Structure and Function in Complex Biological Networks

Joel Hancock, and Jörg Menche, CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

Introduction

The behavior of complex systems that consist of large numbers of interacting components is most often not a direct consequence of any one component or interaction, but arises from the overall architecture of the system. Many biological systems are characterized by such emergent behaviors. Perhaps the most striking example is the brain, whose staggering complexity of behavior arises from the architecture of connections between comparatively simple neurons (**Turkheimer et al., 2019**). Network theory provides tools and concepts for a holistic understanding of such systems. Over the last two decades, numerous methods have been developed to describe and quantify the overall structure of a network, to get insights into the processes that generated a network, and to understand dynamical processes that take place on them.

Parallel to these theoretical advances, rapid technological progress has propelled biology into the genomic era. The mapping of the genome has produced a near complete inventory of the components involved in cellular systems (**ENCODE Project Consortium, 2012**). High-throughput technologies like the yeast two-hybrid method (**Fields and Song, 1989**) and mass spectroscopy (**Gingras et al., 2007**) provided us also with an increasingly detailed map of the interactions between the individual components, in particular proteins (**Huttlin et al., 2017**; **Rolland et al., 2014**; **Venkatesan et al., 2009**). These experimentally determined *protein-protein interaction* (*PPI*) networks can be further complemented by computational predictions (**Jansen et al., 2003**), perturbation studies (**Caldera et al., 2019**; **Ideker et al., 2001**) or comparisons with model organisms (**Zhong et al., 2016**). In view of the resulting, highly interconnected molecular networks, it is clear that genes rarely act in isolation, but exert their cellular functions through their influence on the system as a whole. Understanding a particular gene's function can thus be nearly as challenging as understanding the entire system. Network approaches are ideally suited to disentangle this enormous complexity and we will use PPI networks throughout this review to illustrate biological applications of the introduced network concepts.

Indeed, given that PPIs form the molecular basis of most biological processes, it is little surprising that investigations of their collective network properties have considerably contributed to our understanding of a wide range of important biomedical challenges. For example, poorly-studied genes can be functionally annotated by examining their interaction partners (**Stelzl et al., 2005**). Similarly, genetic mutations can be assessed by their impact on the molecular network (**Caldera et al., 2017**). These findings can be generalized to investigate the impact of pathogens on the host PPI network (**Khamina et al., 2017**), or to reveal the molecular basis for comorbidity patterns between different diseases (**Menche et al., 2015**), to name but a few of the numerous applications of the emerging area of network medicine (**Loscalzo, 2017**).

Formally, a *network* is a collection of objects called *nodes*, which represent the elements of a system, and objects called *edges*, which represent pairwise relationships between the elements, for example similarity or physical interactions. Networks provide an abstract description of the system, often setting aside much of the information at hand. When considering PPI networks, for example, we ignore anything we might know about the structure of the proteins, or details of how, when or where in the body the interactions take place. Despite these simplifications, or in part perhaps because of them, networks provide a powerful platform for investigating complex systems and generating meaningful insights into their behavior at different scales, from the level of individual components to the level of the entire system. A key finding of network science in general, and network biology/medicine in particular, is that *structural* properties of a particular network can often be related to important *functional* characteristics of the system that they represent. This review aims to give a first overview of important structural characteristics and highlights their application in biological networks. We introduce these characteristics in order of scale, beginning with those that describe individual nodes, continuing to their broader neighborhood within the network, and finally introducing global network properties (**Fig. 1**).

Local Network Structures

We start by inspecting the individual nodes in a given network and their immediate *neighborhood*, i.e., the set of directly connected other nodes. The relevance of these immediate interactions is often obvious, for example in PPI networks, where they represent the very basis

https://doi.org/10.1016/B978-0-12-801238-3.11639-3

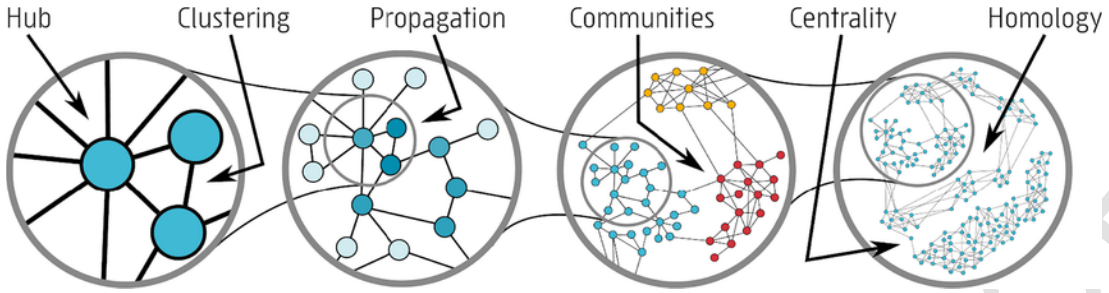

Fig. 1 Networks can be characterized at different scales. From left to right: We start by properties of individual nodes, such as their number of neighbors (defining hubs as highly connected nodes) and small groups of nodes (defining clustering as a tendency to form triangles). The extended neighborhood of a node can be identified using propagation methods. Larger, highly connected sub-networks are referred to as communities. At the level of the entire network, properties like centrality can be defined and structural characteristics can be studied using the concept of homology.

of biological function and whose interruption is often related to disease (**Gonzalez and Kann, 2012**). The total number of directly connected neighbors of a given node is referred to as the node's *degree*. Nodes deemed to be of particularly high degree are referred to as *hubs*. Hub nodes are frequently of particular interest, since the number of interactions that a node takes part in is a crude but generally effective guide to the influence over the rest of the system that the element represented by the node exerts. In the PPI network, for example, it was shown that cancer-genes (**Jonsson and Bates, 2006**) and essential genes (**Sun and Zhao, 2010**) tend to be hubs.

