identifying pathogen interactions was more effective and accurate than using differential regulation alone (McDermott *et al.* 2012; Soto-Girón and García-Vallejo 2012). Ackerman and co-workers proposed a new method of combining host PPI networks with virus–host PPI data to identify host target proteins and demonstrated the method in influenza infection by extracting virus-specific subnetworks (Ackerman *et al.* 2018). This study revealed that network position within the virus–host subnetwork offers an advantage in prioritization of drug targets. Controllability analysis using virus–host networks captured the dynamic properties without the knowledge of experimentally derived concentration parameters and helped identification of antiretroviral targets with higher potential (Ackerman *et al.* 2019).

The current global COVID pandemic has led to a surge in research techniques that can rapidly process the information generated for the newly identified SARS coronavirus 2. The immediate requirement to identify potent remedial solutions and understand the virulence mechanism of rapidly evolving strains to development of vaccines has led researchers towards network biology approaches. Construction and analysis of biomolecular networks in the form of PPINs, transcriptional, and gene co-expression networks have led to rapid assessment of concurrent effects (Nashiry *et al.* 2021), analysing viral host associations (Das *et al.* 2021; Terracciano *et al.* 2021) and predicting miRNAs associated with viral pathogenesis, elucidating neurological manifestations (Prasad *et al.* 2021).

7. Concluding remarks and future prospects

The present review thoroughly evaluates the insightful inferences that can be drawn using graphical networks of biological systems and their integration into different

hybrid network-based modelling techniques to provide additional details about complex system behaviour. With increasing advances in high-throughput technologies in biological research, attempts have been made to incorporate this information into network structures, leading to a continuous update of network biology approaches in the last two decades. New centrality measures like pairwise disconnectivity index, leverage centrality, bottleneckness, bipartivity, etc., along with classical centrality measures like betweenness, Katz centrality, and eigenvector centrality, can be used to predict the influence of genomic regulators on target gene networks and the impact of their deletion on target genes. Advances in topology-based approaches have paved the way to successfully identify perturbation patterns, gene prioritization, and clustering of dynamic networks. The computational advances in terms of amalgamation of machine learning (ML) and artificial intelligence (AI) in using network graph properties have proved beneficial with several applications such as essential gene prediction, drug–disease interaction identification, and prediction of protein abundance from single-cell data. Pertaining to the current knowledge of the available methodological advances in studying these molecular interaction networks, we suggest certain areas where advanced computational approaches incorporating network properties can be developed and applied in future:

Development of a testable hypothesis for disease diagnostics: With a demonstrated application of graph networks in deriving useful inferences about different disease conditions including both infectious disease and cancer, network biology approaches provide the opportunity to elucidate condition-specific network structures depending on the details of the specific disease systems under study. Empirical analysis of the network structure and topology of disease case-specific conditional differences and comparison with the network structure of the normal physiological conditions can help identify prognostic targets and modules for therapeutic benefits.

Identification of gene regulatory targets: Computational analysis of omics and high-throughput data on the translational and post-transcriptional regulators of gene expression has been successful in establishing a cause-and-effect relationship between differential expression of the gene expression regulators and target genes (figure 3C). Graph network analysis is a suitable choice to study these regulatory networks as these approaches can provide conclusions based on holistic analysis of large-scale information. Regulatory targets can be predicted from the analyses that can be further tested through *in vivo* and *in vitro* studies to test their feasibility as therapeutic targets under diseased conditions.

Automated identification of essential genes and prioritized molecular targets: Hybrid network-based models, which that make use of network inferences have paved the way for the development of automated ML and artificial network-based (Bayesian, neural) tools with the ability to predict and identify essential genes and prioritize molecular targets even with limited experimental data availability. These integrative approaches can be continuously updated to improve the quality of prediction with corroboration of additional heterogeneous data. Semi-supervised algorithms can further be improved to increase the prediction accuracy when limited essential gene information as well as GRN data are available.

Continuous improvement in statistical methodologies in the area of deep learning and recurrent neural networks can be effectively applied to improve the use of molecular network data in the context of personalized medicine development. Advances in classification algorithms and improved sensitivity to phenotype-specific classification utilizing network topological information can accelerate research in personalized medicine development and improve our understanding of causal mechanisms in diseases.

Acknowledgements

CM acknowledges the Department of Science and Technology (DST) for the INSPIRE Faculty Fellowship (Award No. IFA19-PH248). RB acknowledges the Council of Scientific and Industrial Research (CSIR) for the Senior Research Fellowship (CSIR File No.: 31/011(1047)/2018-EMR-I dated 26-04-2018).

Funding

Funding was provided by the Department of Science and Technology, Ministry of Science and Technology (Grant No. IFA19-PH248), and Council of Scientific and Industrial Research, India (Grant No. 31/011(1047)/2018-EMR-I dated 26-04-2018).

