frontiers in Genetics

REVIEW
published: 14 September 2021
doi: 10.3389/fgene.2021.701331

Modularity in Biological Networks

Sergio Antonio Alcalá-Corona [1,2], *Santiago Sandoval-Motta* [1,2,3], *Jesús Espinal-Enríquez* [1,2] *and Enrique Hernández-Lemus* [1,2*]

[1] Computational Genomics Division, National Institute of Genomic Medicine, Mexico City, Mexico, [2] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico, [3] National Council on Science and Technology, Mexico City, Mexico

Network modeling, from the ecological to the molecular scale has become an essential tool for studying the structure, dynamics and complex behavior of living systems. Graph representations of the relationships between biological components open up a wide variety of methods for discovering the mechanistic and functional properties of biological systems. Many biological networks are organized into a modular structure, so methods to discover such modules are essential if we are to understand the biological system as a whole. However, most of the methods used in biology to this end, have a limited applicability, as they are very specific to the system they were developed for. Conversely, from the statistical physics and network science perspective, graph modularity has been theoretically studied and several methods of a very general nature have been developed. It is our perspective that in particular for the modularity detection problem, biology and theoretical physics/network science are less connected than they should. The central goal of this review is to provide the necessary background and present the most applicable and pertinent methods for community detection in a way that motivates their further usage in biological research.

Keywords: modularity, community structure, motifs, biological networks, systems biology

OPEN ACCESS

Edited by:
Marco Pellegrini,
Institute of Computer Science and Telematics, Italian National Research Council, Italy

Reviewed by:
Dario Ghersi,
University of Nebraska Omaha, United States
Vinícius Vieira,
Universidade Federal de São João del-Rei, Brazil

****Correspondence:***
Enrique Hernández-Lemus
ehernandez@inmegen.gob.mx

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *27 April 2021*
Accepted: *23 August 2021*
Published: *14 September 2021*

Citation:
Alcalá-Corona SA, Sandoval-Motta S, Espinal-Enríquez J and Hernández-Lemus E (2021) Modularity in Biological Networks. Front. Genet. 12:701331. doi: 10.3389/fgene.2021.701331

1. INTRODUCTION

The field of Systems Biology has many branches that focus on studying networks. It is common to encounter in the literature terms such as metabolic networks, transcriptional networks, protein-protein interaction networks, etc. These networks are graph-theoretical constructs composed of nodes and edges that aim to describe the integrated state of a biological system. Nodes represent the elements of the system, while edges represent the relation between any two of these elements. Depending on the scale of the biological entities at hand, a network can describe systems such as: ecological systems where each node is a biological entity itself; an organism with nodes being organs or groups of organs; tissues or individual cells with genes, proteins, organelles, and metabolites interacting with each other; down even to the level of amino acids interacting to build a protein. Networks facilitate the identification of relevant entities and interactions through the use of theoretical and computational analysis over experimental data. These analyses aim to make predictions, or at least detailed and accurate descriptions of the underlying biological systems. Since one of the most common applications of complex systems in biology is the representation of biological interactions as edges or links of a network, the connectivity or interaction structure of such a network is of utmost importance. This structure is known as the topology of the network and in biological systems it is usually not random. This means that who is connected to whom is relevant, and the distribution of links is arguably related to the particular functionality of such systems.

In biological systems, modularity has been associated with properties such as robustness (Aldana and Cluzel, 2003), mainly derived from the Boolean network approach (Kauffman, 1969). The concept of robustness is related to the ability of a system to withstand perturbations and retain its functionality, whichever it may be (Aldana et al., 2007). Examples of robustness in a biological system can be observed in biochemical networks (Barkai and Leibler, 1997; Morohashi et al., 2002), signaling networks (Igoshin et al., 2007; Espinal et al., 2011; Espinal-Enríquez et al., 2017) and other complex biosystems. For instance, in prokaryotic organisms, sigma factors, despite their structural similarity, regulate different sets of genes, but the regulatory function of a dysfunctional sigma factor can be reassigned to other sigma factors making the organism functional (Torres-Sosa et al., 2012). Another example of modularity arises when a set of genes is regulated by the same transcriptional factor (set known as a *regulon*). It has been proposed that these sets of genes can give rise to functional modules in *Pseudomonas aeruginosa* (Schulz et al., 2015) and that such modules are essential for the adaptation and survival under challenging environments.

One goal of studying biological systems as networks is to understand how the interconnectedness and function of each element derives in a system-level behavior. In order to uncover these features one can look into the *design principles* of the network. This means, to try to uncover the particular patterns present in the network's topology, such as the ways the nodes are connected to each other; the functional groups they belong to; or if nodes with a particular function agglomerate in subgroups. Topological features, of course, are only partially responsible for the actual design principles of biological systems. Connectivity features common of biological networks, such as the approximate scale-free nature of their connectivity distributions, hierarchical and modular organization, set the stage for functional features to emerge. Such functional features are a consequence of the underlying organizational structure of the systems, their physiological setting and environmental constraints. Regarding network connectivity, it is known that the organization patterns of large complex networks are often composed of structural sub-units often called modules or communities (Girvan and Newman, 2002). Communities and modules in the present context are interchangeable terms, however in this manuscript we will use the latter term as we believe it has a similar meaning over a large number of disciplines, with the possible exception of the Social Sciences and Mathematics.

2. MODULARITY IN BIOLOGICAL SYSTEMS

So what is a module? Despite there is still no consensus on what defines a module, a generally accepted notion is that it corresponds to a tightly interconnected set of edges in a network. Intuitively, the density of connections inside any so-called module (*within-connections*) must be significantly higher than the density of connections with other modules (*between-connections*) (Thieffry and Romero, 1999; Girvan and Newman, 2002; Clauset et al., 2004; Palla et al., 2005). Modularity has been helpful in many biological fields and can even be useful in exploratory research (Serban, 2020). In the following sections, we will present and discuss the latest developments of modularity research in biological systems as well as the necessary concepts and formal definitions to understand and promote the usage of several modularity detection algorithms in the biological sciences (Didier et al., 2018; Li et al., 2019).

2.1. Emergence of Modularity

In order to perform their vital functions and at the same time comply with changing environmental conditions, living systems must possess a high degree of internal organization. A likely scheme to attain such a sophisticated degree of organization is through the coupling of diverse biological processes, which creates the needed correlations among their internal and external constraints to perform a certain task. This theory is known as the *networks of processes* (Clarke and Mittenthal, 1992) and suggests that modules can be thought as clusters of coupled elements that work under certain constraints. It also states that organisms can be studied as super-modules (e.g., networks) made up of several interplaying modules that adapt as a whole to changes in their environment. Under this scheme, modularity can be thought of as a very effective way to prioritize and optimize the correct functioning of living systems, which are undoubtedly subject to changing environmental conditions or even to entropic decay.

The question of how modularity emerges in biological networks has no definitive answer yet, either. It has been shown that dynamical networks, which include temporal processes occurring in the whole spatial structure of the network, can give rise to modular behavior when driven by growth, duplication and diversification. These duplication-centered dynamic models emerge from the fact that if some parts of a system undergo duplication, the new system will be more modular than the original (Lorenz et al., 2011). How modularity emerges is closely related to the question of how and why it is preserved across so many biological systems (Kashtan and Alon, 2005; Gibson, 2016). This question has been addressed in evolutionary/developmental biology (evo/devo) and in molecular systems biology as a kind of intersection point between both disciplines. It has been argued that there is indeed a relationship between modularity and controllability (Constantino and Daoutidis, 2019).

Despite underlying mutational mechanisms have been proposed to explain the emergence of modularity, selection and other evolutionary forces have also been part of this discussion (Wagner et al., 2001, 2007; Espinosa-Soto and Wagner, 2010; Clune et al., 2013; Friedlander et al., 2013; Banerjee et al., 2017; Verd et al., 2019; Jaeger and Monk, 2021), as are ecological factors such as spatial distribution and population dynamics (Gilarranz, 2020). Biological modularity arise in the contexts of dynamical process that may even challenge compartmentalization and cause the breakdown of modularity or its rearrangement (Valverde, 2017; Wang et al., 2021).

In the next section, we will discuss the different notions of modularity –particularly those more closely related to the modular organization at the molecular, functional and cellular levels– and their application to a wide diversity of biological phenomena.

2.2. Applications of Network Modularity

One clear example of application of network theory in biology is the study of Gene Regulatory Networks (GRNs) (Davidson and Levin, 2005). These networks can be conceptualized as control systems that drive whole-genome expression patterns (Hernández-Lemus et al., 2019). This coordinated expression is attained through the orchestrated expression of transcription factors and other regulatory molecules like siRNAs, histones, etc. The wider availability of high throughput technologies has sprouted a new wave of modularity research in GRNs. After the completion of the human genome project (HGP), and following the pioneering work of Kauffman (1969) and Britten and Davidson (1969) in the late 1960s, transcriptional regulation module discovery has become an extremely fruitful research field. For instance, it has been demonstrated that modularity can emerge as a consequence of gene co-expression in GRNs; by associating the functions of these genes and their regulators, it has been argued that gene co-expression may confer functional advantages to the organisms, as genes with related functions are likely regulated in a similar manner (Solé et al., 2002; Narula et al., 2010). Gene functionality of several genes with no prior functional description has already been predicted (Segal et al., 2003; Lee et al., 2004; Tanay et al., 2004). Also, by integrating gene expression levels with the modular structure, it was possible to build a comprehensive map of gene regulation for a whole organism (Zhu et al., 2008).

Community structure and modularity in metabolic networks is another important research field. Many biochemical interventions and biotechnological applications depend on modularity, and with the advent of synthetic biology, the use of modules will probably escalate in the near future, driven by the possibility to evolve engineered biological systems (Parter et al., 2007). Modularity in metabolic networks has been extensively explored since the pioneering work by Ravasz et al. (2002) where through the reconstruction of 43 metabolic networks from different organisms, they found that scale free topologies were ubiquitous. Briefly, in these networks the probability distribution of connections on the network (degree-distribution) follows a power law, so that most nodes will end up with few connections and only a few nodes will end up with many. In this case, the studied networks had values of the scaling exponent around 2, and an average clustering coefficients (see section 3) about an order of magnitude larger than expected for scale free networks. This scaling exponent around 2, suggests that these networks are probably under a dynamical regime between that of an ordered system and the one of a chaotic one. This regime is known as *critical* and it has been observed in many different complex systems (Shmulevich et al., 2005). Another important theoretical contribution of this work is the introduction of the *topological overlap matrix* (Ravasz et al., 2002; Cheng et al., 2019).

The **interactome** (Sanchez et al., 1999) is a useful concept related to Protein-protein (physical) interaction (PPI) networks, which are also organized into functional subnetworks or modules. An interactome is defined as a biological network, which encompasses the complete set of molecular interactions in a particular cell. These interactions range from physical (as in PPI networks) to indirect, as is the case of epistatic or gene-gene interactions, and may even include edges defined by regulatory interactions like those of a GRN (Gómez-Romero et al., 2020). Even if interactomes seem to be less clearly defined than other biological networks, they may be used to represent processes that, although not completely understood, may be associated with some specific phenotypes. The *human disease network* (HUDiNE) (Goh et al., 2007) was actually created by using interactomes. HuDiNe, according to its creators is *a network of disorders and disease genes linked by known disorder–gene associations*. The observation that genes linked to similar diseases present a higher likelihood of sharing physical interactions between their products (e.g., PPI) and a higher correlation in their expression profiles, lead to the conclusion that such a network will likely display characteristic disorder-specific functional modules. This fact was corroborated by analyzing the topological structure of the HuDiNe (Goh et al., 2007). Since the release of HUDiNE, interactomes related to disease have been carefully curated and archived in structured databases, thus making possible the discovery of new *co-morbidities* from a molecular rather than epidemiological perspective (Menche et al., 2015).

In the case of human diseases, modular network decomposition has been applied to further our understanding of the interactions driving the emergence of several complex diseases (Sardiu et al., 2017; Tripathi et al., 2019; Lucchetta and Pellegrini, 2020). One good example is the work of De Matos Simoes and collaborators with cancer cells. By using a network modularity analysis, they showed that transmembrane proteins along with ion channel complexes and receptors play a significant role in the pathogenesis of B-cell lymphoma. The authors based their argument on the observation that central and peripheral layers in the modular decomposition of the networks may play different physiological roles. Hierarchical modular separation may then provide clues as to cross-regulatory phenomena in complex phenotypes. Specifically, they noted that these molecules act via the communication disruption between the intracellular regions and the peripheral regions of B cells (de Matos Simoes et al., 2012). In pancreatic cancer, the disruption of intracellular adhesion and cell-division cycles in the tumors were found to be driven by clearly defined transcriptional modules (Long et al., 2016). Also, network communities related to survival have been found in regulatory networks from hepatocellular carcinoma (Xu et al., 2016). Expression activity of the genes in such modules may contribute to timely stratification and tumor staging of liver cancer patients.

Other complex phenotypes have been dissected by analyzing the community structure of their underlying networks. During brain development, for example, it has been shown that the perinatal transition leads to modular reorganization of the brain, which is in turn associated with the development of new functions. This modularization is also correlated with specific gene sets whose expression are synchronously changing, as they share transcriptional regulators (Monzón-Sandoval et al., 2016). Similar methods have allowed the identification of distinctive molecular pathways that differentiate early and late-onset temporal lobe epilepsy in children (Moreira-Filho et al., 2015). These studies have pointed out that differentially expressed modules in early onset epilepsy are related to neural excitability and febrile seizures, whereas no neural excitability gene modules were found for late onset. These findings support the hypothesis that early onset epilepsies, even if accompanied by severe hippocampal damage, may present compensatory effects. This difference may set the basis for differentiated drug treatments.

Community structure in regulatory networks may also be useful to discover potential molecular targets to treat complex diseases (Muraro and Simmons, 2016). In coronary artery disease, for instance, modules associated with the hypertrophic cardiomyopathy pathway and membrane-related functions were detected (Liu et al., 2016). These pathways, the authors suggest, can provide a means to define a set of druggable process-specific targets (Ashrafian et al., 2011). Transcriptional modules associated with the response to allergens leading to seasonal allergic rhinitis have been also identified by Shi and collaborators (Shi et al., 2010). These modules revealed that the MAP kinase, B-cell receptor and toll-like receptor signaling pathways are crucial for the critical stages of allergic rhinitis. Regarding the role of gene regulation on viral pathogenicity and how it has been shaped by modular adaptation, it has been discussed how enhanced redundancy leads to robustness of the infectious phenotypes (Oliveira et al., 2013).

So far we have discussed several examples where finding modules in biological networks lead to a better understanding of the molecular and regulatory processes involved in certain phenotypes and behaviors. A relevant fraction of the modularity finding approaches used in network biology were developed with a particular biological question in mind. The methods thus developed were, in general, efficient to answer that kind of questions but resulted somehow lacking generalizability. We call these methods *ad hoc*, since they have been developed for a special purpose. Most of these methods are indeed quite useful on a case-by-case basis. However, since modularity analysis is a relevant problem in contemporary theoretical biology, it is desirable to have general methods, or at least methods with broad applicability, to help lay the conceptual foundations of biological modularity. We believe that a first step toward this aim consists in applying the general methods developed in graph theory and network science to biological questions and fine-tune them to account for known biological phenomena. In the next section, we will review several necessary concepts and useful methods for modularity detection that come from a more theoretical perspective. As such, these methods were developed to be useful under any, or at least several, quite general circumstances. We have also included a benchmark section, where we discuss how these algorithms stand against each other in the discovery of modules using both real and synthetic datasets. Although the field of modularity detection in biological systems is somewhat young, it has a long history in physics, and thus, many algorithms are already out there making impossible to review all of them. A later section will discuss the most relevant methods separated by the algorithm they are based on in the hopes that the reader will find some of them useful for their research.

