

4

Different Approaches to Community Detection

Martin Rosvall¹, Jean-Charles Delvenne², Michael T. Schaub^{3,4}, and Renaud Lambiotte⁴

¹Umeå University

²Université Catholique de Louvain

³Massachusetts Institute of Technology

⁴University of Oxford

This chapter is an extended version of The many facets of community detection in complex networks, Appl. Netw. Sci. 2: 4 (2017) by the same authors.

4.1 Introduction

A precise definition of what constitutes a community in networks has remained elusive. Consequently, network scientists have compared community detection algorithms on benchmark networks with a particular form of community structure and classified them based on the mathematical techniques they employ. However, this comparison can be misleading because apparent similarities in their mathematical machinery can disguise different reasons for why we would want to employ community detection in the first place. Here we provide a focused review of these different motivations that underpin community detection. This problem-driven classification is useful in applied network science, where it is important to select an appropriate algorithm for the given purpose. Moreover, highlighting the different approaches to community detection also delineates the many lines of research and points out open directions and avenues for future research.

While research related to community detection dates back to the 1970s in mathematical sociology and circuit design [21, 46], Newman's and Girvan's work on modularity in complex

systems just over ten years ago revitalized the field of community detection, making it one of the main pillars of network science research [53, 54]. The promise of community detection, that we can gain a deeper understanding of a system by discerning important structural patterns within a network, has spurred a huge number of studies in network science. However, it has become abundantly clear by now that this problem has no canonical solution. In fact, even a general definition of what constitutes a community is still lacking. The reasons for this are not only grounded in the computational difficulties of tackling community detection. Rather, various research areas view community detection from different perspectives, illustrated by the lack of a consistent terminology: “network clustering”, “graph partitioning”, “community”, “block” or “module detection” all carry slightly different connotations. This jargon barrier creates confusion, as readers and authors have different preconceptions and intuitive notions are not made explicit.

We argue that community detection should not be considered as a well-defined problem, but rather as an umbrella term with many facets. These facets emerge from different goals and motivations for what it is about the network that we want to understand or achieve, and lead to different perspectives on how to formulate the problem of community detection. It is critically important to be aware of these underlying motivations when selecting and comparing community detection methods. Thus, rather than an in-depth discussion of the technical details of different algorithmic implementations [16, 26, 28, 47, 52, 58, 72, 82], here we focus on the conceptual differences between different perspectives on community detection.

By providing a problem-driven classification, however, we do *not* argue that the different perspectives are unrelated. In fact, in some situations, different mathematical problem formulations can lead to similar algorithms and methods, and the different perspectives can offer valuable insights. For example, for undirected networks, optimizing the objective function modularity [54], initially proposed from a clustering perspective, can be interpreted as optimizing both a particular stochastic block model [50] and an auto-correlation measure of a particular diffusion process on the networks [20]. In other situations, however, such relationships disappear.

While some perspectives arguably are more principled than others, we do not assert that there is a particular perspective that is *a priori* better suited for any given network. In fact, as in data clustering [31], no one method can consistently perform the best on all kinds of networks [59]. Community detection is an unsupervised learning task that is blind to a researcher’s intent with the analysis. Accordingly, to understand a particular method’s usefulness, we must take the researcher’s interest in the communities into context [80].

In the following, we unfold different aims underpinning community detection – in a relaxed form that includes assortative as well as disassortative group structures with dense and sparse internal connections, respectively – and discuss how the resulting problem perspectives relate to various applications. We focus on four broad perspectives that have served as motivation for community detection in the literature: (i) the cut-based perspective minimizes a constraint such as the number of links between groups of nodes, (ii) the clustering perspective maximizes internal density in groups of nodes, (iii) the stochastic block model perspective identifies groups of nodes in which nodes are stochastically equivalent, and (iv) the dynamical perspective identifies groups of nodes in which flows stay for a relatively long time such that they form building blocks of dynamics on networks (see Figure 4.1). While this categorization is not unique, we believe that it can help clarify concepts about community detection and serve as a guide to determining the appropriate method for a particular purpose.

