Multi-scale Modularity and Dynamics in Complex Networks

Renaud Lambiotte

1 Introduction

A broad range of systems are made of elements in interaction and can be represented as networks. Important examples include social networks, the Internet, airline routes, and a wide range of biological networks. The study of networks has emerged in the last decade as one of the fundamental building blocks in the wider study of complex systems [8, 15, 39]. Complex network theory emphasizes that the interactions between the individual components of the system are primordial to understand global emergent behavior and give the possibility to analyze systems of a very different nature within a single framework. Despite the fact that nodes and links have a different nature and are driven by different mechanisms for, e.g., the Web and food webs, the identification of similar structures in various complex systems suggests the existence of generic organization principles. A good example is the omnipresent multi-scale modular organization of complex networks, namely, the fact that they are made of modules, also called communities [62]. The precise mathematical definition of what communities are is still the subject of active research, but most definitions agree on the fact that modules, also called communities, are defined as subnetworks that are locally dense even though the network as a whole is sparse [20, 49]. In many complex systems it seems that modularity does not exist only at a single organizational scale but rather that each module can be further partitioned into a set of sub-modules and within each sub-module there may be sub-sub-modules, etc. In other words, many systems have the fractal property of multi-scale modularity. The detection of such community structure can be of importance for the understanding of the interplay between the structural, dynamical, and functional features of the network. This identification

R. Lambiotte (✉)
Department of Mathematics, University of Namur, Namur, Belgium
e-mail: renaud.lambiotte@fundp.ac.be

also has the advantage of providing a coarse-grained representation of the system, thereby allowing to sketch its organization and to identify sets of nodes that are likely to have hidden functions or properties in common. The main purpose of this chapter is to present methods uncovering the multi-scale organization of complex networks. In particular, we will emphasize methods finding modules based on the exploration of the network by a random walker at different time scales. The diffusive process allows for multistep transitions exploring further afield the structure of the graph, which results in the detection of community structure across scales, from finer to coarser. Time thus acts as a resolution parameter, allowing to zoom in and out to uncover the multi-scale structure of the system. As we will argue, this dynamic approach has the further advantages of taking into account the constraints imposed by the network structure on flows and of providing a unifying framework for a broad range of community detection methods.

2   Communities: Why Communities?

For many years, researchers, intrigued by the apparent modularity of many social, technological, and biological systems, have searched for mechanisms driving the evolution of systems towards a modular architecture [38]. One of the earliest and most influential ideas was formulated by Simon [62, 63] who argued that a nearly decomposable system, namely, a system of sparsely interconnected modules, allows faster adaptation of the system in response to changing external conditions. Modular systems can evolve by evolution in one module at a time or by duplication and mutation of modules. Well-adapted modules thus represent stable intermediate states whose stability is not jeopardized by evolution in other modules. This robustness represents a major advantage for any system evolving under fluctuating selection criteria, and this may explain the general prevalence of modular architectures across a very wide range of systems. This idea that a hierarchically modular design will be more rapidly and robustly assembled is illustrated by Simon [62] in a parable about two watchmakers, called Hora and Tempus:

> The watches the men made consisted of about 1,000 parts each. Tempus had so constructed his that if he had one partly assembled and had to put it down to answer the phone say it immediately fell to pieces and had to be reassembled from the elements. The better the customers liked his watches, the more they phoned him, the more difficult it became for him to find enough uninterrupted time to finish a watch.
>
> The watches that Hora made were no less complex than those of Tempus. But he had designed them so that he could put together subassemblies of about ten elements each. Ten of these subassemblies, again, could be put together into a larger subassembly; and a system of ten of the latter subassemblies constituted the whole watch. Hence, when Hora had to put down a partly assembled watch in order to answer the phone, he lost only a small part of his work, and he assembled his watches in only a fraction of the man-hours it took Tempus.

This idea has since been explored further. For instance, in [25], it has been argued that modular networks are optimal for performing tasks in a changing environment.

In situations where different tasks share basic sub-functions, evolutionary pressure produces networks where modules specialize in these sub-functions and where each of the tasks is obtained by a rapid recombination of these building functions. Other types of argument have also been put forward, including:

- Modular network topology is associated with a rich nonlinear dynamical behavior. Modular networks tend to produce time-scale separation, i.e., fast intra-modular processes and slow inter-modular processes [3, 45, 46], which helps at preserving coexistence and diversity in the system [27] and at finding a balance between segregated and integrated activity [47, 59]. The high dynamical complexity [64] leads to complex dynamical states such as transient chimera states [1, 60] where synchronization and desynchronization coexist across the network. Hierarchical modularity also enhances dynamical reconnectability [51], as marginally stable networks can be combined or divided while preserving stability.
- Another plausible mechanism for the formation of modules is coevolution, as the network structure and function coevolve [22]. A broad range of models with adaptive rewiring typically incorporate a reinforcement of links between synchronized units and a pruning of links between asynchronized ones. This feedback between structure and dynamics, similar to synaptic plasticity in neuronal dynamics, naturally drives the emergence of inhomogeneities and modules in networks [31].
- Modular networks have the property of small-worldness which is advantageous in a broad range of systems by combining high clustering and global connectivity [70]. For instance, in social systems, small-worldness balances the relative benefits of social cohesion and brokerage [30]. In neuroscience, the high clustering of connections favors locally segregated processing (with low wiring cost) of specialized functions, while the short path length supports globally integrated processing of more generic functions [65].

