Topology of molecular interaction networks

Wynand Winterbach[*1,2], Piet Van Mieghem[2], Marcel Reinders[2,3,4], Huijuan Wang[2], Dick de Ridder[2,3,4]

[1]Network Architectures and Services Group, Department of Intelligent Systems, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft
[2]Delft Bioinformatics Lab, Department of Intelligent Systems, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft
[3]Netherlands Bioinformatics Center, 6500 HB Nijmegen, The Netherlands
[4]Kluyver Centre for Genomics of Industrial Fermentation, 2600 GA Delft, The Netherlands

Email: Wynand Winterbach*- w.winterbach@tudelft.nl; Piet Van Mieghem - p.f.a.vanmieghem@tudelft.nl; Marcel Reinders - m.j.t.reinders@tudelft.nl; Huijuan Wang - h.wang@tudelft.nl; Dick de Ridder - d.deridder@tudelft.nl;

*Corresponding author

Abstract

Molecular interactions are often represented as network models which have become the common language of many areas of biology. Graphs serve as convenient mathematical representations of network models and have themselves become objects of study. Their topology has been intensively researched over the last decade after evidence was found that they share underlying design principles with many other types of networks.

Initial studies suggested that molecular interaction network topology is related to biological function and evolution. However, further whole-network analyses did not lead to a unified view on what this relation may look like, with conclusions highly dependent on the type of molecular interactions considered and the metrics used to study them. It is unclear whether global network topology drives function, as suggested by some researchers, or whether it is simply a byproduct of evolution or even an artefact of representing complex molecular interaction networks as graphs.

Nevertheless, network biology has progressed significantly over the last years. We review the literature, focusing on two major developments. First, realizing that molecular interaction networks can be naturally decomposed into subsystems (such as modules and pathways), topology is increasingly studied locally rather than globally. Second, there is a move from a descriptive approach to a predictive one: rather than correlating biological network

topology to generic properties such as robustness, it is used to predict specific functions or phenotypes.

Taken together, this change in focus from globally descriptive to locally predictive points to new avenues of research. In particular, multi-scale approaches are developments promising to drive the study of molecular interaction networks further.

---

1 Introduction

Over the last half century, our understanding of life at the molecular level has advanced tremendously. This is made possible by continuously improving technology for measuring the presence or concentrations of molecules at a genome-wide level, such as the microarray (transcriptomics), mass spectrometry (proteomics, metabolomics) and next-generation sequencing (genomics). Perhaps more importantly from a systems biology perspective, similar technology and protocols have been developed to measure interactions among molecules, leading to so-called *interactomics* [1]. Protein-protein interactions are measured using yeast-two-hybrid technology and tandem affinity purification amongst others [2], and stored in a variety of databases [3]; interactions between DNA and proteins, such as histones and transcription factors, are found using yeast-one-hybrid and chromatin immunoprecipitation [4] and deposited in databases such as JASPAR [5] and FactorBook [6]; enzyme-metabolite interactions are measured using enzymatic assays and can be found in for example, BRENDA [7], KEGG [8] and MetaCyc [9]. Besides physical interactions, many indirect interactions have been reported, such as genetic interactions [10], general epistatic interactions [11] and predicted functional interactions [12].

This molecular interaction data is the cornerstone of many computational approaches aiming to analyze, model, interpret and predict biological phenomena, many at a genome-wide scale [13]. Interactions are often thought of as constituting networks, a view already proposed quite early [14] which recently came to full fruition [15]. Networks are now used as vehicles for modeling, storing, reporting, transmitting and interpreting molecular interactions [16]. Often they are represented as graphs, although this is not straightforward for many molecular interactions. For example, metabolic networks, representing physical interactions between enzymes and metabolites as well as conversions between metabolites, are ideally represented by

hypergraphs [17] but are often reduced to simple graphs [18] for further analysis.

Although graphs are convenient representations of molecular interaction networks, it was quickly realized that they could be treated similarly to large systems of interacting particles: small sets of interactions might be difficult to understand, but statistical properties relating to all interactions could contain valuable information [19]. This led to **network biology** [20]: a combination of systems biology, graph theory and computational and statistical analyses in which the topology of the graphs representing molecular interaction networks themselves became the subject of study. In subsequent work, statistically maintained properties, such as scale-freeness, were found in molecular networks of different types. In similar analyses, graphs were mined for statistically overrepresented network motifs [21], small subgraphs, suggesting that certain interaction patterns are common to many networks [22].

Despite their apparent universality, it proved difficult to derive biological conclusions from the patterns discovered in these initial global statistical analyses of molecular interaction networks. They may therefore be labeled as *descriptive*, pointing at generic underlying properties rather than leading to verifiable hypotheses. In time, molecular interactions networks were studied more locally, leading to more tangible biological insights. For example, clustering was used to discover significant biological modules and their interconnection patterns, which shed some light on evolutionary constraints of organisms [23]. Ranking of nodes by topological features (such as degree) was shown to relate to biological importance of a gene or protein and may for example be used to prioritize targets for development of pharmaceuticals [24]. We label such approaches *suggestive*. Finally, by studying networks even more locally, typically neighborhoods surrounding a few nodes, it has become possible to derive *predictive* results from molecular interaction networks. A typical approach is to compute a topological fingerprint of the neighborhood around a node; nodes are found to be functionally similar when their fingerprints are similar [25].

Over the past decade, network biology has thus transformed from being an initially descriptive approach to a predictive tool that is routinely applied to discover biologically relevant facts. In this survey, we chart this progression, showing that it corresponds well to a focus change from global to local. Many reviews of developments in network biology have appeared over the last years; here we list those most closely related to ours. Pržulj [26] reviews the use of protein interaction networks in network biology, touching on some of the techniques discussed throughout this review and calling for more integration of biological knowledge with network theory. A review of network theory from the perspective of data mining may be found in Pavlopoulos *et al.* [27]. This review covers a variety of network metrics with an especially strong focus on

clustering and node centrality. Likewise, Cho *et al.* [13] review several data-mining approaches applicable to molecular networks. A related topic is that of random molecular networks, which serve as benchmarks against which data mining results are measured. Such networks are generally produced through processes mimicking evolution, several of which are reviewed by Foster *et al.* [28] and Sun & Kim [29]. Finally, many recent reviews focus on the use of network biology in diagnosing disease [30–32], in particular network-based disease markers.

Our review adds to the existing literature by taking a high-level view of network biology as moving from descriptive to predictive, and by maintaining a clear focus on research exploiting the topology of molecular interaction graphs. The remainder of the paper is organized as follows: in Section 2, a brief overview of relevant biological and mathematical theory is presented. Sections 3-5 then give a chronological overview of research on the graph topology of molecular interaction networks, moving from descriptive to suggestive and predictive. We end with a conclusion and outlook in Section 6.

2 Network Biology

For the purposes of this review, we define network biology to be the study of the topology of graph representations of molecular interaction networks, both to describe such networks and as a tool to make biological predictions. We briefly review graph theory and discuss graph representations of molecular interaction networks.

2.1 Graph Theory

Graph theory is the study of **graphs**: structures representing relationships between pairs of objects. The set $\mathcal{N}$ of objects in a graph G are called **nodes**; the relationships between the objects are captured by a set $\mathcal{L}$ of node pairs called **links**. When nodes u and v are linked (i.e. $\{u, v\} \in \mathcal{L}$), u is said to be a **neighbor** of v and vice-versa. In **directed graphs**, used for modeling non-symmetric relationships such as activation or repression, each link is directed and has a source node (origin) and a target node (destination). The number of neighbors of a node u is called its **degree**. Figure 2 shows examples of directed graphs. **Weighted graphs** model non-binary relations by associating scalars or **weights** with links. An example is the affinity with which proteins bind to one another. Box 1 lists some metrics often used to study graphs. Many more metrics in the context of network biology are covered in [27].

Metric types	Metric descriptions
Degree Distribution	The statistical distribution followed by the degrees of the nodes in a network. Many real-world networks have degree distributions that depart sharply from those of classical random network models (Box 3).
Path Metrics	In an unweighted graph G, the shortest path between nodes u and v is the minimum number of links one must traverse to move from u to v. If G is weighted, the shortest path is that with the minimal sum of link weights. The **average shortest path** or **characteristic path length** is the average length of all shortest paths (between all node pairs) in a network.
Centrality Metrics	A centrality metric gives a ranking of nodes according to their "importance". The simplest measure is **degree centrality** – the degree of a node specifies its importance. **Closeness centrality** is the reciprocal of the sum of the shortest paths to all other nodes (i.e. a node whose closeness centrality is high is close to many nodes). **Betweenness centrality** is the fraction of shortest paths passing through a node. **Eigenvector centrality** and **Pagerank** are measures of how frequently one arrives at a node when performing a random walk on a network.

Table 1: Graph metrics reduce structural properties of network to (vectors of) real numbers, facilitating the comparison of different networks.

An **induced subgraph** G' of G is a subset of the nodes of G, along with all links whose endpoint nodes are both in G'. In a **bipartite graph**, the nodes can be split into two sets such that no two vertices in the same set are adjacent. A complete bipartite graph in which all nodes from the first set are connected to all nodes in the second is said to be **complete**.