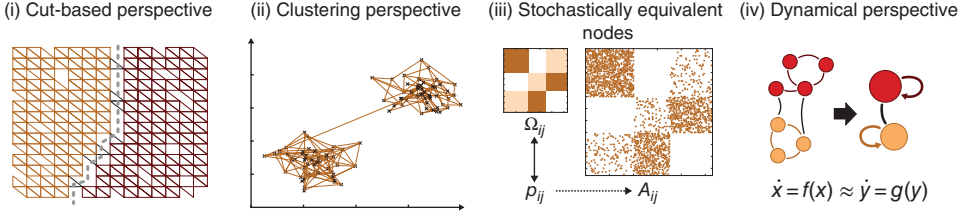


Figure 4.1 Schematic of four different approaches to community detection. (i) The cut-based perspective aims at minimizing the number of links between groups of nodes, independently of their intrinsic structure. (ii) The clustering perspective produces groups of densely connected nodes. (iii) The stochastic equivalence perspective looks for groups in which nodes are stochastically equivalent, typically inferred through a generative statistical network model. (iv) The dynamical perspective focuses on the impact of communities on dynamical processes and searches for dynamically relevant coarse-grained descriptions.

4.2 Minimizing Constraint Violations: the Cut-based Perspective

An early network partitioning application was circuit layout and design [3, 26]. This application spurred development of the now classical Kernighan–Lin algorithm [39] and the work by Donath and Hoffmann [21, 22], who were among the first to suggest the use of eigenvectors for network partitioning. For example, we might be confronted with a network that describes the signal flows between different components of a circuit. To design the circuit in an efficient way, our goal is now to partition the network into a fixed number of approximately equally sized groups for balanced load with a small number of edges between those groups for minimal communication overhead. The edges that run between the groups are commonly denoted as the cut. To design the most efficient circuit, our aim is thus to minimize this cut with more or less balanced groups.

To make this more precise, let us consider one specific variant of this scheme, known as *ratio cut* [32]. Let us denote the adjacency matrix of an undirected network \mathcal{N} with n nodes by A , where $A_{uv} = 1$ if there is a connection from node u to node v , and $A_{uv} = 0$ if there is no connection. We can now write the problem of optimizing the ratio cut for a bipartition of all nodes \mathcal{V} into two communities \mathcal{V}_1 and $\mathcal{V}_2 = \mathcal{V} \setminus \mathcal{V}_1$ as follows [32, 79]:

$$\min_{\mathcal{V}_1} \text{RatioCut}(\mathcal{V}_1, \mathcal{V}_2) := \min_{\mathcal{V}_1} \sum_u \frac{\text{cut}(\mathcal{V}_1, \mathcal{V} \setminus \mathcal{V}_1)}{|\mathcal{V}_1|}, \quad (4.1)$$

where $\text{cut}(\mathcal{V}_1, \mathcal{V}_2) := \sum_{u \in \mathcal{V}_1, v \in \mathcal{V}_2} (A_{uv} + A_{vu})/2$ is the sum of the possibly weighted edges between the two vertex sets $\mathcal{V}_1, \mathcal{V}_2$. Related problem formulations also occur in the context of parallel computations and load scheduling [64, 77], where approximately equally sized portions of work are to be sent to different processors, while keeping the dependencies between those tasks minimal. Further applications include scientific computing [64, 77], where partitioning algorithms divide the coordinate meshes when discretizing and solving partial differential equations. Image segmentation problems may also be phrased in terms of cut-based measures [76, 79].