3 Combinatorial Approaches to Community Detection

3.1 Community Detection

Community detection aims at uncovering the modular organization of large-scale networks in an automatic fashion [20, 35, 49]. Most community detection methods find a partition of the nodes into communities, where most of the links are concentrated within the communities. Each node is assigned to one and only one community, i.e., partitions are not compatible with overlapping communities [2, 16, 44]. At the heart of a partitioning method, there is a mathematical definition for the quality of any partition. Once a quality function has been defined, different types of heuristics can be used in order to find, approximatively, its optimal partition, i.e.,

to find the partition having the highest value of the quality function. In a majority of cases, this quality function is based on the number of links within and/or across the communities, i.e., this is a combinatorial approach where quality is measured by counting certain motifs in the graph.

Modularity and Its Limitations

The most popular quality function for community detection is Newman–Girvan modularity, which we will describe in detail in this section. In the following, we focus on unweighted networks for the sake of clarity, but the results are directly applicable to weighted networks. Let A be the adjacency matrix of the network, where A_{ij} determines the presence of a link going from j to i. The in-degree and out-degree of node i are defined as $k_i^{\text{in}} \equiv \sum_j A_{ij}$ and $k_i^{\text{out}} \equiv \sum_j A_{ji}$, respectively; $L \equiv \sum_{i,j} A_{ij}$ is the total number of links in the network. If the network is undirected, the adjacency matrix is symmetric $A_{ij} = A_{ji}$, $k_i^{\text{in}} = k_i^{\text{out}} = k_i$, and the number of undirected links $m = L/2$.

Modularity is a function of the adjacency matrix A and of the partition $\mathcal{P}$ of the nodes into communities. It measures if links are more abundant within communities than would be expected on the basis of chance

$$Q = (\text{fraction of links within communities}) - (\text{expected fraction of such links}) \tag{1}$$

and reads

$$Q = \frac{1}{L} \sum_{C \in \mathcal{P}} \sum_{i,j \in C} \left[A_{ij} - P_{ij} \right] \tag{2}$$

where $i, j \in C$ is a summation over pairs of nodes i and j belonging to the same community C of $\mathcal{P}$ and therefore counts intra-community links. The null hypothesis is an extra ingredient in the definition and is incorporated in the matrix P_{ij}. P_{ij} is the expected number of links between nodes i and j over an ensemble of random networks with certain constraints. These constraints correspond to known information about the network organization, i.e., its total number of links and nodes, which has to be taken into account when assessing the relevance of an observed topological feature.

Let us first consider undirected networks. Two standard choices of null model are

$$P_{ij} = \langle k \rangle^2 / 2m, \text{ then } Q \equiv Q_{\text{unif}} \tag{3}$$

where $\langle k \rangle = 2m/N$ is the average degree and the only constraint is thus the total number of links in the network, and

$$P_{ij} = k_i k_j / 2m, \text{ then } Q \equiv Q_{\text{conf}}. \tag{4}$$

where randomized networks now preserve the degree of each node. The latter null model is usually preferred because it takes into account the degree heterogeneity of the network [42]. More complicated null models can in principle be constructed in order to preserve other properties of the network under consideration [5, 43]. For instance, in the case of directed networks, it is common to use the null model

$$P_{ij} = k_i^{in} k_j^{out}/L, \ \text{then } Q \equiv Q_{\text{dir}}. \tag{5}$$

preserving the in- and out-degrees of each node.

In the case of undirected networks, again, it is interesting to note that Q_{unif} and Q_{conf} are naturally related to the combinatorial Laplacian $L_{ij}^{(C)} = A_{ij} - k_i \delta_{ij}$ and the (normalized) Laplacian $L_{ij} = A_{ij}/k_j - \delta_{ij}$, respectively,[1] and, more generally, to the dynamics induced by these operators, as we will discuss more in detail in Sect. 4. For Q_{conf}, this relation is particularly clear after expressing modularity in terms of the (right) eigenvectors v_α of L_{ij}, i.e., v_α satisfy $\sum_j L_{ij} v_{\alpha,j} = \lambda_\alpha v_{\alpha,i}$. Without loss of generality, we assume that $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_\alpha \geq \ldots \geq \lambda_N$. The dominant eigenvector v_1 of eigenvalue $\lambda_1 = 0$ is given by $v_{1;i} = k_i/2m$ and is unique if the network is connected. By using a spectral decomposition of L_{ij}, one finds [13]

$$Q_{\text{conf}} = \sum_{\alpha=2}^{N} \frac{\lambda_\alpha + 1}{2m} \sum_{C} \sum_{i,j \in C} v_{\alpha;i} v_{\alpha;j} \tag{6}$$

where the contribution of the dominant eigenvector v_1 and the null model have cancelled each other out.