References

Ackerman EE, Alcorn JF, Hase T and Shoemaker JE 2019 A dual controllability analysis of influenza virus–host

protein–protein interaction networks for antiviral drug target discovery. *BMC Bioinform.* **20** 297

Ackerman EE, Kawakami E, Katoh M, *et al*. 2018 Network-guided discovery of influenza virus replication host factors. *mBio* **9** 6

Alvarez-Ponce D, Feyertag F and Chakraborty S 2017 Position matters network centrality considerably impacts rates of protein evolution in the human protein–protein interaction network. *Genome Biol. Evol.* **9** 1742–1756

Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, *et al*. 2018 A systematic survey of centrality measures for protein–protein interaction networks. *BMC Syst. Biol.* **12** 80

Bagler G and Sinha S 2005 Network properties of protein structures. *Phys. A Stat. Mech. Appl.* **346** 27–33

Bagler G and Sinha S 2007 Assortative mixing in Protein Contact Networks and protein folding kinetics. *Bioinformatics* **23** 1760–1767

Banf M and Rhee SY 2017 Computational inference of gene regulatory networks approaches, limitations and opportunities. *Biochim. Biophys. Acta Gene Regul. Mech.* **1860** 41–52

Barabási A-L and Oltvai ZN 2004 Network biology understanding the cell's functional organization. *Nat. Rev. Genet.* **5** 101–113

Bardini R, Politano G, Benso A and Di Carlo S 2017 Multi-level and hybrid modelling approaches for systems biology. *Comput. Struct. Biotechnol. J.* **15** 396–402

Beguerisse-Díaz M, Bosque G, Oyarzún D, Picó J and Barahona M 2018 Flux-dependent graphs for metabolic networks. *NPJ Syst. Biol. Appl.* **4** 32

Bergen V, Lange M, Peidli S, Wolf FA and Theis FJ 2020 Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38** 1408–1414

Bhavani DS, Savarnavani K and Sinha S 2011 Mining of protein contact maps for protein fold prediction. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* **1** 362–368

Bidkhori G, Benfeitas R, Elmas E, *et al*. 2018 Metabolic network-based identification and prioritization of anti-cancer targets based on expression data in hepatocellular carcinoma. *Front. Physiol.* **9** 916

Brin S and Page L 1998 The anatomy of a large-scale hypertextual web search engine. *Comput. Networks* **30** 107–117

Brockman IM and Prather KLJ 2015 Dynamic metabolic engineering: new strategies for developing responsive cell factories. *Biotechnol. J.* **10** 1360–1369

Buffard M, Naldi A, Freiss G, *et al*. 2021 Comparison of syk signaling networks reveals the potential molecular determinants of its tumor-promoting and suppressing functions. *Biomolecules* **11** 308

Campos TL, Korhonen PK, Sternberg PW, Gasser RB and Young ND 2020 Predicting gene essentiality in *Caenorhabditis elegans* by feature engineering and machine-learning. *Comput. Struct. Biotechnol. J.* **18** 1093–1102

Chakrabarty B, Naganathan V, Garg K, Agarwal Y and Parekh N 2019 NAPS update network analysis of molecular dynamics data and protein–nucleic acid complexes. *Nucleic Acids Res.* **47** W462–W470

Chakrabarty B and Parekh N 2016 NAPS Network analysis of protein structures. *Nucleic Acids Res.* **44** W375–W382

Charitou T, Bryan K and Lynn DJ 2016 Using biological networks to integrate, visualize and analyze genomics data. *Genet. Sel Evol.* **48** 27

Chen L, Kulasiri D and Samarasinghe S 2018 A novel data-driven Boolean model for genetic regulatory networks. *Front. Physiol.* **9** 1328

Cheng Y and Church GM 2000 Biclustering of expression data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8** 93–103

Chi Y, Song X, Zhou D, Hino K and Tseng BL 2007 Evolutionary spectral clustering by incorporating temporal smoothness. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 153–162

Chowdhury HA, Bhattacharyya DK and Kalita JK 2020 (Differential) Co-expression analysis of gene expression: a survey of best practices. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17** 1154–1173

Chowdhury S, Pradhan RN and Sarkar RR 2013 Structural and logical analysis of a comprehensive hedgehog signaling pathway to identify alternative drug targets for glioma, colon and pancreatic cancer. *PLoS One* **8** e69132

Chowdhury S and Sarkar RR 2019 Exploring notch pathway to elucidate phenotypic plasticity and intra-tumor heterogeneity in gliomas. *Sci. Rep.* **9** 9488

Chuang HY, Lee E, Liu YT, Lee D and Ideker T 2007 Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3** 140

Cohen IR and Harel D 2007 Explaining a complex living system Dynamics, multi-scaling and emergence. *J. R. Soc. Interface* **4** 175–182

Crawford J and Milenković T 2018 ClueNet Clustering a temporal network based on topological similarity rather than denseness. *PLoS One* **13** e0195993

Cui Z, Gao Y-L, Liu J-X, *et al*. 2019 The computational prediction of drug-disease interactions using the dual-network L2,1-CMF method. *BMC Bioinform.* **20** 5

da Costa WLO, de Araújo CL, *et al*. 2018 Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. *PLoS One* **13** 1–28

da Silva JPM, Acencio ML, Mombach JCM, *et al*. 2008 In silico network topology-based prediction of gene essentiality. *Phys. A Stat. Mech. Appl.* **387** 1049–1055