Investigating these types of problems has led to many important contributions to partitioning networks, in particular in relation to spectral methods. The connection between spectral algorithms and cut-based problem formulations arises naturally by considering relaxations of the original, combinatorially hard discrete optimization problems, such as Equation (4.2), or other

related objective functions such as the average or normalized cuts. This can be best seen when rewriting the above optimization problem as follows:

$$\min_f f^T L f \quad (4.2)$$

$$\text{subject to } f \perp \mathbf{1} \quad \|f\| = \sqrt{n} \quad (4.3)$$

$$\text{where } f_u := \begin{cases} -\sqrt{|\mathcal{V}_2|/|\mathcal{V}_1|} & \text{if } u \in \mathcal{V}_1 \\ \sqrt{|\mathcal{V}_1|/|\mathcal{V}_2|} & \text{if } u \in \mathcal{V}_2 \end{cases} \quad (4.4)$$

Here the Laplacian matrix of the network has been defined as $L = D - A$, where D is the diagonal degree matrix with $D_{uu} = \sum_v A_{uv}$. Fiedler realized already in the 1970s that the second smallest eigenvalue of the Laplacian matrix is associated with the connectivity of the network, and that the associated eigenvector thus can be used to compute spectral bi-partitions [24, 25]. Such spectral ideas led to many influential algorithms and methods; see, for example, von Luxburg [79] for a tutorial on spectral algorithms.

In this cut-based problem formulation, there is no specification as to how the identified groups in the partition should be connected internally. While the implicit constraint is that the groups must not split into groups with an even smaller cut, there is no specification that the groups of nodes should be densely connected internally. Indeed, the type of networks considered in the context of cut-based partitions are often of a mesh- or grid-like form, for which several guarantees can be given in terms of the quality of the partitions obtained by spectral algorithms [77]. While such non-dense groupings emerging from the analysis of non-clique structures [73] can also be dynamically relevant (see section 4.5), they are likely missed when employing a community notion that focuses on finding dense groupings, as discussed next.

4.3 Maximizing Internal Density: the Clustering Perspective

A different motivation for community detection arises in the context of data clustering. We use the term clustering, which can have many definitions, in the following sense: for a set of given data points in a possibly high-dimensional space, the goal is to partition the points into a number of groups such that points within a group are close to or similar to each other in some sense, and points in different groups are more distant from each other. To achieve this goal, one often constructs a proximity or similarity network between the points and tries to group together nodes that are closer to each other than they are to the rest of the network. This approach results in a form of community detection problem where the closeness between nodes is described by the presence and weight of the edges between them.

Although minimizing the cut size and maximizing the internal number of links are closely related, there are differences pertaining to the typical constraints and search space associated with these objective functions. First, when employing a clustering perspective, there is normally no *a priori* information about the number of groups we are looking for. Second, we do not necessarily require the groups to be balanced in any way; rather we would like to find an optimal split into densely knit groups irrespective of their relative sizes.

Unsurprisingly, finding an optimal clustering is a computationally difficult problem. Further, as Kleinberg has shown [40], there are no clustering algorithms that satisfy a certain set of intuitive properties we might require from a clustering algorithm in continuous spaces. Similar problems also arise in the discrete setting for clustering of networks [13].

Nevertheless, there exists a large number of methods that follow a clustering-like paradigm and separate the nodes of a network into cohesive groups of nodes, often by optimizing a quality function. An important clustering metric in this context is the so-called conductance [4, 37, 41, 78]. Optimizing the global conductance was introduced as a way to produce a global bi-partition similarly to the two-way ratio-cut. However, this quantity has been successfully employed more recently as a local quality function to find localized clusters around one or more seed nodes. The local conductance of a set of nodes $\mathcal{V}_q \subset \mathcal{V}$ can be written

$$\phi(\mathcal{V}_q) := \frac{\sum_{u \in \mathcal{V}_q, v \notin \mathcal{V}_q} A_{uv}}{\min\{\text{vol}(\mathcal{V}_q), \text{vol}(\mathcal{V} - \mathcal{V}_q)\}}, \quad (4.5)$$

where $\text{vol}(\mathcal{V}_q) := \sum_{u \in \mathcal{V}_q} \sum_v A_{uv}$ is the total degree of the nodes in set \mathcal{V}_q , commonly called its volume in analogy with geometric objects. Interestingly, it has been shown that, in specific contexts, the conductance can be a good predictor of some latent group structures in real-world applications [84].