Considering the network in its entirety, we can then define the *degree distribution*, giving the number of nodes in the network with a particular degree *k*. The degree distribution of a network heavily influences many other structural properties and thus needs to be considered when evaluating the significance of any structural property measured on a particular network. Moreover, given the interpretation of degree as a measure of influence of the individual elements on a system, the degree distribution is key to understanding how information propagates through the network (**Teschendorff et al., 2015**), as well as its robustness under random failure or targeted attack of specific nodes (**Albert et al., 2000**).

It is instructive to compare the measured degree distribution to standard probability distributions. For example, we may find that the degree distribution of a network is well described by the familiar binomial distribution. By far the most studied degree distribution is the *power-law distribution*, due to its frequent appearance in real world systems (**Barabasi and Albert, 1999**). Networks with a power-law degree distribution are also called *scale-free networks*. By definition, a network whose degrees follow a power-law distribution will have number of nodes of degree *k* proportional to $k^{-\alpha}$, where α is a network specific constant (**Fig. 2**).

A power-law degree distribution implies that there is a non-negligible fraction of the nodes with dramatically higher degrees than a typical node (**Fig. 2**A). The parameter α describes the extent of this imbalance, with a larger value of α in the range 3 and above indicating a comparatively balanced distribution of the degrees, while a small value α only slightly larger than 1 (the theoretical minimum) indicates that the network is entirely dominated by the nodes of very high degrees.

In practice, the degree distribution rarely follows an exact power-law. Instead, it is commonly observed that the power-law distribution only holds for nodes of degrees above a certain threshold k_{min} (**Fig. 2**B). In order to get an intuition as to whether a network's empirical degree distribution might be described by a power-law, the cumulative histogram of the nodes' degrees should be plotted on doubly logarithmic axes (**Fig. 2**C). If the original distribution follows a power-law of degree α, then the histogram should follow a line with slope α–1, for values above k_{min}.

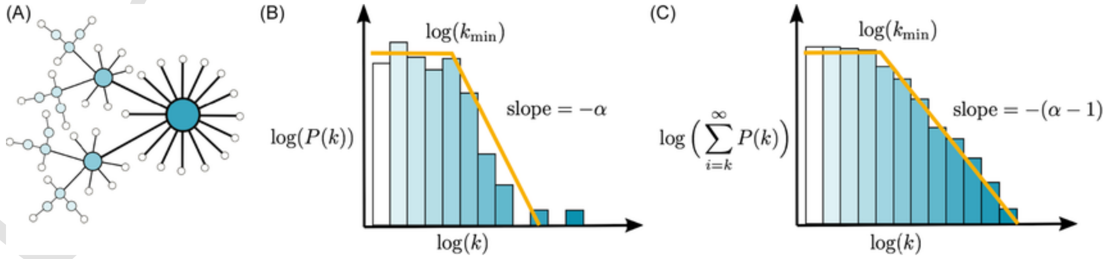

Fig. 2 The degree distribution of a network. (A) Small network illustrating that nodes may have very different numbers of neighbors, or degree *k*. (B) An important characteristic of a network is the histogram of the degrees across all nodes. Many real-world networks exhibit degree distributions $P(k)$ that follow a power-law, i.e., $P(k) \sim k^{-\alpha}$ for values above a certain threshold k_{min}. Plotted on a log-log scale, this leads to a straight line with slope $-\alpha$. (C) A more accurate visual assessment of a potential power-law can be achieved by plotting the cumulative distribution histogram instead of the probability distribution, as it smooths out the curve, especially at higher values of *k*.

The maximum likelihood value of α for a network of size n with apparent power-law distribution with lower threshold $k_{\min}$ can be determined from

$$1 + n\left(\sum_i \ln \frac{k_i}{k_{\min} - \frac{1}{2}}\right)^{-1}.$$

There is no equivalent simple formula for estimating $k_{\min}$, so in practice different values should be tried out to determine the combination of $k_{\min}$ and α that best approximates a true power-law, for example by examining the Kolmogorov–Smirnov distance. It has been reported that these parameters can be fairly accurately determined for a network with as few as 50 nodes (**Clauset et al., 2009**).

Power-laws are not the only distributions that can describe the prevalence of few highly connected hubs that characterizes many real world networks. In **Stumpf and Ingram (2005)**, the authors compared the degree distributions of the PPI networks of six model organisms to three distributions with heavy tails (power-law, log-normal, stretched exponential) and three distributions with exponential decay (Poisson, exponential, and gamma distribution). For the *C. elegans* and *D. melanogaster* PPI networks—the most complete available at the time—the power-law distribution provided the best fit and also for other species with less complete PPI networks, the heavy-tailed nature of the degree distribution was still apparent. Similar results were obtained in a systematic investigation of the degree distributions of networks of metabolic reactions mapped out for 43 organisms from across the tree of life (**Jeong et al., 2000**).