Dai X, Xu F, Wang S, Mundra PA and Zheng J 2021 PIKE-R2P Protein–protein interaction network-based knowledge embedding with graph neural network for single-cell RNA to protein prediction. *BMC Bioinform.* **22** 139

Das JK, Chakraborty S and Roy S 2021 A scheme for inferring viral-host associations based on codon usage patterns identifies the most affected signaling pathways during COVID-19. *J. Biomed. Inform*. **118** 103801

de Lorenzo V, Prather KL, Chen G-Q, *et al*. 2018 The power of synthetic biology for bioproduction, remediation and pollution control The UN's Sustainable Development Goals will inevitably require the application of molecular biology and biotechnology on a global scale. *EMBO Rep.* **19** e45658

Dhasmana A, Uniyal S, Anukriti VK, *et al*. 2020 Topological and system-level protein interaction network (PIN) analyses to deduce molecular mechanism of curcumin. *Sci. Rep.* **10** 12045

Díaz J 2020 SARS-CoV-2 molecular network structure. *Front. Physiol.* **11** 870

Di Paola L, Platania CBM, Oliva G, *et al*. 2015 Characterization of protein–protein interfaces through a protein contact network approach. *Front. Bioeng. Biotechnol.* **3** 170

Dobon B, Montanucci L, Peretó J, Bertranpetit J and Laayouni H 2019 Gene connectivity and enzyme evolution in the human metabolic network. *Biol. Direct.* **14** 17

Dusad V, Thiel D, Barahona M, Keun HC and Oyarzún DA 2021 Opportunities at the interface of network science and metabolic modeling. *Front. Bioeng. Biotechnol.* **8** 1570

Ebert BE, Lamprecht A-L, Steffen B and Blank LM 2012 Flux-p automating metabolic flux analysis. *Metabolites* **2** 872–890

Estrada E and Rodríguez-Velázquez JA 2005 Subgraph centrality in complex networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys*. **71** 056103

Freeman LC 1977 A set of measures of centrality based on betweenness. *Sociometry* **40** 35–41

Freeman LC 1978 Centrality in social networks conceptual clarification. *Soc. Netw.* **1** 215–239

Gaiha GD, Rossin EJ, Urbach J, *et al*. 2019 Structural topology defines protective CD8+ T cell epitopes in the HIV proteome. *Science.* **364** 480–484

García-Ruiz S, Gil-Martínez AL, Cisterna A, *et al*. 2021 CoExp: A web tool for the exploitation of co-expression networks. *Front. Genet*. **12** 630187

Gordon DE, Jang GM, Bouhaddou M, *et al*. 2020 A SARS-CoV-2 protein interaction map reveals targets for drug repurposing *Nature* **583** 459–468

Gottlieb A, Stein GY, Ruppin E and Sharan R 2011 PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **7** 496

Grennan KS, Chen C, Gershon ES and Liu C 2014 Molecular network analysis enhances understanding of the biology of mental disorders. *BioEssays* **36** 606–616

Grimes T, Potter SS and Datta S 2019 Integrating gene regulatory pathways into differential network analysis of gene expression data. *Sci. Rep.* **9** 5479

Hage P and Harary F 1995 Eccentricity and centrality in networks. *Soc. Networks* **17** 57–63

Hahn MW and Kern AD 2005 Comparative genomics of centrality and essentiality in three eukaryotic protein–interaction networks. *Mol. Biol. Evol.* **22** 803–806

Han H, Lee S and Lee I 2019 NGSEA network-based gene set enrichment analysis for interpreting gene expression phenotypes with functional gene sets. *Mol. Cells* **42** 579–588

Han J-DJ 2008 Understanding biological functions through molecular networks. *Cell. Res.* **18** 224–237

Hari K, Sabuwala B, Subramani BV, *et al.* 2020 Identifying inhibitors of epithelial–mesenchymal plasticity using a network topology-based approach. *NPJ Syst. Biol. Appl*. **6** 15

Hasan MA and Lonardi S 2020 DeeplyEssential: A deep neural network for predicting essential genes in microbes. *BMC Bioinform.* **21** 367

Hawe JS, Theis FJ and Heinig M 2019 Inferring interaction networks from multi-omics data. *Front. Genet.* **10** 535

Held P, Krause B and Kruse R 2016 Dynamic clustering in social networks using louvain and infomap method. *Proc. 2016 3rd Eur. Netw. Intel. Conf. ENIC 2016*

Huang Z, Han Z, Wang Resource T, *et al.* 2021 TSUNAMI translational bioinformatics tool suite for network analysis and mining. *Genom. Proteomics Bioinform.* **S1672–0229** 00054–00061

Huynh-Thu VA, Irrthum A, Wehenkel L and Geurts P 2010 Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5** e12776

Hwang YC, Lin CC, Chang JY, *et al.* 2009 Predicting essential genes based on network and sequence analysis. *Mol. Biosyst.* **5** 1672–1678

Hwang W, Kim T, Ramanathan M and Zhang A 2008 Bridging centrality graph mining from element level to group level. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 336–344