Moreover, a local perspective on community detection has two appealing properties. First, the definition of a cluster does not depend on the global network structure but only on the relative local density. Second, only a portion of a network needs to be accessed, which is advantageous if there are computational constraints in using large networks, or we are only interested in a particular subsystem. In such cases, we would like to avoid having to apply a method to the whole network in order to find, for example, the cluster containing a particular node in the network.

The Newman–Girvan modularity [53, 54] is arguably one of the most common clustering measures used in the literature and was originally proposed from the clustering perspective discussed here. It is a global quality function and aims to find the community structure of the network as a whole. Given a partition $\mathbf{C} = \{\mathcal{V}_1, \dots, \mathcal{V}_k\}$ of a network into k groups, the modularity of \mathbf{C} can be written as:

$$Q(\mathbf{C}) := \frac{1}{2m} \sum_{q=1}^k \sum_{u,v \in \mathcal{V}_q} \left[A_{uv} - \frac{d_u d_v}{2m} \right], \quad (4.6)$$

where $d_u = \sum_v A_{uv}$ is the degree of node u and $2m = \sum_u d_u$ is the total weight of all edges in the network. By optimizing the modularity measure over the space of all partitions, one aims to identify groups of nodes that are more densely connected to each other than one would expect from a statistical null model of the network. This statistical null model is commonly chosen to be the configuration model with preserved degree sequence.

However, a by-product of this choice of a global null-model is the tendency of modularity to balance the size of the groups in terms of their total connectivity. While different variants of modularity aim to account for this effect [26], it means modularity can be interpreted as a trade-off between a cut-based measure and an entropy [20]. Modularity is typically optimized with spectral or greedy algorithms [11, 26, 51]. While there are problems with modularity, such as its resolution limit [27] and other spurious effects [27, 29, 30, 44], the general idea has triggered researchers to develop a plethora of algorithms that follow a similar strategy [26]. Several works have addressed some of the shortcomings, by incorporating a resolution parameter, for example, or by explicitly accounting for the density inside each group [14, 15]. In practice, however, less seems to beat more and the original formulation of modularity remains the most widely used.

4.4 Identifying Structural Equivalence: the Stochastic Block Model Perspective

By grouping *similar* nodes that link to similar nodes within communities, we constrain ourselves to finding *assortative* group structure [28]. While we may also have hierarchical clusters with clusters of clusters, etc., such an assortative structural organization is too restrictive if we want to define groups based on more general connectivity patterns that include disassortative communities with weaker interactions within rather than between communities.

In social network analysis, a common goal is to identify nodes within a network that serve a similar structural role in terms of their connectivity profile. Accordingly, nodes are similar if they share the same kind of connection patterns to other nodes [46]. This idea is captured in concepts such as *regular equivalence*, which states that nodes are regularly equivalent if they are equally related to equivalent others [23, 33]. The first algorithms for identifying groups of “approximately equivalent” nodes were deterministic and permuted adjacency matrices to reveal block structures in so-called block models [6, 81].

A relaxation of regular equivalence is *stochastic equivalence* [34], where nodes are equivalent if they connect to equivalent nodes with equal probability. The stochastic formulation generalizes observations and forms generative models, which can be used for prediction. Because of this advantage over non-stochastic formulations, we focus on stochastic equivalence.

One of the most popular techniques to model and detect stochastically equivalent relationships in network data is to use stochastic block models (SBMs) [34, 56] and associated inference techniques. These models have their roots in the social networks literature [5, 34], and provide a flexible framework for modeling block structures within a network. When considering block models, we are interested in identifying node groups such that nodes within a community connect to nodes in other communities in an “equivalent way” [28].