Modularity has become an essential element of a large number of clustering methods for large-scale networks. These methods aim at optimizing the modularity of a graph, i.e., finding the partition having the maximal value of Q. An exhaustive optimization of Q is impossible because of the explosion in the number of ways to partition a graph, when its size increases. It has been shown that the optimization of modularity is an NP-complete problem [9]. For this reason, several heuristics have been proposed to find high-quality partitions [7, 11, 20, 42, 48]. The optimization of modularity has the advantage of being performed without a priori specifying the number of modules nor their size. This procedure has been shown to produce useful and relevant partitions in a number of systems [41]. Unfortunately, it has also been shown that modularity suffers from several limitations [19, 23], partly because modularity optimization produces one single partition, which is not satisfactory when dealing with multi-scale systems. Related to this issue, there is the so-called resolution limit of modularity [19], namely, the fact that modularity is blind to modules smaller than a certain scale. This point originates from the bias of modularity

---

[1]Strictly speaking, the normalized Laplacian of a network is $L'_{ij} = A_{ij}/(k_i^{1/2} k_j^{1/2}) - \delta_{ij}$, but L and L' are equivalent by similarity as $L'_{ij} = k_i^{-1/2} L_{ij} k_j^{1/2}$.

towards modules having a certain scale which might not be compatible with the system architecture [66]. This incompatibility also makes modularity inefficient in practical contexts as it may lead to a high degeneracy of its landscape [21], i.e., the existence of several distinct partitions having a modularity close to the optimum, which implies that approximate solutions of the optimization problem are very dissimilar and that a partition derived from modularity optimization has to be considered with caution.

Multi-scale Methods

Different methods have been proposed to go beyond modularity optimization. A first set of methods looks for local maxima of the modularity landscape in order to uncover partitions at different scales [56]. A good example is the so-called Louvain method, which is a greedy method taking advantage of the hierarchical organization of complex networks in order to facilitate the optimization of modularity [7]. This heuristic performs the optimization in a multi-scale way: by comparing the communities first of adjacent nodes, then of adjacent groups of nodes found in the first round, etc. It has been shown in several examples that modularity estimated by this method is close to the optimal value obtained with slower methods but also that intermediate partitions are meaningful and correspond to communities at intermediate resolutions [37]. This approach has the advantage of being fast, but it lacks theoretical foundations and is not able to uncover coarser partitions than those obtained by modularity optimization. Moreover, it may produce hierarchies even when the system is single scale or, worse, completely random [23] (see [37] for a discussion of how to deal with this issue).

Another class of methods is based on multi-scale quality functions. These quality functions incorporate a resolution parameter allowing to tune the characteristic size of the modules in the optimal partition and aim at uncovering modules at the true scale of organization of a network, i.e., not at a scale imposed by modularity optimization. The two most popular multi-scale quality functions are ad hoc, parametric generalizations of modularity. A first quantity is the parametric modularity introduced by Reichardt and Bornholdt [50]:

$$Q_\gamma = \frac{1}{2m} \sum_{C \in \mathcal{P}} \sum_{i,j \in C} \left[A_{ij} - \gamma P_{ij} \right], \tag{7}$$

which is usually defined for the configuration null model $P_{ij} = k_i k_j / 2m$ and mainly consists in changing the effective size of the system $m_{\text{eff}} = m/\gamma$. The optimization of Q_γ leads to larger and larger communities in the optimal partition when γ is decreased. This approach makes use of the size dependence of modularity: because of the factor $1/2m$ in the null model, modularity depends on the total size

of the network and not only on its local properties.[2] Decreasing m_{eff} (increasing γ) increases the expected number of links γP_{ij} between i and j, which makes it less advantageous to assign i and j to the same community (because $A_{ij} - \gamma P_{ij}$ decreases).

An alternative approach proposed by Arenas et al. [4] keeps modularity unchanged but modifies the network by adding self-loops to the original network. This approach therefore consists in optimizing

$$Q_r = Q(A_{ij} + rI_{ij}). \tag{8}$$

As expected, increasing r has a tendency to decrease the size of the communities and the optimal partition of Q_∞ is made of single nodes. Even if increasing γ and r has, qualitatively, the same effect on the characteristic size of the communities, one should keep in mind that Q_γ and Q_r are in general optimized by different partitions, except if the network is regular and the resolution parameters verify $\gamma = 1 + r/\langle k \rangle$. It is also interesting to note that the quality function Eq. (8) was first proposed in order to preserve the eigenvectors of the adjacency matrix, as the eigenvectors of $A_{ij} + rI_{ij}$ and A_{ij} are obviously the same. From a partitioning viewpoint, however, the eigenvectors of A_{ij} do not matter as much as the eigenvectors of the combinatorial Laplacian $L_{ij}^{(C)}$ [18] and the normalized Laplacian L_{ij} [61]. Moreover, modularity is related to the eigenvectors of the Laplacian and not of the adjacency matrix; see Eq. (6). These observations suggest to adapt the unfitting quality function Eq. (8) and to optimize the modularity of a modified adjacency matrix preserving the eigenvectors of L_{ij}. This can readily be done by adding degree-dependent self-loops to the nodes

$$A'_{ij} = A_{ij} + r \frac{k_i}{\langle k \rangle} \delta_{ij}, \tag{9}$$

and by optimizing the quality function

$$Q'_r \equiv Q(A_{ij} + r \frac{k_i}{\langle k \rangle} \delta_{ij}). \tag{10}$$

This quality function is equivalent, up to a linear transformation, to Q_γ for any network, i.e., not only for regular networks, with $\gamma = 1 + r/\langle k \rangle$, thereby providing two alternative interpretations to resolution parameters.

Before closing this section, we should emphasize that many other types of methods have also been proposed to detect the multi-scale modular organization

---

[2]In a nutshell, this size dependence originates from a choice of null model where each pair of nodes i and j can be connected, given a certain number of available links m in the system, whatever the distance between i and j in the network. Local null models where pairs of nodes are randomly connected only within a finite radius of interaction are expected to hinder this effect.

of networks, e.g., the hierarchical extension of the Map formalism [55] whose single-scale version is described below, local algorithms [34], or the modeling of the system by hierarchical random graphs [10].