Imam S, Schäuble S, Brooks AN, Baliga NS and Price ND 2015 Data-driven integration of genome-scale regulatory and metabolic network models. *Front. Microbiol.* **6** 409

Jalili M, Salehzadeh-Yazdi A, Asgari Y, *et al*. 2015 Centiserver a comprehensive resource, web-based application and R package for centrality analysis. *PLoS One* **10** e0143111

Jalili M, Salehzadeh-Yazdi A, Gupta S, *et al*. 2016 Evolution of centrality measurements for the detection of essential proteins in biological networks. *Front. Physiol.* **7** 375

Janjić V and Pržulj N 2012 Biological function through network topology: a survey of the human diseasome. *Brief. Funct. Genom.* **11** 522–532

Jarada TN, Rokne JG and Alhajj R 2021 SNF-NN computational method to predict drug–disease interactions using similarity network fusion and neural networks. *BMC Bioinform.* **22** 28

Jiang P, Wang H, Li W, *et al*. 2015 Network analysis of gene essentiality in functional genomics experiments. *Genome Biol.* **16** 239

Joyce KE, Laurienti PJ, Burdette JH and Hayasaka S 2010 A new measure of centrality for brain networks. *PLoS One* **5** e12200

Junker BH and Schreiber F 2007 Signal transduction and gene regulation networks; in *Analysis of Biological Networks* (Wiley) pp 181–286

Kabir MH, Patrick R, Ho JWK and O'Connor MD 2018 Identification of active signaling pathways by integrating gene expression and protein interaction data. *BMC Syst. Biol.* **12** 120

Kadkhoda S, Darbeheshti F and Tavakkoly-Bazzaz J 2020 Identification of dysregulated miRNAs-genes network in ovarian cancer: an integrative approach to uncover the molecular interactions and oncomechanisms. *Cancer Rep*. **3** e1286

Kairov U, Karpenyuk T, Ramanculov E and Zinovyev A 2012 Network analysis of gene lists for finding reproducible prognostic breast cancer gene signatures. *Bioinformation* **8** 773–776

Kandhari N and Sinha S 2017 Complex network analysis of thermostable mutants of *Bacillus subtilis* Lipase A. *Appl. Netw. Sci.* **2** 18

Kang X, Hajek B and Hanzawa Y 2020 From graph topology to ODE models for gene regulatory networks. *PLoS One* **15** 1–26

Kanter I, Yaari G and Kalisky T 2021 Applications of community detection algorithms to large biological datasets; in *Deep Sequencing Data Analysis* (ed.) N Shomron (Springer) pp. 59–80

Kim EY, Ashlock D and Yoon SH 2019 Identification of critical connectors in the directed reaction-centric graphs of microbial metabolic networks. *BMC Bioinform.* **20** 328

Koh HWL, Fermin D, Vogel C, *et al*. 2019 iOmicsPASS network-based integration of multiomics data for predictive subnetwork discovery. *NPJ Syst. Biol. Appl.* **5** 22

Koschützki D, Junker BH, Schwender J and Schreiber F 2010 Structural analysis of metabolic networks based on flux centrality. *J. Theor. Biol.* **265** 261–269

Koschützki D, Schwöbbermeyer H and Schreiber F 2007 Ranking of network elements based on functional substructures. *J. Theor. Biol.* **248** 471–479

Koutrouli M, Karatzas E, Paez-Espino D and Pavlopoulos GA 2020 A guide to conquer the biological network era using graph theory. *Front. Bioeng. Biotechnol.* **8** 34

Kumar N, Mishra B, Mehmood A, Athar M and Mukhtar MS 2020 Integrative network biology framework elucidates molecular mechanisms of SARS-CoV-2 pathogenesis. *iScience* **23** 101526

Kumar T, Blondel L and Extavour CG 2020 Topology-driven protein–protein interaction network analysis detects genetic sub-networks regulating reproductive capacity. *eLife* **9** e54082

La Manno G, Soldatov R, Zeisel A, *et al*. 2018 RNA velocity of single cells. *Nature* **560** 494–498

Lancichinetti A and Fortunato S 2012 Consensus clustering in complex networks. *Sci. Rep.* **2** 336

Langfelder P and Horvath S 2008 WGCNA An R package for weighted correlation network analysis. *BMC Bioinform.* **9** 559

Lee D, Jayaraman A and Kwon JS 2020 Development of a hybrid model for a partially known intracellular signaling pathway through correction term estimation and neural network modeling. *PLoS Comput. Biol.* **16** e1008472

Lee J, Shah M, Ballouz S, Crow M and Gillis J 2020b CoCoCoNet conserved and comparative co-expression across a diverse set of species. *Nucleic Acids Res.* **48** W566–W571

Lee YS, Hwang SG, Kim JK, *et al*. 2015 Topological network analysis of differentially expressed genes in cancer cells with acquired gefitinib resistance. *Cancer Genom. Proteom.* **12** 153–166

Lemoine GG, Scott-Boyer MP, Ambroise B, Périn O and Droit A 2021 GWENA gene co-expression networks analysis and extended modules characterization in a single Bioconductor package. *BMC Bioinform.* **22** 267