2.2 Molecular Interaction Networks

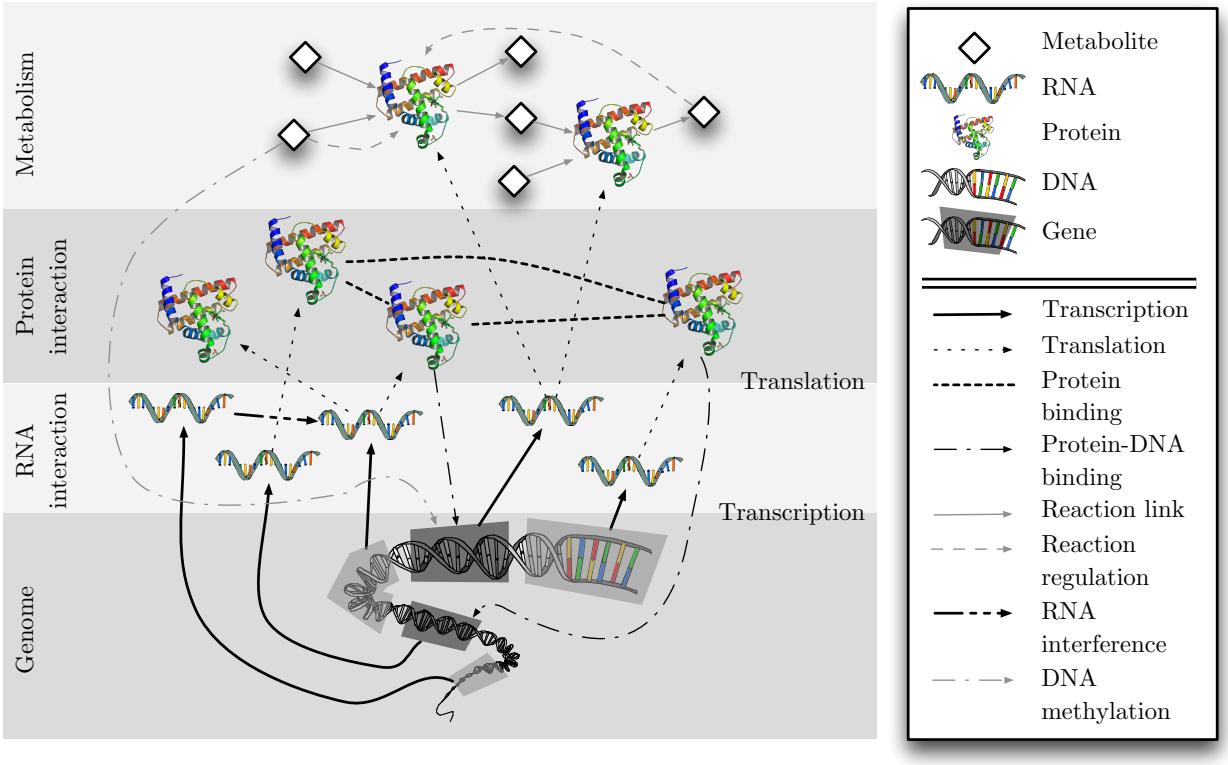
Molecular biology is the study of all cellular processes involving DNA, RNA, proteins and metabolites. A simplified overview of common interactions between these molecules is shown in Figure 1 (a). Although simplified, models such as Figure 1 (a) are still complex. Researchers generally study models with fewer molecules and interactions, such as the signaling pathway model in Figure 1 (b).

Both Figures 1 (a) and (b) focus on interactions and can therefore be represented as networks. But neither is a graph, since Figure 1 (b) contains non-pairwise relationships and Figure 1 (a) contains multiple types of relationships while both contain multiple types of nodes. Complex interaction models that distinguish between node and link types are useful when the focus of study is on a small molecular subsystem but a hindrance when the aim is the discovery of *interaction patterns* across large sets of interactions. When pattern discovery is the aim, networks are reduced to graphs by including only links and nodes modeling

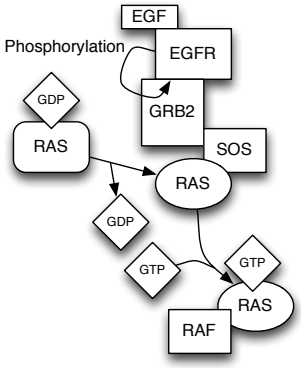
Type of network	Network description
Association networks	Association networks model *any* kind of relation between molecules (e.g. binding, co-expression and structural similarities). Examples of association networks are **gene co-expression networks** and **protein similarity networks**.
Functional networks	Functional networks model functional relations between pairs of molecules (usually genes or proteins). A link implies that both are involved in the same function, process or phenotype. **Genetic interaction networks** represent interactions where a double mutation leads to an epistatic effect, i.e., worse or better than expected based on the single mutation.
Protein-protein Interaction Networks (PPI Networks)	Protein-protein interaction networks are undirected networks that model protein binding. PPI networks are derived from high-throughput experiments using techniques such as yeast two-hybrid screening, mass spectrometry and tandem affinity purification [2]. **Signaling networks** are related to protein interaction networks, but their links are directed according to the flow of molecular signals.
Transcription-regulatory Networks (TR Networks)	Transcription-regulatory networks are bipartite networks with one set of nodes representing genes and the other representing transcription factors (TFs). TFs are products of genes (modeled by gene-TF links) whilst genes are regulated by TFs (modeled by TF-gene links). Data for such networks is derived through the process of chromatin immunoprecipitation (ChIP) [33]. **Gene regulatory (GR) networks** are related to TR networks but contain only genes. Their links represent indirect regulatory relationships.
Metabolic Networks	Metabolic Networks are bipartite networks that model the relationships between the chemical reactions that occur in cells and the substrates involved in the reactions (the solid gray lines in Figure 1 (a)). Reduced, non-bipartite metabolic networks containing only metabolites or only reactions are also often studied.

Table 2: Commonly studied molecular interaction networks.

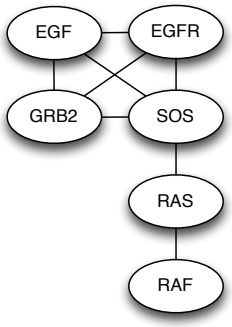
one or two concepts and by converting non-pairwise links to pairwise links. The graph in Figure 1 (c) is one possible simplification of the pathway in Figure 1 (b). While **network** and **graph** are thus two distinct concepts, we will henceforth use the term **network** to refer to both concepts. Box 2 lists several such networks commonly studied.



(a)

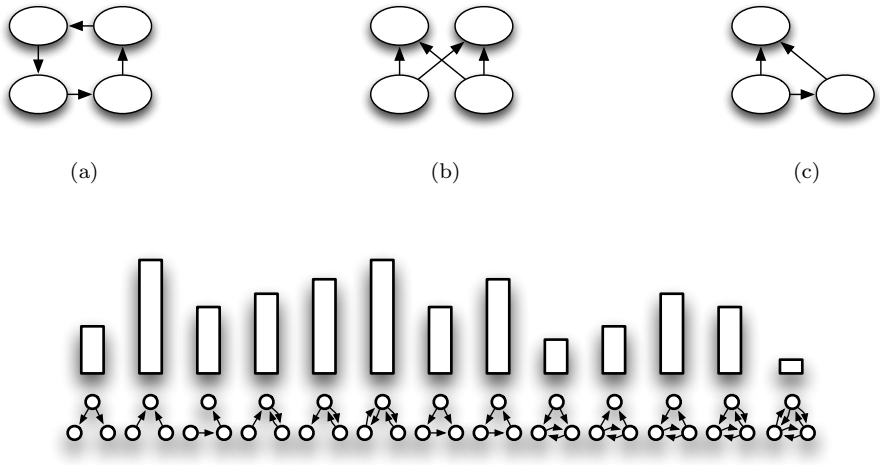


(b)



(c)

Figure 1: From biological models to networks. **(a)** Simple overview of molecular interactions in the cell. **(b)** Part of the MAPK/ERK pathway modeled as a network. **(c)** Homogenous protein interaction graph representation of part of the MAPK/ERK pathway.



(a) (b) (c)

(d)

Figure 2: Some motifs thought to be overrepresented in molecular interaction networks. Arrowheads indicate link directionality. **(a)** A four-node feed-back motif. **(b)** A four-node bi-fan motif. **(c)** A three-node feed-forward motif. **(d)** Three-node motif signature for a network.

3 Descriptive Analysis

During the 1990's, researchers in various scientific fields started studying macro-scale systems in which individual entities locally interact in simple ways, leading to complex behavior emerging at a global scale. Examples include telecommunications networks [19, 42], social relationship structures [35] and biological interactions from the molecular to the ecological scale [21].

The structure of the above networks departed significantly from the random network models – the Erdős-Renyí model [34] and the Watts-Strogatz model [35] – commonly used in that day to model large networks (see Box 3). Real-world networks had short average path lengths and degree distributions approximating power laws [19]. The slopes of the degree distributions, when plotted on log-log axes, tended to fall within a narrow range, regardless of the numbers of nodes in these networks. This independence of scale or **scale-freeness** was thought be indicative of networks formed through gradual growth processes based on **preferential attachment**: every time a node is added to a network, it is linked to existing nodes with probabilities proportional to the degrees of those nodes [19, 20].