3. NETWORK THEORY

In order to better understand the modularity detection methods that will follow, we will briefly define/recall a few important network properties. For a deeper coverage of these and several other properties we suggest the reader to look, for instance, at the review by Newman (2010). For an introductory lecture on the importance of networks in biology and their main applications besides modularity detection we suggest the review by Green et al. (2018).

3.1. Complex Networks: Concepts and Definitions

For the sake of clarity, we will briefly introduce some well-known definitions of network theoretical concepts.

DEFINITION 1. *A* ***network*** *is formally defined as a graph* $G(V, E)$ *over two sets: a set of nodes or vertices,* $v_i \in V$*, (e.g., bio-reactants), and a set of edges or links connecting such vertices* ($e_i \in E$) *(e.g., chemical reactions). The connectivity of the network is often represented by the* ***adjacency matrix*** $\mathbb{A} = A_{i,j}$*, where* $A_{i,j} \neq 0$ *implies an existing interaction between nodes* v_i *and* v_j*.*

DEFINITION 2. *The degree-distribution of a network refers to the distribution of the number of connections per node, and is defined as the number of connections a given node has to other nodes (called the* degree *of the node). Thus,* ***the degree distribution*** *is defined as the probability distribution of the degrees of all the nodes of the network. This measure is often used as an indicator of the relative importance of a particular node (Barabasi and Oltvai, 2004).*

Mathematically: Let v_i^m *be the set of vertices connected to a given vertex (a.k.a. node) m (i.e.,* $A_{i,m} \neq 0; \quad \forall v_i \in v_i^m$*). We call* v_i^m *the* ***neighborhood*** *of vertex m. The size, or cardinality, of this set* $C(e_i^m) = k_m$ *is called the* ***degree*** *or* ***connectivity*** *of vertex m, also written as* $deg(v_m)$*.*

DEFINITION 3. *A* ***Network motif*** *is defined by a group of connected nodes (a sub-graph) that is prevalent in a network or in several networks. Each motif is thus associated with a particular pattern of interconectedness between vertices, and may reflect a framework in which particular functions are achieved efficiently. These patterns describe arrangements of interconnection that are present with a significantly higher frequency than in networks where nodes are randomly connected (Milo et al., 2002).*

DEFINITION 4. *Intuitively,* ***network modularity*** *consists in associating network nodes to different categories or subsets of the network. Assignment is based on connectivity patterns within the graph, rather than on some inherent node features. The formal definition of network modularity is still controversial, but we believe that by giving some enclosing definitions from graph theory, we can gain a deeper understanding of this concept and methods described below.*

DEFINITION 5. ***Full/Overlapping partition.*** *We may consider a set Z of disjoint subsets of a network* $Z(V, E)$ *so that* $Z = Z_1 \dot{\bigcup} Z_2 \dot{\bigcup} \ldots \dot{\bigcup} Z_k$*. This is called a* full partition *of the network.*

If, on the other hand, we allow a non-empty intersection between the subsets $Z_i \bigcap Z_j \neq \emptyset$, *we have* $Z = \hat{Z}_1 \bigcup \hat{Z}_2 \bigcup \ldots \bigcup \hat{Z}_k$ *which is called an* overlapping partition *of the network.*

DEFINITION 6. ***Incomplete/Modular Partition.*** *We can also consider an incomplete partition of Z, i.e., one in which not every vertex in V is assigned to a subset. In this case we call* $M \subset Z$ *a* modular partition *of the network,* $M = M_1 \bigcup M_2 \bigcup \cdots \bigcup M_k \subset Z$. *The subsets* M_i *(which may or may not be overlapping) are called the* modules *of Z. There are several ways in which a network can be partitioned. Here lies the difficulty in defining modularity in complex networks: different definitions of modularity may induce different modular partitions of the network, which leads to different modularity measures.*

DEFINITION 7. *The* ***clustering coefficient*** $CC(i)$ *for a particular vertex i in a network is given by:*

$$CC(i) = \frac{\textit{number of triangles connected to i}}{\textit{number of possible triangles connected to i}} \qquad (1)$$

Here, a triangle is a set of three fully interconnected nodes. Since $0 \leq CC(i) \leq 1$. *Equation (1) can be rewritten as:*

$$CC(i) = \frac{2\,E_i}{k_i\,(k_i - 1)} \qquad (2)$$

Where E_i *is the number of triangles centered in vertex i and* k_i *is the degree of that vertex.*

Once we have an operative definition of clustering coefficient, its mean value is the average over all nodes i.

$$\langle CC \rangle = \frac{1}{N} \sum_{i}^{N} CC(i) \qquad (3)$$

$\langle CC \rangle$ *is a probabilistic measure of the abundance of triangles (not necessarily triads, but also higher order motifs) in the network.*

Global measures such as the $\langle CC \rangle$ are computationally cheap (Fortunato, 2010). However, their utility is mostly restricted to the case of hierarchic modularity scenarios (modules within modules). Hierarchic modularity was originally defined as the property of self similarity in the module distribution in a large scale network, evidenced by a power-law behavior of the clustering coefficient $C(k) \sim k^{-1}$. This relation in turns involves the coexistence of a hierarchy of nodes with different degrees of *node-modularity* –as measured by the node-specific clustering coefficient–. In brief, under such assumptions, the higher a node connectivity k is, the smaller its clustering coefficient, which in the asymptotic regime gives rise to the inverse law, $1/k$.

3.2. Network Models: Types and Approaches

3.2.1. Weighted Networks

A weighted network is defined by the assignment of a weight for each of the edges of the network. These weights are established based on the type and strength of the interaction at hand. Interestingly, weighted networks have proven to further increase the reliability of the modules proposed. For instance, the weighted overlap measure (WOM) is a similarity measure that calculates the overlap between two sets weighted by their relative contribution to the overall (joint set) (Smith, 1985). The WOM has been used to define gene modules that are more cohesive than those obtained through unweighted networks though this is not always the case. Here a more *cohesive* module means that the average value of the inter-module clustering coefficient is higher than the average value of the network's clustering coefficient. Since its proposal, the WOM has been used to recover experimentally validated functional gene modules in cancer cells and in yeast (Zhang and Horvath, 2005). More importantly, it has been shown that modularity affects biological functions as the dynamics of the whole network is determined by the organizational patterns generated by the modules themselves. For example, bi-stable switches, where weighted edges are essential for bi-stability, are known to enhance regulatory feedback and feed-forward loops, which in turn are related to the ability of an organism to adapt to changing environments (Kashtan et al., 2009; Gyorgy and Del Vecchio, 2014).

The functional role of regulatory modules has proved to go beyond that of loops and motifs. By studying a transcriptional network of myeloid cells, Alcalá-Corona and coworkers showed that modules are consistently associated at the pathway level to sets of biological functions (Alcalá-Corona et al., 2016). Community structure has also proven to affect the dynamical behavior of the network (Qi and Ge, 2006). By analyzing simple models of gene regulation, Xu and Wang were able to fully decompose a complex network in terms of independent functional modules (Xu and Wang, 2010). Although clear cut decompositions are not likely to occur in a real biological networks due to pleiotropy, decompositions make possible to observe modular effects in an idealized way. For instance, they have been used to study the effects of the free scale topology and of hierarchical modularity on the large scale structure of GRNs (Zhan, 2007). When network structural properties are supplemented with appropriate dynamic behavior, robustness is enhanced (Aldana et al., 2007). This increase in robustness has been shown to be due to the presence of large attractor basins that lead to stable gene expression patterns (Sevim and Rikvold, 2008).

3.2.2. Multi-Level Networks

The advancement of graph theory along with interactomes gave rise to the concept of multi-layered networks. Multi-layered networks encompass several types of interactions and node types. However, in this *multiplex framework* interactions are integrated into different network layers and therefore more information about the real underlying phenomena can be retained (Didier et al., 2015). Adding extra dimensions to a graph can make the associated mathematical analyses more intricate and hinder the application of common topological approaches to study modularity. Nevertheless, it has been shown that real modules encountered in curated networks are better recovered with modular algorithms applied to multilayered networks, compared with the same algorithms applied to single-layer networks. A detailed mathematical framework for multilayer networks—introductory, though not elementary—is found in the comprehensive paper by De Domenico et al. (2013).

In addition to the multiple molecular levels of description of a phenomenon, multi-layered networks can be adapted to include multiple species which can be useful in disciplines such as in comparative genomics. This extended approach also has more robust scalability features than mono-layered networks (Ritchie et al., 2016). Multi-layered networks have enforced the development of new theoretical approaches need for discovering modularity such as the *Multiplex PageRank algorithm* (Iacovacci and Bianconi, 2016).

Another important feature of multi-layered networks is that they allow a direct analysis of the functional features of their subjacent modules (e.g., pathway-based strategies). This approach is useful for studying phenotypes that are naturally multi-layered, like those associated with genetic regulation where multiple different sources (e.g., transcription factors, chromatin, methylation, etc.) are responsible for the phenotype. For instance, through the use of a multi-layered network of transcription factors and microRNA co-targeting, along with protein-protein interaction and gene co-expression (Cantini et al., 2015) were able to find a set of cancer driver genes associated with the community structure of the network.

A related issue to that of multilayer networks is *multiscale modularity*. Despite highly connected nodes, or hubs, are often labeled as the most important nodes of a network, recent studies in the modular structure of the regulatory networks of *Escherichia coli*, *Saccharomyces cerevisiae*, and *Staphylococcus aureus* revealed an unexpected relevance for low degree metabolites. By using flux balance analysis and graph theoretical methods, Samal et al. (2006) were able to discover connected clusters of low-degree metabolites. These large clusters of low degree nodes turned out to be over-represented in these metabolic networks so that a majority of the essential metabolic reactions could be characterized by just a few low degree metabolites. In this study, reactions whose fluxes were strongly correlated formed well-defined communities in metabolic networks of the organism. The large scale community structure, that is, the network modules conforming relatively large subnetworks, and the small scale modularity (partitions of small motifs), represent a complex interplay that has been shown to play an important role in metabolism under the assumption of hierarchical network organization (Gao et al., 2016). By introducing the concept of multiscale modularity, they propose that network community structure may be defined in several organizational levels, taking into account high and low degree nodes.

4. MODULARITY DETECTION ALGORITHMS

From the perspective of the statistical physics, computer science, computational sociology, network science and complex systems communities, there has been a significant amount of work devoted to solve the modular partition or community detection problem. Unlike what happened with biological networks, these methods aimed at reaching formal and theoretically-founded results with wide applicability. It is important to note that there is the possible drawback of losing some interpretability of the results in the quest for generality. However, it is our belief that these methods will prove useful for the biological community, as these approaches remain largely unknown and offer complementary views of the same problem. With this in mind, the following sections will be focused on introducing this second perspective to the community detection problem.

Classification of community detection algorithms depends on their approach to the graph partition problem. Although there is a wide variety of methods and algorithms to approach the problem of graph partitioning and network modularity detection, they often fall in one of five (quite general and sometimes overlapping) possible categories:

1. Methods based on data clustering
2. Methods based on optimization of the modular partition
3. Methods based on the spectral properties of the adjacency matrix
4. Random walk based and other dynamical algorithm methods
5. Stochastic block models

As we will see, there are advantages, disadvantages and limitations in all types of models. For this reason, it is wise to consider the features, applicability and benchmark performance before opting-in for a certain model.

4.1. Data Clustering-Based Methods

There are several methods based on measuring some significant statistical similarity or distance over the biological data. Some techniques have been developed to ascertain whether a set of proposed modules adequately represents the whole set of molecular determinants of a single disease, or closely related diseases.

For instance, in Menche et al. (2015), a topological method was devised in order to locate disease-related communities within the interactome (whole set of interactions in a particular cell). This method uses the overlap among communities of different pathologies to predict disease-disease associations. Although simple, this method has proved very useful and further improvements have been made to the initial algorithm, in particular on relation to the establishment of endo-phenotype models as discussed in Ghiassian et al. (2015) and Ghiassian et al. (2016).

One important limitation of clustering based methods rely on the challenge to determine the optimal number of clusters. The problem of an optimal number of clusters/modules is actually an open challenge in theoretical computer science and graph theory. Even approximate solutions often depend on the specifics of the algorithm used. Some methods as the ones based on spectral bisection have conditions to define an a priori number of clusters, while other methods like those based on structural properties, on dynamical process over the networks and those which have a stochastic component; may determine a number of clusters, based on their large and local structure of the network, an approach some consider to be more *natural*.

One relevant method for disease module detection is DIAMOND (Ghiassian et al., 2015). The theoretical ground for DIAMOND is that in incomplete interactomes *"diseases cannot be associated with topologically dense network communities"*, rather, the statistical significance of an interaction, meaning the weight of the link, is the relevant quantity used to characterize such modules. This highlights the impact of the node/link ratio in the establishment of interacting structure and then in biological function. By extending the ideas of the DIAMOND/HuDiNe approaches it is possible to analyze the relationship between drug targets and disease-proteins through a topological *proximity measure*. This measure quantifies the interactions between drugs and disease-proteins in the human disease interactome (Guney et al., 2016) and can be used as a proxy for therapeutic effect. This can be useful for establishing a basis for drug screening and repositioning and evaluation strategies. Another approach to detect modularity in the interactome was based on identifying joint patterns of gene expression and drug response (Chen and Zhang, 2016). This was done to gain further insight into the biochemical mechanisms of drug action that may drive the development of new therapeutic targets in cancer. Interactome modularity has allowed *de novo* design of therapeutic strategies in cancer and also allowed the creation of methods for drug repositioning analysis (Chen et al., 2016). Such methods are aimed at detecting multi-targeted drug candidates that may disable malignant cellular functions.

Several methods have been proposed to analyze community structure in PPI networks. Feature selection by clustering has been applied to real and synthetic interaction data revealing modules with increased biological significance for *E. coli* and yeast networks (Henriques and Madeira, 2016). A similar approach was used in the `NCMine` method (Tadaka and Kinoshita, 2016) which is implemented as a plug-in for the popular network visualization and analysis suite `Cytoscape` (Adamcsek et al., 2006; Su et al., 2010; van Dongen and Abreu-Goodger, 2012) and is based on a technique called near-clique mining that distinguishes nodes in a network as either "core" or "peripheral" to a given subnetwork. Topological Data analysis (TDA) has also been used to detect topological network modules in protein interaction networks. TDA encompasses several statistical methods like clustering and perturbation analysis to find structure in data. By deleting protein complexes of the *S. cerevisiae* INO80 protein interaction network and performing TDA, isolated modules that contain proteins with shared biological functions were discovered to belong to the same module, even if they mapped to distinct locations of the network (Sardiu et al., 2017).