Lewis NE, Nagarajan H and Palsson BO 2012 Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10** 291–305

Li M, Wang J, Wang H and Pan Y 2010 Essential proteins discovery from weighted protein interaction networks. *Proceedings of the 6th international conference on Bioinformatics Research and Applications* pp. 89–100

Li M, Zhang H, Wang J and Pan Y 2012 A new essential protein discovery method based on the integration of protein–protein interaction and gene expression data. *BMC Syst. Biol.* **6** 15

Li Z-C, Huang M-H, Zhong W-Q, *et al*. 2016 Identification of drug–target interaction from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics* **32** 1057–1064

Liang L, Zhu K, Tao J and Lu S 2021 ORN Inferring patient-specific dysregulation status of pathway modules in cancer with OR-gate network. *PLoS Comput. Biol*. **17** e1008792

Liseron-Monfils C and Ware D 2015 Revealing gene regulation and associations through biological networks. *Curr. Plant. Biol.* **3–4** 30–39

Liu D, Mannan AA, Han Y, Oyarzún DA and Zhang F 2018 Dynamic metabolic control towards precision engineering of metabolism. *J. Ind. Microbiol. Biotechnol.* **45** 535–543

Liu E, Zhang ZZ, Cheng X, Liu X and Cheng L 2020 SCNrank Spectral clustering for network-based ranking to

reveal potential drug targets and its application in pancreatic ductal adenocarcinoma. *BMC Med. Genom.* **13** 50

Luo H, Wang J, Li M, *et al*. 2016 Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics* **32** 2664–2671

Lv Q, Ma W, Liu H, *et al*. 2015 Genome-wide protein–protein interactions and protein function exploration in cyanobacteria. *Sci. Rep.* **5** 15519

Lysenko A, Boroevich KA and Tsunoda T 2017 Arete—candidate gene prioritization using biological network topology with additional evidence types. *BioData Min.* **10** 22

Ma T and Zhang A 2019 Integrate multi-omics data with biological interaction networks using Multi-view Factorization AutoEncoder (MAE). *BMC Genom.* **20** 944

Mabonga L and Kappo AP 2019 Protein–protein interaction modulators advances, successes and remaining challenges. *Biophys. Rev.* **11** 559–581

Magnusson R and Gustafsson M 2020 LiPLike: Towards gene regulatory network predictions of high certainty. *Bioinformatics* **36** 2522–2529

Mahadevan R and Palsson BO 2005 Properties of metabolic networks: Structure versus function. *Biophys. J.* **88** L07–L09

Margolin AA, Nemenman I, Basso K, *et al.* 2006 ARACNE An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7** S7

Martínez V, Navarro C, Cano C, Fajardo W and Blanco A 2015 DrugNet: Network-based drug–disease prioritization by integrating heterogeneous data. *Artif. Intell. Med.* **63** 41–49

Masoomy H, Askari B, Tajik S, Rizi AK and Jafari GR 2021 Topological analysis of interaction patterns in cancer-specific gene regulatory network persistent homology approach. *Sci. Rep.* **11** 16414

May P, Barthel S and Koch I 2004 PTGL—A web-based database application for protein topologies. *Bioinformatics* **20** 3277–3279

McDermott JE, Diamond DL, Corley C, *et al*. 2012 Topological analysis of protein co-abundance networks identifies novel host targets important for HCV infection and pathogenesis. *BMC. Syst. Biol.* **6** 28

Meena C, Rungta PD and Sinha S 2017 Threshold-activated transport stabilizes chaotic populations to steady states. *PLoS One* **12** e0183251

Meena C, Rungta PD and Sinha S 2020a Resilience of networks of multi-stable chaotic systems to targetted attacks. *Eur. Phys. J. B* **93** 210

Meena C, Hens C, Acharyya S, *et al*. 2020 Emergent stability in complex network dynamics. *arXiv:2007.04890v5*

Mei S and Zhu H 2015 A simple feature construction method for predicting upstream/downstream signal flow in human protein–protein interaction networks. *Sci. Rep.* **5** 17983

Messina F, Giombini E, Agrati C, *et al*. 2020 COVID-19 viral-host interactome analyzed by network based-approach model to study pathogenesis of SARS-CoV-2 infection. *J. Transl. Med.* **18** 233

Mistry D, Wise RP and Dickerson JA 2017 DiffSLC A graph centrality method to detect essential proteins of a protein–protein interaction network. *PLoS One* **12** e0187091

Mochida K, Koda S, Inoue K and Nishii R 2018 Statistical and machine learning approaches to predict gene regulatory networks from transcriptome datasets. *Front. Plant. Sci.* **9** 1770

Mulder NJ, Akinola RO, Mazandu GK and Rapanoel H 2014 Using biological networks to improve our understanding of infectious diseases. *Comput. Struct. Biotechnol. J.* **11** 1–10

Naderi Yeganeh P, Richardson C, Saule E, Loraine A and Taghi Mostafavi M 2020 Revisiting the use of graph centrality models in biological pathway analysis. *BioData* **13** 5