In biology, initial studies on molecular interaction networks matched the topologies observed in other real-

Type of network	Network description
Erdős-Renyí (ER) [34]	The oldest class of random networks. To construct a graph instance, links are added between each pair of nodes with probability p (a parameter).
Watts-Strogatz (WS) [35]	A kind of generalization of ER networks in which links of a regular lattice are rewired. Characterized by high clustering coefficients and short average path lengths.
Barabási-Albert (BA) [19]	A class of random networks constructed one node at a time, with new nodes preferentially attaching to existing high-degree nodes. These networks are scale-free (i.e. hub-like) and more closely resemble molecular interaction network networks than ER or WS networks.
Duplication-divergence	These networks, inspired by gene duplication and subsequent divergence (in sequence, interaction and function) [36] are generated by duplicating nodes and randomly removing/adding links. Architecturally, duplication-divergence networks are similar to Barabási-Albert networks [37, 38]
Fixed node degrees	Random networks characterized by their specific node degree sequences that are generated either by randomly rewiring the links of an existing network [39] or through the configuration model [40, 41].

Table 3: In graph theory, topological characteristics of a network are often compared to those of instances of random network models. Listed are a few widely used random network models in which nodes represent a single concept; these are generally unsuitable for generating networks in which nodes correspond to multiple concepts (e.g. metabolites and reactions in metabolic networks) since additional structural constraints apply to their connectivity.

world networks. Gene co-expression networks [43], protein-protein interaction networks [44], metabolic networks [45] and transcription regulation networks [20] all contain aspects of scale-free networks. Nevertheless, although various random network models reproduce some salient properties of molecular networks, each has been criticized for not being consistent with other important aspects of molecular networks [46–49].

Molecular networks are often also highly clustered, implying **modular** design (see Box 4) and supporting the idea that biological systems are modular at all levels [50]. An early study on the *S. cerevisiae* PPI network showed proteins with similar functional annotations to be highly connected, strongly suggesting modularity [25]. Similarly, in the yeast TR network, highly co-expressed genes were found to be clustered [51]. Evidence for hierarchical modularity was found in a PPI network [52] and in the metabolic networks of several organisms [53]. In general, molecular interaction networks were increasingly thought to consist of modules, linked through connector or linker nodes [54]. In other words, molecular networks are networks of networks that can tolerate disruptions to individual modules but whose functions are sensitive to disruptions module of connectors.

Although early attempts at understanding molecular interaction networks took a top-down approach, char-

Network decomposition	Decomposition description
Modules	are induced subgraphs whose link density is high in comparison to the rest of the graph. This definition is deliberately vague, as what constitutes a module depends on the context and the algorithm used to discover modules.
Motifs	are small subgraphs, usually of 3 or 4 nodes, whose over- or underrepresentation may indicate that their structures are important or detrimental to the system [21]. Usually, all distinct motifs in a network are counted, yielding a motif signature for the network that may then be compared to signatures obtained by sampling from an appropriate random network null model (see Box 3) to determine over- or underrepresentation. A signature for all motifs on 3 nodes is shown in Figure 2 (d). Motif signatures can be used to characterize networks.
Graphlets	are similar to motifs but always fully connected. As with motifs, graphlets are used to construct signatures that capture the local characteristics of a network [55].

Table 4: Modules, motifs and graphlets. These concepts are used to decompose networks into smaller units that are easier to study.

acterizing networks using global metrics such as their degree distributions, it was soon suggested that global behavior of the cell could be the result of local features [56], a bottom-up view. One view was that behavior of molecular interaction networks emerges from the interactions of many small subgraphs or **motifs** (see Box 2), in the same way that the behavior of a computer results from the interactions of simple logic circuits [21]. Statistical overrepresentation of a motif is thought to be evidence that the motif offers a functional advantage to its host organism. Such motifs – feed-back loops, feed-forward loops and bi-fan motifs (see Figure 2) – all have analogues in the electronic world [21]. This fitted well with the increasing popularity of systems biology [57] that advocated an engineering-inspired approach to study biology. Simple motifs may act as sign-sensitive delay mechanisms or as input response-accelerators, depending on their mix of activators and repressors [22]. More complex motifs may even act as logic circuits, switches and memory states, making them interesting building blocks for synthetic biology [58].

Motifs can also be used to characterize networks more globally. Global motif signatures were found to be unique for different types of networks [21] but conserved between organisms [59], providing further evidence that motifs embody underlying design principles in different types of molecular interaction networks, that are preserved across evolution [22].

The global, module and motif views led to the idea that molecular networks are organized at multiple levels of complexity [60]. At the local level, motifs act as small control circuits or building blocks. Motifs aggregate

into modules that, through the interactions of their motifs, implement more complex biological processes. At the global level, modules are connected to each other – and may thus exchange information or molecules – through a small number of linker nodes. The fact that certain topological features, such as scale-free degree distributions, are common among molecular networks suggests that the designs of these networks are shaped at all levels by evolutionary mechanisms.

The case for an architecture based on a hierarchy of motifs, modules and global properties was strong and it appeared to be universal, so that its presence came to be assumed. At the local level, overrepresented motifs were used to filter spurious links from noisy high-throughput networks by rejecting links that did not form part of motif structures [61]. At the global level, the assumption of power-law degree distributions led researchers to propose the evolutionary processes of duplication and divergence as leading to preferential attachment in the formation of molecular networks [36].

3.1 Limits to the Descriptive Approach

Details of the multi-layered view were increasingly disputed as data quality improved and as researchers revisited interpretations of older findings. At the global level, the most contested trait was that of scale-freeness, a property found to arise under many circumstances, challenging its significance [62]. Careful examination of molecular interaction data showed that some non-scale-free distributions fit degree distributions of molecular networks as well as scale-free distributions [63, 64]. More contentious was the suggestion that some global features are modeling artifacts. The hub-like architecture of protein interaction networks was questioned, since no protein can realistically bind to the number of proteins suggested by hub nodes; hub nodes are more likely to represent groups of proteins that only appear to be individuals owing to experimental limitations [46]. Likewise, metabolic networks do not display short average path lengths when metabolite paths are traced; shortest path algorithms on metabolic networks do not take into account the requirement that all metabolites be present for a reaction to occur and their direct application to these networks is meaningless [17].

At the module level, it was found that modules are less clearly delineated than previously assumed. There appeared to be many connections between modules, making it difficult to distinguish linker nodes [65]. Without linker nodes, assignment of nodes to modules is more difficult, leading to "fuzzy" modules. Motifs were also criticized. The bi-fan motif, found to be overrepresented in molecular networks [21] and assumed to be functionally important, was shown to have no characteristic behavior when considered as a dynamic

system [66]. If motifs lack characteristic behavior, aggregates of motifs, such as motif clusters, cannot be assumed to implement specialized biological functions. Motif signatures (Box 4 and Figure 2 (d)) of networks were argued to be by-products of simple evolutionary mechanisms (such as gene duplication and divergence) [67]. Evolution may thus not be driven by motifs; rather, motifs may be the inevitable result of the self-organizing effects of evolution.

Although there is less universal structure in molecular networks than once thought, the original multi-layered model is still useful, albeit with some modifications. There is much evidence that molecular networks are not scale-free, but they are generally heavy-tailed [64], meaning that they have a few hubs and many low-degree nodes. Motifs may not be simple biological circuits [21], but they established the idea that local structure is important; one way in which this was later exploited was to compute node signatures for use in function prediction in molecular networks [55] and alignment of molecular networks [68]. Perhaps the most important contribution of the layered view was the idea that molecular networks are organized at multiple levels; the molecular organization of the cell cannot be understood at one scale only.

3.2 Topological Features as Target or By-product of Evolution

The global approach was not meant to be purely descriptive: its original goal was the discovery of universal architectural features. Universality suggests that organisms are selected *because* they posses such features and would provide clues about the topological requirements that are essential to life.

One property thought to emerge from natural selection is *robustness*, the ability to maintain function under perturbations [69]. Network biologists have sought to explain robustness in terms of topological characteristics. In PPI networks, the number of interaction partners of nodes initially appeared to correlate with their essentiality [56]: robustness may come from the fact that PPI networks have few hubs and many low-degree nodes. In metabolic networks, almost the opposite is true, with networks being susceptible to disruption of low-degree linker nodes that connect metabolic modules [70]. However, in both cases the systems are resilient to most perturbations but susceptible to targeted attacks, a property known as *highly optimized tolerance* [71].

After-the-fact attempts to match topology to properties such as robustness were eventually called into question. *In silico* evolution experiments with simple gene-regulatory networks showed that many such structural features emerge from network dynamics rather than selective pressure [72]. Other such network

evolution experiments suggested that the drivers could be simple processes such as reuse, genetic drift and mutation [67, 73, 74]. Even higher-level organization such as modularity is thought to arise from such simple processes [23]. A study comparing a metabolic network to a network of atmospheric chemical reactions found large topological similarities and concluded that many large-scale topological features have no functional nor evolutionary significance, the so-called **neutral theory of chemical reaction networks** [75]. In bacteria, horizontal gene transfer is thought to play an important role in module formation, as cells adopt clusters of foreign genetic material wholesale in reaction to environmental variability [76]. Nevertheless, the extent of this influence was recently questioned, stressing possible interplay between variability and gene transfer [77, 78].