Consider a network composed of n nodes divided into k classes. The standard SBM is defined by a set of node class labels and the affinity matrix Ω . More precisely, the link probability between two nodes u, v belonging to class c_u and c_v is given by:

$$p_{uv} := \mathbb{P}(A_{uv}) = \Omega_{c_u c_v}.$$

Under an SBM, nodes within the same class share the same probability of connecting to nodes of another class. This is the mathematical formulation of having stochastically equivalent nodes within each class. Finding the latent groups of nodes in a network now amounts to inferring the model parameters that provide the best fit for the observed network. That is, find the SBM with the highest likelihood of generating the data.

The standard SBM assumes that the expected degree of each node is a Poisson binomial random variable, a binomial random variable with possibly non-identical success probabilities in each trial. Because inferring the most likely SBM typically results in grouping nodes based on their degree in empirical networks with broad degree distributions, it can be advantageous to include a degree-correction into the model. In the degree corrected SBM [38], the probability p_{uv} that a link will appear between two nodes u, v depends both on their class labels c_u, c_v and their respective degree parameters d_i, d_j (each entry A_{ij} might be a Bernoulli or a Poisson random variable such as in [38]):

$$p_{uv} \sim d_u d_v \Omega_{c_u c_v}.$$

Thus, while edges in real-world networks tend to be correlated with effects such as triadic closure [26], by construction edges are conditionally independent random variables in SBMs.

Moreover, most common SBMs are defined for unweighted networks or networks with integer weights by modeling the network as a multi-graph. Though generalizations are available [2, 61], this is still a less studied area.

In contrast to the notions of community considered above, with stochastic equivalence we are no longer interested in maximizing some internal density or minimizing a cut. To see this, consider a bipartite network that from a cut- or density-based perspective contains no communities. From the stochastic equivalence perspective, however, we would say that this network contains two groups because nodes in each set only connect to nodes in the other set. When adopting an SBM to detect such structural organization of the links, we explicitly adopt a statistical model for the networks. The network is essentially an instance of an ensemble of possible networks generated from such a model.¹

This model-based approach comes with several advantages. First, by defining the model, we effectively declare what is signal and what is noise in the data under the SBM. We can thus provide a statistical assessment of the observed data with, for example, p -values under the SBM. In other words, we can identify patterns that cannot be reasonably explained from density fluctuations of edges inherent to any realization of the model. Second, we are able, for example, to generate new networks from our model with a similar group structure, or predict missing edges and impute data. Third, we can make strong statements about the detectability of groups within a network. For example, precise criteria specify when any algorithm can recover the planted group structure for a network created by an SBM [18, 49]. By fitting an SBM to an observed adjacency matrix, it is possible to recover such a planted group structure down to its theoretical limit [48, 49]. These criteria apply to networks generated with SBMs and not real networks in general, in which case we do not know what kind of process created the network [59]. It is nevertheless a remarkable result since it highlights the fact that there are networks with undetectable block patterns.

Moreover, this model-based approach also offers ways to estimate the number of communities from the data by some form of model selection, including hypothesis testing [10], spectral techniques [42, 70], the minimum description length principle [60], or Bayesian inference [83] (see Chapter 11).

Finally, the generative nature of SBMs also makes them well suited for constructing benchmark networks. As a consequence, many benchmark networks proposed in the literature, such as the commonly used Lancichinetti–Fortunato–Radicchi (LFR) benchmarks [45], are specific types of SBMs. Results on these benchmark networks should therefore be taken for what they are: the ability to recover the underlying group structure of specific types of SBM-generated networks. For example, sparse networks without any underlying group structure still can contain meaningful dynamical building blocks.