Nandi S, Ganguli P and Sarkar RR 2020 Essential gene prediction using limited gene essentiality information—an integrative semi-supervised machine learning strategy. *PLoS One* **15** e0242943

Narang P, Khan S, Hemrom AJ and Lynn AM 2014 Consortium OSDD. MetaNET - a web-accessible interactive platform for biological metabolic network analysis. *BMC Syst. Biol.* **8** 130

Nashiry MA, Sumi SS, Sharif Shohan MU, *et al*. 2021 Bioinformatics and system biology approaches to identify the diseasome and comorbidities complexities of SARS-CoV-2 infection with the digestive tract disorders. *Brief Bioinform.* **2** bbab126

Navlakha S, Gitter A and Bar-Joseph Z 2012 A network-based approach for predicting missing pathway interactions. *PLOS Comput. Biol.* **8** 1–13

Nazarieh M and Helms V 2019 TopControl: A tool to prioritize candidate disease-associated genes based on topological network features. *Sci. Rep.* **9** 19472

Newman MEJ 2006 Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103** 8577–8582

Newman MEJ and Clauset A 2016 Structure and inference in annotated networks. *Nat. Commun.* **16** 7

Nguyen LK, Matallanas D, Croucher DR, Von Kriegsheim A and Kholodenko BN 2013 Signalling by protein phosphatases and drug development: a systems-centred view. *FEBS J.* **280** 751–765

Nguyen P and Braun R 2018 Semi-supervised network inference using simulated gene expression dynamics. *Bioinformatics* **34** 1148–1156

Niu B, Liang C, Lu Y, *et al.* 2020 Glioma stages prediction based on machine learning algorithm combined with

protein–protein interaction networks. *Genomics* **112** 837–847

Niu X, Zhang J, Zhang L, *et al*. 2019 Weighted gene co-expression network analysis identifies critical genes in the development of heart failure after acute myocardial infarction. *Front. Genet.* **10** 1214

Novac N 2013 Challenges and opportunities of drug repositioning. *Trends. Pharmacol. Sci.* **34** 267–272

Oh M, Ahn J and Yoon Y 2014 A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. *PLoS One* **9** e111668

Oldham S, Fulcher B, Parkes L, *et al*. 2019 Consistency and differences between centrality measures across distinct classes of networks. *PLoS One* **14** 1–23

Oughtred R, Stark C, Breitkreutz B-J, *et al*. 2019 The BioGRID interaction database 2019 update. *Nucleic Acids Res.* **47** D529–D541

Oyelade J, Isewon I, Oladipupo F, *et al*. 2016 Clustering algorithms their application to gene expression data. *Bioinform. Biol. Insights* **10** 237–253

Ozturk K, Dow M, Carlin DE, Bejar R and Carter H 2018 The emerging potential for network analysis to inform precision cancer medicine. *J. Mol. Biol.* **430** 2875–2899

Panditrao G, Ganguli P and Sarkar RR 2021 Delineating infection strategies of leishmania donovani secretory proteins in human through host–pathogen protein interactome prediction. *Pathog. Dis.* **79** 8

Pavlopoulos GA, Kontou PI, Pavlopoulou A, *et al*. 2018 Bipartite graphs in systems biology and medicine: a survey of methods and applications. *Gigascience* **7** 1–31

Pavlopoulos GA, Secrier M, Moschopoulos CN, *et al*. 2011 Using graph theory to analyze biological networks. *BioData Min.* **4** 10

Peel L, Larremore DB and Clauset A 2017 The ground truth about metadata and community detection in networks. *Sci. Adv.* **3** e1602548

Plaimas K, Eils R and König R 2010 Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Syst. Biol.* **4** 56

Ponnapalli SP, Saunders MA, van Loan CF and Alter O 2011 A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. *PLoS One* **6** e28072

Potapov AP, Goemann B and Wingender E 2008 The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks. *BMC Bioinform.* **9** 227

Prasad K, AlOmar SY, Alqahtani SAM, Malik MZ and Kumar V 2021 Brain disease network analysis to elucidate the neurological manifestations of COVID-19. *Mol. Neurobiol.* **58** 1875–1893

Prifti E, Zucker JD, Clément K and Henegar C 2010 Interactional and functional centrality in transcriptional co-expression networks. *Bioinformatics* **26** 3083–3089

Proctor CH and Loomis CP 1951 Analysis of sociometric data. *Res. Methods Social Relat.* **2** 561–585

Rahmatallah Y, Emmert-Streib F and Glazko G 2014 Gene sets net correlations analysis (GSNCA): a multivariate differential coexpression test for gene sets. *Bioinformatics* **30** 360–368

Rai A, Pradhan P, Nagraj J, *et al.* 2017 Understanding cancer complexome using networks, spectral graph theory and multilayer framework. *Sci. Rep.* **7** 41676

Ramadan E, Alinsaif S and Hassan MR 2016 Network topology measures for identifying disease-gene association in breast cancer. *BMC Bioinform.* **17** 274

Rancati G, Moffat J, Typas A and Pavelka N 2018 Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **19** 34–49