Clustering using genetic algorithms has been also applied with certain success (Ramadan et al., 2016). In brief, an objective function is built for exclusive clustering (nodes belonging to a unique module) and overlapping clustering (a particular node or set of nodes can be as indicated by spectral clustering methods, see section 4.2). This function is then optimized by a replication/mutation/recombination genetic algorithm in order to detect modular components of the network identified as protein complexes. One approach to detect such modularity in GRNs is through phylogenetic profiling. This approach is based on the idea that the joint presence or joint absence of two traits across various species is used to infer a meaningful biological connection, such as involvement of two different proteins in the same biological pathway.

As it was mentioned, sometimes approaches made use of hybrid methods, such is the case, for instance, of the work by Servis and Clark (2021) that perform a cluster identification strategy by using modularity optimization to analyze chemical heterogeneity in complex solutions. We will abound on modularity optimization in the next subsection.

4.2. Methods Based on *Modularity* Optimization

Unlike the methods based on similarity of data, most of the methods take into account the large-scale structure of the network itself, defined by the edges between nodes, regardless of the source of the data (Newman, 2012). Such as the case of the methods based on and supported by some class of Modularity optimization (see Definition 8).

In order to categorize different modularity measures, we must distinguish between *local* and *global* methods that quantify and assess network modularity. Measures of local modularity emphasize scoring specific clusters or partitions of the network. This score considers the number of modules that are dense or sparsely connected in a given assignment (Reichardt and Bornholdt, 2006). The more dense connections are within a module and the more sparse the connections are from within a module to outside vertices, the higher the modularity score will be. The local modularity of a network is usually given as the score of the highest-scoring partition. Finding the best partition and evaluating its score solves the modularity problem completely, but it relies on comprehensive enumeration of partitions, a problem that often carries computationally prohibitive combinatorial burdens (Fortunato, 2010).

The case of *global* modularity of a network is different in the sense that global measures usually are computed without *a priori* computing the network partitions. Instead, this measure relies on other network properties such as the *average clustering coefficient* $\langle CC \rangle$. The rationale is that vertices that form a module should have adjacent neighbors, as they increase the modular density and induce the formation of *triangles* in the graph.

An important family of local modularity measures is based on the concept of *edge-betweenness*, a concept introduced to generalize the node-associated betweenness centrality measure. Edge betweenness is then defined as the number of shortest paths between pairs of nodes that run along a given edge. The more paths traverse pairs of nodes traversed by an edge, the more *central* the edge is for the global connectivity structure of the network (Freeman, 1977). The first algorithm that used this concept was proposed by Girvan and Newman (Newman and Girvan, 2004) and is a paradigmatic example of the application of local modularity measures. The method consists in disconnecting sets of vertices by removing edges with larger betweenness. This algorithm was applied to several simulated networks as well as a number of real networks with an *a priori* known modular structure with good overall performance. More importantly, Newman and Girvan also provided a formal measure of network modularity.

DEFINITION 8. *Given a network modular partition we have the following:*

$$Q = \sum_i (e_{ii} - a_i^2) = Tr(\mathbb{E}) - ||\mathbb{E}^2|| \tag{4}$$

Here, e_{ij} is the matrix element –from the modularity matrix $\mathbb{E}$– whose entries are defined as the fraction of all the edges in the network that connects nodes in the i module to the nodes in the module j, $a_i = \sum_j e_{ij}$. Notice that, for an arbitrary matrix $\mathbb{X}$, a norm is defined as $||\mathbb{X}|| = \sum_i \sum_j x_{ij}$.

Q is called the *Girvan-Newman modularity* of a network partition, or sometimes just the *Modularity*. Q measures the fraction of edges in the network connecting vertices within the same module or *community* (or *intra-community edge* ratio) and then subtracts form this fraction its expected value in a network with the same partition scheme over randomly connected nodes. $Q = 0$ implies that the partition's modularity is not better than random, whereas $Q = 1$ is indicative of a strong modular structure.

Modularity can also be rewritten (Clauset et al., 2004) as:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(C_i, C_j) \tag{5}$$

Where m is the total number of edges in the network. k_i is the degree for node i. A_{ij} is the adjacency matrix. C is an indicator function such that $C_i = C_j$ implies that nodes i and j belong to the same community, δ is Kronecker's delta function. This way, if two nodes i and$j$ belong to the same community $\delta(C_i, C_j) = 1$, otherwise $\delta(C_i, C_j) = 0$.

There is yet another (equivalent) way to represent the modularity Q that may result even more useful in practice (Fortunato and Barthelemy, 2007; Porter et al., 2009):

$$Q = \sum_{s=1}^{M} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right] \tag{6}$$

The sum, over all M modules of the partition, l_s is the number of edges inside community s. L is the number of edges in the network and d_s is the total degree of nodes in module s.

These important ideas lead to the establishment of *Community Detection* as one of the foundational problems of Network Science (Newman and Girvan, 2003; Newman, 2004a; Kovács and Barabási, 2015). Maximization of modularity Q has been proposed as a central idea in several optimal network partition algorithms (Clauset et al., 2004; Newman, 2004b, 2006b). However, modularity optimization, also known as Q_{max} algorithms, are constrained by a resolution limit that depends on the overall size of the network and on the interconnection density of the modules, which may lead to failure of Q_{max} methods due to sub-optimal optimization caused by the presence of a multitude of local minima on the modularity function (Fortunato and Barthelemy, 2007).

A related issue with respect to large networks is that calculating the modularity score Q (see Equation 6) belongs to the family of `NP-Hard` or non-deterministic polynomial-time problems. The main characteristic of these problems is that they cannot be solved in polynomial-time, so they are computationally and time consuming, precluding its direct use on extremely large networks. Several heuristic approaches have been proposed to deal with this problem (Danon et al., 2005; Duch and Arenas, 2005; Guimera and Amaral, 2005; Newman, 2006b; Von Luxburg, 2007; Brandes et al., 2008). One particularly useful technique is known as the Louvain method (Blondel et al., 2008). This approach is based on a two-step heuristic: (1) a maximal modularity full partition is obtained by merging nodes in order to maximize modularity through a greedy method, (2) then a network is formed in which nodes are the modules from the first step. This stage is continued recursively until no further improvement in modularity can be obtained.

A whole new family of methods was developed after the introduction of the modularity measure Q. Most of these methods aimed to maximize either Q itself or some proper function of Q under the rationale that if one is able to find a partition that maximizes Q, the induced community structure would be optimal. In this family we can find the original works by Newman (2004b) as well as later refinements of his method, either by himself (Clauset et al., 2004; Newman, 2006b) or by others (Guimera et al., 2004; Duch and Arenas, 2005; Blondel et al., 2008; De Leo et al., 2013). However, since maximization of the Q-measure has a resolution limit that depends on the size of the network and the degree of interconnection between the modules, the method is not fail-safe (Fortunato and Barthelemy, 2007; Lancichinetti and Fortunato, 2011). Some recent implementations, however, have been developed to improve the results obtained under Q-optimization as is the case of the works by Medus and Dorso (2009), Khadivi et al. (2011), Gong et al. (2011), and (Bettinelli et al., 2012).

4.3. Spectral Graph Theory

Another family of algorithms is based on *Spectral graph theory*, which uses the analysis of the eigenvalues of the *adjacency matrix* or the *Laplacian matrix* of a graph. It consists in a transformation of the set of nodes into a set of points in a space whose coordinates are elements of eigenvectors, then the set of points can be clustered via standard techniques (Fortunato, 2010). The change of representation induced by the eigenvectors makes the cluster properties much more evident (Donath and Hoffman, 1972; Fiedler, 1973).

The analysis of the spectrum of the **Laplacian matrix** $\mathbb{L}$, is the most used approach in spectral clustering. This matrix can be derived from the adjacency matrix $\mathbb{A}$ of a network and it is constructed by reversing the signs of the non-diagonal entries and replacing the diagonal entries with the degree of the corresponding node (See **Figure 1**).

The Laplacian matrix can be written in block-diagonal form, that is, the nodes can be ordered in such a way that the Laplacian displays k square blocks along the diagonal, with some entries different from zero, and all other elements vanish. Each block is the Laplacian of the corresponding subgraph, so it has the trivial eigenvector $\vec{1}$ with components $(1, 1, 1, ..., 1, 1)$. Therefore, there are k degenerate eigenvectors with equal non-vanishing components in correspondence with the nodes of a block, whereas all other components are zero. In this way, from the components of the eigenvectors, it is possible to identify the connected components of the graph, and then based on this property, it is possible to find highly connected groups of nodes and the expected number of modules in which the network may be partitioned.

Since the values of the eigenvector components are close for nodes in the same community, it is possible to use them as coordinates, such that vertices turn into points in a metric space. So, for M eigenvectors, the nodes can embed in an M-dimensional space. Thus, modules appear as groups of points well-separated from each other (Donetti and Muñoz, 2004). Also, it is possible to use the Laplacian matrix property, in which, if the graph has g connected components, the largest g eigenvalues are equal to 1, with eigenvectors characterized by having equal-valued components for nodes belonging to the same component. Thus, the modules can be found by inspecting the components of the eigenvectors with eigenvalue 1 (Capocci et al., 2005).

Furthermore, in the context of *Spectral clustering*, there is a remarkable relationship introduced by Newman (Newman, 2006b), between *Modularity optimization* and the spectral properties of the *adjacency matrix* known as *Spectral optimization*. We can rewrite the Q optimization in terms of finding the spectrum of a particular matrix as we will see below.

Starting from Equation (5), it is possible to define the *modularity matrix* B_{ij} as:

$$\mathbb{B} = B_{ij} = \left(A_{ij} - \frac{k_i k_j}{2m} \right)$$

Now, let us suppose a particular *a partition* of a network into just **two** modules. Thus we can assign to each node, a quantity s_i, such as:

$$s_i = \begin{cases} +1, & \text{if a node } i \text{ belongs to group 1} \\ -1, & \text{if vertex } i \text{ belongs to group 2} \end{cases}$$

Thus, Q can conveniently be written in matrix form:

$$Q = \frac{1}{4m} \sum_{ij} B_{ij} s_i s_j = \frac{1}{4m} \vec{s}^T \mathbb{B} \vec{s} \tag{7}$$

where $\vec{s}$ is a column vector whose elements are s_i.

Then, in order to optimize this form of Q it is possible to perform the so-called *relaxation method* (that is, allowing its entries to take continuous values and retaining the norm of the vector), which is one of the standard methods for the approximate solution of vector optimization problems such as this one. Thus, by differentiating and imposing the constraint $|s| = \sqrt{n}$ or equivalently:

$$\sum_i s_i^2 = n$$

The modularity maximization problem is now straightforward. We now have a maximization problem with this norm as a constraint, or equivalently, $(n - \sum_i s_i^2) = 0$. This is done by introducing a *Lagrange multiplier* λ, and taking the partial derivative with respect to the components of the vector (one at a time) of the following expression:

$$\frac{\partial}{\partial s_{i=k, j=k}} \left[\sum_i \sum_j B_{ij} s_i s_j + \lambda \left(n - \sum_i s_i^2 \right) \right] = 0 \tag{8}$$

to obtain:

$$\left[\sum_i B_{ik} s_i + \sum_j B_{kj} s_j - 2\lambda s_k \right] = 0 \tag{9}$$

which leads to:

$$\sum_j B_{kj} s_j - \lambda s_k = 0$$

$$\sum_j B_{kj} s_j = \lambda s_k$$

for all k.

Which is in a matrix form an eigenvalue problem for the *modularity matrix*:

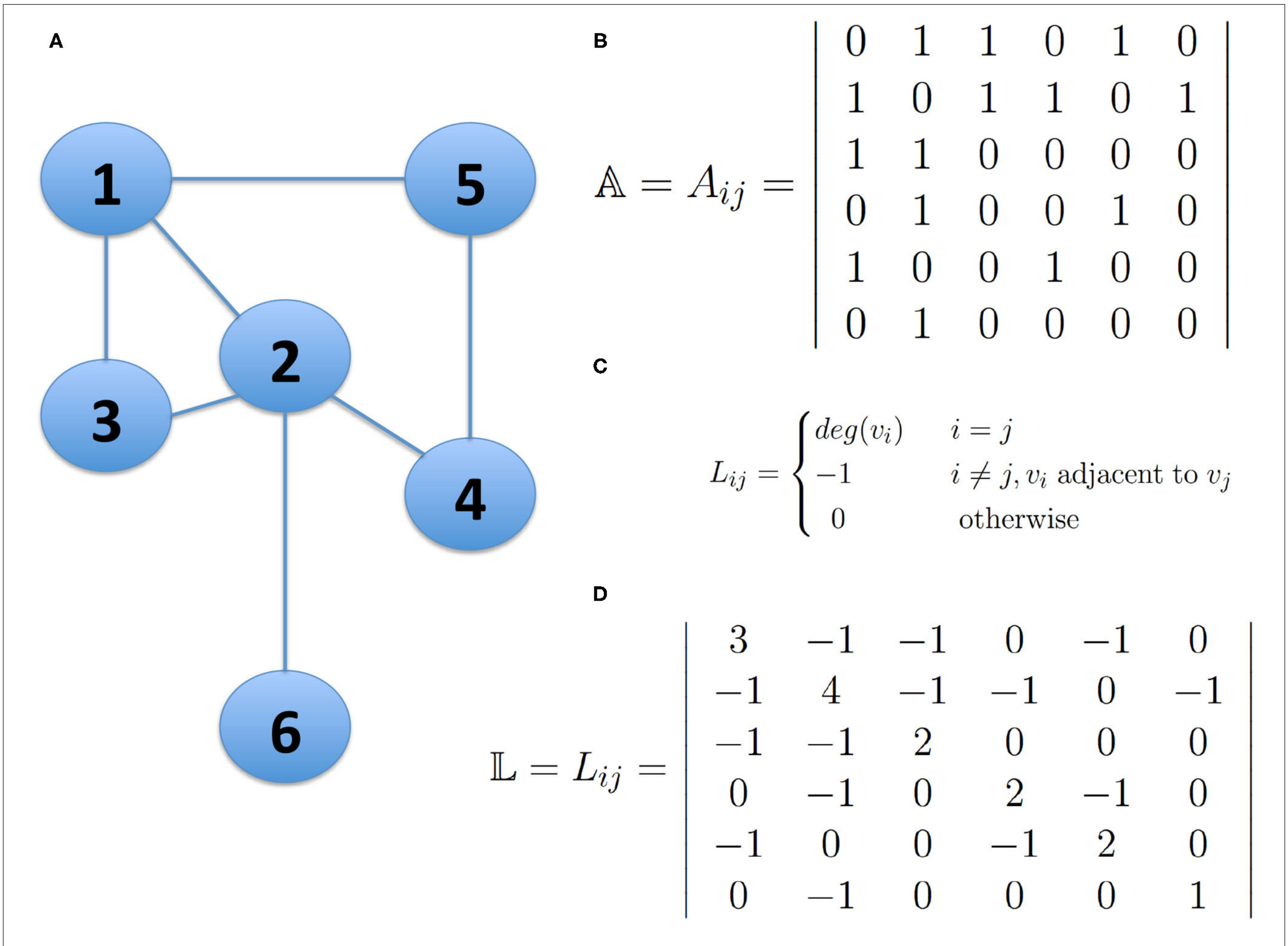
$$\mathbb{B} \vec{s} = \lambda \vec{s} \tag{10}$$

The value of λ that maximizes Q is the largest possible one, that is the dominant eigenvalue of the matrix $\mathbb{B}$.