4.5 Identifying Coarse-grained Descriptions: the Dynamical Perspective

Let us now consider a fourth alternative motivation for community detection, focusing on the processes that take place on the network. All notions of community outlined above are effectively structural in the sense that they are mainly concerned with the composition of

¹ This ensemble assumption is also reflected in the modularity formalism, where the observed network is compared to a null model.

the network itself or its representation as an adjacency matrix. However, in many cases one of the main reasons to apply tools from network science is to understand the *behavior* of a system. While the topology of a system puts constraints on the dynamics that can take place on the network, the network topology alone cannot explain the system behavior. For example, instead of finding a coarse-grained description of the adjacency matrix, we might be interested in finding a coarse-grained description of the dynamics acting on top of the network with multi-step paths beyond the nearest neighbors.

Take air traffic as an example. An airline network, with weighted links connecting cities according to the number of flights between them, can offer some interesting insights about air traffic. For instance, in the US air traffic network based on the number of flying passengers, Las Vegas and Atlanta form two major hubs. However, if we focus instead on the passenger flows based on actual multi-leg itineraries, the two cities show very different behaviors: Las Vegas is a tourist destination and typically the final destination of itineraries, whereas Atlanta is often a transfer hub to other final destinations [62, 69]. Thus, these airports play dynamically quite different roles in the network. Focusing on interconnection patterns alone can give an incomplete picture if we are interested in the dynamical behavior of a system, for which additional dynamical information should be taken into account. Conversely, a concentration of edges with high impact on the dynamics may arise just from a statistical fluctuation, if the network is seen as a realization of a particular random network model. In this way, structural and dynamical approaches can offer complementing information.

In general, however, they are blocks of nodes with different identities that trap the flow or channel it in specific directions. That is, they form reduced models of the dynamics where blocks of nodes are aggregated to single meta nodes with similar *dynamical function* with respect to the rest of the network. In this view, the goal of community detection is to find effective coarse-grained system descriptions of how the dynamics take place on the network structure.

To induce multi-step paths and couple also non-neighboring nodes, the dynamical approach to community detection has primarily focused on modeling the dynamics with Markovian diffusion processes [19, 43, 65], though the work of topological scales and synchronization share the same common ground [7]. Interestingly, for simple diffusion dynamics such as a random walk on an undirected network, which is essentially determined by the spectral properties of the network's Laplacian matrix, this perspective is tightly connected to the clustering perspective discussed in section 4.3. This is because the presence of densely knit groups within the network can introduce a time-scale separation in the diffusion dynamics: a random walker traversing the network will initially be trapped for a significant time inside a community corresponding to the fast time-scale, before it can escape and explore the larger network corresponding to a slower time-scale. However, this connection between link density and dynamical behavior breaks down for directed networks, even for a simple diffusion process [43, 65, 70]. This apparent relationship breaks down completely when focusing on longer pathways, possibly with memory effects in the dynamics [69, 71].

A dynamical perspective is useful especially in applications in which the network itself is well defined, but the emergent dynamics are hard to grasp. For instance, consider the nervous system of the roundworm *C. elegans*, for which there exists a distinct network. A basic generative network model, such as a Barabasi–Albert network or an SBM, might be too simple to capture the complex architecture of the network, and sampling alternative networks from such a model will not create valid alternative roundworm connectomes. Indeed, some more complicated network generative models have been proposed to model the structure of the network [55], and

may be used to assess the significance of individual patterns compared to the background of the assumed model. However, if we are interested instead in assessing the dynamical implications of the evolutionary conserved network structure, it may be fruitful to engineer differences in the actual network and investigate how they affect the dynamical flows in the system. For instance, one can replicate experimental node ablations *in silico* and assess their dynamical impact [8].

In the dynamical perspective, we are typically interested in how short-term dynamics integrate into long-term behavior of the system and seek a coarse-grained description of the dynamics occurring on a given network. That is, the network itself represents the true structure, save for empirical imperfections. Therefore, in the dynamical perspective, model selection is in general not about comparing competing models [60, 83] but about comparing coarse-grained descriptions of the dynamics on resampled realizations of the observed network with, for example, the bootstrap [66] or cross-validation [63]. Nevertheless, it is possible to formulate generative statistical models for empirically observed pathways [62]. However, whereas the generative approach in, for example, [62] explicitly models the underlying state space of trajectories, we may simply be interested in effectively compressing the long-term behavior of the system [63].