Razaghi-Moghadam Z and Nikoloski Z 2020 Supervised learning of gene-regulatory networks based on graph distance profiles of transcriptomics data. *NPJ Syst. Biol. Appl.* **6** 21

Ren C, Li M, Zheng Y, *et al*. 2021 Identification of diagnostic genes and vital microRNAs involved in rheumatoid arthritis based on data mining and experimental verification. *PeerJ*. **9** e11427

Resendis-Antonio O, Hernández M, Mora Y and Encarnación S 2012 FUnctional modules, structural topology, and optimal activity in metabolic networks. *PLoS Comput. Biol.* **8** 1–13

Rowe E, Palsson BO and King ZA 2018 Escher-FBA a web application for interactive flux balance analysis. *BMC Syst. Biol.* **12** 84

Roy S 2012 Systems biology beyond degree, hubs and scale-free networks the case for multiple metrics in complex networks. *Syst. Synth. Biol.* **6** 31–34

Rubanova N, Pinna G, Kropp J, *et al*. 2020 MasterPATH Network analysis of functional genomics screening data. *BMC Genom.* **21** 632

Ruhnau B 2000 Eigenvector-centrality—a node-centrality? *Soc. Networks* **22** 357–365

Rungta PD, Meena C and Sinha S 2018 Identifying nodal properties that are crucial for the dynamical robustness of multistable networks. *Phys. Rev. E* **98** 022314.

Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, *et al*. 2014 Protein–protein interaction networks (PPI) and complex diseases. *Gastroenterol. Hepatol. Bed Bench.* **7** 17–31

Saha S, Sengupta K, Chatterjee P, Basu S and Nasipuri M 2018 Analysis of protein targets in pathogen–host interaction in infectious diseases a case study on *Plasmodium falciparum* and Homo sapiens interaction network. *Brief. Funct. Genom.* **17** 441–450

Saint-Antoine MM and Singh A 2020 Network inference in systems biology recent developments, challenges, and applications. *Curr. Opin. Biotechnol.* **63** 89–98

Santolini M and Barabási AL 2018 Predicting perturbation patterns from the topology of biological networks. *Proc. Natl. Acad. Sci. USA* **115** E6375–E6383

Sauer U, Lasko DR, Fiaux J, *et al.* 1999 Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.* **181** 6679–6688

Schreiber G 2021 Protein–protein interaction interfaces and their functional implications; in *Protein–Protein Interaction Regulators* (The Royal Society of Chemistry) pp 1–24

Sekula M, Gaskins J and Datta S 2020 A sparse Bayesian factor model for the construction of gene co-expression networks from single-cell RNA sequencing count data. *BMC Bioinform.* **21** 361

Sharma A and Ali HH 2017 Analysis of clustering algorithms in biological networks; in *Proc. 2016 IEEE Int. Conf. Bioinform. Biomed. BIBM 2016*, 2303–2305

Simko GI and Csermely P 2013 Nodes having a major influence to break cooperation define a novel centrality measure game centrality. *PLoS One* **8** e67159

Singh P, Amir M, Chaudhary U, *et al*. 2020 Identification of robust genes in transcriptional regulatory network of *Mycobacterium tuberculosis*. *IET Syst. Biol.* **14** 292–296

Sinha S 1997 Modelling biological systems. *Curr. Sci.* **72** 903–907

Sinha N, Chowdhury S and Sarkar RR 2020a Molecular basis of drug resistance in smoothened receptor: an in silico study of protein resistivity and specificity. *Proteins. Struct. Funct. Bioinform.* **88** 514–526

Sinha S, Jones BM, Traniello IM, *et al*. 2020b Behavior-related gene regulatory networks: a new level of organization in the brain. *Proc. Natl. Acad. Sci. USA* **117** 23270–23279

Soto-Girón MJ and García-Vallejo F 2012 Changes in the topology of gene expression networks by human immunodeficiency virus type 1 (HIV-1) integration in macrophages. *Virus. Res.* **163** 91–97

Soyer OS, Salathé M and Bonhoeffer S 2006 Signal transduction networks Topology, response and biochemical processes. *J. Theor. Biol.* **238** 416–425

Srivastava A and Sinha S 2014 Thermostability of in vitro evolved *Bacillus subtilis* Lipase A: a network and dynamics perspective. *PLoS One* **9** e102856

Stéphanou A and Volpert V 2016 Hybrid modelling in biology: a classification review. *Math. Model. Nat. Phenom.* **11** 37–48

Sun MW, Moretti S, Paskov KM, *et al*. 2020 Game theoretic centrality a novel approach to prioritize disease candidate genes by combining biological networks with the Shapley value. *BMC Bioinform.* **21** 356

Szalay KZ and Csermely P 2013 Perturbation centrality and turbine a novel centrality measure obtained using a versatile network dynamics tool. *PLoS One* **8** e78059

Szklarczyk D, Gable AL, Nastou KC, *et al*. 2021 The STRING database in 2021 customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49** D605–D612