The simplest network characteristic that moves beyond pairs of connected nodes is the so called *clustering*, which quantifies the tendency of three nodes to form a connected triangle. It was first introduced in the context of social networks to describe the phenomenon that two friends of a particular individual are often also friends with each other (**Holland and Leinhardt, 1971**). The clustering coefficient of a node is thus given by the number of its neighbors that are connected to one another, divided by the number of all possible pairs of neighbors that could be connected (**Newman, 2003**). In PPI networks, high clustering coefficients can aid in the identification of protein complexes (**Zaki et al., 2013**).

More complex patterns of connectivity among a small number of nodes can be studied using so-called network *motifs* (**Milo et al., 2002**). They were first introduced to identify recurrent patterns in *S. cerevisiae* and *E. coli* gene regulatory networks, by comparing how often a particular small subgraph was actually observed to the respective frequency among randomly rewired control networks. This allowed for the identification of several motifs that can be interpreted as logical operations on the state of the nodes in analogy to an electrical circuit, such as the “feed-forward loop” or the “bi-fan” (**Alon, 2007**). The feed-forward loop can serve as both a filter for reducing intermittent noisy signaling, or a response accelerator, depending on whether the edge connecting the initial and final nodes is inhibitory or excitatory. The occurrence of these and other motifs across various biological networks of different sizes and for different species suggests that they may play an important role in the evolution of biological networks (**Conant and Wagner, 2003**).

Expanded Network Neighborhoods

Perhaps the most fundamental benefit that the network viewpoint provides is that it allows us to generalize the context of elements beyond their immediate neighborhood of direct connections and also take into account indirect influences from their broader context within the system. In biological networks, this expanded network neighborhood often contains functionally related nodes, for example gene products that jointly perform a certain task, or genes associated with the same disease (**Barabási et al., 2011**).

There are numerous methods for identifying the relevant expanded neighborhood around a given node or set of nodes. We can broadly categorize these methods into connectivity-based methods and dynamic methods. Connectivity-based methods prioritize the broader neighborhood of a given node (or node set) by evaluating connectivity patterns such as edge density (**Erten et al., 2011**) or significance (**Ghiassian et al., 2015**). A prominent biological application is to use PPI networks to identify genes that may be implicated in a specific disease. The DIAMOnD method (**Ghiassian et al., 2015**), for example, starts from a given set of disease associated seed genes and iteratively expands it by appending the gene whose neighbors form the most statistically significant overlap with the seed set. The order in which the previously unannotated genes are added can be interpreted as a rank according to their proximity to the seed gene set.

A prominent class of dynamic methods for the identification of the expanded neighborhood is based on *random walks* as a method of propagation. These methods are conceptualized as observing the pattern of movement of a set of “walkers” that begin at a particular seed node (or set of seed nodes), and propagate through the network by moving from node to node along a randomly selected edge at each time-step. Intuitively, those nodes that are closer to the starting nodes of the walkers will be visited earlier and more frequently. Another frequently employed metaphor is the spread of heat, electrical potential or fluid when initially deposited on the starting nodes—the mathematical treatment is the same in all cases. Note that given enough time, the random walkers will distribute themselves evenly throughout the entire network, so that the probability of visiting a given node will be simply proportional to its degree, unaffected by the starting point of the walker. To prevent this, the walkers are typically given some fixed probability of instantly returning to the seed set at any given time point. This is referred to as a *random walk with restart*, and its effect is to balance the impact of degree on the visitation probability with the desired influence of proximity to the seed set.

Random walks represent a powerful tool with many biological applications, see **Cowen et al. (2017)** for a comprehensive review. Here, we will briefly showcase some examples highlighting their versatile usage in the context of disease gene prioritization: In **Kim et al. (2011)** a method was outlined for identifying causal mutations in glioblastoma multiforme using a random walk on a PPI network whose edges are weighted based on co-expression data. A random walk with restart beginning at each of the candidate mutations was used to evaluate a mutation’s potential effect to each of the differentially expressed genes. This method, along with some further network-based filtering steps, allowed for a more significant portion of known glioblastoma multiforme causal mutations to be recovered than simply selecting those associated with expression changes across the cancer cohort. The HotNet2 algorithm introduced in **Leiserson et al. (2015)** is also based on a PPI network and employs a random walk with restart from each node to build a directed weighted network. From this secondary network, densely interconnected node sets are extracted, which are then shown to correspond to pathways that are enriched with mutated genes for individual cancers. Yet another network type was considered in **Wang et al. (2014)** to aggregate different sources of tumor data for five cohorts of cancer patients provided by the TCGA (**Tomczak et al., 2015**). For each cohort, three patient-patient similarity networks were constructed, based on DNA methylation, mRNA expression, and miRNA expression, respectively. The authors then combined pairs of networks by repeatedly taking the similarity scores of one network and applying a random walk using the similarity scores of one of the other networks. This was then averaged across pairs to obtain a consensus similarity network. The prediction of survival based on communities (see below) in this network was far superior than predictions based on clusters identified in networks constructed from the individual data sources.