Two methods that exploit the long-term dynamics of the system by identifying communities with long flow persistence are the Markov stability [20] and the map equation [65]. Whereas the Markov stability detailed in Chapter 12 takes a statistical approach and favors communities inside which a random walker is more likely to remain after a given time t than expected at infinite time, the map equation reveals modular regularities by compressing the dynamics. It is an information-theoretic approach that uses the duality between compressing data and finding regularities in the data [65, 75]. It measures the quality of communities by how much they can compress a modular description of the dynamics. The shorter description, the more detected regularities, such that the shortest description captures the most regularities. Given module assignments \mathbf{C} of all nodes in the network, the map equation measures the description length $L(\mathbf{C})$ of a random walker that moves within and between modules from node to node by following the links between the nodes [68]:

$$L(\mathbf{C}) = q_{\circlearrowleft} H(\mathcal{Q}) + \sum_{q=1}^k p_{\circlearrowright}^q H(\mathcal{P}^q) \quad (4.7)$$

Here the entropy $H(\mathcal{Q})$ measures the average per-step description length of movements between modules derived from module-enter rates \mathcal{Q} of all k modules and $H(\mathcal{P}^q)$ measures the average per-step description length of movements within module q derived from node-visit and module-exit rates \mathcal{P}^q . The description lengths are weighted by their rate of use, q_{\circlearrowleft} and p_{\circlearrowright}^q , respectively. The visit rates can be obtained by first calculating the PageRank of links and nodes or directly from the data if they represent flow themselves. In any case, finding the optimal partition of the network by assigning each node to one or more modules corresponds to testing different node assignments and picking the one that minimizes the map equation. This simple formulation allows for straightforward generalizations to coarse-grained hierarchical [67] descriptions of dynamics in memory [69] and multilayer [17] networks.

As the air traffic example above illustrates, it can be crucial to go beyond standard network abstractions and consider memory and higher-order effects in multi-step pathways to better understand system behavior. For example, higher-order abstractions, such as memory and multilayer networks, provide principled means to reveal highly overlapping modular organization

in complex systems: link clustering [1] and clique percolation [57] methods can be interpreted as trying to account for second-order Markov dynamics (see Supplementary Note 3 in [69]).

Compared to some of the other perspectives, the dynamical viewpoint has received somewhat less attention and has been confined mainly to diffusion dynamics (see Chapter 12). A key challenge is to extend this perspective to other types of dynamics and link it more formally to approaches of model order reduction considered in control theory. In light of the recently growing interest in the control of complex systems, this could help us better understanding complex systems.

4.6 Discussion

Community detection can be viewed through a range of different lenses. Rather than looking at community detection as a generic tool that is supposed to work in a generic context, considering the application in mind is important when choosing between or comparing different methods. Each of the perspectives outlined above has its own particularities, which may or may not be suitable for the problem of interest.

We emphasize the different perspectives in the following example. Given a real-world network generated by a possibly complex random assignment of edges, we assume that we are interested in some particular dynamics taking place on this network, such as epidemic spreading. We also assume that the network is structured such that the dynamics exhibit a time-scale separation. If, for instance, we want to coarse-grain an epidemic and identify critical links that should be controlled to confine the epidemic, then it does not matter whether or not random fluctuations generated the modules that induce the time-scale separation. In any case, these modules will be relevant for the dynamics.

Assume now that the same network encodes interdependency of tasks in a load-scheduling problem. In such a circumstance, a cut-based approach will find a relevant community structure, in that it will allow an optimally balanced assignment of tasks to processors that minimizes communication between processors. These communities may be different from the ones attached to the epidemic-spreading example.