Terkelsen T, Russo F, Gromov P, *et al*. 2020 Secreted breast tumor interstitial fluid microRNAs and their target genes are associated with triple-negative breast cancer, tumor grade, and immune infiltration. *Breast Cancer Res.* **22** 73

Terracciano R, Preianò M, Fregola A, *et al*. 2021 Mapping the SARS-CoV-2–host protein–protein interactome by affinity purification mass spectrometry and proximity-dependent biotin labeling: a rational and straightforward route to discover host-directed anti-SARS-CoV-2 therapeutics. *Int. J. Mol. Sci.* **22** 532

Tew KL, Li XL and Tan SH 2007 Functional centrality detecting lethality of proteins in protein interaction networks. *Genome Inform.* **19** 166–177

Tomkins JE and Manzoni C 2021 Advances in protein–protein interaction network analysis for Parkinson's disease. *Neurobiol. Dis.* **155** 105395

Toubiana D, Puzis R, Wen L, *et al*. 2019 Combined network analysis and machine learning allows the prediction of metabolic pathways from tomato metabolomics data. *Commun. Biol.* **2** 214

van Dam S, Võsa U, van der Graaf A, Franke L and de Magalhães JP 2018 Gene co-expression analysis for functional classification and gene-disease predictions. *Brief. Bioinform.* **19** 575–592

Wagner A and Fell DA 2001 The small world inside large metabolic networks. *Proc. R. Soc B. Biol. Sci.* **268** 1803–1810

Walker ML, Holt KE, Anderson GP, *et al*. 2014 Elucidation of pathways driving asthma pathogenesis development of a systems-level analytic strategy. *Front. Immunol.* **5** 447

Wang J, Wang W, Yan C, Luo J and Zhang G 2021 Predicting drug-disease association based on ensemble strategy. *Front. Genet.* **12** 666575

Wang L, Tu Z and Sun F 2009 A network-based integrative approach to prioritize reliable hits from multiple genome-wide RNAi screens in *Drosophila*. *BMC Genom.* **10** 220

Wang W, Yang S, Zhang X and Li J 2014 Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* **30** 2923–2930

Warde-Farley D, Donaldson SL, Comes O, *et al*. 2010 The GeneMANIA prediction server biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38** W214–W220

Wu G, Liu J and Wang C 2017 Predicting drug–disease interactions by semi-supervised graph cut algorithm and three-layer data integration. *BMC Med. Genom.* **10** 79

Xiao X, Moreno-Moral A, Rotival M, Bottolo L and Petretto E 2014 Multi-tissue analysis of co-expression networks by higher-order generalized singular value decomposition identifies functionally coherent transcriptional modules. *PLoS Genet.* **10** e1004006

Xu B, Guan J, Wang Y and Wang Z 2019 Essential protein detection by random walk on weighted protein–protein

interaction networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **16** 377–387

Yan W, Yu C, Chen J, Zhou J and Shen B 2020 ANCA: A web server for amino acid networks construction and analysis. *Front. Mol. Biosci*. **7** 582702

Yan W, Zhou J, Sun M, *et al*. 2014 The construction of an amino acid network for understanding protein structure and function. *Amino Acids* **46** 1419–1439

Yin T, Chen S, Wu X and Tian W 2017 GenePANDA-a novel network-based gene prioritizing tool for complex diseases. *Sci. Rep.* **46** 1419–1439

Younis H, Anwar MW, Khan MUG, Sikandar A and Bajwa UI 2021 A new sequential forward feature selection (SFFS) algorithm for mining best topological and biological features to predict protein complexes from protein–protein interaction networks (PPINs). *Interdiscip. Sci. Comput. Life. Sci.* **13** 371–388

Yu H, Lu L, Chen M, *et al.* 2019 KDDANet-a novel computational framework for systematic uncovering hidden gene interactions underlying known drug-disease associations. *bioRxiv* 749762 https://doi.org/10.1101/749762v3

Zambrana C, Xenos A, Böttcher R, Malod-Dognin N and Pržulj N 2021 Network neighbors of viral targets and differentially expressed genes in COVID-19 are drug target candidates. *Sci. Rep.* **11** 18985

Zand M and Ruan J 2020 Network-based single-cell RNA-seq data imputation enhances cell type identification. *Genes* **11** 377

Zhang X, Acencio ML and Lemke N 2016 Predicting essential genes and proteins based on machine learning and network topological features: a comprehensive review. *Front. Physiol.* **7** 75

Zhang X, Xiao W and Xiao W 2020 DeepHE Accurately predicting human essential genes based on deep learning. *PLoS. Comput. Biol*. **16** e1008229

Zhong J, Tang C, Peng W, *et al.* 2021 A novel essential protein identification method based on PPI networks and gene expression data. *BMC Bioinform.* **22** 248

Zhou J, Yan W, Hu G and Shen B 2014 Amino acid network for the discrimination of native protein structures from decoys. *Curr. Protein. Pept. Sci.* **15** 522–528

Zhu S, Bing J, Min X, Lin C and Zeng X 2018 Prediction of drug–gene interaction by using Metapath2vec. *Front. Genet.* **9** 248

Corresponding editor: Mohit Kumar Jolly