Not all network features emerge through network dynamics. Selection pressure does seem necessary for the fine-tuning of topological features and may in some cases be responsible for the difference between a robust and fragile network [79]. In simulations of metabolic network evolution, hubs emerge when networks are selected for their ability to grow [80]. In models of GR network evolution, sparsity (i.e. low link counts) emerges when selectional stability (which models energy minimization of the mutation process) is enforced [81]. Even modularity may rely on selection pressure, albeit in a more subtle form. When networks are evolved and selected for their ability to prosper in varying conditions, modularity is found to emerge and, crucially, to be maintained [82]. A similar result was obtained by subjecting randomly generated metabolic networks (i.e., not generated by a procedure mimicking evolution) to a range of environments and assessing the amount of biomass they produced [83].

4 Suggestive Analysis

Since the early days of network biology, data mining was used to discover unexpected (ir)regularities in molecular interaction networks. Some findings were already discussed in Section 3 (the use of clustering to discover functional annotation, the existence of hub proteins). While data mining techniques shed light on aspects of biological function, they do not necessarily lead to directly testable hypotheses. In this sense, we call the methods in this section "suggestive". We describe four strategies for extracting network regularities: significant feature detection, clustering, central and hub node discovery and network homology.

Significant Feature Detection The idea behind this strategy is that unlikely patterns in molecular networks are indicative of underlying "design" processes (such as evolution). The likelihood of a feature is determined

by considering its distribution in network instances generated using a random network model (see Box 3). In early work, PPI networks were rewired (link pairs were shuffled) to generate random networks [39]. The connections between high-degree nodes in the original protein interaction network were found to be statistically unlikely in rewired networks, leading to the hypothesis that interactions between high-degree proteins are suppressed in evolution, perhaps to control cross-talk in the cell. Modules and motifs [21] can also be considered as significant features. Some of the clustering algorithms mentioned earlier in this section explicitly assess cluster significance as a function of its likelihood [84].

Such significant features can sometimes be biologically interpreted. Statistical analysis of miRNA targets in a human signaling network found that miRNAs tend to target proteins that are part of positive feedback motifs [85]. Similarly, cancer genes tend to be part of positive feedback motifs whilst genes that are highly methylated tend to be part of negative feedback motifs [86]. In both of these cases, the motifs are interpreted as amplification or dampening circuits, analogous to electronic circuits. An interesting recent view is that individual motifs are not necessarily significant but that large clusters of positive or negative feedback motifs act as stochastic amplifiers or dampers, respectively [87].

The advantage of significant feature detection lies in its simplicity: existing techniques are used to analyze and compare the input network and networks derived from a random model. But this is also its main drawback: choosing an incorrect random network model can make features appear significant when they are not.

Clusters Modules in complex systems tend to be highly internally connected whilst sharing only a few connections with the outside world. Graph clustering is an approach to discover such modules by decomposing a network into a number of subnetworks or **clusters** that are internally highly connected. The "big data" era has inspired development of clustering algorithms that efficiently deal with large datasets.

In network biology, general clustering algorithms have been used to discover functional modules in gene co-expression networks [88] and genomic cooccurence networks [89]. Since proteins in complexes highly interact with one another, graph clustering has also been used to discover protein complexes in PPI networks [54]. Here we mention a few of such general clustering algorithms; the interested reader is referred to [90] for a more thorough overview. Most modern clustering algorithms are based on physical models, data mining techniques or spatial partitioning. Physics-inspired approaches include spin models [91, 92], random walk models [93,94] and synchronization models [95]. Data mining approaches treat cluster discovery as a problem

of significant feature discovery. A few clustering algorithms discussed below are (at least partially) based on this idea. Spatial partitioning approaches associate distance metrics on pairs of nodes that are then clustered using approaches such as k-means clustering. A number of such distance metrics are discussed later in the context of "neighborhood homology" later in this review.

Whilst general algorithms can be applied to molecular networks, clustering algorithms that exploit the specific structure of molecular networks may achieve better results. MCODE is a heuristic algorithm developed to detect complexes in protein interaction networks [96]. Other examples include Restricted Neighborhood Search Clustering [97] and CODENSE, an algorithm for finding dense subgraphs [98]. A number of algorithms based on local neighborhood statistics were proposed as well, for example to find subgraphs of PPI networks that are active according to high-throughput measurements (ActiveModules [99] and MATISSE [100]). More generally, a likelihood score for the density of a subgraph can be used in (greedy) optimization algorithms to mine dense subgraphs, such as in CEZANNE, which finds functional modules in gene co-expression networks [100].

Besides fully connected clusters, clusters that resemble bi-cliques (complete bi-partite subgraphs, see Section 2.1) have been shown to be common and biologically relevant in protein interaction networks [101]. Furthermore, clusters in bipartite networks such as TR and metabolic networks are also manifested as bi-clique-like networks. Algorithms have been proposed to mine such (bi-)clique clusters [102,103]. Specialized algorithms for bipartite networks have also been developed, such as SAMBA, that integrates additional biological data to discover modules [104].

A still-difficult problem is the discovery of overlapping clusters. Many molecules are components of multiple modules (e.g. proteins are part of multiple protein complexes, metabolites are inputs to multiple metabolic reactions) whilst most existing clustering algorithms place each molecule in exactly one cluster. A relatively simple approach is to group molecules in topics and to apply node-based clustering on the topics; a node that belongs to topics in different clusters would be a member of (at least) two clusters. Recent research uses the more restricted case of edge clustering (which is equivalent to topic clustering on topics of two nodes each) with good success [105–107].

Clustering is a useful technique to gain understanding of the modular construction of a molecular network, but caution is required. Recovered clusters may not reflect actual biological modules; inaccurate clustering can arise from badly chosen clustering criteria (in particular from criteria unrelated to biological

constraints) [108]. Algorithms that produce overlapping clusters may assign nodes to too many or too few clusters and rigorous techniques for handling such problems are still lacking.

Central Nodes and Hubs Early findings in network biology suggested that some nodes are more important or *central* [109] (see Box 1) in molecular interaction networks. This manifestation of highly optimized tolerance entails that the survival of an organism depends more on the presence of a few central nodes than on most other, less central nodes. First, it was found that disrupting the highly connected, "hub-like" p53 gene in the human signaling leads to cancer [110]. It was subsequently shown that the number of interaction partners of a protein (i.e., *degree centrality*) in the *S. cerevisiae* protein interaction network is correlated with its lethality [56]. Research on protein interaction networks [111], co-expression networks [112] and synthetic genetic interaction networks [113] showed similar correlations. Furthermore, the number of interaction partners was shown to be negatively correlated with the rate of evolution in protein interaction networks [114], metabolic networks [115] and transcription-regulatory networks [116], further supporting the idea that central nodes are important.

Closeness centrality was used to find central metabolites in metabolic networks [117]. Betweenness centrality was used to identify bottleneck nodes – nodes of low degree whose removal is fatal to the organism [118]. Both of these metrics fit the interpretation of central nodes as being chemical flow routers. In signaling networks, disruption of central nodes has been linked to cancer, suggesting that they act as information co-ordinators/routers [119,120]. However, not all centrality measures can be easily related to routing, examples of which include subgraph centrality [121], coreness centrality [122], bipartivity (the fraction of closed loops including the node that are of even length) [123] and node hierarchy [124].

In spite of the initial positive findings, further experiments on *S. cerevisiae* showed little correlation between protein degree and essentiality [125], a finding strengthened by computer simulations of gene expression [126]. This cast doubt on the use of centrality measures alone to predict node functionality. Some researchers have sought to refine the notion of centrality by considering interaction patterns of central nodes: those that interact with many interaction partners simultaneously are called "party" hubs whilst those that interact with a few of their partners at a time are called "date" hubs [127]. Party hubs are thought to be global coordinators that connect components within network modules whilst date hubs may be local coordinators that connect network modules [127]. However, this distinction has been challenged with the availability of new data that does not show such clear distinctions between central nodes [128].

Even if node centrality is not as well correlated with node function as hoped, research in this field has shown that hubs do tend to be more essential than non-hubs. Furthermore, subversion of central nodes has been implicated in the formation of cancer [119, 129], suggesting possibly useful drug targets.

It has been suggested that a simple explanation for the essentiality of high degree nodes is that they are more likely to interact with essential complexes and their removal breaks such complexes [125]. The implication is that local topology is a deciding factor in essentiality. Indeed, versions of existing centrality measures modified to take more local information into account are better at predicting which nodes are essential [130]. However, it is important not to conflate node essentiality, a concept tied to survivability, with the influence that a node exerts on a network. The latter concept is discussed in the next section in the guise of "controllability".

Global Homology The principle of homology states that biological systems related by evolution are structurally similar. Its converse – structural similarities imply common heritage – is often used to predict the function of unknown proteins and genes. In networks, topological similarity can likewise be used to infer functional similarity. Using this approach, metabolic networks of 43 organisms were found to display hierarchical modularity [53]; these modules were found to center around core metabolites [131]. In the same vein, the connectivity of a protein in a PPI network was shown to be proportional to its age. In a study on three species, common proteins are likely to be older than those present in only a single species [132].

The approaches above focus on high-level similarities between networks without attempting to match individual nodes in the networks. By performing such alignments, clustering and significant feature detection applied across species can lead to more insight. In an early example, the glycolytic pathways of 17 organisms were aligned [133] and revealed many interesting differences between species in this essential part of metabolism. Alignment of the *E. coli* metabolic network to those of other organisms identified enzymes whose genes were candidates for horizontal gene transfer [38]. The average degree of these candidates is higher than that of other enzymes, implying that they are central to metabolism. Thus, ancestors to *E. coli* replaced their central enzymes with better functioning enzymes from other species.