It is worth mentioning, that similarly to this approach, the **spectral bisection method** (Barnes, 1982), uses the spectrum of the Laplacian matrix, to find partitions of a graph by dividing it recursively into two groups. Every partition of a graph with n nodes in two groups can be represented by an index vector $\vec{s}$, whose component s_i is $+1$ if a node i is in one group and $a1$ if it is in the other group. Then the cut size R of the partition of the graph in the two groups can be written as:

$$R = \frac{1}{4} \vec{s}^T \mathbb{L} \vec{s} \tag{11}$$

Finally, the *Modularity optimization* approach can be extended to a more than two modules, by writing an additional contribution



$$\mathbb{A} = A_{ij} = \begin{vmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$L_{ij} = \begin{cases} deg(v_i) & i = j \\ -1 & i \neq j, v_i \text{ adjacent to } v_j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{L} = L_{ij} = \begin{vmatrix} 3 & -1 & -1 & 0 & -1 & 0 \\ -1 & 4 & -1 & -1 & 0 & -1 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 2 & -1 & 0 \\ -1 & 0 & 0 & -1 & 2 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \end{vmatrix}$$

FIGURE 1 | The Laplacian Matrix of a network. Panel **(A)** presents a small undirected network; Panel **(B)** shows the Adjacency Matrix $\mathbb{A}$ describing the network connectivity of the network in **(A)**; Panel **(C)** shows the definition of the Laplacian Matrix of a Network and panel **(D)** shows the Laplacian Matrix $\mathbb{L}$ of the network in **(A)**. The bold numbers represent the degree of node i, whenever i=j. This figure is intended for illustrative purposes, no *actual results* are presented.

ΔQ to the modularity upon further dividing a group g of size n_g in two as:

$$\Delta Q = \frac{1}{4m} \sum_{i,j \in g} \left[B_{ij} - \delta_{ij} \sum_{k \in g} B_{ik} \right] s_i s_j \tag{12}$$

$$\Delta Q = \frac{1}{4m} \vec{s}^T \mathbb{B}^{(g)} \vec{s} \tag{13}$$

where δ_{ij} is Kronecker's δ, and $\mathbb{B}^{(g)}$ is the $n_g \times n_g$ matrix with elements indexed by the labels i, j of nodes within group g. Because Equation (13) has the same form as Equation (7) it is possible to apply the spectral approach to this generalized *modularity matrix*, just as before, to maximize ΔQ.

In addition, the *modularity matrix* $\mathbb{B}$ also has always the trivial eigenvector $\vec{1}$ with eigenvalue zero (like the *Laplacian matrix*), because the sum of the elements of each row/column of the matrix vanishes. Thus, it is also possible to optimize modularity on bipartitions via *spectral bisection*, by replacing the Laplacian matrix with the modularity matrix (Newman, 2006a,b).

4.4. Random Walk Based Models

The use of random walks to find modules on a network is based on the somehow intuitive premise that a random walker moving on the network will spent more time inside modules—due to the high density of edges, thus many possible trajectories—than hoping from one module to another. A first approach to this problem was addressed by Zhou (2003) who used random walks to define a *distance* between pairs of nodes, assuming that there is a high likelihood that *closer* nodes—under this measure of distance—belong to the same module. Such distance was used to define global and local *attractor nodes* used to detect modules, i.e., minimal distance subnetworks. A different but related approach was taken by Pons and Latapy (2006) on a method called *Walktrap*. Here, distance is calculated via the probability that a random walk moves from one module to another on a fixed number of steps, then grouping nodes via hierarchical clustering.

A method based on the application of the Markov property of node-to-node walks called Markov Cluster algorithm (MCL) was developed by Van Dongen (2001). MCL simulates a diffusive process in the network. A *stochastic matrix* is obtained by dividing every entry of the adjacency matrix A_{ij} by the corresponding degree of node i. This stochastic matrix is used to calculate transition probabilities on a Markov random field. This method is quite elegant and comparatively easy to implement, however, its large computational complexity makes it difficult to apply in practice for real (large) networks (even in sparse cases).

As already mentioned, for large sparse networks also the standard versions of spectral based algorithms are suboptimal, in the sense that in some cases these fail to detect communities even when other algorithms such as belief propagation can do so. Efforts to improve these spectral theory methods have been made by resorting again to random walk dynamics, mainly through implementing non-backtracking random walks (the random walker cannot move backwards) over the network (Krzakala et al., 2013; Newman, 2013; Zhang and Newman, 2015). Other methods in the literature are built on ideas borrowed from non-linear dynamic processes, such as spin-coupling models with nearest neighbor interaction (Reichardt and Bornholdt, 2004), synchronized oscillators (Arenas et al., 2006; Arenas and Diaz-Guilera, 2007), as well as generalized random walks (Van Dongen, 2001; Zhou and Lipowsky, 2004; Pons and Latapy, 2005). Among this plethora of models, INFOMAP has been shown to be quite reliable and computationally efficient (Rosvall and Bergstrom, 2007, 2008).

The INFOMAP algorithm is founded on a clever combination of random walk dynamics and information theory. The main idea is to reach optimal compression of the information needed to describe the diffusion process of a set of random walkers. This is achieved by using the random walk *itself* as a proxy for the diffusion process via a sequential enumeration algorithm and the use of tools of information theory and computational linguistics.

In a nutshell, the approach is quite similar to the way we imprint location information on geographic maps of cities: you can map a large number of close-to-each-other streets into a neighborhood ("a module," with its own description) and a series of close-by neighborhoods into a town. The larger the scale of these *urban modules*, the smaller the total amount of information needed for their description. In a similar way, the INFOMAP algorithm looks up for the *minimal description length* for the modular partition of a network. The best partition is the one that can be described with the minimal information.

In brief, the *description length* is a measure of the complexity of a given process. By using the description length is possible to characterize the trajectory of a random walk (or the trajectories for an ensemble of random walkers), in the form of the *map equation*:

$$L(M) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} q_{\curvearrowright} H(\mathcal{P}_i) \tag{14}$$

Here, $L(M)$ is the description length of an ensemble of random walkers moving through a given modular partition M. The first term $q_{\curvearrowright} H(\mathcal{Q})$ represents the average number of bits needed to describe the movements from nodes in one module of the partition to nodes in another module, whereas the second term represents the information for the intramodule walks. Since by the coding theorem (Knuth, 1985), the information needed to characterize inside module walks is smaller, a minimal description length implies that most of the time walkers move inside modules of a given partition, thus optimizing modularity, allowing however for the presence of a number of intermodule hops. This method uses a *greedy* algorithm, so it can be applied quite efficiently even to large networks, directed or undirected. There are also INFOMAP implementations to find hierarchic modular structure (Rosvall and Bergstrom, 2011) and overlapped modules (Esquivel and Rosvall, 2011).

4.5. Stochastic Block Models

Statistical inference provides a powerful set of methodological tools useful in modularity detection. The usual way to proceed is by adjusting a *generative network model* to the experimental data. A stochastic block model (SBM) is by far, the most used model to generate networks with a modular structure. The essentials of the SBM are as follows:

The stochastic block model generates a number n of vertices of the network; the algorithm makes a partition of the vertex set $\{1, \ldots, n\}\{1, \ldots, n\}$ into q disjoint subsets $C_1, \ldots, C_q$ i.e., the modules. By starting with a symmetric $q \times q$ matrix P containing edge probabilities for all the possible connections. These probabilities must be known a priori. Then the SBM is generated by randomly sampling this edge set as follows: any two vertices $u \in C_i$ and $v \in C_j$ are connected by an edge with probability P_{ij}.

Modularity detection works out by optimizing the unnormalized log-likelihood that a given partition g of a graph G in q modules will be reproduced by the SBM (Karrer and Newman, 2011).

$$\mathcal{L}(G|g) = \sum_{i,j=1}^{q} e_{ij} \log\left(\frac{e_{ij}}{n_i n_j}\right) \tag{15}$$

Here $\mathcal{L}(G|g)$ is the log-likelihood for a partition g of a given network G to be produced by the standard SBM. e_{ij} is the number of edges connecting module i with module j of the partition, and n_i, n_j are the number of nodes in modules i and j respectively. The sum includes the case $i = j$. The strongest drawback of the method is that it requires *a priori* knowledge of the number q of modules in which the network has to be partitioned, although this limitation has been recently overcome by using a Bayesian formulation (Peixoto, 2018).

General SBM models (i.e., non-Bayesian) have been demonstrated to be formally equivalent to modularity optimization approaches that do not usually require a fixed number of modules for the partition (Newman, 2013). Despite this and the fact that maximum likelihood exact estimation is an *NP* problem—so all solutions are approximate—SBM models are still popular in statistics and machine learning algorithms.

As we have discussed in this section, topology based methods for modularity detection are robust, general and intelligible. They can also be benchmarked with experimentally available modular partitions. Such validation uses robust statistics, such as the ones given by normalized mutual information measures. The strength of these methods is that they do not rely *a priori* on any non-topological information, as they are based on the (weighted or un-weighted, directed or un-directed) connectivity as given by adjacency matrices. This is the basis of their generality and broad applicability, in particular to complex biological problems.

The fact that these methods do not need any prior knowledge—aside from the connectivity structure—does not preclude us to incorporate such information when available, to enhance our intuition and empower our predictions when applied to real large scale biological networks. For this reason we strongly believe that the popularization of these approaches within the computational and systems biology research settings will prove to be highly beneficial for both, the construction of more general approaches to study modularity in biology and for the further development of analytic methodologies in the theory of complex networks.

5. BENCHMARKING AND PERFORMANCE TESTS

Whenever several methods perform a similar task, benchmarking becomes necessary. However, as described in Tripathi et al. (2016), a large heterogeneity among different community structure discovery methods is often found. As many of the available methods for module discovery have been developed as *ad-hoc* solutions, they often lack reliability when applied to other biological systems. Also, the intrinsic complexity of biological modularity makes it hard for a single method to describe all types of modules correctly. Nevertheless, in the following section we will show how by resorting to theoretically sound and rigorous methods of comparison that do not rely on the specifics of a given biological system, one can attain precise measurements of performance for any module detection method.

5.1. Testing Performance and Scoring Measurements

Benchmarking community detection algorithms using real biological networks is not optimal, as it is not clear what the ideal partition is. However, real networks such as the social network of bottle-nose dolphins from Doubtful Sound (New Zealand) built and studied by Lusseau (2007), as well as the network of college football teams obtained by Girvan and Newman (Girvan and Newman, 2002) have been used for this purpose. Real biological network communities (also called *ground-truth communities*) are often inferred from non-topological studies carried out by network curators, which based on experimental observations (e.g., protein-protein interactions) define the network itself. As these methods rely only on observed data, it is possible that the resulting network is either incomplete or has spurious interactions. So how can one find these modules and relate them to particular functionalities, especially when such functionalities are unknown? One general approach is to use random network methods to test if the community or modular structure in our networks is valid and significant (Sah et al., 2014). One common approach consists in generating network models that satisfy the constraints imposed by the real networks (such as the connectivity, the number of nodes, etc.) and keep a graph structure that is as random as possible. These network realizations allow the use of a large set of tools already available to analyze the topology of random networks. In particular, they are useful for creating *null-models* that serve as a baseline to which we can compare the significance of our partition model. As such null models have been established, they can be used to test biological functional hypotheses. This generation of null models serves directly to generate scoring metrics that allow the comparison and selection of the best network partitions. These null-model networks may be generated synthetically, and this way we could test to what extent the algorithm is able to found the a-priori known communities.

There are two classic and widely used performance tests for community detection algorithms: the GN and the LFR (Fortunato, 2010), both of which belong to a class of methods generated under the *planted l-partition model* (Condon and Karp, 2001).

DEFINITION 9. *In the* ***planted l-partition model*** *a network with $n = g \cdot l$ nodes, is partitioned into l groups of g nodes each. Nodes in the same group are linked with a fixed probability p_{in}, whereas nodes in different groups are linked with probability p_{out}. Each module is then a random Erdös-Rényi network with $p = p_{in}$ and if every module were a node, the whole network would also be an Erdös-Rényi graph with $p = p_{out}$.*

For a subgraph representing a module or community C, the average connectivity degree will be given as $\langle k \rangle_{in} = p_{in}(g-1)$ and the average external degree would be $\langle k \rangle_{out} = g \cdot p_{out}(l-1)$ (recall that for an Erdös-Rényi graph connected with probability p, the average degree is given as $\langle k \rangle = p(n-1)$). If these conditions hold, the average degree for the whole network is

$$\langle k \rangle = p_{in}(g-1) + g \cdot p_{out}(l-1) \tag{16}$$

This way, if $\langle k \rangle_{in} > \langle k \rangle_{out}$ (i.e., if the intra-module average degree is greater than the inter-module average degree), then the network will have well-defined community structure. This is equivalent to the intuitive definition of modularity, namely $p_{in} > p_{out}$.

The GN test was designed by Girvan and Newman (Girvan and Newman, 2002) to test their community detection algorithm. It is a particular case of the *planted l-partition model* where the authors fixed $l = 4$ and $g = 32$ to get a network composed of 128 nodes forming 4 modules with 32 nodes each and an average degree of $\langle k \rangle = 16$. Within this framework link-density is adjusted by scanning the values of the average in-degree $\langle k \rangle_{in}$ and out-degree $\langle k \rangle_{out}$ to choose specific values to change the community structure for each network provided that $\langle k \rangle = \langle k \rangle_{in} + \langle k \rangle_{out} = 16$.

Under this model it is possible to have explicit expressions for the average in- and out- degrees, namely: $\langle k \rangle_{in} = p_{in}(g-1) = 31\,p_{in}$ and $\langle k \rangle_{out} = g \cdot p_{out}(l-1) = 96\,p_{out}$. By varying the values of p_{in} and p_{out} it is then possible to simulate networks with a stronger or weaker modularity. For instance, a clearly defined community structure is induced if $p_{in} \simeq 0.5$ or larger, whereas a value of $p_{in} \simeq 0.25$ or lesser precludes the existence of well-defined modules.

For this benchmark communities are well-defined for $\langle k\rangle_{in} > 8$. One of the advantages of the GN test is that by varying a single parameter in a pretty simple network it is possible to contrast different network partition methods. In order to test a particular method via the GN test one has to calculate a *similarity measure* between the partition of the GN network as given by this method against the natural partition of the network in four modules of the same size. A highly used similarity measure—proposed by Newman and Girvan (Girvan and Newman, 2002)—is the fraction of edges correctly classified, though a more objective measure can be the normalized mutual information between partitions (see Equation 17) (Arenas et al., 2008).

In spite of its simplicity and mathematical rigor, the GN test presents a couple of important shortcomings derived from unrealistic assumptions. First, all the nodes are expected to have the same degree. Second, all the communities must be of the same size. Clearly real complex networks, such as those encountered in biology, are characterized by long-tailed degree distributions or power law-like ones, and also by heterogeneous community sizes. Some improved versions of the GN method have been developed such as the one presented in Fan et al. (2007) where different weights are assigned to *inner* and *outer* edges, regarding their position in the communities.