4 Communities: Dynamics and Function

4.1 Linear Dynamics

The behavior of complex systems is determined not only by the topological organization of their interconnections but also by the dynamical processes taking place among their constituents [6]. A faithful modeling of the dynamics is essential because different dynamical processes may be affected very differently by network topology. A full characterization of such systems thus requires a formalization that encompasses both aspects simultaneously, rather than relying only on the topological adjacency matrix [32]. In the simple case of linear dynamics alone, a broad range of qualitatively different processes can be defined on the same graph, e.g., with the same adjacency matrix A_{ij}. Let us consider the class of linear processes defined by the equation

$$x_{i;t+\tau} = \sum_j B_{ij} x_{j;t} \tag{11}$$

where the evolution of a quantity x_i, associated to node i, is driven by B_{ij}, a matrix somehow related to the adjacency matrix A_{ij}.

In the subclass of *random walk processes* alone, modeling the diffusion of some quantity or information between nodes, Eq. (11) takes the form

$$p_{i;t+\tau} = \sum_j T_{ij} p_{j;t} \tag{12}$$

where $p_{j;t}$ is the probability to observe a walker on node i at time t and where T is the transition matrix whose entry T_{ij} represents the probability to jump from j to i in a time interval of duration τ. T_{ij} is a nonnegative matrix verifying the condition $\sum_i T_{ij} = 1$ to ensure the conservation of probability. Except for those constraints, it is arbitrarily defined on a given graph. Important cases include:

- Discrete-time, unbiased random walk, where

$$T_{ij} = A_{ij}/k_j^{out} \tag{13}$$

- Biased random walks, where

$$T_{ij}^{(\alpha)} = \frac{\alpha_i A_{ij}}{\sum_k \alpha_k A_{kj}}, \tag{14}$$

and where α is an attribute biasing the motion of random walkers towards certain nodes.

- Continuous-time random walks, where walkers perform their jumps asynchronously. In the case of exponentially distributed waiting times between jumps, the transition matrix is

$$T_{ij} = \left(e^{-\tau L}\right)_{ij} \tag{15}$$

and relation (12) can be seen as the formal solution of the rate equation

$$\dot{p}_i = \sum_j \left(\frac{A_{ij}}{k_j^{out}} - \delta_{ij}\right) p_j \equiv -\sum_j L_{ij}\, p_j. \tag{16}$$

A Taylor expansion of (15) in terms of τ clearly shows that the transition matrix accounts for paths of any length on the graph, i.e., each path of length k contributes proportionally to the probability for a walker to perform k jumps in a time interval τ.

It is interesting to note that to each random walk process corresponds a consensus process, in which nodes imitate their neighbors such as to reach a coordinated behavior. In its simplest form, consensus is implemented by the so-called agreement algorithm [67]. Each node i is endowed with a scalar value x_i which evolves as

$$x_{i;t+\tau} = \frac{1}{k_i^{in}} \sum_j A_{ij}\, x_{j;t}. \tag{17}$$

The matrix driving the dynamics now verifies the constraint $\frac{1}{k_i^{in}} \sum_j A_{ij} = 1$, i.e., the value on a node is updated by computing a weighted average of the values on its neighbors. Similarly to the case of random walks, the process (17) can be generalized either by introducing a bias in the weighted average or by introducing a rate at which nodes update their state.

4.2 Flow-Based Approaches

The multi-resolution quality functions defined in Sect. 3 have been successfully tested on multi-scale benchmark and empirical networks [17, 50, 52]. They have the further advantage of being mathematically very similar to modularity and of being optimized by modularity optimization algorithms with minimum code development. Unfortunately, the introduction of a resolution parameter, γ or r, feels like a trick and lacks theoretical ground. In order to define a resolution parameter in a more satisfying way and, as we will see, to provide a more solid foundation to Q_γ and Q_r', we look at communities from a different angle; instead from a combinatorial point of view, where intra-community links are counted as in (2), we investigate

from a dynamical point of view accounting for the interplay between structure and dynamics. In the following, we describe two such quality functions and emphasize how their dynamical nature helps at clarifying the concept of resolution parameter and at solving some of the issues of the aforementioned combinatorial quality functions. In each case, the starting point is the following: a flow taking place on a network is expected to be trapped for long times in good communities before being able to escape [13, 53, 68]. This argument suggests to measure the quality of a partition in terms of the persistence of flows of random walkers on the network.

From now on, we consider a random walk process at equilibrium. By doing so, we assume that the process is ergodic, i.e., any initial configuration asymptotically reaches the unique stationary solution. If this is not the case, standard tricks, e.g., teleportation, can be introduced to make the dynamics ergodic [33].

Stability

A first quality function, called stability [13], consists in estimating the probability that a random walker stays in a community during a certain time interval

$$R(\tau) = \text{(probability for a random walker to be in the}$$

$$\text{same community initially and at time } \tau)$$

$$- \text{(probability for two independent random}$$

$$\text{walkers to be in the same community)} \tag{18}$$

when the system is at equilibrium. To clarify this concept, let us focus on the continuous-time random walk Eq. (15) on an undirected network.[3] By definition, the corresponding stability of a partition is

$$R(\tau) = \sum_{C} \sum_{i,j \in C} \left[\left(e^{-\tau L} \right)_{ij} \pi_j - \pi_i \pi_j \right], \tag{19}$$

where the probability to find a walker on node i at equilibrium is $\pi_i = k_i/2m$ because of the undirectedness of the links. This expression clearly shows that stability depends on time. The quality of a partition is thus measured differently at different time scales and is, in general, maximized by different partitions when time is tuned, thereby leading to a sequence of optimal partitions.