Data Mining in Biological Networks Suggests Biological Findings Data mining techniques have been successfully applied in network biology to suggest biological functions for genes and proteins. The common theme is that instead of considering global properties of biological networks, they focus on subnetworks, from individual nodes to neighborhoods and features shared between networks. This increased focus allows the derivation

of more tangible biological results. However, when analyses are based on comparisons to random network models (Box 3), such as in significant feature detection, the problem of telling these apart from evolutionary by-products remains.

5 Predictive Analysis

The data mining approaches discussed in Section 4 reveal the large-scale organization of molecular networks in some detail but do not, in general, yield testable biological hypotheses. Approaches that do give such results tend to be based on network generalizations of existing principles in molecular biology: guilt-by-association, homology and differential analysis.

Guilt-by-association The principle of guilt-by-association is based on the observation that if most of the interaction partners of a molecule are associated with some property (such as a specific biological process or molecular function [134]), the molecule itself is also likely to be associated with that property [135]. Guilt-by-association has been used to assign functions to proteins with unknown roles based on the functions shared by the majority of their direct neighbors (i.e. interaction partners) in protein interaction networks [25]. The properties shared by the majority of a node's neighbors do not necessarily yield the best annotations [136] and more sophisticated approaches, such as Markov random fields trained on node neighborhoods [137], have been developed as alternatives.

By only taking direct interactions into account, the above applications of guilt-by-association ignore the impact of potentially informative indirect interactions. So-called n-hop features have been used to predict disease associations of proteins in PPI networks [138]. Another technique for incorporating indirect neighbors is graph diffusion, an idea derived from the study of diffusion in physical systems. Here, properties of nodes are diffused across links in a network; properties that diffuse in high quantities to nodes with unknown roles are used to annotate these nodes [139]. In both n-hop methods and graph diffusion, interaction strength between nodes depends on the path structure between the nodes.

Path structure need not be the only determinant of interaction strength. Nodes that are members of the same biological module may have similar functions [25]. Thus, a node whose role is unknown can be annotated with the functions appearing most frequently in the module(s) to which it belongs. Whilst we do not know what the biological modules are, we can compute approximate modules through clustering. Such an approach has been used to annotate unknown proteins in *S. cerevisiae* protein interaction networks [102].

Guilt-by-association is a simple and effective technique that extends naturally to networks. However, it is only effective when the roles of the majority of molecules in a network are known, limiting the technique to well-studied organisms.

Neighborhood Homology Since the use of homology is pervasive in biology, we expect the principle to extend to networks. Indeed, in Section 4 it was already discussed how networks found in different organisms have similar structural properties. Predictive approaches use topological and possibly biological similarity to match similar nodes across different networks. Once nodes are aligned, the function of a protein or gene whose role is unknown can be predicted, if the function of its matched node in the other network is known.

The first network alignment algorithms operated at a local level, attempting to match only small parts of entire networks to one another [68, 140]. Global alignment is more difficult, because networks to be aligned generally differ in size. Moreover, homology is not a one-to-one relation: many nodes may align to many nodes. There are two main approaches for performing global alignment:

1. Cluster the nodes in each network and compute topological matching scores on the clusters [141, 142] ("matching clusters").

2. Select groups of nodes in different networks that are pairwise similar in local neighborhoods and possibly biological labels [143, 144] ("clustering matches").

The first type of algorithm has the disadvantage that the clustering step precedes matching and thus ignores potentially useful information. Many algorithms of the second type associate feature vectors of topological (and possibly biological) attributes with nodes that are then used to compute node similarity. Various metrics have been used [145]. The Jaccard coefficient, a measure of overlap between sets of binary attributes, has been widely used, an example of which was the prediction of protein function in human PPI networks [146]. The h-confidence metric [147] is a data-mining tool for discovering associations and has been used in protein function prediction. Specialized metrics, such as the graphlet distance (tailored to graphlet signatures [55]) have been used to discover genes implicated in cancer [148].

Variations of clustering algorithms, looking for dense subgraphs within one network, have been proposed to mine subgraphs similar in two networks. For example, the PathBlast algorithm combined a statistical score for protein similarity and probability of a reported protein interaction to mine pathways or complexes occurring in PPI networks of different species [140]. Similar approaches were applied to assign functions to

proteins [149] and to align metabolic pathways [150].

Differential Analysis Diagnosis of many diseases (such as cancer) is based on the fact they influence the regulation programs of cells. Traditionally, this involved finding changed expression of marker genes, or specific gene mutations, i.e. focusing on the nodes in the network. Network biology allows additional focus on node relations, making it possible to diagnose molecular diseases that cannot be well characterized by the traditional techniques [151]. This so-called differential analysis, finding changes in network structure [31], is currently complicated by the fact that construction of high-quality molecular networks requires considerable time and resources. One common way around this is to use an existing high-quality network, typically a PPI or TR network, as a scaffold onto which noisy high-throughput patient data (typically gene expression or methylation data) is overlaid. If multiple measurements are available for each patient, gene coexpression/comethylation values can be computed and overlaid as link weights on PPI links.

Expression changes of genes/proteins linked to central nodes in molecular networks have been proven to be reliable markers of disease. Differential expression around topologically central nodes in protein interaction networks has been used to diagnose cancer [152,153]. Disease central nodes (i.e., nodes implicated in disease) have been similarly used in the diagnosis of breast cancer and leukemia [154]. More recently, co-expression changes around biologically central nodes, such as signaling hubs, have shown to be even more reliable disease markers [155, 156].

More elaborate differential approaches consider changes in expression patterns of subnetworks, instead of only central nodes. Automatic extraction of such subnetworks based on topology and measurements such as gene expression has revealed subnetworks associated with cancer (in which differential gene/protein expression could be used for diagnosis of the disease) [86, 157] as well as subnetworks that are implicated in heart failure [158]. An alternative to automatic extraction is to use biological modules based on theoretical knowledge; such an approach has been used in cancer prognosis [159].

Differential diagnosis, despite its relative newness has quickly grown to a large field. Our discussion is necessarily limited by the scope of this review; the interested reader is referred to recent reviews that consider the discipline in much more depth [31,32,160].

Relating Topology to Biological Properties Leads to Predictive Power The data mining techniques discussed in Section 4 are mostly based on topological information. In contrast, the predictive approaches discussed above depend on additional biological information. This approach to network biology clearly yields more

testable hypotheses than the suggestive and descriptive approaches.

Since we do not, in general, have good models of biological function at large scales, predictive approaches are most often applied to small groups of nodes or subnetworks. There are exceptions with metabolic networks being the most prominent. Flux balance analysis (FBA) [161,162] is a framework for computing steady-state reaction rates in such networks based on reaction stoichiometry, assuming the cell attempts to achieve some objective such as maximum growth. FBA is often used in a predictive way, but has also been applied in a "suggestive" setting, e.g. to study robustness of metabolic networks [70]. FBA allows one to take additional physical constraints into account, such as thermodynamic interactions [163] or responses to signaling [164]; for an extensive overview see [165].

The biggest problem with incorporating additional biological knowledge into existing models is that, for any given biological attribute, we seldom have complete data. Two recent ideas, "controllability" and "observability", potentially allow to use partial (local) knowledge to predict global state. Controllability refers to "driver" nodes that have a large influence on the state of a system [166]; observability is almost complementary, focusing on a small set of appropriately chosen observation nodes whose properties allow reconstruction of the global system state [167]. These techniques promise to allow associating local information with driver/observation nodes and to predict global properties from limited available data.

6 Conclusion and outlook

In this review, we have summarized common research themes in the field of network biology. We find a slow movement from global to local analysis, arguing that this trend emerged from a need to draw more concrete biological knowledge from networks.

The survey findings seem to suggest that one must either choose between untestable abstract hypotheses about large-scale topological patterns or small-scale results that neglect large-scale topology. But the successes of local techniques lie not in their focus on the local but because *they tightly couple topological observations to biological knowledge*. From this starting point, we see two broad research directions for improving the explanatory power of large-scale topology patterns. The first approach is theoretical and is aimed at making descriptive and suggestive techniques more predictive, whilst the second approach is practical and extends the predictive techniques to work at larger topological scales.

The theoretical research direction entails the improvement of network evolution models in order that they

reproduce as much of the topological aspects of real molecular networks as possible. Better models of network evolution can better reveal the topological features that are by-products of evolution, permitting researchers to concentrate on explaining topological results that cannot be explained by the models. An additional benefit is that these models could themselves lead to biological insight.

In the practical direction, we propose the application of predictive techniques to various "resolutions" of molecular networks, that is, multi-resolution analysis. Lower resolution versions of a network are typically obtained by grouping subnetworks into meta-nodes (by analogy, the entire street network of a city is represented by a single city node in national road maps). How nodes are grouped depends on the topological properties that must be maintained in low-resolution network versions. Node clustering techniques from Section 4 can be used to produce low-resolution networks by grouping node clusters into meta-nodes. Another promising technique that aims to maintain random-walk properties is **spectral coarse graining** [168].

The two research directions outlined above are by no means the only possible paths for developing network biology. Rather, they show this young field still has much potential for development; we expect that future researchers will bring us unexpected biological insights with the help of network biology.

7 Authors contributions

WW performed literature research and drafted the manuscript together with DR. PVM, HW and MR helped draft the manuscript. All authors read and approved the final manuscript.