Network Properties at the "Mesoscale"

We turn now to a yet higher level of organization in networks – the partition of a network into sizable chunks, called *communities* or *modules*. The general idea behind these only loosely defined terms is that each partition should contain a non-negligible portion of the nodes in the network, and that nodes within one partition should be more strongly connected to each other than to the rest of the network. Community structure has been shown to have functional relevance in networks across many domains of science, including biology (**Girvan and Newman, 2002**). This type of *mesoscale* structure is one of the most intuitive ways of describing network architecture. The intuitive appeal and obvious importance of this idea, however, does not translate into one universally accepted method of describing it mathematically or algorithmically detecting communities (**Newman, 2012**). Not only are there numerous competing notions about what exactly constitutes a community of nodes, but the different definitions can each have multiple methods of approximating those communities which may yield varying results (**Fortunato and Hric, 2016**).

A first mathematical formulation for the concept of network communities was introduced in **Newman (2012)** in terms of network *modularity*. The modularity M_S quantifies the number of connections between the members of a set of nodes S, relative to the expected number if the respective nodes were connected completely at random to the rest of the network. It is defined as

$$M_S = \sum_{i,j \in S} \left(a_{ij} - \frac{k_i k_j}{2m} \right), M_S = \sum_{i,j \in S} \left(a_{ij} - \frac{k_i k_j}{2m} \right),$$

where a_{ij} denotes the adjacency matrix of the network, k_i and k_j the degrees of nodes i and j, respectively, and m the total number of links in the network. The identification of communities within a network can then be rephrased as the problem of finding a *maximum modularity* partition, that is, a partition of the network where the modularity for each of the putative communities has the highest sum to make the highest possible total. This method for finding communities is fast, intuitive, and naturally incorporates the heterogenous degree distributions observed in many real world networks. On the downside, naive modularity optimization has a strong bias towards identifying communities of a specific size, specifically the square root of the size of the total network (**Newman, 2012**).

An efficient heuristic for identifying a network partition with high modularity is given by the Louvain algorithm (**Blondel et al., 2008**), named from the authors' institution at the time of its discovery. The Louvain algorithm was shown to be effective at recovering community structure from toy models and ranges among the computationally fastest algorithms (**Yang et al., 2016**). Moreover, in addition to returning modular network partitions, it provides a hierarchical clustering of communities, from which the user may select the scale which is of most interest (**Fig. 3**). This circumvents the aforementioned bias of modularity maximization towards specific community sizes, and also accommodates the fact the scale of the communities of interest may vary depending on the problem at hand. The algorithm picks one node at a time and evaluates whether merging it into a community with any of its neighbors improves the modularity score. When no further merges can improve the score, a new aggregated network is constructed, in which nodes represent the clusters identified in the last iteration. This procedure is repeated, thus producing a hierarchy of clusters, until a final network emerges, typically consisting of very few nodes, whose modularity score cannot be improved anymore. This final partition was shown to produce good results at identifying ground-truth communities in synthetic benchmark networks, but has a tendency to underestimate the number of clusters when the clusters are not well separated and contain more than about a thousand nodes (**Yang et al., 2016**).

Many real world networks are characterized by a hierarchical organization of functional modules, including biological networks, such as the metabolic networks of multiple organisms (**Ravasz et al., 2002**). This has spurred the development of numerous methods for identifying such structures. For example, **Ahn et al. (2010)** introduced a method that iteratively agglomerates edges, rather than nodes and identified functional modules and their hierarchical organization in PPI and metabolic networks. In **Lewis et al. (2010)**, the authors used a modularity maximization technique to dissect the functional communities in the PPI network at multiple scales. An entirely different approach was introduced in **Reichardt and Bornholdt (2006)** and is based on spin-glasses, a model for certain types of magnets from the area of condensed matter physics. The authors showed that a free parameter in the model can be used to tune the size of the communities that the user would like to detect and identified a spectrum of functional communities of varying size.

Community detection is a powerful tool to uncover parts of the network that share a common biological function, and in some cases, also genes associated with the same disease form densely connected communities (**Oti et al., 2006**). More generally, however, disease associated genes are only sparsely connected among each other and thus rarely enriched in large communities. It was found in **Ghiassian et al. (2015)** that only 15% of all diseases were enriched in one of the communities detected on the PPI network even when using the