By looking at limiting values of τ, one can show that time acts as a resolution parameter [13, 28]. As time grows, the characteristic size of the communities is thus adjusted to reveal the possible multi-scale organization of the system. In the limit $\tau \to 0$, keeping linear terms in t in the expansion of $R(\tau)$ leads to

$$R(\tau) \approx (1 - \tau) R(0) + \tau Q_{\text{conf}} \equiv Q(t), \tag{20}$$

---

[3]Ergodicity is ensured if the underlying network is non-bipartite and connected.

which is equivalent up to a linear transformation to Q_γ and Q'_r when $P_{ij} = k_i k_j / 2m$ (with $\tau = 1/\gamma$, $\tau = \langle k \rangle / (r + \langle k \rangle)$). The latter multi-resolution quality functions can therefore be seen as a simple linear approximation of $R(\tau)$, which provides a physical interpretation to the resolution parameter r and γ, i.e., the inverse of the time used to explore the network. It is also interesting to note that the configuration null model naturally emerges from the definition of stability and from the dynamics (15). Interestingly, other null models, including the uniform null model, are associated to other random walk processes [28]. We should also stress that this connection with modularity only exists for undirected networks and that stability is radically different from modularity when the network is directed. This difference stems from the fact that detailed balance is not verified at equilibrium for directed networks. In the limit $\tau \to \infty$, making use of the spectral decomposition of L, stability simplifies as

$$R(\tau) \approx \frac{1}{2m} e^{\tau \lambda_2} \sum_C \sum_{i,j \in C} v_{2;i} v_{2;j} \tag{21}$$

where it is assumed that the second dominant eigenvalue λ_2 of L is not degenerate and v_2 is its corresponding (right) eigenvector. $R(\tau)$ is therefore maximized by a partition into two communities in accordance with the normalized Fiedler eigenvector [61].

The Map Equation

An alternative way to measure the trapping of walkers in communities is to adopt a coding perspective and to search for a compact description of trajectories on a network in terms of its communities. To do so, the Map equation method [53, 54] relies on a compression of the description length of a random walk inside and between communities. The underlying principle is that for a strongly modular network, the code for the transitions of the random walker can be efficiently compressed by capitalizing on the presence of the community structure. In this formalism, the movement of the walker is described in terms of two features: first, within each community the movement is encoded assigning a unique code word for each node and a particular exit code word for the community. These are stored in an index codebook that is specific to that community. Second, there is intercommunity codebook with unique code words that describe the movements between different communities. The argument is that a walker will rarely leave a good community and a strong community structure leads to a reduction of the code length since short code words can be reused within each community codebook.

Let us focus again on a random walk process in equilibrium, and consider a partition of the network into communities. The probability of leaving a particular community C at equilibrium is denoted by $q_{C\curvearrowright} = \sum_{i \in C} \sum_{j \notin C} T_{ji} \pi_i$, with $\pi_i = k_i/2m$ if the graph is undirected. The Map equation for this partition is

$$L(\mathrm{M}) = q_\curvearrowright H(\mathcal{Q}) + \sum_C p_\circlearrowleft^C H(\mathcal{C}^C),$$ (22)

where H is the Shannon entropy. The first term of this equation is the weighted entropy associated with the intercommunity movement of the random walker, where the weighting factor $q_\curvearrowright = \sum_C q_{C\curvearrowright}$ is the overall probability of leaving a community. For the Map coding scheme $H(\mathcal{Q})$ is the minimal average per-step code length to describe the transition of the walker between different communities

$$H(\mathcal{Q}) = -\sum_C \frac{q_{C\curvearrowright}}{q_\curvearrowright} \log_2 \left(\frac{q_{C\curvearrowright}}{q_\curvearrowright} \right).$$ (23)

In the second term, each term $p_\circlearrowleft^C H(\mathcal{C}^C)$ is the weighted average per-step code length needed to describe the movement of the random walker within (and leaving) community C. The entropy $H(\mathcal{C}^C)$ is given analogously by

$$H(\mathcal{C}^C) = -\frac{q_{C\curvearrowright}}{p_\circlearrowleft^C} \log_2 \left(\frac{q_{C\curvearrowright}}{p_\circlearrowleft^C} \right) - \sum_{i \in C} \frac{\pi_i}{p_\circlearrowleft^C} \log_2 \left(\frac{\pi_i}{p_\circlearrowleft^C} \right),$$ (24)

where $p_\circlearrowleft^C = q_{C\curvearrowright} + \sum_{j \in C} \pi_j$ is the associated weighting factor, describing the probability to use a code word from the codebook of community C. The optimal partition of a network is found by minimizing the Map equation, e.g., by finding the partition minimizing the description length of a random walk on the graph. The Map equation is usually defined for the discrete-time unbiased random walk (13) and is thus based on one-step transitions on the graph. However, it has been shown that the corresponding quality function suffers from some limitations, such as the fact that it is not able to uncover communities in multi-scale networks and that it exhibits a field-of-view limit that can result in undesirable over-partitioning when communities are highly structured [57]. This issue can be addressed by adopting a similar approach as in the previous section, namely, in introducing time explicitly by letting the random walker explore the network over a tunable amount of time [58]. The resulting Markov time sweeping induces a dynamical zooming across scales that can reveal community structure at different scales and circumvent field-of-view limit. In practice, defining the Map equation for the continuous-time random walk (15) leads to time-dependent leaving probabilities $q_{C\curvearrowright}(\tau)$ given as

$$q_{C\curvearrowright}(\tau) = \sum_{i \in C} \sum_{j \notin C} \left(e^{-\tau L} \right)_{ji} \pi_i.$$ (25)