8 Acknowledgments

References

1. Cusick M, Klitgard N, Vidal M, Hill D: **Interactome: Gateway into systems biology**. *Human Molecular Genetics* 2005, **14**:R171–181.
2. De Las Rivas J, Fontanillo C: **Protein–protein interactions essentials: Key concepts to building and analyzing interactome networks**. *PLoS Computional Biology* 2010, **6**(6):e1000807.
3. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman F, Cesareni G, Chatr-aryamontri A, Chautard E, Chen C, Dumousseau M, Goll J, Hancock R, Hannick L, Jurisica I, Khadake J, Lynn D, Mahadevan U, Perfetto L, Raghunath A, Ricard-Blum S, Roechert B, Salwinski L, Stumpflen V, Tyers M, Uetz P, Xenarios I, Hermjakob H: **Protein interaction data curation: The International Molecular Exchange (IMEx) consortium**. *Nature Methods* 2012, **9**:345–350.
4. Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, Rani V: **DNA-protein interactions: Methods for detection and analysis**. *Molecular and Cellular Biochemistry* 2012, **365**:279–299.
5. Portales-Casamar E, Thongjuea S, Kwon A, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman W, Sandelin A: **JASPAR 2010: The greatly expanded open-access database of transcription factor binding profiles**. *Nucleic Acids Research* 2010, **38**:D105–110.
6. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, Rando OJ, Birney E, Myers RM, Noble WS, Snyder M, Weng Z: **Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors**. *Genome Research* 2012, **22**(9):1798–1812.
7. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D: **BRENDA, the enzyme information system in 2011**. *Nucleic Acids Research* 2011, **39**:D670–676.
8. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and interpretation of large-scale molecular datasets**. *Nucleic Acids Research* 2012, **40**:D109–114.
9. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al.: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases**. *Nucleic Acids Research* 2012, **40**:D742–753.
10. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AHY, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C: **The genetic landscape of a cell**. *Science* 2010, **327**(5964):425–431.
11. Snitkin ES, Segrè D: **Epistatic interaction maps relative to multiple metabolic phenotypes**. *PLoS Genetics* 2011, **7**(2):e1001294.
12. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen L, von Mering C: **The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored**. *Nucleic Acids Research* 2011, **39**:D561–568.
13. Cho DY, Kim YA, Przytycka T: **Chapter 5: Network biology approach to complex diseases**. *PLoS Computional Biology* 2012, **8**(12):e1002820.
14. Rashevsky N: **Topology and life: In search of general mathematical principles in biology and sociology**. *Bulletin of Mathematical Biology* 1954, **16**(4):317–348.
15. Ideker T, Lauffenburger D: **Building with a scaffold: Emerging strategies for high- to low-level cellular modeling**. *Trends in Biotechnology* 2003, **21**(6):255–262.
16. Albert R: **Network inference, analysis, and modeling in systems biology**. *The Plant Cell* 2007, **19**(11):3327–3338.
17. Arita M: **The metabolic world of *Escherichia coli* is not small**. *Proceedings of the National Academy of Sciences of the USA* 2004, **101**(6):1543–1547.
18. Aittokallio T, Schwikowski B: **Graph-based methods for analysing networks in cell biology**. *Briefings in Bioinformatics* 2006, **7**(3):243–255.

19. Barabási AL, Albert R: **Emergence of scaling in random networks**. *Science* 1999, **286**(5439):509–512.
20. Barabási AL, Oltvai ZN: **Network biology: Understanding the cell's functional organization**. *Nature Reviews Genetics* 2004, **5**(2):101–113.
21. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: Simple building blocks of complex networks**. *Science* 2002, **298**(5594):824–827.
22. Alon U: **Network motifs: Theory and experimental approaches**. *Nature Reviews Genetics* 2007, **8**(6):450–461.
23. Takemoto K: **Metabolic network modularity arising from simple growth processes**. *Physical Review E* 2012, **86**:036107.
24. Chen J, Aronow B, Jegga A: **Disease candidate gene identification and prioritization using protein interaction networks**. *BMC Bioinformatics* 2009, **10**:73.
25. Schwikowski B, Uetz P, Fields S: **A network of protein-protein interactions in yeast**. *Nature Biotechnology* 2000, **18**(12):1257–1261.
26. Pržulj N: **Protein-protein interactions: Making sense of networks via graph-theoretic modeling**. *Bioessays* 2011, **33**(2):115–123.
27. Pavlopoulos G, Secrier M, Moschopoulos C, Soldatos T, Kossida S, Aerts J, Schneider R, Bagos P: **Using graph theory to analyze biological networks**. *BioData Mining* 2011, **4**:10.
28. Foster DV, Kauffman SA, Socolar JES: **Network growth models and genetic regulatory networks**. *Physical Review E* 2006, **73**:031912.
29. Sun M, Kim P: **Evolution of biological interaction networks: From models to real data**. *Genome Biology* 2011, **12**(12):235.
30. Barabási AL, Gulbahce N, Loscalzo J: **Network medicine: A network-based approach to human disease**. *Nature Reviews Genetics* 2011, **12**:56–68.
31. Ideker T, Krogan NJ: **Differential network biology**. *Molecular Systems Biology* 2012, **8**.
32. Gonzalez MW, Kann MG: **Chapter 4: Protein interactions and disease**. *PLoS Computational Biology* 2012, **8**(12):e1002819.
33. Aparicio O, Geisberg JV, Sekinger E, Yang A, Moqtaderi Z, Struhl K: *Chromatin Immunoprecipitation for Determining the Association of Proteins with Specific Genomic Sequences In Vivo*, John Wiley & Sons, Inc. 2005 [http://dx.doi.org/10.1002/0471142727.mb2103s69].
34. Bollobás B: *Random graphs*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2^{nd} edition 2001.
35. Watts DJ, Strogatz SH: **Collective dynamics of 'small-world' networks.** *Nature* 1998, **393**(6684):440–442.
36. Rzhetsky A, Gomez SM: **Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome**. *Bioinformatics* 2001, **17**(10):988–996.
37. Teichmann S, Babu M: **Gene regulatory network growth by duplication**. *Nature Genetics* 2004, **36**(5):492–496.
38. Light S, Kraulis P, Elofsson A: **Preferential attachment in the evolution of metabolic networks**. *BMC Genomics* 2005, **6**:159.
39. Maslov S, Sneppen K: **Specificity and stability in topology of protein networks**. *Science* 2002, **296**(5569):910–913.
40. Newman M: **Random graphs as models of networks**. In *Handbook of graphs and networks: From the genome to the internet*. Edited by Bornholdt S, Schuster HG, Wiley-VCH, Berlin 2003:35–68.
41. Catanzaro M, Boguñá M, Pastor-Satorras R: **Generation of uncorrelated random scale-free networks**. *Physical Review E* 2005, **71**:027103.
42. Albert R, Jeong H, Barabási AL: **Internet: Diameter of the World-Wide Web**. *Nature* 1999, **401**(6749):130–131.
43. Carter SL, Brechbühler CM, Griffin M, Bond AT: **Gene co-expression network topology provides a framework for molecular characterization of cellular state**. *Bioinformatics* 2004, **20**(14):2242–2250.

44. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P: **The protein–protein interaction map of *Helicobacter pylori***. *Nature* 2001, **409**(6817):211–215.

45. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL: **The large-scale organization of metabolic networks**. *Nature* 2000, **407**(6804):651–654.

46. Tsai CJ, Ma B, Nussinov R: **Protein-protein interaction networks: How can a hub protein bind so many different partners?** *Trends in Biochemical Sciences* 2009, **34**(12):594–600.

47. Lima-Mendez G, van Helden J: **The powerful law of the power law and other myths in network biology.** *Molecular Biosystems* 2009, **5**(12):1482–1493.

48. Samal A, Martin OC: **Randomizing genome-scale metabolic networks**. *PLoS ONE* 2011, **6**(7):e22295.

49. Basler G, Ebenhöh O, Selbig J, Nikoloski Z: **Mass balanced randomization of metabolic networks**. *Bioinformatics* 2011.

50. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology**. *Nature* 1999, **402**(6761 Suppl):C47–C52.

51. Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N: **Revealing modular organization in the yeast transcriptional network**. *Nature Genetics* 2002, **31**(4):370–377.

52. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM: **A protein interaction map of *Drosophila melanogaster***. *Science* 2003, **302**(5651):1727–1736.

53. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL: **Hierarchical organization of modularity in metabolic networks**. *Science* 2002, **297**(5586):1551–1555.

54. Rives AW, Galitski T: **Modular organization of cellular networks**. *Proceedings of the National Academy of Sciences of the USA* 2003, **100**(3):1128–1133.

55. Milenković T, Pržulj N: **Uncovering biological network function via graphlet degree signatures.** *Cancer Informatics* 2008, **6**:257–273.

56. Jeong H, Mason SP, Barabási AL, Oltvai ZN: **Lethality and centrality in protein networks.** *Nature* 2001, **411**(6833):41–42.

57. Kitano H: **Computational systems biology**. *Nature* 2002, **420**(6912):206–210.

58. Entus R, Aufderheide B, Sauro H: **Design and implementation of three incoherent feed-forward motif based biological concentration sensors**. *Systems and Synthetic Biology* 2007, **1**:119–128. [10.1007/s11693-007-9008-6].

59. Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U: **Superfamilies of evolved and designed networks**. *Science* 2004, **303**(5663):1538–1542.