The fact that the planted l-partition model generates mutually-interconnected Erdös-Renyi random graphs implies that all the nodes will have almost the same degree and all the communities will have exactly the same size. Of course, these two features do not match with what is observed in real networks. To tackle this problem, Lancichinetti et al. proposed the *LFR Benchmark test* (Lancichinetti et al., 2008). The LFR test assumes that the node degree distribution and the module size distribution follow a—more realistic—power law behavior. Each node shares a fraction $1-\mu$ of its edges with nodes within its community and a fraction μ with nodes in other communities. Hence $0 \leq \mu \leq 1$ the mixing parameter is equivalent to a normalized version of the $\langle k\rangle_{out}$ used in the GN test. The LFR test was devised for undirected, unweighted networks, but there are implementations for directed, weighted graphs including the possibility to have overlapping communities (Lancichinetti and Fortunato, 2009a). Aside from purely computational costs, the main performance test for network community detection algorithms must establish a clear criterion to compare the degree of *similarity* between the modules discovered (i.e., the specific partition) by an algorithm and the real (in the test, a priori known) partition. There are several proposals in the complex network literature as how to measure similarity between different partitions (Meilă, 2007), some of them based on pair recounting and group coincidence counts (Fortunato, 2010).

Additionally, two widely used measures are the fraction of correctly classified edges and the normalized mutual information between partitions. The former was proposed by Girvan and Newman to test their algorithm, but can be generalized to other benchmark tests. The criteria for the correct classification is as follows: Each of the modules A_i of the partition found by the given algorithm is compared to all of the *actual* modules B_i, known a priori from the real network partition. When more than half of the nodes in one of these A_i correspond to those of a community B_i then A_i is considered to be correctly classified and no more comparisons between A_i and the rest of the B_is are carried out. In the contrary case (less than half corresponding nodes) or when the community A_i is smaller than half the size of the given B_i, then the module is compared to the rest of the B_i's until exhaustion. This criterion is quite stringent since there are cases in which one may consider that some of the nodes have been correctly classified by the algorithm but the measure (total node count divided by the size of the network to give a number between 0 and 1) rules them out.

DEFINITION 10. *The* ***normalized mutual information between partitions*** *(NMIBP) was proposed by Danon et al. as a similarity measure (Danon et al., 2005) built on ideas proposed by Ana and Jain (2003), Kuncheva and Hadjitodorov (2004).*

The rationale is that if two partitions are similar, very little information is needed to infer one partition given the other. One is able to calculate the mutual information between two partitions A and B by building a confusion matrix $\mathbb{N}$ where rows correspond to the actual modules and columns correspond to the modules found by the given algorithm. The N_{ij}-th element of $\mathbb{N}$ is the number of nodes in a real (known a priori) community i that are also present in the community j detected by the algorithm. Since the partitions under comparison may have a different number of groups (the modules or communities), $\mathbb{N}$ is not necessarily a square matrix. This way the similarity between two partitions A and B is given by the normalized mutual information measure (NMI) as follows:

$$NMI(A,B) = \frac{-2\sum_{i=1}^{C_A}\sum_{j=1}^{C_B} N_{ij}\, log\left(\frac{N_{ij}N}{N_{i.}N_{.j}}\right)}{\sum_{i=1}^{C_A} N_{i.}\, log\left(\frac{N_{i.}}{N}\right) + \sum_{j=1}^{C_B} N_{.j}\, log\left(\frac{N_{.j}}{N}\right)} \qquad (17)$$

Here, the number of actual modules (partition A) is denoted by C_A, the number of modules found by the algorithm (partition B) is C_B, the sum over the row i of the matrix $\mathbb{N} = N_{ij}$ is $N_{i.}$ and the sum over column j is $N_{.j}$ and N is the total number of nodes. If the partitions A and B are identical, then $NMI(A,B) = 1$, whereas completely dissimilar partitions give $NMI(A,B) = 0$.

This measure is highly used in the performance tests for community detection algorithms since it is highly sensitive as it quantifies explicitly the amount of information recovered by the algorithm from the original topological structure of the network (Lancichinetti and Fortunato, 2009b; Lancichinetti et al., 2011; Tripathi et al., 2016). The NMIBP measure can be used in the GN and LFR performance tests, both in standard and overlapping partitions (Lancichinetti and Fortunato, 2009a).

More recently there have been some other approaches that propose new benchmarks that provide actual techniques to determine which is the most suited algorithm in most circumstances based on observable properties of the network under consideration. Also considering the use of the mixing parameter μ and the Normalized Mutual Information measure (NMI) (Yang et al., 2016). There are also benchmarks based on novel methods that generate networks with topological properties found in empirical biological networks (Sah et al., 2014; Gilbert, 2015).

Despite the high performance of algorithms and methods shown on the artificial networks generated by benchmarks and its test with the μ (mixing factor), for example on the LFR test, an open question is, whether the methods with good results on benchmarks necessarily find meaningful modules in actual networks (Jebabli et al., 2018; Cherifi et al., 2019).

It may happen that the community structure found by some methods with high performance in benchmarks, does not necessarily correspond to correct ground-truth community structure—that is, the one based on real known node groups, or derived from some metadata or even identified by the node attributes—and vice versa. There could be a substantial difference between structural communities and metadata groups (Orman et al., 2012; Hric et al., 2014; Jebabli et al., 2018).

So, for a fair assessment of the performance of some methods, it is necessary to have a good match between the detected partition and the attribute-based partitioning for considering that a method is reliable. Both tests are complementary, and we recommend applying both of them to perform a complete and accurate assessment of an actual community structure.

Nonetheless, to overcome these limitations, exploiting the topological features of the so-called "*community graphs*" (where the nodes are the communities and the links represent their interactions) has been proposed to evaluate the algorithms; in contrast with metrics defined at node level that are fairly insensitive to the variation of the overall community structure. Thus, if the ground-truth community structure is available, it is possible to compare it vs. the one discovered by these algorithms by using these clustering-based metrics as has been proposed by some authors (Orman et al., 2012; Hric et al., 2014; Jebabli et al., 2018; Cherifi et al., 2019), where more emphasis has been put on the topology of the community structure.

In this direction, some modifications to the LFR benchmarks have been proposed to make generated networks more realistic (Orman et al., 2012). In this work, authors studied generated networks in terms of community-centered topological properties to evaluate some methods, they used such properties to compare community structures to rank the tested community detection algorithms. As well, recently da Fonseca Vieira et al. (2020) tested some representative state-of-the-art methods for overlapping community detection (Cherifi et al., 2019) with synthetic and real-world benchmark *Ground-Truth networks* showing that, although the methods can identify modular communities, they often miss many structural properties of the communities.

5.2. Good Performance Methods Commonly Applied to Biological Networks

Beyond presenting the benchmarking for the performance of the different algorithms, it is important to point out which methods we think are good for finding modules, given the biological question under consideration. The question of which algorithm is the best for biological networks is not easy to answer, it will depend on the context of the research question and the data on which the network is built.

However, two of these graph-theoretically-grounded, general purpose algorithms have been widely applied in biological networks with good and significant results, such methods are the **Louvain** (Blondel et al., 2008) and **Infomap** (Rosvall and Bergstrom, 2008). Both methods have good performance and accuracy scores, as we can see from the several artificial network bencharmking analyses (Lancichinetti et al., 2008, 2009; Lancichinetti and Fortunato, 2009a; Sah et al., 2014; Gilbert, 2015; Yang et al., 2016), as well as in *Ground-Truth networks* and also in terms of *community-centered topological properties* (Orman et al., 2012; Hric et al., 2014; Jebabli et al., 2018). In addition, both methods show good results and performance in biological networks, even in comparison with more recent methods (Mall et al., 2017b; Debnath et al., 2021). Furthermore, they also have been proved as standard methods to identify biologically meaningful modules in biological networks (Zheng et al., 2021) and even for evaluating significant topological differences between networks (Mall et al., 2017a). In addition, they have been incorporated on different Bioinformatic analysis suites and tools, as well as implemented in different programming languages widely used today, such as R, Python, MatLab, and C++ and incorporated into standard widely network analysis libraries such as `igraph`.

The *Louvain method* (Blondel et al., 2008) is by far the most widely used method in biological networks, showing significant results and meaningful modules (Praneenararat et al., 2011) even compared with newer methods in recent studies (Şen et al., 2014; Bennett et al., 2015; Rahiminejad et al., 2019; Calderer and Kuijjer, 2021). The method is indeed still widely used nowadays, for example, in the context of SARS-COV-2 analyses (Zheng et al., 2020). The efficiency and high performance of this method lie on its taking into account the whole structure of the network and searching for the best partition in an algorithmic greedy fashion. In addition, this method has been extended and applied to bipartite biological networks (Pesantez-Cabrera and Kalyanaraman, 2016; Calderer and Kuijjer, 2021) as well as to multilayer and multiplex biological networks (Mucha et al., 2010; Didier et al., 2015; Mittal and Bhatia, 2018).

On the other hand, *Infomap* is accepted as a very well-known method in module detection (Acharya et al., 2012) and even as a method for comparing the performance and accuracy of novel methods in biological networks (Lecca and Re, 2015), and has been incorporated in some bioinformatic layouts as a standard community detection framework (Aldecoa and Marín, 2014; Zhou and Xia, 2018; Farage et al., 2021). Moreover, has been widely adapted and extended by its authors in several ways to different kinds of networks and problems in community detection, for example, hierarchical module detection (Rosvall and Bergstrom, 2011), bipartite networks (Kheirkhahzadeh et al., 2016) and multilayer networks (De Domenico et al., 2015). In addition, these extensions have proved to give meaningful results in the context of biological networks as ecological networks (Pilosof et al., 2020; Farage et al., 2021), multiplex genetic datasets

(Mittal and Bhatia, 2018) and breast cancer networks (Alcalá-Corona et al., 2018a). The efficiency and high performance of Infomap lie in how information flow in a network can reveal the structure of it (Esquivel and Rosvall, 2011; Aslak et al., 2018; Eriksson et al., 2021), combined with a strategy of optimizing partitions such as the *Louvain method*, which make it one of the most robust and applicable methods for all kinds of networks and giving meaningful results (Kawamoto and Rosvall, 2015; Emmons and Mucha, 2019).

Finally, it is worth mentioning that other three methods have been demonstrated to be efficient and reliable in the context of biological networks in comparison with Infomap and Louvain: the **Spinglass Method** (Reichardt and Bornholdt, 2004, 2006), **OSLOM** (Lancichinetti et al., 2011), and **Label Propagation approach** (Garza and Schaeffer, 2019).

Thus, we can suggest **as a general strategy for community detection in biological networks to apply both Louvain and Infomap, in addition to one of these three latter methods and then consensing the partition by the Consensus Clustering approach** (Lancichinetti and Fortunato, 2012) to compute a unique community structure.

6. APPLICATION EXAMPLE: COMMUNITY DETECTION METHODS FOR CANCER NETWORKS

Network approaches have been extensively used for instance, to observe structural differences between cancer and non-cancer related networks (Reyna et al., 2020; Wang et al., 2020). These differences, often carry functional features that may help to understand such complex phenotypes (Miecznikowski et al., 2016; Drago-García et al., 2017; de Anda-Jáuregui et al., 2019; Dorantes-Gilardi et al., 2020).

Finding functional modules in cancer has been a matter of intense research. A common method to infer such modules resorts to the so-called *Weighted gene co-expression network analysis (WGCNA)* (Zhang and Horvath, 2005; Langfelder et al., 2008). In this method, Pearson correlation is used to evaluate pairwise gene co-expression. Such co-expression network can be decomposed into modules by using different methods.

For instance, in Ai et al. (2020), the authors used the dynamic tree cut method (Langfelder and Horvath, 2008) to infer modules in a microarray-based colorectal cancer (CRC) gene co-expression network. This method improves the classic hierarchical clustering that sets a fixed cutoff value. A dynamic branch cutting depending on the dendrogram shape is implemented. With this approach, Ai and cols., found that GUCA2A, GUCA2B, and CDH3 genes were highly correlated with the occurrence of CRC.

Along similar lines, WGCNA was used to analyze 182 CRC and 54 normal samples (Qiu et al., 2020). There, a k-means clustering was used to find modules, and the hub genes from those modules were separated into samples with high and low expression. The authors identified that overexpression of MYL9, MYLK, and CNN1 genes was associated with poorer outcome in CRC patients.

In breast cancer, efforts have been made to observe modules that may be underlying functional processes (Wilkinson and Huberman, 2004; Zhu et al., 2008; Cantini et al., 2015). It is widely known that breast cancer is a highly heterogeneous disease. This heterogeneity can be traced down to the genetic level (Alcalá-Corona et al., 2017).

Molecular subtyping provides a helpful tool to classify tumors by identifying common patterns in their genetic expression. One of the most used classification methods is PAM50 (Sørlie et al., 2001). Samples are grouped based on the molecular signature. With this method, breast cancer can be divided into four main differentiated subtypes: Luminal A, Luminal B, HER2+, and Basal-like. Each subtype has a different clinical and histopathological manifestation.

Network approaches to identify modules in breast cancer molecular subtypes has been a matter of intense research. For instance, the `infomap` algorithm has been used to reveal functional modules in HER2+ breast cancer transcriptional network (Alcalá-Corona et al., 2018b). Additionally, it has been observed that in the HER2+ tumors related network, a hierarchical modular structure appears (Alcalá-Corona et al., 2018a).

In basal-like breast cancer, network modularity has been used to observe functional modules and discern whether or not those modules are shared between the cancer and the non-cancer network (de Anda-Jáuregui et al., 2019). It has been observed that the basal breast cancer has a different distribution of module size between cancer and non-cancer networks (de Anda-Jáuregui et al., 2019). Additionally, those modules are composed of different genes.

In all those cases, cancer networks are formed by small connected same-chromosome gene components. Often, said components coincide with modules independent of the community detection method. However, this is not always the case. For example, in García-Cortés et al. (2021), for Luminal A breast cancer, an RNA-Seq-derived gene co-expression network was decomposed into communities by using four different methods: Fast greedy (Clauset et al., 2004), Infomap (Rosvall and Bergstrom, 2008), Leading eigenvector (Newman, 2006b) and Louvain (Blondel et al., 2008).

The aforementioned methods have different postulates and different approaches to detect communities. In that work (García-Cortés et al., 2021) it was demonstrated that, independent of the algorithm used to detect communities, the results were very similar in terms of the number of detected communities and the nature of the genes observed in each community.

Despite modules being quite similar, independently of the method to detect them (Jaccard indexes between modules obtained by the different methods, are larger than 0.95), the algorithm with optimal modularity was the Louvain method. Interestingly, Modularity is larger in the case of Luminal A network than the healthy network, for all methods.

An additional effect observed when comparing cancer and non-cancer derived networks, is a high proportion of same-chromosome gene-gene interactions in cancer phenotypes. On the other hand, healthy tissue-derived networks are composed of interactions between genes from any chromosome in a homogeneous fashion. This phenomenon has been called *loss of long-distance co-expression in cancer* (Espinal-Enríquez et al., 2017). This abrupt change has been reported for different tissues such as breast cancer (Espinal-Enríquez et al., 2017; de Anda-Jáuregui et al., 2019), each breast cancer molecular subtype (García-Cortés et al., 2020), clear cell renal carcinoma (Zamora-Fuentes et al., 2020), lung adenocarcinoma and lung sqamous cell carcinoma (Andonegui-Elguera et al., 2021). It is worth noticing that modularity has been used as an indirect measure of coordinated gene function (Solé et al., 2002; Segal et al., 2003; Lee et al., 2004; Tanay et al., 2004; Zhu et al., 2008). In this case, modules do not always represent gene function, but often act as a proxy for *spatial clustering* between genes from the same chromosome.