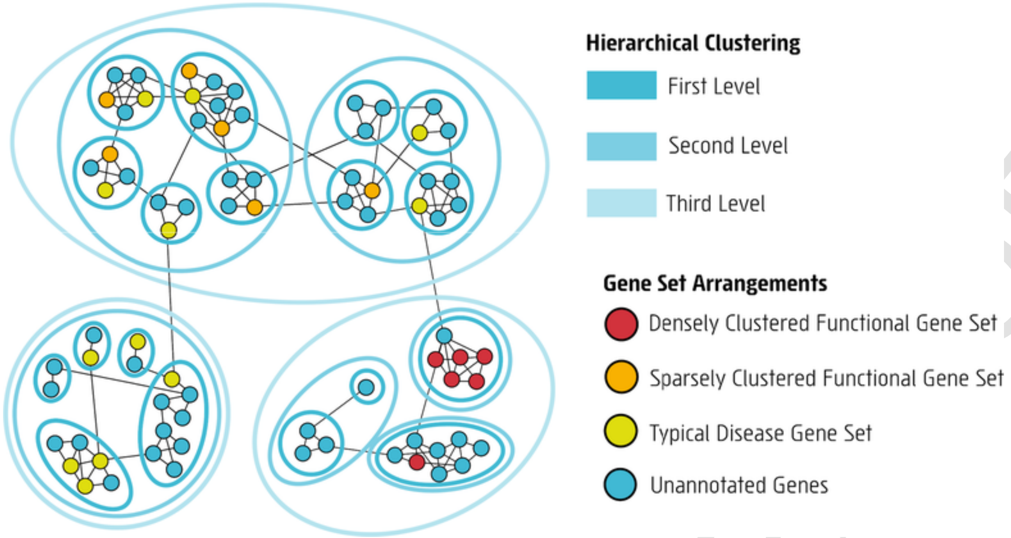

Fig. 3 Hierarchical community organization. Many networks are characterized by the presence of communities, representing groups of densely interconnected nodes. These communities may form a hierarchy, offering different levels of resolution for studying their functional implications. In PPI networks, functionally related proteins often form small, highly connected clusters, whereas proteins associated with the same disease tend to be more loosely connected.

most suitable community detection algorithms. As a consequence, algorithms that are not based on connection density, such as the ones introduced above in the context of expanded network neighborhoods, generally perform better in the identification of disease associated network neighborhoods.

Global Network Characteristics

Several of the concepts introduced above concern at least in part the global structure of a network, such as the degree distribution or the hierarchical organization into communities. In this last section, we will explicitly address network characteristics that can only be defined in the context of the entire network.

We start with the *diameter* of a network, first described in the context of social networks, where the "small world phenomenon" was popularized as early as 1967 (**Travers and Milgram, 1967**). The diameter of a network is defined as the largest distance between any node pair in the network. The distance between two nodes, in turn, is given by the length of the shortest path that can be constructed between them. In classical random networks (**Erdös et al., 1959**; **Gilbert, 1959**), where edges are distributed completely at random among the nodes, the diameter scales logarithmically with the number of nodes in the network. This is dramatically shorter than the diameters of regular lattices, which scale as a power-law of the number of nodes (with the exponent being determined by the dimensionality of the lattice). In a seminal paper, Watts and Strogatz showed that by introducing a small fraction of random links into a regular lattice, one can obtain networks that are both highly clustered and exhibit a small diameter, thus recapitulating two important features of many real world networks (**Watts and Strogatz, 1998**). Shortly after, it was found that scale-free networks can have dramatically slower growth of their diameter as a function of network size: For power-law exponents $\alpha < 3$, the diameter scales with the logarithm of the logarithm of the number of nodes (**Cohen and Havlin, 2003**). Many biological networks exhibit small diameters: for example, **Jeong et al. (2000)** found that the diameter of metabolic networks is fixed across organisms and network size, which they claim implies a fundamental constraint on metabolic networks.

Another network characteristic that was first introduced in social network analysis is the so-called *centrality*, which quantifies how important a particular node or edge is in the context of the entire network (**Newman, 2003**). As with network communities, there are numerous variations of centrality that approach the concept in different ways. The most studied of them, *betweenness centrality* of a node, is defined as the fraction of shortest paths between all pairs of nodes in the networks that pass through the node in question (**Freeman, 1977**). Intuitively, this determines what fraction of the flow of information would be disturbed by the deletion of the node, or how much information it "controls." The nodes in a network with relatively high betweenness are often referred to as *bottlenecks*. Early research in sociology found that people's centrality within social networks was correlated with a number of behavioral and social traits (**Wasserman et al., 1994**). Centrality was also shown to be important in biological networks, for example correlating with gene essentiality in PPI networks in yeast (**Jeong et al., 2001**) and humans (**Blomen et al., 2015**), cancer related genes (**Sun and Zhao, 2010**), and pathogen fitness in host–pathogen interactomes (**Crua Asensio et al., 2017**). Note that while betweenness centrality and node degree are typically strongly correlated, they often convey different information. In gene regulatory networks, for example, it was shown that betweenness centrality is a better measure of gene essentiality than degree (**Yu et al., 2007**), suggesting a role of bottleneck nodes in forming essential connections between different pathways, as opposed to non-bottleneck hubs which tend to connect protein complexes of similar function.

Recently, an entirely new set of tools and concepts for the global characterization of networks has emerged that may offer novel insights beyond long established measures like diameter, centrality and others. These tools are borrowed and adapted from a branch of mathematics called *homology*. Homology was originally developed to study the deformation-invariant properties of surfaces, for example the number of holes and inner voids. A particular application of homology, *persistent homology* (**Edelsbrunner and Harer, 2008**), uses these invariants to classify the global architecture of a set of points with pairwise similarities, thus allowing to generalize the technique to investigate networks and quantify its global similarity to a circle, sphere, or other elementary shapes and surfaces using a fast implementation of linear-algebra methods (**Zomorodian and Carlsson, 2005**). The application of these techniques to networks depends on choosing the optimal similarity measure for nodes in unweighted networks, which is still an open problem. Successful applications of homology-based data analysis techniques include the construction and analysis of similarity networks between patients using noisy features like health records and mutational profiles, for example for stratifying diabetes patients into subtle subclusters that were shown to have different patterns of comorbidity and were enriched in different SNPs (**Li et al., 2015**). A related technique was used in **Nicolau et al. (2011)** to construct a similarity network of breast cancer tissues based on their mutational profile, which lead to the identification of a previously uncharacterized tumor subtype.