60. Babu M, Luscombe NM, Aravind L, Gerstein M, Teichmann SA: **Structure and evolution of transcriptional regulatory networks.** *Current Opinion in Structural Biology* 2004, **14**(3):283–291.

61. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Statistical Applications in Genetics and Molecular Biology* 2005, **4**.

62. Fox Keller E: **Revisiting "scale-free" networks**. *BioEssays* 2005, **27**(10):1060–1068.

63. Khanin R, Wit E: **How scale-free are biological networks.** *Journal of Computational Biology* 2006, **13**(3):810–818.

64. de Lomana ALG, Beg QK, de Fabritiis G, Villà-Freixa J: **Statistical analysis of global connectivity and activity distributions in cellular networks**. *Journal of Computational Biology* 2010.

65. Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hurst LD, Tyers M: **Still stratus not altocumulus: Further evidence against the date/party hub distinction**. *PLoS Biology* 2007, **5**(6):e154+.

66. Ingram PJ, Stumpf MP, Stark J: **Network motifs: Structure does not determine function**. *BMC Genomics* 2006, **7**:108+.

67. Kuo PD, Banzhaf W, Leier A: **Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence**. *Biosystems* 2006, **85**(3):177–200.

68. Berg J, Lässig M: **Local graph alignment and motif search in biological networks.** *Proceedings of the National Academy of Sciences of the USA* 2004, **101**(41):14689–14694.

69. Waddington CH: **Canalization of development and the inheritance of acquired characters**. *Nature* 1942, **150**:563–565.

70. Smart AG, Amaral LAN, Ottino JM: **Cascading failure and robustness in metabolic networks**. *Proceedings of the National Academy of Sciences of the USA* 2008, **105**(36):13223–13228.

71. Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H: **Structure of protein interaction networks and their implications on drug design**. *PLoS Computational Biology* 2009, **5**(10):e1000550.

72. Bergman A, Siegal ML: **Evolutionary capacitance as a general feature of complex gene networks**. *Nature* 2003, **424**(6948):549–552.

73. Maslov S, Krishna S, Pang TY, Sneppen K: **Toolbox model of evolution of prokaryotic metabolic networks and their regulation**. *Proceedings of the National Academy of Sciences of the USA* 2009, **106**(24):9743–9748.

74. Pang TY, Maslov S: **A toolbox model of evolution of metabolic pathways on networks of arbitrary topology**. *PLoS Computational Biology* 2011, **7**(5):e1001137.

75. Lee SH, Bernhardsson S, Holme P, Kim BJ, Minnhagen P: **Neutral theory of chemical reaction networks**. *New Journal of Physics* 2012, **14**(3):033032.

76. Kreimer A, Borenstein E, Gophna U, Ruppin E: **The evolution of modularity in bacterial metabolic networks**. *Proceedings of the National Academy of Sciences of the USA* 2008, **105**(19):6976–6981.

77. Zhou W, Nakhleh L: **Convergent evolution of modularity in metabolic networks through different community structures**. *BMC Evolutionary Biology* 2012, **12**:181.

78. Takemoto K: **Does habitat variability really promote metabolic network modularity?** *PLoS ONE* 2013, **8**(4):e61348.

79. Ciliberti S, Martin OC, Wagner A: **Robustness can evolve gradually in complex regulatory gene networks with varying topology**. *PLoS Computional Biology* 2007, **3**(2):e15+.

80. Pfeiffer T, Soyer OS, Bonhoeffer S: **The evolution of connectivity in metabolic networks**. *PLoS Biology* 2005, **3**(7):e228.

81. Leclerc RD: **Survival of the sparsest: Robust gene networks are parsimonious**. *Molecular Systems Biology* 2008, **4**.

82. Kashtan N, Alon U: **Spontaneous evolution of modularity and network motifs**. *Proceedings of the National Academy of Sciences of the USA* 2005, **102**(39):13773–13778.

83. Samal A, Wagner A, Martin O: **Environmental versatility promotes modularity in genome-scale metabolic networks**. *BMC Systems Biology* 2011, **5**:135.

84. Sharan R, Ulitsky I, Shamir R: **Network-based prediction of protein function.** *Molecular Systems Biology* 2007, **3**.

85. Cui Q, Yu Z, Purisima EO, Wang E: **Principles of microRNA regulation of a human cellular signaling network**. *Molecular Systems Biology* 2006, **2**.

86. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, Purisima EO, Wang E: **A map of human cancer signaling**. *Molecular Systems Biology* 2007, **3**.

87. Kittisopikul M, Süel GM: **Biological role of noise encoded in a genetic network motif**. *Proceedings of the National Academy of Sciences of the USA* 2010, **107**(30):13300–13305.

88. Ben-Dor A, Shamir R, Yakhini Z: **Clustering gene expression patterns**. *Journal of Computational Biology* 1999, **6**(3-4):281–297.

89. Snel B, Bork P, Huynen MA: **The identification of functional modules from the genomic association of genes**. *Proceedings of the National Academy of Sciences of the USA* 2002, **99**(9):5890–5895.

90. Fortunato S: **Community detection in graphs**. *Physics Reports* 2010, **486**(3-5):75 – 174.

91. Blatt M, Wiseman S, Domany E: **Superparamagnetic Clustering of Data**. *Physical Review Letters* 1996, **76**:3251–3254.

92. Son SW, Jeong H, Noh JD: **Random field Ising model and community structure in complex networks**. *The European Physical Journal B - Condensed Matter and Complex Systems* 2006, **50**(3):431–437.

93. van Dongen S: **Graph clustering by flow simulation**. *PhD thesis*, University of Utrecht 2000.

94. E W, Li T, Vanden-Eijnden E: **Optimal partition and effective dynamics of complex networks**. *Proceedings of the National Academy of Sciences of the USA* 2008, **105**(23):7907–7912.

95. Pluchino A, Latora V, Rapisarda A: **Changing opinions in a changing world: A new perspective in sociophysics**. *International Journal of Modern Physics C* 2005, **16**(04):515–531.

96. Bader G, Hogue C: **An automated method for finding molecular complexes in large protein interaction networks**. *BMC Bioinformatics* 2003, **4**:2+.

97. King AD, Pržulj N, Jurisica I: **Protein complex prediction via cost-based clustering**. *Bioinformatics* 2004, **20**(17):3013–3020.

98. Hu H, Yan X, Huang Y, Han J, Zhou XJ: **Mining coherent dense subgraphs across massive biological networks for functional discovery**. *Bioinformatics* 2005, **21**(suppl 1):i213–i221.

99. Ideker T, Ozier O, Schwikowski B, Siegel AF: **Discovering regulatory and signalling circuits in molecular interaction networks**. *Bioinformatics* 2002, **18**(suppl 1):S233–S240.

100. Ulitsky I, Shamir R: **Identifying functional modules using expression profiles and confidence-scored protein interactions**. *Bioinformatics* 2009, **25**(9):1158–1164.

101. Thomas A, Cannings R, Monk N, Cannings C: **On the structure of protein–protein interaction networks**. *Biochemical Society Transactions* 2003, **31**:1491–1496.

102. Bu D, Zhao Y, Cai L, Xue H, Zhu X, Lu H, Zhang J, Sun S, Ling L, Zhang N, Li G, Chen R: **Topological structure analysis of the protein–protein interaction network in budding yeast**. *Nucleic Acids Research* 2003, **31**(9):2443–2450.

103. Liu H, Liu J, Wang L: **Searching maximum quasi-bicliques from protein-protein interaction network**. *Journal of Biomedical Science and Engineering* 2008, **1**(3):200–203.

104. Tanay A, Sharan R, Kupiec M, Shamir R: **Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data**. *Proceedings of the National Academy of Sciences of the USA* 2004, **101**(9):2981–2986.

105. Solava RW, Michaels RP, Milenković T: **Graphlet-based edge clustering reveals pathogen-interacting proteins**. *Bioinformatics* 2012, **28**(18):i480–i486.

106. Ahn YY, Bagrow JP, Lehmann S: **Link communities reveal multiscale complexity in networks**. *Nature* 2010, **466**(7307):761–764.

107. Becker E, Robisson B, Chapple CE, Guénoche A, Brun C: **Multifunctional proteins revealed by overlapping clustering in protein interaction network**. *Bioinformatics* 2012, **28**:84–90.

108. Fortunato S, Barthélemy M: **Resolution limit in community detection**. *Proceedings of the National Academy of Sciences of the USA* 2007, **104**:36–41.

109. Freeman LC: **Centrality in social networks conceptual clarification**. *Social Networks* 1978-1979, **1**(3):215–239.

110. Vogelstein B, Lane D, Levine AJ: **Surfing the p53 network**. *Nature* 2000, **408**(6810):307–310.

111. Yu H, Greenbaum D, Lu HX, Zhu X, Gerstein M: **Genomic analysis of essentiality within protein networks**. *Trends in Genetics* 2004, **20**(6):227 – 231.

112. Bergmann S, Ihmels J, Barkai N: **Similarities and Differences in Genome-Wide Expression Data of Six Organisms**. *PLoS Biology* 2003, **2**:e9.

113. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Ménard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C: **Global mapping of the yeast genetic interaction network**. *Science* 2004, **303**(5659):808–813.

114. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW: **Evolutionary rate in the protein interaction network**. *Science* 2002, **296**(5568):750–752.