The studies just mentioned are just a handful instances, illustrating how network modularity determination is a becoming an essential approach to biological discovery.

7. CONCLUDING REMARKS

As we have already discussed, complexity in biological systems can be understood partially by using network approaches. Modularity is often an inherent component of complex biological networks. However relevant, network modularity discovery (or community detection, as is also called) is a daunting task. Its importance in theoretical biology, to describe the emergence of functional behaviors in biological systems, as well as its use in understanding the underlying principles behind such functionality make it a worthy tool in biology.

In the past years, a number of relevant approaches to this problem have been developed in the computational and systems biology settings. Most of these approaches, although extremely informative are built upon *Ad Hoc* assumptions and are thus not easy to generalize. Hence, they provide useful information, but are too specific. On then other hand, the network science and statistical physics research communities have been developing a series of quite general modularity detection algorithms. Here we present some of them, organized as *families* of methods, depending on their methodological foundations: (i) clustering algorithms, (ii) modularity optimization methods, (iii) methods based on the spectral properties of adjacency matrices, (iv) methods based on random walks and (v) methods based on stochastic block models. These broad families of methods along with the benchmarks that have been developed to evaluate their performance may constitute a relevant toolbox for the analysis of biological systems from a more general perspective. We argue that by resorting to these methods (freed from the design constraints typical of *Ad Hoc* methods) will allow to focus on the actual biology rather than on the method's specificities.

The problem of modularity and the discovery of functional communities in biological networks is an important emerging field of research. Omic high throughput technologies and the rise of computing power as well as the development of novel analytical algorithms have allowed the generation of bio-molecular network models at an unprecedented pace. This has led us with the need to develop theoretical and computational tools to extract biologically useful (e.g., functional or mechanistic) information from such large scale models. A wide variety of biological questions that can be answered—at least partially—by knowing the modular structure of the underlying networks, are being added to the current research scenario in the systems biology and genomics communities. A number of powerful mathematical and computational schemes to deal with modularity are also currently under development.

In the preceding review, we have discussed both, the biological problems and the computational approaches to the problem of modularity in complex bio-molecular networks. It is our sincere desire that works like this will stimulate the discussion between researchers in all the involved fields. A discussion that may in turn strengthen the ties of collaboration and ultimately leads to fruitful cross-fertilized scientific discoveries.

AUTHOR CONTRIBUTIONS

SA-C and EH-L: conceived the idea, contributed to the writing of the manuscript, and revised the manuscript. SS-M and JE-E: contributed to the writing of the manuscript and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by CONACYT (grant no. 179431/2012), as well as by federal funding from the National Institute of Genomic Medicine (Mexico). Additional support has been granted by the National Laboratory of Complexity Sciences (grant no. 232647/2014 CONACYT). SS-M appreciates the support from CONACyT through the Catedras-CONACyT program. EH-L also acknowledges additional support from the 2016 Marcos Moshinsky Research Chair in the Physical Sciences. The funders have no role in the design or development of this work.

REFERENCES

Acharya, L., Judeh, T., and Zhu, D. (2012). "A survey of computational approaches to reconstruct and partition biological networks," in *Statistical and Machine Learning Approaches for Network Analysis*, eds M. Dehmer and S. C. Basak (New Jersey: Wiley), 1. doi: 10.1002/9781118346990.ch1

Adamcsek, B., Palla, G., Farkas, I. J., Derényi, I., and Vicsek, T. (2006). Cfinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021–1023. doi: 10.1093/bioinformatics/btl039

Ai, D., Wang, Y., Li, X., and Pan, H. (2020). Colorectal cancer prediction based on weighted gene co-expression network analysis and variational auto-encoder. *Biomolecules* 10:1207. doi: 10.3390/biom10091207

Alcalá-Corona, S. A., de Anda-Jáuregui, G., Espinal-Enriquez, J., and Hernández-Lemus, E. (2017). Network modularity in breast cancer molecular subtypes. *Front. Physiol*. 8:915. doi: 10.3389/fphys.2017.00915

Alcalá-Corona, S. A., de Anda-Jáuregui, G., Espinal-Enriquez, J., Tovar, H., and Hernández-Lemus, E. (2018a). "Network modularity and hierarchical structure in breast cancer molecular subtypes," in *International Conference on Complex Systems* (Cham: Springer), 352–358. doi: 10.1007/978-3-319-96661-8_36

Alcalá-Corona, S. A., Espinal-Enriquez, J., De Anda Jáuregui, G., and Hernandez-Lemus, E. (2018b). The hierarchical modular structure of HER2+ breast cancer network. *Front. Physiol*. 9:1423. doi: 10.3389/fphys.2018.01423

Alcalá-Corona, S. A., Velázquez-Caldelas, T. E., Espinal-Enriquez, J., and Hernández-Lemus, E. (2016). Community structure reveals biologically functional modules in MEF2C transcriptional regulatory network. *Front. Physiol*. 7:184. doi: 10.3389/fphys.2016.00184

Aldana, M., Balleza, E., Kauffman, S., and Resendiz, O. (2007). Robustness and evolvability in genetic regulatory networks. *J. Theoret. Biol*. 245, 433–448. doi: 10.1016/j.jtbi.2006.10.027

Aldana, M., and Cluzel, P. (2003). A natural class of robust networks. *Proc. Natl. Acad. Sci. U.S.A*. 100, 8710–8714. doi: 10.1073/pnas.1536783100

Aldecoa, R., and Marin, I. (2014). Surpriseme: an integrated tool for network community structure characterization using surprise maximization. *Bioinformatics* 30, 1041–1042. doi: 10.1093/bioinformatics/btt741

Ana, L., and Jain, A. K. (2003). "Robust data clustering," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Madison, WI: IEEE), 2–128. doi: 10.1109/CVPR.2003.1211462

Andonegui-Elguera, S. D., Zamora-Fuentes, J. M., Espinal-Enriquez, J., and Hernández-Lemus, E. (2021). Loss of long distance co-expression in lung cancer. *Front. Genet*. 12:625741. doi: 10.3389/fgene.2021.625741

Arenas, A., and Diaz-Guilera, A. (2007). Synchronization and modularity in complex networks. *Eur. Phys. J. Spcl. Top*. 143, 19–25. doi: 10.1140/epjst/e2007-00066-2

Arenas, A., Díaz-Guilera, A., and Pérez-Vicente, C. J. (2006). Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett*. 96:114102. doi: 10.1103/PhysRevLett.96.114102

Arenas, A., Fernandez, A., Fortunato, S., and Gomez, S. (2008). Motif-based communities in complex networks. *J. Phys. A Math. Theoret*. 41:224001. doi: 10.1088/1751-8113/41/22/224001

Ashrafian, H., McKenna, W. J., and Watkins, H. (2011). Disease pathways and novel therapeutic targets in hypertrophic cardiomyopathy. *Circ. Res*. 109, 86–96. doi: 10.1161/CIRCRESAHA.111.242974

Aslak, U., Rosvall, M., and Lehmann, S. (2018). Constrained information flows in temporal networks reveal intermittent communities. *Phys. Rev. E* 97:062312. doi: 10.1103/PhysRevE.97.062312

Banerjee, K., Kolomeisky, A. B., and Igoshin, O. A. (2017). Accuracy of substrate selection by enzymes is controlled by kinetic discrimination. *J. Phys. Chem. Lett*. 8, 1552–1556. doi: 10.1021/acs.jpclett.7b00441

Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet*. 5, 101–113. doi: 10.1038/nrg1272

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. *Nature* 387, 913–917. doi: 10.1038/43199

Barnes, E. R. (1982). An algorithm for partitioning the nodes of a graph. *SIAM J. Alg. Disc. Meth*. 3, 541–550.

Bennett, L., Kittas, A., Muirhead, G., Papageorgiou, L. G., and Tsoka, S. (2015). Detection of composite communities in multiplex biological networks. *Sci. Rep*. 5, 1–12. doi: 10.1038/srep10345

Bettinelli, A., Hansen, P., and Liberti, L. (2012). Algorithm for parametric community detection in networks. *Phys. Rev. E* 86:016107. doi: 10.1103/PhysRevE.86.016107

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp*. 2008:P10008. doi: 10.1088/1742-5468/2008/10/P10008

Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., et al. (2008). On modularity clustering. *IEEE Trans. Knowl. Data Eng*. 20, 172–188. doi: 10.1109/TKDE.2007.190689

Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science* 165, 349–357. doi: 10.1126/science.165.3891.349

Calderer, G., and Kuijjer, M. L. (2021). Community detection in large-scale bipartite biological networks. *Front. Genet*. 12:520. doi: 10.3389/fgene.2021.649440

Cantini, L., Medico, E., Fortunato, S., and Caselle, M. (2015). Detection of gene communities in multi-networks reveals cancer drivers. *Sci. Rep*. 5:17386. doi: 10.1038/srep17386

Capocci, A., Servedio, V. D., Caldarelli, G., and Colaiori, F. (2005). Detecting communities in large networks. *Phys. A Stat. Mech. Appl*. 352, 669–676. doi: 10.1016/j.physa.2004.12.050

Chen, H.-R., Sherr, D. H., Hu, Z., and DeLisi, C. (2016). A network based approach to drug repositioning identifies plausible candidates for breast cancer and prostate cancer. *BMC Med. Genomics* 9:1. doi: 10.1186/s12920-016-0212-7

Chen, J., and Zhang, S. (2016). Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data. *Bioinformatics* 32, 1724–1732. doi: 10.1093/bioinformatics/btw059

Cheng, L., Liu, P., Wang, D., and Leung, K.-S. (2019). Exploiting locational and topological overlap model to identify modules in protein interaction networks. *BMC Bioinformatics* 20:23. doi: 10.1186/s12859-019-2598-7

Cherifi, H., Palla, G., Szymanski, B. K., and Lu, X. (2019). On community structure in complex networks: challenges and opportunities. *Appl. Network Sci*. 4, 1–35. doi: 10.1007/s41109-019-0238-9

Clarke, B. S., and Mittenthal, J. E. (1992). Modularity and reliability in the organization of organisms. *Bull. Math. Biol*. 54, 1–20. doi: 10.1016/S0092-8240(05)80173-9

Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E* 70:066111. doi: 10.1103/PhysRevE.70.066111

Clune, J., Mouret, J.-B., and Lipson, H. (2013). The evolutionary origins of modularity. *Proc. R. Soc. B Biol. Sci*. 280:20122863. doi: 10.1098/rspb.2012.2863

Condon, A., and Karp, R. M. (2001). Algorithms for graph partitioning on the planted partition model. *Random Struct. Algorithms* 18, 116–140. doi: 10.1002/1098-2418(200103)18:2<116::AID-RSA1001>3.0.CO;2-2

Constantino, P. H., and Daoutidis, P. (2019). A control perspective on the evolution of biological modularity. *IFAC Pap. Online* 52, 172–177. doi: 10.1016/j.ifacol.2019.09.136

da Fonseca Vieira, V., Xavier, C. R., and Evsukoff, A. G. (2020). A comparative study of overlapping community detection methods from the perspective of the structural properties. *Appl. Network Sci*. 5, 1–42. doi: 10.1007/s41109-020-00289-9

Danon, L., Diaz-Guilera, A., Duch, J., and Arenas, A. (2005). Comparing community structure identification. *J. Stat. Mech. Theory Exp*. 2005:P09008. doi: 10.1088/1742-5468/2005/09/P09008

Davidson, E., and Levin, M. (2005). Gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A*. 102, 4935–4935. doi: 10.1073/pnas.0502024102

de Anda-Jáuregui, G., Alcalá-Corona, S. A., Espinal-Enriquez, J., and Hernández-Lemus, E. (2019). Functional and transcriptional connectivity of communities in breast cancer co-expression networks. *Appl. Network Sci*. 4, 1–13. doi: 10.1007/s41109-019-0129-0

De Domenico, M., Lancichinetti, A., Arenas, A., and Rosvall, M. (2015). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* 5:011027. doi: 10.1103/PhysRevX.5.011027

De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., et al. (2013). Mathematical formulation of multilayer networks. *Phys. Rev. X* 3:041022. doi: 10.1103/PhysRevX.3.041022

De Leo, V., Santoboni, G., Cerina, F., Mureddu, M., Secchi, L., and Chessa, A. (2013). Community core detection in transportation networks. *Phys. Rev. E* 88:042810. doi: 10.1103/PhysRevE.88.042810

de Matos Simoes, R., Tripathi, S., and Emmert-Streib, F. (2012). Organizational structure and the periphery of the gene regulatory network in B-cell lymphoma. *BMC Syst. Biol*. 6:1. doi: 10.1186/1752-0509-6-38

Debnath, S., Rakshit, S., Sengupta, K., and Plewczynski, D. (2021). "Biomolecular clusters identification in linear time complexity for biological networks," in *Proceedings of International Conference on Frontiers in Computing and Systems* (Singapore: Springer), 611–622. doi: 10.1007/978-981-15-7834-2_57

Didier, G., Brun, C., and Baudot, A. (2015). Identifying communities from multiplex biological networks. *PeerJ* 3:1042. doi: 10.7717/peerj.1525

Didier, G., Valdeolivas, A., and Baudot, A. (2018). Identifying communities from multiplex biological networks by randomized optimization of modularity. *F1000Research* 7:1042. doi: 10.12688/f1000research.15486.1

Donath, W. E., and Hoffman, A. J. (1972). Algorithms for partitioning of graphs and computer logic based on eigenvectors of connection matrices. *IBM Tech. Disclosure Bull*. 15, 938–944.