Outlook

Network science has generated a vast body of knowledge since the appearance of the two seminal papers on small-world (**Watts and Strogatz, 1998**) and scale-free networks (**Barabasi and Albert, 1999**) a little over two decades ago. The application of tools and concepts to biology and medicine played no small part in the success of this dynamic field. In this short review, we introduced only the most basic concepts of how networks can be characterized. There are many extensions and generalizations of the introduced concepts, for example for networks that contain directed and/or weighted edges. We will conclude this review by highlighting a few areas of active research that may be of particular interest to the study of biological systems.

Biological systems span many orders of magnitude in both space and time, ranging from the molecular scale to the global ecosystem. The interactions within and between these different layers of organization can be investigated using so called *multilayer networks* that combine several layers, each may contain qualitatively different types of nodes and interactions (**De Domenico et al., 2013**). The dynamics on multilayer networks are generally richer, often displaying counterintuitive effects on diffusion (**De Domenico et al., 2016**), and other dynamic processes (**Boccaletti et al., 2014**).

There is a vast body of literature aiming to understand how different structural features of networks will affect the *dynamics* of processes on the network. Above we only superficially touched on diffusive processes like random walks and naturally, these ideas can be studied rigorously. Many other important processes like synchronization (**Arenas et al., 2008**) have been left out. Another approach to understanding dynamics on networks is through *controllability*, that is, studying minimal sets of nodes that can be varied so as to guide the evolution of a dynamic process on the network (**Liu et al., 2011**).

The study of dynamics on networks is particularly relevant to metabolic networks, for which a large body of analytical, computational, as well as applied work exists (**Voit, 2000**). In metabolic networks, nodes and edges represent metabolites and biochemical reactions, respectively, and the dynamics are the changing concentrations of the reactants. Under some broad assumptions, we can model the metabolic dynamics as what is known as an *S-system*, for which the steady-states can be derived algebraically (**Savageau, 1988**). This formalism is subsumed within the more general *Biochemical Systems Theory* (BST), which also allows for metabolites to control the rates of reactions without participating in them (**Voit, 2013**), modeling the rates of reactions as proportional to a power-law of the abundance of the participating or regulatory metabolites.

Lastly, one may also consider *temporal networks*, whose structure changes over time, possibly in response to processes taking place on it. In metabolic networks for example, the edges, reactions in this case, only occur in response to specific conditions and hence are time-dependent (**Almaas et al., 2004**). The temporal properties of networks have a strong influence on controllability, and have been claimed to have superior controllability properties (**Li et al., 2017**).

References

Ahn, Y.-Y., Bagrow, J.P., Lehmann, S., 2010. Link communities reveal multiscale complexity in networks. Nature 466, 761–764.

Albert, R., Jeong, H., Barabasi, A.L., 2000. Error and attack tolerance of complex networks. Nature 406, 378–382.

Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z.N., Barabási, A.-L., 2004. Global organization of metabolic fluxes in the bacterium Escherichia coli. Nature 427, 839–843.

Alon, U., 2007. Network motifs: Theory and experimental approaches. Nature Reviews. Genetics 8, 450–461.

Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., Zhou, C., 2008. Synchronization in complex networks. Physics Reports 469, 93–153.

Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.

Barabási, A.-L., Pósfai, M., 2016. Network Science. Cambridge University Press.

Barabási, A.-L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: A network-based approach to human disease. Nature Reviews. Genetics 12, 56–68.

Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al., 2015. Gene essentiality and synthetic lethality in haploid human cells. Science 350, 1092–1096.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics 2008, P10008.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., Zanin, M., 2014. The structure and dynamics of multilayer networks. Physics Reports 544, 1–122.

Caldera, M., Buphamalai, P., Müller, F., Menche, J., 2017. Interactome-based approaches to human disease. Current Opinion in Systems Biology 3, 88–94.

Caldera, M., Müller, F., Kaltenbrunner, I., Licciardello, M.P., Lardeau, C.-H., Kubicek, S., Menche, J., 2019. Mapping the perturbome network of cellular perturbations. Nature Communications 10, 1–14.

Clauset, A., Shalizi, C., Newman, M., 2009. Power-law distributions in empirical data. SIAM Review 51, 661–703.

Cohen, R., Havlin, S., 2003. Scale-free networks are ultrasmall. Physical Review Letters 90, 058701.

Conant, G.C., Wagner, A., 2003. Convergent evolution of gene circuits. Nature Genetics 34, 264–266.

Cowen, L., Ideker, T., Raphael, B.J., Sharan, R., 2017. Network propagation: A universal amplifier of genetic associations. Nature Reviews. Genetics 18, 551–562.

Crua Asensio, N., Muñoz Giner, E., de Groot, N.S., Torrent Burgas, M., 2017. Centrality in the host–pathogen interactome is associated with pathogen fitness during infection. Nature Communications 8, 14092.

De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S., Arenas, A., 2013. Mathematical formulation of multilayer networks. Physical Review X 3, 041022.