115. Vitkup D, Kharchenko P, Wagner A: **Influence of metabolic network structure and function on enzyme evolution**. *Genome Biology* 2006, **7**(5):R39.

116. Wang Y, Franzosa EA, Zhang XS, Xia Y: **Protein evolution in yeast transcription factor subnetworks**. *Nucleic Acids Research* 2010, **38**(18):5959–5969.

117. Ma HW, Zeng AP: **The connectivity structure, giant strong component and centrality of metabolic networks**. *Bioinformatics* 2003, **19**(11):1423–1430.

118. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M: **The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics.** *PLoS Computational Biology* 2007, **3**(4):e59+.

119. Li L, Tibiche C, Fu C, Kaneko T, Moran MF, Schiller MR, Li SSC, Wang E: **The human phosphotyrosine signaling network: Evolution and hotspots of hijacking in cancer**. *Genome Research* 2012, **22**(7):1222–1230.

120. Wang E: **Understanding genomic alterations in cancer genomes using an integrative network approach**. *Cancer letters* 2013, [http://linkinghub.elsevier.com/retrieve/pii/S0304383512007124?showall=true].

121. Estrada E, Rodríguez Velázquez JA: **Subgraph centrality in complex networks**. *Physical Review E* 2005, **71**(5):056103+.

122. Wuchty S, Almaas E: **Peeling the yeast protein network**. *Proteomics* 2005, **5**(2):444–449.

123. Estrada E: **Protein bipartivity and essentiality in the yeast protein-protein interaction network.** *Journal of Proteome Research* 2006, **5**(9):2177–2184.

124. Bhardwaj N, Kim PM, Gerstein MB: **Rewiring of transcriptional regulatory networks: Hierarchy, rather than connectivity, better reflects the importance of regulators**. *Science Signaling* 2010, **3**(146):ra79.

125. He X, Zhang J: **Why do hubs tend to be essential in protein networks?** *PLoS Genetics* 2006, **2**(6):e88.

126. Siegal M, Promislow D, Bergman A: **Functional and evolutionary inference in gene networks: Does topology matter?** *Genetica* 2007, **129**:83–103.

127. Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M: **Evidence for dynamically organized modularity in the yeast protein–protein interaction network**. *Nature* 2004, **430**(6995):88–93.

128. Agarwal S, Deane CM, Porter MA, Jones NS: **Revisiting date and party hubs: Novel approaches to role assignment in protein interaction networks**. *PLoS Computional Biology* 2010, **6**(6):e1000817+.

129. Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA: **Molecular signaling network complexity is correlated with cancer patient survivability**. *Proceedings of the National Academy of Sciences of the USA* 2012, **109**(23):9209–9212.

130. Park K, Kim D: **Localized network centrality and essentiality in the yeast–protein interaction network**. *Proteomics* 2009, **9**(22):5143–5154.

131. Holme P, Huss M, Jeong H: **Subnetwork hierarchies of biochemical pathways**. *Bioinformatics* 2003, **19**(4):532–538.

132. Eisenberg E: **Preferential attachment in the protein network evolution**. *Physical Review Letters* 2003, **91**(13).

133. Dandekar T, Schuster S, Snel B, Huynen M, Bork P: **Pathway alignment: Application to the comparative analysis of glycolytic enzymes.** *The Biochemical journal* 1999, **343 Pt 1**.

134. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G: **Gene ontology: Tool for the unification of biology**. *Nature Genetics* 2000, **25**:25–29.

135. Oliver S: **Proteomics: Guilt-by-association goes global**. *Nature* 2000, **403**(6770):601–603.

136. Gillis J, Pavlidis P: **"Guilt by Association" is the exception rather than the rule in gene networks**. *PLoS Computational Biology* 2012, **8**(3):e1002444.

137. Kourmpetis YAI, van Dijk ADJ, Bink MCAM, van Ham RCHJ, ter Braak CJF: **Bayesian Markov random field analysis for protein function prediction based on network data**. *PLoS ONE* 2010, **5**(2):e9293.

138. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M: **Towards a proteome-scale map of the human protein-protein interaction network**. *Nature* 2005, **437**(7062):1173–1178.

139. Qi Y, Suhail Y, Lin Yy, Boeke JD, Bader JS: **Finding friends and enemies in an enemies-only network: A graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions**. *Genome Research* 2008, **18**(12):1991–2004.

140. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T: **PathBLAST: A tool for alignment of protein interaction networks**. *Nucleic Acids Research* 2004, **32**(suppl 2):W83–W88.

141. Phan HTT, Sternberg MJE: **PINALOG: A novel approach to align protein interaction networks—implications for complex detection and function prediction**. *Bioinformatics* 2012, **28**(9):1239–1245.

142. Shih YK, Parthasarathy S: **Scalable global alignment for multiple biological networks**. *BMC Bioinformatics* 2012, **13**(Suppl 3):S11.

143. Kuchaiev O, Milenković T, Memišević V, Hayes W, Pržulj N: **Topological network alignment uncovers biological function and phylogeny**. *Journal of The Royal Society Interface* 2010, **7**(50):1341–1354.

144. Patro R, Kingsford C: **Global network alignment using multiscale spectral signatures**. *Bioinformatics* 2012, **28**(23):3105–3114.

145. Pandey G, Manocha S, Atluri G, Kumar V: **Enhancing the functional content of protein interaction networks**. *arXiv preprint arXiv:1210.6912* 2012.

146. Sardiu ME, Cai Y, Jin J, Swanson SK, Conaway RC, Conaway JW, Florens L, Washburn MP: **Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics**. *Proceedings of the National Academy of Sciences of the USA* 2008, **105**(5):1454–1459.

147. Pandey G, Steinbach M, Gupta R, Garg T, Kumar V: **Association analysis-based transformations for protein interaction networks: A function prediction case study**. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, New York, NY, USA: ACM 2007:540–549.

148. Milenković T, Memišević V, Ganesan AK, Pržulj N: **Systems-level cancer gene identification from protein interaction network topology applied to melanogenesis-related functional genomics data**. *Journal of The Royal Society Interface* 2010, **7**(44):423–437.

149. Bandyopadhyay S, Sharan R, Ideker T: **Systematic identification of functional orthologs based on protein network comparison**. *Genome Research* 2006, **16**(3):428–435.

150. Li Y, de Ridder D, de Groot M, Reinders M: **Metabolic pathway alignment between species using a comprehensive and flexible similarity measure**. *BMC Systems Biology* 2008, **2**:111.

151. Ideker T, Sharan R: **Protein networks in disease**. *Genome Research* 2008, **18**(4):644–652.

152. Wachi S, Yoneda K, Wu R: **Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues**. *Bioinformatics* 2005, **21**(23):4205–4208.

153. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL: **Dynamic modularity in protein interaction networks predicts breast cancer outcome**. *Nature Biotechnology* 2009, **27**(2):199–204.

154. Tuck D, Kluger H, Kluger Y: **Characterizing disease states from topological properties of transcriptional regulatory networks**. *BMC Bioinformatics* 2006, **7**:236.

155. Yao C, Li H, Zhou C, Zhang L, Zou J, Guo Z: **Multi-level reproducibility of signature hubs in human interactome for breast cancer metastasis**. *BMC Systems Biology* 2010, **4**:151.

156. Li J, Lenferink AEG, Deng Y, Collins C, Cui Q, Purisima EO, O'Connor-McCourt MD, Wang E: **Identification of high-quality cancer prognostic markers and metastasis network modules**. *Nature Communications* 2010, **1**:34.

157. Chuang HYY, Lee E, Liu YTT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Molecular Systems Biology* 2007, **3**.

158. Lin CC, Hsiang JT, Wu CY, Oyang YJ, Juan HF, Huang HC: **Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy**. *BMC Systems Biology* 2010, **4**:138.

159. Efroni S, Schaefer CF, Buetow KH: **Identification of key processes underlying cancer phenotypes using biologic pathway analysis**. *PLoS ONE* 2007, **2**(5):e425.

160. Cho DY, Kim YA, Przytycka TM: **Chapter 5: Network biology approach to complex diseases**. *PLoS Computational Biology* 2012, **8**(12):e1002820.

161. Ibarra RU, Edwards JS, Palsson BO: ***Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth**. *Nature* 2002, **420**(6912):186–189.

162. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, Konig R, Smith RD, Palsson BO: **Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models**. *Molecular Systems Biology* 2010, **6**.

163. Hoppe A, Hoffmann S, Holzhutter HG: **Including metabolite concentrations into flux balance analysis: Thermodynamic realizability as a constraint on flux distributions in metabolic networks**. *BMC Systems Biology* 2007, **1**:23.

164. Covert MW, Xiao N, Chen TJ, Karr JR: **Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli***. *Bioinformatics* 2008, **24**(18):2044–2050.

165. Lewis NE, Nagarajan H, Palsson BO: **Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods**. *Nature Reviews Microbiology* 2012, **10**(4):291–305.

166. Liu YY, Slotine JJ, Barabasi AL: **Controllability of complex networks**. *Nature* 2011, **473**(7346):167–173.

167. Liu YY, Slotine JJ, Barabási AL: **Observability of complex systems**. *Proceedings of the National Academy of Sciences of the USA* 2013, **110**(7):2460–2465.

168. Gfeller D, De Los Rios P: **Spectral coarse graining of complex networks**. *Physical Review Letters* 2007, **99**:038701.