Donetti, L., and Munoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. *J. Stat. Mech. Theory Exp*. 2004:P10012. doi: 10.1088/1742-5468/2004/10/P10012

Dorantes-Gilardi, R., Garcia-Cortés, D., Hernández-Lemus, E., and Espinal-Enriquez, J. (2020). Multilayer approach reveals organizational principles disrupted in breast cancer co-expression networks. *Appl. Network Sci*. 5, 1–23. doi: 10.1007/s41109-020-00291-1

Drago-García, D., Espinal-Enriquez, J., and Hernández-Lemus, E. (2017). Network analysis of emt and met micro-rna regulation in breast cancer. *Sci. Rep*. 7:13534. doi: 10.1038/s41598-017-13903-1

Duch, J., and Arenas, A. (2005). Community detection in complex networks using extremal optimization. *Phys. Rev. E* 72:027104. doi: 10.1103/PhysRevE.72.027104

Emmons, S., and Mucha, P. J. (2019). Map equation with metadata: varying the role of attributes in community detection. *Phys. Rev. E* 100:022301. doi: 10.1103/PhysRevE.100.022301

Eriksson, A., Carletti, T., Lambiotte, R., Rojas, A., and Rosvall, M. (2021). Flow-based community detection in hypergraphs. *arXiv preprint arXiv:2105.04389.*

Espinal, J., Aldana, M., Guerrero, A., Wood, C., Darszon, A., and Martinez-Mekler, G. (2011). Discrete dynamics model for the speract-activated ca 2+ signaling network relevant to sperm motility. *PLoS ONE* 6:e22619. doi: 10.1371/journal.pone.0022619

Espinal-Enríquez, J., Fresno, C., Anda-Jáuregui, G., and Hernández-Lemus, E. (2017). RNA-Seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. *Sci. Rep*. 7:1760. doi: 10.1038/s41598-017-01314-1

Espinal-Enriquez, J., Priego-Espinosa, D. A., Darszon, A., Beltrán, C., and Martinez-Mekler, G. (2017). Network model predicts that catsper is the main ca 2+ channel in the regulation of sea urchin sperm motility. *Sci. Rep*. 7, 1–14. doi: 10.1038/s41598-017-03857-9

Espinosa-Soto, C., and Wagner, A. (2010). Specialization can drive the evolution of modularity. *PLoS Comput. Biol*. 6:e1000719. doi: 10.1371/journal.pcbi.1000719

Esquivel, A. V., and Rosvall, M. (2011). Compression of flow can reveal overlapping-module organization in networks. *Phys. Rev. X* 1:021025. doi: 10.1103/PhysRevX.1.021025

Fan, Y., Li, M., Zhang, P., Wu, J., and Di, Z. (2007). Accuracy and precision of methods for community identification in weighted networks. *Phys. A Stat. Mech. Appl*. 377, 363–372. doi: 10.1016/j.physa.2006.11.036

Farage, C., Edler, D., Eklöf, A., Rosvall, M., and Pilosof, S. (2021). Identifying flow modules in ecological networks using infomap. *Methods Ecol. Evol*. 12, 778–786. doi: 10.1111/2041-210X.13569

Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J*. 23, 298–305. doi: 10.21136/CMJ.1973.101168

Fortunato, S. (2010). Community detection in graphs. *Phys. Rep*. 486, 75–174. doi: 10.1016/j.physrep.2009.11.002

Fortunato, S., and Barthelemy, M. (2007). Resolution limit in community detection. *Proc. Natl. Acad. Sci. U.S.A*. 104, 36–41. doi: 10.1073/pnas.0605965104

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry* 40, 35–41. doi: 10.2307/3033543

Friedlander, T., Mayo, A. E., Tlusty, T., and Alon, U. (2013). Mutation rules and the evolution of sparseness and modularity in biological systems. *PLoS ONE* 8:e70444. doi: 10.1371/journal.pone.0070444

Gao, S., Chen, A., Rahmani, A., Zeng, J., Tan, M., Alhajj, R., et al. (2016). Multi-scale modularity and motif distributional effect in metabolic networks. *Curr. Protein Peptide Sci*. 17, 82–92. doi: 10.2174/1389203716666150923104603

García-Cortés, D., de Anda-Jáuregui, G., Fresno, C., Hernandez-Lemus, E., and Espinal-Enriquez, J. (2020). Gene co-expression is distance-dependent in breast cancer. *Front. Oncol*. 10:1232. doi: 10.3389/fonc.2020.01232

García-Cortés, D. E., Hernandez-Lemus, E., and Espinal-Enriquez, J. (2021). Luminal a breast cancer co-expression network: structural and functional alterations. *Front. Genet*. 12:514. doi: 10.3389/fgene.2021.629475

Garza, S. E., and Schaeffer, S. E. (2019). Community detection with the label propagation algorithm: a survey. *Phys. A Stat. Mech. Appl*. 534:122058. doi: 10.1016/j.physa.2019.122058

Ghiassian, S. D., Menche, J., and 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 Comput. Biol*. 11:e1004120. doi: 10.1371/journal.pcbi.1004120

Ghiassian, S. D., Menche, J., Chasman, D. I., Giulianini, F., Wang, R., Ricchiuto, P., et al. (2016). Endophenotype network models: common core of complex diseases. *Sci. Rep*. 6:27414. doi: 10.1038/srep27414

Gibson, G. (2016). On the evaluation of module preservation. *Cell Syst*. 3, 17–19. doi: 10.1016/j.cels.2016.07.009

Gilarranz, L. J. (2020). Generic emergence of modularity in spatial networks. *Sci. Rep*. 10, 1–8. doi: 10.1038/s41598-020-65669-8

Gilbert, J. P. (2015). *A probabilistic model for the evaluation of module extraction algorithms in complex biological networks* (Ph.D. thesis). University of Nottingham, Nottingham, United Kingdom.

Girvan, M., and Newman, M. E. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A*. 99, 7821–7826. doi: 10.1073/pnas.122653799

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabási, A.-L. (2007). The human disease network. *Proc. Natl. Acad. Sci. U.S.A*. 104, 8685–8690. doi: 10.1073/pnas.0701361104

Gómez-Romero, L., López-Reyes, K., and Hernández-Lemus, E. (2020). The large scale structure of human metabolism reveals resilience via extensive signaling crosstalk. *Front. Physiol*. 11:1667. doi: 10.3389/fphys.2020.588012

Gong, M., Fu, B., Jiao, L., and Du, H. (2011). Memetic algorithm for community detection in networks. *Phys. Rev. E* 84:056101. doi: 10.1103/PhysRevE.84.056101

Green, S., Şerban, M., Scholl, R., Jones, N., Brigandt, I., and Bechtel, W. (2018). Network analyses in systems biology: new strategies for dealing with biological complexity. *Synthese* 195, 1751–1777. doi: 10.1007/s11229-016-1307-6

Guimera, R., and Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895–900. doi: 10.1038/nature03288

Guimera, R., Sales-Pardo, M., and Amaral, L. A. N. (2004). Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* 70:025101. doi: 10.1103/PhysRevE.70.025101

Guney, E., Menche, J., Vidal, M., and Barábasi, A.-L. (2016). Network-based *in silico* drug efficacy screening. *Nat. Commun*. 7:10331. doi: 10.1038/ncomms10331

Gyorgy, A., and Del Vecchio, D. (2014). Modular composition of gene transcription networks. *PLoS Comput. Biol*. 10:e1003486. doi: 10.1371/journal.pcbi.1003486

Henriques, R., and Madeira, S. C. (2016). Bicnet: Flexible module discovery in large-scale biological networks using biclustering. *Algorithms Mol. Biol*. 11:1. doi: 10.1186/s13015-016-0074-8

Hernández-Lemus, E., Reyes-Gopar, H., Espinal-Enriquez, J., and Ochoa, S. (2019). The many faces of gene regulation in cancer: a computational oncogenomics outlook. *Genes* 10:865. doi: 10.3390/genes10110865

Hric, D., Darst, R. K., and Fortunato, S. (2014). Community detection in networks: structural communities versus ground truth. *Phys. Rev. E* 90:062805. doi: 10.1103/PhysRevE.90.062805

Iacovacci, J., and Bianconi, G. (2016). Extracting information from multiplex networks. *Chaos* 26:065306. doi: 10.1063/1.4953161

Igoshin, O. A., Brody, M. S., Price, C. W., and Savageau, M. A. (2007). Distinctive topologies of partner-switching signaling networks correlate with their physiological roles. *J. Mol. Biol*. 369, 1333–1352. doi: 10.1016/j.jmb.2007.04.021

Jaeger, J., and Monk, N. (2021). Dynamical modules in metabolism, cell and developmental biology. *Interface Focus* 11:20210011. doi: 10.1098/rsfs.2021.0011

Jebabli, M., Cherifi, H., Cherifi, C., and Hamouda, A. (2018). Community detection algorithm evaluation with ground-truth data. *Phys. A Stat. Mech. Appl*. 492, 651–706. doi: 10.1016/j.physa.2017.10.018

Karrer, B., and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* 83:016107. doi: 10.1103/PhysRevE.83.016107

Kashtan, N., and Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13773–13778. doi: 10.1073/pnas.0503610102

Kashtan, N., Parter, M., Dekel, E., Mayo, A. E., and Alon, U. (2009). Extinctions in heterogeneous environments and the evolution of modularity. *Evol. Int. J. Organ. Evol.* 63, 1964–1975. doi: 10.1111/j.1558-5646.2009.00684.x

Kauffman, S. (1969). Homeostasis and differentiation in random genetic control networks. *Nature* 224, 177–178. doi: 10.1038/224177a0

Kawamoto, T., and Rosvall, M. (2015). Estimating the resolution limit of the map equation in community detection. *Phys. Rev. E* 91:012809. doi: 10.1103/PhysRevE.91.012809

Khadivi, A., Rad, A. A., and Hasler, M. (2011). Network community-detection enhancement by proper weighting. *Phys. Rev. E* 83:046104. doi: 10.1103/PhysRevE.83.046104

Kheirkhahzadeh, M., Lancichinetti, A., and Rosvall, M. (2016). Efficient community detection of network flows for varying Markov times and bipartite networks. *Phys. Rev. E* 93:032309. doi: 10.1103/PhysRevE.93.032309

Knuth, D. E. (1985). Dynamic huffman coding. *J. Algorithms* 6, 163–180. doi: 10.1016/0196-6774(85)90036-7

Kovács, I. A., and Barabási, A.-L. (2015). Network science: destruction perfected. *Nature* 524, 38–39. doi: 10.1038/524038a

Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., et al. (2013). Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20935–20940. doi: 10.1073/pnas.1312486110

Kuncheva, L. I., and Hadjitodorov, S. T. (2004). "Using diversity in cluster ensembles," in *IEEE International Conference on Systems, Man and Cybernetics, 2004* (The Hague: IEEE), 1214–1219. doi: 10.1109/ICSMC.2004.1399790

Lancichinetti, A., and Fortunato, S. (2009a). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E* 80:016118. doi: 10.1103/PhysRevE.80.016118

Lancichinetti, A., and Fortunato, S. (2009b). Community detection algorithms: a comparative analysis. *Phys. Rev. E* 80:056117. doi: 10.1103/PhysRevE.80.056117

Lancichinetti, A., and Fortunato, S. (2011). Limits of modularity maximization in community detection. *Phys. Rev. E* 84:066122. doi: 10.1103/PhysRevE.84.066122

Lancichinetti, A., and Fortunato, S. (2012). Consensus clustering in complex networks. *Sci. Rep.* 2, 1–7. doi: 10.1038/srep00336

Lancichinetti, A., Fortunato, S., and Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *N. J. Phys.* 11:033015. doi: 10.1088/1367-2630/11/3/033015

Lancichinetti, A., Fortunato, S., and Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* 78:046110. doi: 10.1103/PhysRevE.78.046110

Lancichinetti, A., Radicchi, F., Ramasco, J. J., and Fortunato, S. (2011). Finding statistically significant communities in networks. *PLoS ONE* 6:e18961. doi: 10.1371/journal.pone.0018961

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. *Bioinformatics* 24, 719–720. doi: 10.1093/bioinformatics/btm563

Lecca, P., and Re, A. (2015). Detecting modules in biological networks by edge weight clustering and entropy significance. *Front. Genet.* 6:265. doi: 10.3389/fgene.2015.00265

Lee, I., Date, S. V., Adai, A. T., and Marcotte, E. M. (2004). A probabilistic functional network of yeast genes. *Science* 306, 1555–1558. doi: 10.1126/science.1099511

Li, Y., Liu, B., Li, J., and Li, G. (2019). Mimod: a new algorithm for mining biological network modules. *IEEE Access* 7, 49492–49503. doi: 10.1109/ACCESS.2019.2909946

Liu, J., Jing, L., and Tu, X. (2016). Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. *BMC Cardiovasc. Disord.* 16:1. doi: 10.1186/s12872-016-0217-3

Long, J., Liu, Z., Wu, X., Xu, Y., and Ge, C. (2016). Screening for genes and subnetworks associated with pancreatic cancer based on the gene expression profile. *Mol. Med. Rep.* 13, 3779–3786. doi: 10.3892/mmr.2016.5007

Lorenz, D. M., Jeng, A., and Deem, M. W. (2011). The emergence of modularity in biological systems. *Phys. Life Rev.* 8, 129–160. doi: 10.1016/j.plrev.2011.02.003

Lucchetta, M., and Pellegrini, M. (2020). Finding disease modules for cancer and covid-19 in gene co-expression networks with the core&peel method. *Sci. Rep.* 10, 1–18. doi: 10.1038/s41598-020-74705-6

Lusseau, D. (2007). Evidence for social role in a dolphin social network. *Evol. Ecol.* 21, 357–366. doi: 10.1007/s10682-006-9105-0

Mall, R., Cerulo, L., Bensmail, H., Iavarone, A., and Ceccarelli, M. (2017a). Detection of statistically significant network changes in complex biological networks. *BMC Syst. Biol.* 11:32. doi: 10.1186/s12918-017-0412-6

Mall, R., Ullah, E., Kunji, K., D'Angelo, F., Bensmail, H., and Ceccarelli, M. (2017b). "Differential community detection in paired biological networks," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Boston, MA, 330–339. doi: 10.1145/3107411.3107418

Medus, A., and Dorso, C. (2009). Alternative approach to community detection in networks. *Phys. Rev. E* 79:066111. doi: 10.1103/PhysRevE.79.066111

Meilă, M. (2007). Comparing clusterings? An information based distance. *J. Multivariate Anal.* 98, 873–895. doi: 10.1016/j.jmva.2006.11.013

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., et al. (2015). Uncovering disease-disease relationships through the incomplete interactome. *Science* 347:1257601. doi: 10.1126/science.1257601

Miecznikowski, J. C., Gaile, D. P., Chen, X., and Tritchler, D. L. (2016). Identification of consistent functional genetic modules. *Stat. Appl. Genet. Mol. Biol.* 15, 1–18. doi: 10.1515/sagmb-2015-0026

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827. doi: 10.1126/science.298.5594.824

Mittal, R., and Bhatia, M. (2018). "Analyzing the structures of clusters in multi-layer biological networks," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (Jalandhar: IEEE), 502–507. doi: 10.1109/ICSCCC.2018.8703271

Monzón-Sandoval, J., Castillo-Morales, A., Urrutia, A. O., and Gutierrez, H. (2016). Modular reorganization of the global network of gene regulatory interactions during perinatal human brain development. *BMC Dev. Biol.* 16:1. doi: 10.1186/s12861-016-0111-3

Moreira-Filho, C. A., Bando, S. Y., Bertonha, F. B., Iamashita, P., Silva, F. N., da Fontoura Costa, L., et al. (2015). Community structure analysis of transcriptional networks reveals distinct molecular pathways for early-and late-onset temporal lobe epilepsy with childhood febrile seizures. *PLoS ONE* 10:e0128174. doi: 10.1371/journal.pone.0128174

Morohashi, M., Winn, A. E., Borisuk, M. T., Bolouri, H., Doyle, J., and Kitano, H. (2002). Robustness as a measure of plausibility in models of biochemical networks. *J. Theoret. Biol.* 216, 19–30. doi: 10.1006/jtbi.2002.2537

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328, 876–878. doi: 10.1126/science.1184819

Muraro, D., and Simmons, A. (2016). An integrative analysis of gene expression and molecular interaction data to identify dys-regulated sub-networks in inflammatory bowel disease. *BMC Bioinformatics* 17:1. doi: 10.1186/s12859-016-0886-z