De Domenico, M., Granell, C., Porter, M.A., Arenas, A., 2016. The physics of spreading processes in multilayer networks. Nature Physics 12, 901–906.

Edelsbrunner, H., Harer, J., 2008. Persistent homology—A survey. Contemporary Mathematics 453, 257–282.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Erdös, P., Rényi, A., et al., 1959. On random graphs. Universitatis Debreceniensis 6, 290–297.

Erten, S., Bebek, G., Ewing, R.M., Koyutürk, M., 2011. DADA: Degree-aware algorithms for network-based disease gene prioritization. BioData Mining 4, 19.

Fields, S., Song, O.-K., 1989. A novel genetic system to detect protein--protein interactions. Nature 340, 245–246.

Fortunato, S., Hric, D., 2016. Community detection in networks: A user guide. Physics Reports 659, 1–44.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 40, 35–41.

Ghiassian, S.D., Menche, J., Barabási, A.-L., 2015. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. PLoS Computational Biology 11.

Gilbert, E.N., 1959. Random graphs. Annals of Mathematical Statistics 30, 1141–1144.

Gingras, A.-C., Gstaiger, M., Raught, B., Aebersold, R., 2007. Analysis of protein complexes using mass spectrometry. Nature Reviews. Molecular Cell Biology 8, 645–654.

Girvan, M., Newman, M.E.J., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99, 7821–7826.

Gonzalez, M.W., Kann, M.G., 2012. Chapter 4: Protein interactions and disease. PLoS Computational Biology 8, e1002819.

Holland, P.W., Leinhardt, S., 1971. Transitivity in structural models of small groups. Comparative Group Studies 2, 107–124.

Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al., 2017. Architecture of the human interactome defines protein communities and disease networks. Nature 545, 505–509.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., Hood, L., 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., Gerstein, M., 2003. A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science 302, 449–453.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic networks. Nature 407, 651–654.

Jeong, H., Mason, S.P., Barabási, A.-L., Oltvai, Z.N., 2001. Lethality and centrality in protein networks. Nature 411, 41–42.

Jonsson, P.F., Bates, P.A., 2006. Global topological features of cancer proteins in the human interactome. Bioinformatics 22, 2291–2297.

Khamina, K., Lercher, A., Caldera, M., Schliehe, C., Vilagos, B., Sahin, M., Kosack, L., Bhattacharya, A., Májek, P., Stukalov, A., et al., 2017. Characterization of host proteins interacting with the lymphocytic choriomeningitis virus L protein. PLoS Pathogens 13, e1006758.

Kim, Y.-A., Wuchty, S., Przytycka, T.M., 2011. Identifying causal genes and dysregulated pathways in complex diseases. PLoS Computational Biology 7, e1001095.

Leiserson, M.D.M., Vandin, F., Wu, H.-T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al., 2015. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nature Genetics 47, 106–114.

Lewis, A.C.F., Jones, N.S., Porter, M.A., Deane, C.M., 2010. The function of communities in protein interaction networks at multiple scales. BMC Systems Biology 4, 100.

Li, L., Cheng, W.-Y., Glicksberg, B.S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E.P., Dudley, J.T., 2015. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Science Translational Medicine 7, 311ra174.

Li, A., Cornelius, S.P., Liu, Y.-Y., Wang, L., Barabási, A.-L., 2017. The fundamental advantages of temporal networks. Science 358, 1042–1046.

Liu, Y.-Y., Slotine, J.-J., Barabási, A.-L., 2011. Controllability of complex networks. Nature 473, 167–173.

Loscalzo, J., 2017. Network Medicine. Harvard University Press.

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., Barabási, A.-L., 2015. Uncovering disease-disease relationships through the incomplete interactome. Science 347, 1257601.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: Simple building blocks of complex networks. Science 298, 824–827.

Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Review 45, 167–256.

Newman, M.E.J., 2012. Communities, modules and large-scale structure in networks. Nature Physics 8, 25–31.

Newman, M., 2018. Networks. Oxford University Press.

Nicolau, M., Levine, A.J., Carlsson, G., 2011. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proceedings of the National Academy of Sciences of the United States of America 108, 7265–7270.

Oti, M., Snel, B., Huynen, M.A., Brunner, H.G., 2006. Predicting disease genes using protein-protein interactions. Journal of Medical Genetics 43, 691–698.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555.

Reichardt, J., Bornholdt, S., 2006. Statistical mechanics of community detection. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics 74, 016110.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al., 2014. A proteome-scale map of the human interactome network. Cell 159, 1212–1226.

Savageau, M.A., 1988. Introduction to S-systems and the underlying power-law formalism. Mathematical and Computer Modelling 11, 546–551.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al., 2005. A human protein-protein interaction network: A resource for annotating the proteome. Cell 122, 957–968.

Stumpf, M.P.H., Ingram, P.J., 2005. Probability models for degree distributions of protein interaction networks. EPL 71, 152.

Sun, J., Zhao, Z., 2010. A comparative study of cancer proteins in the human protein-protein interaction network. BMC Genomics 11 (Suppl 3), S5.

Teschendorff, A.E., Banerji, C.R.S., Severini, S., Kuehn, R., Sollich, P., 2015. Increased signaling entropy in cancer requires the scale-free property of protein interaction networks. Scientific Reports 5, 9646.