With increasing time, this leaving probability and the associated cost for encoding communities increase. Looking at limiting values of τ suggests again that time acts as a resolution parameter [58]. In the limit of vanishing times, when the leaving probabilities go to zero, the Map equation is minimized setting each node in its own community. In the limit $\tau \to \infty$, in contrast, the Map equation finds a partition

made of one single module due to the mean-field nature of the dynamics. For intermediate times, the partitions optimizing the Map equation are expected to be made of modules of varying size, as confirmed by numerical simulations [58].

Discussion

As we have discussed in the previous subsections, flow-based approaches have the interesting property to incorporate a natural resolution parameter, time, allowing to explore the graph through paths of different length. However, flow-based approaches offer other advantages that make them an interesting alternative to combinatorial approaches. First, stability is based on flows of probability on the graph and therefore captures how the global structure of the system constrain patterns of flows, while quantities such as modularity focus on pairwise interactions and are blind to such patterns, thereby neglecting important aspects of the network architecture [53]. This difference is particularly marked when the network is directed, when the equilibrium solution of the process depends on the global organization of the process [28]. Flow-based approaches also offer the flexibility to chose a random process properly modeling how, e.g., information or energy flows on the graph and thus to tailor the quality function according to the network nature. For instance, this can be done by using the biased random walk process (14) in systems where unbiased random walks are not realistic models. Important examples include the Internet and traffic networks, where a bias is necessary to account for local search strategies and navigation rules [69].

Partitions at different values of τ are found independently by optimizing either stability or the Map equation, thereby producing a sequence of partitions that are optimal at different scales. However, one expects that only a small number of these partitions are significant, which raises another question: how can one select the most significant partitions, or equivalently the most significant scales of description of the network? It is ironical to note that we are thus confronted with a problem similar to the one that initially led to the definition of modularity.[4] In order to address this problem, it has recently been proposed to look for robust partitions [17, 24, 29, 37]. In practice, the problem is slightly modified, by changing either the resolution parameter, the graph, or the optimization algorithm, and variability among the uncovered partitions is considered. In each case, robustness is related to the ruggedness of the quality function landscape [21]. Lack of robustness corresponds to high degeneracy, namely, to the existence of incompatible partitions that are local maxima of the quality function such that partitions are strongly altered by a slight modification of the optimization problem. Significant partitions are uncovered by identifying values of the resolution parameter where these measures of robustness are significantly low. In the case of the Map equation, an alternative

---

[4]Modularity was first proposed to find the best partition in a nested hierarchy of possible community divisions [40].

approach is based on an information theoretic indicator for the reliability of the Map equation [58], by measuring the gap between the optimal code compression and the compression achieved by the Map coding strategy. Relevant values of the resolution parameter are then signaled by a low compression gap.

Before concluding, let us mention recent works where the ideas developed in this chapter have been successfully tested on empirical networks of different nature and on benchmark networks [12–14, 28, 29, 57, 58]. We should also stress that the multi-scale methods described here are expected to fail when the system does not exhibit any scale of description, e.g., when the size of the communities is heterogeneously distributed [35]. In that case, other types of methods might reveal more successful, for instance, local methods in contrast with global ones [36].

5 Conclusion

In this chapter, we have focused on the detection of nonoverlapping modules in multi-scale networks. These networks are made of different levels of organization and are typically (but not necessarily) hierarchical, in the sense that the system is made of modules, which themselves are made of sub-modules, etc. In order to account for the presence of multiple levels of organization in the system, multi-scale methods have been introduced that incorporate a resolution parameter allowing to zoom in and out and to focus on the appropriate level of resolution. A class of popular combinatorial quality functions incorporate a resolution parameter to Newman–Girvan modularity in order to adjust the characteristic size of the modules and to uncover the true modular organization of a network. Unfortunately, these multi-resolution quality functions exhibit the same type of limitation as modularity when the resolution parameter is fixed [26]. Moreover, the introduction of a resolution parameter lacks a sound mathematical justification. Here, we argue for the use of flow-based approaches instead of combinatorial ones for several reasons. We have shown that using the trajectories of random walker on the graph defines quality functions with a natural resolution parameter, *time*. As *time* increases, the diffusive process involves longer trajectories and explores further afield the structure of the graph, resulting in the detection of the modular structure across scales. As we have shown, this general scheme can be used in a variety of methods, such as stability and the Map equation. Flow-based approaches have the further advantage of properly taking into account dynamical flows taking place on the graph. This property reveals crucial to provide a faithful characterization of the system whenever complex interdependences between network subunits are generated by patterns of flow, e.g., information in social networks or passengers in airline networks.