Narula, J., Smith, A. M., Gottgens, B., and Igoshin, O. A. (2010). Modeling reveals bistability and low-pass filtering in the network module determining blood stem cell fate. *PLoS Comput. Biol.* 6:e1000771. doi: 10.1371/journal.pcbi.1000771

Newman, M. (2010). *Networks: An Introduction*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199206650.003.0001

Newman, M. E. (2004a). Detecting community structure in networks. *Eur. Phys. J. B* 38, 321–330. doi: 10.1140/epjb/e2004-00124-y

Newman, M. E. (2004b). Fast algorithm for detecting community structure in networks. *Phys. Rev. E* 69:066133. doi: 10.1103/PhysRevE.69.066133

Newman, M. E. (2006a). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74:036104. doi: 10.1103/PhysRevE.74.036104

Newman, M. E. (2006b). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8577–8582. doi: 10.1073/pnas.0601602103

Newman, M. E. (2012). Communities, modules and large-scale structure in networks. *Nat. Phys.* 8, 25–31. doi: 10.1038/nphys2162

Newman, M. E. (2013). Spectral methods for community detection and graph partitioning. *Phys. Rev. E* 88:042822. doi: 10.1103/PhysRevE.88.042822

Newman, M. E., and Girvan, M. (2003). "Mixing patterns and community structure in networks," in *Statistical Mechanics of Complex Networks* (Heidelberg: Springer), 66–87. doi: 10.1007/978-3-540-44943-0_5

Newman, M. E., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69:026113. doi: 10.1103/PhysRevE.69.026113

Oliveira, J. V., de Brito, A. F., Braconi, C. T., de Melo Freire, C. C., Iamarino, A., and de Andrade Zanotto, P. M. (2013). Modularity and evolutionary constraints in a baculovirus gene regulatory network. *BMC Syst. Biol.* 7:1. doi: 10.1186/1752-0509-7-87

Orman, G. K., Labatut, V., and Cherifi, H. (2012). Comparative evaluation of community detection algorithms: a topological approach. *J. Stat. Mech. Theory Exp*. 2012:P08001. doi: 10.1088/1742-5468/2012/08/P08001

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 814–818. doi: 10.1038/nature03607

Parter, M., Kashtan, N., and Alon, U. (2007). Environmental variability and modularity of bacterial metabolic networks. *BMC Evol. Biol*. 7:169. doi: 10.1186/1471-2148-7-169

Peixoto, T. P. (2018). Nonparametric weighted stochastic block models. *Phys. Rev. E* 97:012306. doi: 10.1103/PhysRevE.97.012306

Pesantez-Cabrera, P., and Kalyanaraman, A. (2016). "Detecting communities in biological bipartite networks," in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Seattle, WA, 98–107. doi: 10.1145/2975167.2975177

Pilosof, S., Alcala-Corona, S. A., Wang, T., Kim, T., Maslov, S., Whitaker, R., et al. (2020). The network structure and eco-evolutionary dynamics of crispr-induced immune diversification. *Nat. Ecol. Evol*. 4, 1650–1660. doi: 10.1038/s41559-020-01312-z

Pons, P., and Latapy, M. (2005). "Computing communities in large networks using random walks," in *International Symposium on Computer and Information Sciences* (Poznan: Springer), 284–293. doi: 10.1007/11569596_31

Pons, P., and Latapy, M. (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl*. 10, 191–218. doi: 10.7155/jgaa.00124

Porter, M. A., Onnela, J.-P., and Mucha, P. J. (2009). Communities in networks. *Notices AMS* 56, 1082–1097.

Praneenararat, T., Takagi, T., and Iwasaki, W. (2011). Interactive, multiscale navigation of large and complicated biological networks. *Bioinformatics* 27, 1121–1127. doi: 10.1093/bioinformatics/btr083

Qi, Y., and Ge, H. (2006). Modularity and dynamics of cellular networks. *PLoS Comput. Biol*. 2:e174. doi: 10.1371/journal.pcbi.0020174

Qiu, X., Cheng, S.-H., Xu, F., Yin, J.-W., Wang, L.-Y., and Zhang, X.-Y. (2020). Weighted gene co-expression network analysis identified MYL9 and CNN1 are associated with recurrence in colorectal cancer. *J. Cancer* 11:2348. doi: 10.7150/jca.39723

Rahiminejad, S., Maurya, M. R., and Subramaniam, S. (2019). Topological and functional comparison of community detection algorithms in biological networks. *BMC Bioinformatics* 20:212. doi: 10.1186/s12859-019-2746-0

Ramadan, E., Naef, A., and Ahmed, M. (2016). Protein complexes predictions within protein interaction networks using genetic algorithms. *BMC Bioinformatics* 17:481. doi: 10.1186/s12859-016-1096-4

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555. doi: 10.1126/science.1073374

Reichardt, J., and Bornholdt, S. (2004). Detecting fuzzy community structures in complex networks with a Potts model. *Phys. Rev. Lett*. 93:218701. doi: 10.1103/PhysRevLett.93.218701

Reichardt, J., and Bornholdt, S. (2006). Statistical mechanics of community detection. *Phys. Rev. E* 74:016110. doi: 10.1103/PhysRevE.74.016110

Reyna, M. A., Haan, D., Paczkowska, M., Verbeke, L. P., Vazquez, M., Kahraman, A., et al. (2020). Pathway and network analysis of more than 2500 whole cancer genomes. *Nat. Commun*. 11, 1–17. doi: 10.1038/s41467-020-14367-0

Ritchie, S. C., Watts, S., Fearnley, L. G., Holt, K. E., Abraham, G., and Inouye, M. (2016). A scalable permutation approach reveals replication and preservation patterns of network modules in large datasets. *Cell Syst*. 3, 71–82. doi: 10.1016/j.cels.2016.06.012

Rosvall, M., and Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proc. Natl. Acad. Sci. U.S.A*. 104, 7327–7331. doi: 10.1073/pnas.0611034104

Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A*. 105, 1118–1123. doi: 10.1073/pnas.0706851105

Rosvall, M., and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6:e18209. doi: 10.1371/journal.pone.0018209

Sah, P., Singh, L. O., Clauset, A., and Bansal, S. (2014). Exploring community structure in biological networks with random graphs. *BMC Bioinformatics* 15:220. doi: 10.1186/1471-2105-15-220

Samal, A., Singh, S., Giri, V., Krishna, S., Raghuram, N., and Jain, S. (2006). Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinformatics* 7:118. doi: 10.1186/1471-2105-7-118

Sanchez, C., Lachaize, C., Janody, F., Bellon, B., Röder, L., Euzenat, J., et al. (1999). Grasping at molecular interactions and genetic networks in drosophila melanogaster using flynets, an internet database. *Nucleic Acids Res*. 27, 89–94. doi: 10.1093/nar/27.1.89

Sardiu, M. E., Gilmore, J. M., Groppe, B., Florens, L., and Washburn, M. P. (2017). Identification of topological network modules in perturbed protein interaction networks. *Sci. Rep*. 7, 1–13. doi: 10.1038/srep43845

Schulz, S., Eckweiler, D., Bielecka, A., Nicolai, T., Franke, R., Dötsch, A., et al. (2015). Elucidation of sigma factor-associated networks in *Pseudomonas aeruginosa* reveals a modular architecture with limited and function-specific crosstalk. *PLoS Pathog*. 11:e1004744. doi: 10.1371/journal.ppat.1004744

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., et al. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet*. 34, 166–176. doi: 10.1038/ng1165

Şen, F., Wigand, R. T., Agarwal, N., Mete, M., and Kasprzyk, R. (2014). "Focal structure analysis in large biological networks," in *3rd International Conference on Environment, Energy and Biotechnology (ICEEB 2014)*, Bangkok.

Serban, M. (2020). Exploring modularity in biological networks. *Philos. Trans. R. Soc. B* 375:20190316. doi: 10.1098/rstb.2019.0316

Servis, M. J., and Clark, A. E. (2021). Cluster identification using modularity optimization to uncover chemical heterogeneity in complex solutions. *J. Phys. Chem. A* 125, 3986–3993. doi: 10.1021/acs.jpca.0c11320

Sevim, V., and Rikvold, P. A. (2008). Chaotic gene regulatory networks can be robust against mutations and noise. *J. Theoret. Biol*. 253, 323–332. doi: 10.1016/j.jtbi.2008.03.003

Shi, Z., Derow, C. K., and Zhang, B. (2010). Co-expression module analysis reveals biological processes, genomic gain, and regulatory mechanisms associated with breast cancer progression. *BMC Syst. Biol*. 4:1. doi: 10.1186/1752-0509-4-74

Shmulevich, I., Kauffman, S. A., and Aldana, M. (2005). Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proc. Natl. Acad. Sci. U.S.A*. 102, 13439–13444. doi: 10.1073/pnas.0506771102

Smith, E. P. (1985). Statistical comparison of weighted overlap measures. *Trans. Am. Fish. Soc*. 114, 250–257. doi: 10.1577/1548-8659(1985)114<250:SCOWOM>2.0.CO;2

Solé, R. V., Salazar-Ciudad, I., and Garcia-Fernández, J. (2002). Common pattern formation, modularity and phase transitions in a gene network model of morphogenesis. *Phys. A Stat. Mech. Appl*. 305, 640–654. doi: 10.1016/S0378-4371(01)00580-5

Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A*. 98, 10869–10874. doi: 10.1073/pnas.191367098

Su, G., Kuchinsky, A., Morris, J. H., Meng, F., et al. (2010). Glay: community structure analysis of biological networks. *Bioinformatics* 26, 3135–3137. doi: 10.1093/bioinformatics/btq596

Tadaka, S., and Kinoshita, K. (2016). NCMine: core-peripheral based functional module detection using near-clique mining. *Bioinformatics* 32, btw488. doi: 10.1093/bioinformatics/btw488

Tanay, A., Sharan, R., Kupiec, M., and Shamir, R. (2004). Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc. Natl. Acad. Sci. U.S.A*. 101, 2981–2986. doi: 10.1073/pnas.0308661100

Thieffry, D., and Romero, D. (1999). The modularity of biological regulatory networks. *Biosystems* 50, 49–59. doi: 10.1016/S0303-2647(98)00087-2

Torres-Sosa, C., Huang, S., and Aldana, M. (2012). Criticality is an emergent property of genetic networks that exhibit evolvability. *PLoS Comput. Biol*. 8:e1002669. doi: 10.1371/journal.pcbi.1002669

Tripathi, B., Parthasarathy, S., Sinha, H., Raman, K., and Ravindran, B. (2019). Adapting community detection algorithms for disease module identification in heterogeneous biological networks. *Front. Genet*. 10:164. doi: 10.3389/fgene.2019.00164

Tripathi, S., Moutari, S., Dehmer, M., and Emmert-Streib, F. (2016). Comparison of module detection algorithms in protein networks and investigation of the biological meaning of predicted modules. *BMC Bioinformatics* 17:1. doi: 10.1186/s12859-016-0979-8

Valverde, S. (2017). Breakdown of modularity in complex networks. *Front. Physiol*. 8:497. doi: 10.3389/fphys.2017.00497

van Dongen, S., and Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. *Bacterial Mol. Netw. Methods Protoc*. 804, 281–295. doi: 10.1007/978-1-61779-361-5_15

Van Dongen, S. M. (2001). *Graph clustering by flow simulation* (Ph.D. thesis), Utrecht.

Verd, B., Monk, N. A., and Jaeger, J. (2019). Modularity, criticality, and evolvability of a developmental gene regulatory network. *eLife* 8:e42832. doi: 10.7554/eLife.42832

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput*. 17, 395–416. doi: 10.1007/s11222-007-9033-z

Wagner, G., Mezey, J., and Calabretta, R. (2001). *Modularity: Understanding the Development and Evolution of Complex Natural Systems. Natural Selection and the Origin of Modules*. Cambridge, MA: MIT Press.

Wagner, G. P., Pavlicev, M., and Cheverud, J. M. (2007). The road to modularity. *Nat. Rev. Genet*. 8, 921–931. doi: 10.1038/nrg2267

Wang, H., Ye, M., Fu, Y., Dong, A., Zhang, M., Feng, L., et al. (2021). Modeling genome-wide by environment interactions through omnigenic interactome networks. *Cell Rep*. 35:109114. doi: 10.1016/j.celrep.2021.109114

Wang, J., Yi, Y., Chen, Y., Xiong, Y., and Zhang, W. (2020). Potential mechanism of rrm2 for promoting cervical cancer based on weighted gene co-expression network analysis. *Int. J. Med. Sci*. 17:2362. doi: 10.7150/ijms.47356

Wilkinson, D. M., and Huberman, B. A. (2004). A method for finding communities of related genes. *Proc. Natl. Acad. Sci. U.S.A*. 101(Suppl 1), 5241–5248. doi: 10.1073/pnas.0307740100

Xu, H., and Wang, S. (2010). "Research on functional modules of gene regulatory network," in *Advancing Computing, Communication, Control and Management* ed Q. Luo, (Heidelberg: Springer), 264–271. doi: 10.1007/978-3-642-05173-9_34

Xu, X., Zhou, Y., Miao, R., Chen, W., Qu, K., Pang, Q., et al. (2016). Transcriptional modules related to hepatocellular carcinoma survival: coexpression network analysis. *Front. Med*. 10, 183–190. doi: 10.1007/s11684-016-0440-4

Yang, Z., Algesheimer, R., and Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Sci. Rep*. 6, 1–18. doi: 10.1038/srep30750

Zamora-Fuentes, J. M., Hernández-Lemus, E., and Espinal-Enriquez, J. (2020). Gene expression and co-expression networks are strongly altered through stages in clear cell renal carcinoma. *Front. Genet*. 11:1232. doi: 10.3389/fgene.2020.578679

Zhan, M. (2007). Deciphering modular and dynamic behaviors of transcriptional networks. *Genomic Med*. 1, 19–28. doi: 10.1007/s11568-007-9004-7

Zhang, B., Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol*. 4:1128. doi: 10.2202/1544-6115.1128

Zhang, X., and Newman, M. (2015). Multiway spectral community detection in networks. *Phys. Rev. E* 92:052808. doi: 10.1103/PhysRevE.92.052808

Zheng, F., Zhang, S., Churas, C., Pratt, D., Bahar, I., and Ideker, T. (2020). Decoding of persistent multiscale structures in complex biological networks. *bioRxiv*. 92, 1–8. doi: 10.1186/s13059-020-02228-4

Zheng, F., Zhang, S., Churas, C., Pratt, D., Bahar, I., and Ideker, T. (2021). HiDeF: identifying persistent structures in multiscale 'omics data. *Genome Biol*. 22, 1–15.

Zhou, G., and Xia, J. (2018). OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Res*. 46, W514–W522. doi: 10.1093/nar/gky510

Zhou, H. (2003). Network landscape from a Brownian particle's perspective. *Phys. Rev. E* 67:041908. doi: 10.1103/PhysRevE.67.041908

Zhou, H., and Lipowsky, R. (2004). "Network Brownian motion: a new method to measure vertex-vertex proximity and to identify communities and subcommunities," in *International Conference on Computational Science* (Krakow: Springer), 1062–1069. doi: 10.1007/978-3-540-24688-6_137

Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., et al. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat. Genet*. 40, 854–861. doi: 10.1038/ng.167

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.