Tomczak, K., Czerwińska, P., Wiznerowicz, M., 2015. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. Contemporary Oncology 19, A68–A77.

Travers, J., Milgram, S., 1967. The small world problem. Psychology Today 1, 61–67.

Turkheimer, F.E., Hellyer, P., Kehagia, A.A., Expert, P., Lord, L.-D., Vohryzek, J., De Faria Dafflon, J., Brammer, M., Leech, R., 2019. Conflicting emergences. Weak vs. strong emergence for the modelling of brain function. Neuroscience & Biobehavioral Reviews 99, 3–10.

Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.-I., et al., 2009. An empirical framework for binary interactome mapping. Nature Methods 6, 83–90.

Voit, E.O., 2000. Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists. Cambridge University Press.

Voit, E.O., 2013. Biochemical systems theory: A review. ISRN Biomathematics 2013.

Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., Goldenberg, A., 2014. Similarity network fusion for aggregating data types on a genomic scale. Nature Methods 11, 333–337.

Wasserman, S., Faust, K., Stanley (University of Illinois Wasserman, Urbana-Champaign), 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of “small-world” networks. Nature 393, 440–442.

Yang, Z., Algesheimer, R., Tessone, C.J., 2016. A comparative analysis of community detection algorithms on artificial networks. Scientific Reports 6, 30750.

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

Zaki, N., Efimov, D., Berengueres, J., 2013. Protein complex detection using interaction reliability assessment and weighted clustering coefficient. BMC Bioinformatics 14, 163.

Zhong, Q., Pevzner, S.J., Hao, T., Wang, Y., Mosca, R., Menche, J., Taipale, M., Taşan, M., Fan, C., Yang, X., et al., 2016. An inter-species protein-protein interaction network across vast evolutionary distance. Molecular Systems Biology 12, 865.

Zomorodian, A., Carlsson, G., 2005. Computing persistent homology. Discrete & Computational Geometry 33, 249–274.

Further Reading

Barabási, A.-L., Pósfai, M., 2016. Network Science. Cambridge University Press, A broadly accessible introduction to network science.

Loscalzo, J., Barabási, A.-L., Silverman, E.K. (Eds.), 2017. Network Medicine: Complex Systems in Human Disease and Therapeutics. Harvard University Press, A textbook covering a wide range of applications of network science in medicine.

Newman, M.E.J., 2018. Networks. Oxford University Press, A detailed textbook on the mathematical techniques of network theory.

Relevant Websites

http://networksciencebook.com/—Website of the book Network Science.
http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie—A source for downloading PPI information.
http://snap.stanford.edu/—An inventory of tools for analyzing complex networks.
https://icon.colorado.edu/#!/—A repository of complex networks for download.

Glossary

Adjacency matrix A network of n nodes can be represented by an $(n \times n)$ matrix a with values $a_{ij} = 1$ if two nodes i and j are connected, and $a_{ij} = 0$ otherwise.

Centrality Family of related measures that quantify the importance of a node or edge in a network.

Clustering Measure for the tendency of nodes to form connected triangles.

Community The heuristic notion of a group of nodes in a network that are more closely linked to one another than they are to the rest of the network.

Degree The number of edges connecting a node to the rest of the network.

Degree distribution Histogram over the degrees of all nodes in a network.

Edge Networks are defined as a set of objects (nodes) and their pairwise relationships (edges). Edges are often also referred to as links.

Homology A mathematical technique for quantifying the large-scale topology of a geometrical shape, network or dataset.

Modularity A network measure that quantifies the tendency of a group of nodes to be densely interconnected among themselves, compared to the rest of the network.

Neighbors The set of nodes connected to a given node.

Network A collection of nodes and edges between them.

Node Networks are defined as a set of objects (nodes) and their pairwise relationships (edges). Nodes are often also referred to as vertices.

Power Law A distribution in which the probability P of a value k is inversely proportional to some power α of its magnitude: $P(k) \sim k^{-\alpha}$.

Protein-protein interaction (PPI) An experimentally observed physical binding between two proteins.

Random walk A process in which some hypothetical agents move from node to node by selecting outbound edges at random and moving along them.

Scale-free network A network in which the degree distribution is well-approximated by a power law.

Biography

Jörg Menche is a theoretical and computational physicist by training. During his PhD at the Max-Planck-Institute for Colloids and Interfaces in Potsdam (Germany), he specialized in network science and afterwards went to work with one of the world's leading experts in this field, Albert-László Barabási at Northeastern University in Boston (USA). Collaborating closely with Joseph Loscalzo from Harvard Medical School and Marc Vidal from Dana Farber Cancer Institute, he laid out the basic theoretical framework for how the interactome can be understood as a map to study human disease. Since 2015, he is a principal investigator at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences in Vienna (Austria). Major research areas of his group are network-based approaches to rare diseases, understanding the basic principles of how perturbations of biological systems influence each other and developing novel Virtual Reality (VR) based technologies for analyzing large genomic data.

Joel Hancock, MMath, studied mathematics at Oxford University (UK) as a Scholar at St. Catherine's College, graduating with a Master's degree in 2017 and thereafter joining Jörg Menche's lab as a PhD student. His research interests include combinatorial graph theory, image analysis, and network-based data analysis.