References

[1] D.M. Abrams, R. Mirollo, S.H. Strogatz, D.A. Wiley, Solvable model for chimera states of coupled oscillators. Phys. Rev. Lett. **101**, 084103 (2008)

[2] Y.Y. Ahn, J.P. Bagrow, S. Lehmann, Communities and hierarchical organization of links in complex networks. Nature **466**, 761 (2010)

[3] A. Arenas, A. Díaz-Guilera, C.J. Pérez-Vicente, Synchronization reveals topological scales in complex networks. Phys. Rev. Lett. **96**, 114102 (2006)

[4] A. Arenas, A. Fernández, S. Gómez, Analysis of the structure of complex networks at different resolution levels. New J. Phys. **10**, 053039 (2008)

[5] M.J. Barber, Modularity and community detection in bipartite networks. Phys. Rev. E **76**, 066102 (2007)

[6] M. Batty, K.J. Tinkler, Symmetric structure in spatial and social processes. Env. Plan. B **6**, 3 (1979)

[7] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. J. Stat. Mech. P10008 (2008)

[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics. Phys. Rep. **424**, 175–308 (2006)

[9] U. Brandes, D. Delling, M. Gaertler, R. Goerke, M. Hoefer, Z. Nikoloski, D. Wagner, Maximizing modularity is hard. arXiv:physics/0608255

[10] A. Clauset, C. Moore, M.E.J. Newman, Hierarchical structure and the prediction of missing links in networks. Nature **453**, 98–101 (2008)

[11] L. Danon, J. Duch, A. Diaz-Guilera, A. Arenas, Comparing community structure identification. J. Stat. Mech. P09008 (2005)

[12] A. Delmotte, E.W. Tate, S.N. Yaliraki, M. Barahona, Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin-myosin light chain interaction. Phys. Biol. **8**, 055010 (2011)

[13] J.-C. Delvenne, S. Yaliraki, M. Barahona, Stability of graph communities across time scales. Proc. Natl. Acad. Sci. USA **107**, 12755–12760 (2010)

[14] J.-C. Delvenne, M.T. Schaub, S.N. Yaliraki, M. Barahona, The stability of a graph partition: a dynamics-based framework for community detection, in *Dynamics on and of Complex Networks, vol. 2: Applications to Time-Varying Dynamical Systems*, ed. by A. Mukherjee, M. Choudhury, F. Peruani, N. Ganguly, B. Mitra (Springer, New York, 2013)

[15] T.S. Evans, Complex networks. Contemp. Phys. **45**, 455 (2004)

[16] T.S. Evans, R. Lambiotte, Line graphs, link partitions and overlapping communities. Phys. Rev. E **80**, 016105 (2009)

[17] D.J. Fenn, M.A. Porter, M. McDonald, S. Williams, N.F. Johnson, N.S. Jones, Dynamic communities in multichannel data: an application to the foreign exchange market during the 2007–2008 credit crisis. Chaos **19**, 033119 (2009)

[18] M. Fiedler, A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. Czech. Math. J. **25**, 619–633 (1975)

[19] S. Fortunato, M. Barthélemy, Resolution limit in community detection. Proc. Natl. Acad. Sci. USA **104**, 36 (2007)

[20] S. Fortunato, Community detection in graphs. Phys. Rep. **486**, 75–174 (2010)

[21] B.H. Good, Y.-A. de Montjoye, A. Clauset, The performance of modularity maximization in practical contexts. Phys. Rev. E **81**, 046106 (2010)

[22] T. Gross, B. Blasius, Adaptive coevolutionary networks: a review. J. R. Soc. Interface **5**, 259–271 (2008)

[23] R. Guimerá, M. Sales, L.A.N. Amaral, Modularity from fluctuations in random graphs and complex networks. Phys. Rev. E **70**, 025101 (2004)

[24] B. Karrer, E. Levina, M.E.J. Newman, Robustness of community structure in networks. Phys. Rev. E **77**, 046119 (2008)

[25] N. Kashtan, U. Alon, Spontaneous evolution of modularity and network motifs. Proc. Natl. Acad. Sci. USA **102**, 13773–13778 (2005)

[26] J.M. Kumpula, J. Saramäki, K. Kaski, J. Kertész, Limited resolution in complex network community detection with Potts model approach. Eur. Phys. J. B **56**, 41–45 (2007)

[27] R. Lambiotte, M. Ausloos, J.A. Holyst, Majority model on a network with communities. Phys. Rev. E **75**, 030101 (2007)

[28] R. Lambiotte, J.-C. Delvenne, M. Barahona, Laplacian dynamics and multiscale modular structure in networks. arXiv:0812.1770

[29] R. Lambiotte, Multi-scale modularity in complex networks, in *Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 546–553, 2010

[30] R. Lambiotte, P. Panzarasa, Communities, knowledge creation and information diffusion. J. Informetrics **3**, 180–190 (2009)

[31] R. Lambiotte, J.C. Gonzalez-Avella, On co-evolution and the importance of initial conditions. Phys. A **390**, 392–397 (2011)

[32] R. Lambiotte, R. Sinatra, J.-C. Delvenne, T.S. Evans, M. Barahona, V. Latora, Flow graphs: interweaving dynamics and structure. Phys. Rev. E **84** 017102 (2011)

[33] R. Lambiotte, M. Rosvall, Ranking and clustering of nodes in networks with smart teleportation. Phys. Rev. E **85**, 056107 (2012)

[34] A. Lancichinetti, S. Fortunato, J. Kertész, Detecting the overlapping and hierarchical community structure of complex networks. New J. Phys. **11**, 033015 (2009)

[35] A. Lancichinetti, S. Fortunato, Community detection algorithms: a comparative analysis. Phys. Rev. E **80**, 056117 (2009)

[36] A. Lancichinetti, F. Radicchi, J.J. Ramasco, S. Fortunato, Finding statistically significant communities in networks. PLoS ONE **6**, e18961 (2011)

[37] D. Meunier, R. Lambiotte, A. Fornito, K.D. Ersche, E.T. Bullmore, Hierarchical modularity in human brain functional networks. Front. Neuroinform. **3**, 37 (2009)

[38] D. Meunier, R. Lambiotte, E.T. Bullmore, Modular and hierarchically modular organization of brain networks. Front. Neurosci. **4**, 200 (2010)

[39] M.E.J. Newman, The structure and function of complex networks. SIAM Rev. **45**, 167 (2003)

[40] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks. Phys. Rev. E **69**, 026113 (2004)

[41] M.E.J. Newman, Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA **103**, 8577 (2006)

[42] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E **74**, 036104 (2006)

[43] V. Nicosia, G. Mangioni, V. Carchiolo, M. Malgeri, Extending the definition of modularity to directed graphs with overlapping communities. J. Stat. Mech. P03024 (2009)

[44] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society. Nature **435**, 814–818 (2005)

[45] R.K. Pan, S. Sinha, Modular networks with hierarchical organization: The dynamical implications of complex structure. Pramana J. Phys. **71**, 331–340 (2008)

[46] R.K. Pan, S. Sinha, Modularity produces small-world networks with dynamical time-scale separation. Europhys. Lett. **85**, 68006 (2009)

[47] R.K. Pan, N. Chatterjee, S. Sinha, Mesoscopic organization reveals the constraints governing Caenorhabditis elegans nervous system. PLoS ONE **5**, e9240 (2010)

[48] P. Pons, M. Latapy, Computing communities in large networks using random walks. J. Graph Algorithm Appl. **10**, 191 (2006)

[49] M.A. Porter, J.-P. Onnela, P.J. Mucha, Communities in networks. Not. Am. Math. Soc. **56**, 1082–1097 (2009)

[50] J. Reichardt, S. Bornholdt, Statistical mechanics of community detection. Phys. Rev. E **74**, 016110 (2006)

[51] P.A. Robinson, J.A. Henderson, E. Matar, P. Riley, R.T. Gray, Dynamical reconnection and stability constraints on cortical network architecture. Phys. Rev. Lett. **103**, 108104 (2009)

[52] P. Ronhovde, Z. Nussinov, Multiresolution community detection for megascale networks. Phys. Rev. E **80**, 016109 (2009)

[53] M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. USA **105**, 1118 (2008)

[54] M. Rosvall, D. Axelsson, C.T. Bergstrom, The map equation. Eur. Phys. J. Spec. Top. **178**, 13–23 (2009)

[55] M. Rosvall, C.T. Bergstrom, Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PLoS ONE **6**, e18209 (2011)

[56] M. Sales-Pardo, R. Guimerá, A. Moreira, L.A.N. Amaral, Extracting the hierarchical organization of complex systems. Proc. Natl. Acad. Sci. USA **104**, 15224 (2007)

[57] M.T. Schaub, J.-C. Delvenne, S.N. Yaliraki, M. Barahona, Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit. PLoS ONE **7**, e32210 (2012)

[58] M.T. Schaub, R. Lambiotte, M. Barahona, Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. Phys. Rev. E **86**, 026112. Published 21 August 2012

[59] M. Shanahan, Dynamical complexity in small-world networks of spiking neurons. Phys. Rev. E **78**, 041924 (2008)

[60] M. Shanahan, Metastable chimera states in community-structured oscillator networks. Chaos **20**, 013108 (2010)

[61] J. Shi, J. Malik, Normalized cuts and image segmentation. IEEE Trans. Patt. Anal. Mach. Intell. **22**, 888 (2000)

[62] H.A. Simon, The architecture of complexity. Proc. Am. Phil. Soc. **106**, 467–482 (1962)

[63] H.A. Simon, Near-decomposability and complexity: How a mind resides in a brain, in *The Mind, the Brain, and Complex Adaptive Systems*, ed. by H. Morowitz, J. Singer (Addison-Wesley, Reading, 1995), pp. 25–43

[64] O. Sporns, G. Tononi, G. Edelman, Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. Cereb Cortex **10**, 127–141 (2000)

[65] O. Sporns, D. Chialvo, M. Kaiser, C.C. Hilgetag, Organization, development and function of complex rain networks. Trends Cognit. Sci. **8**, 418–425 (2004)

[66] V.A. Traag, P. Van Dooren, Y. Nesterov, Narrow scope for resolution-limit-free community detection. Phys. Rev. E **84**, 016114 (2011)

[67] J.N. Tsitsiklis, Problems in decentralized decision making and computation. Ph.D. thesis, MIT (1984)

[68] S. Van Dongen, Graph clustering via a discrete uncoupling process. SIAM. J. Matrix Anal. Appl. **30**, 121–141 (2008)

[69] W.X. Wang, B.-H. Wang, C.-Y. Yin, Y.-B. Xie, T. Zhou, Traffic dynamics based on local routing protocol on a scale-free network. Phys. Rev. E **73**, 026111 (2006)

[70] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks. Nature **393**, 440–442 (1998)