If we instead assume that the links represent friendships, we may want to identify densely knit groups irrespective of their relative sizes. Accordingly, taking the clustering perspective and maximizing the internal density can give yet another set of communities.

In these three cases, we considered a single realization of the network with the goal of extracting useful information about its structure, independently of the possible mechanisms that generated it.

Let us finally consider the same network from a stochastic equivalence perspective, and assume for simplicity that the network is a particular realization of an Erdős–Rényi network. In this case, an approach based on the SBM is expected to declare that there is no significant pattern to be found here at all, as the encountered structural variations can already be explained by random fluctuations rather than by hidden class labels. Thus, communities in the SBM picture are defined via the latent variables within the statistical model of the network structure, and not via their impact on the behavior of the system. In this way, different motivations for community detection can find different answers even for the very same network.

To illustrate that different motivations can give different answers for the same network, we use an example from [65]. The directed, weighted network is formed as a ring of rings such that

each internal ring captures flows for a relatively long time despite the stronger links between the rings (see Figure 4.2). For example, a random walker takes on average three steps within a ring highlighted as a cluster in Figure 4.2a before exiting. In contrast, a random walker takes on average only 2.4 steps within a cluster in Figure 4.2b. A method that seeks to coarse-grain the dynamics will therefore identify the flow modules in Figure 4.2a rather than the clusters with high internal density in Figure 4.2b. For example, the modular description quantified by the map equation is almost twice as efficient with the flow modules as it is with the clusters with high internal density. The opposite is true for a method that highlights structural regularity and high internal density: the modularity score is twice as large for the clustering in Figure 4.2b. While this example only illustrates the fundamental difference between two methods applied to a schematic network, methods from different perspectives will give different answers for real networks as well [36].

In addition to the differences *between* these perspectives, there are also variations *within* each perspective. For instance, distinct plausible generative models such as the standard SBM or the degree-corrected SBM will, for a given network, lead to different inferred community structure. Similar variations exist in the dynamical paradigm as well: distinct natural assumptions for the dynamics, such as dynamics with or without memory, uniform across nodes or edges, etc., applied to a given network will lead to different partitions. Also different balancing criteria (see section 4.2) or different concepts of high internal density (see section 4.3) will be valid in different contexts.

In fact, some of the internal variations make the perspectives overlap in particular scenarios. For instance, one can compare all the algorithms on simple, undirected LFR benchmark networks [45]. However, the LFR benchmark clearly imposes a density-based notion of communities. Similarly, for simple undirected networks, optimizing modularity corresponds to the inference of a particular SBM [50] or may be reinterpreted as a diffusion process on a network [20]. Nevertheless, this overlap of concepts, typically present in unweighted, undirected networks, is only partial, and breaks down, for example, in directed networks or for more complex dynamics.

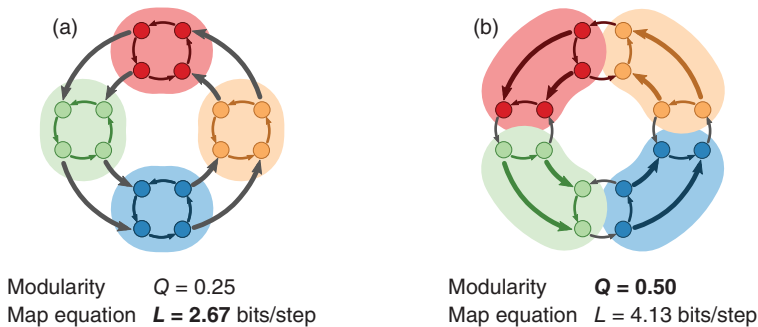


Figure 4.2 Communities that highlight different aspects of networks. Identifying coarse-graining flows in groups, here illustrated by the map equation, and densely connected groups, here illustrated by modularity, highlights different aspects of structure in directed and weighted networks. Each shaded area represents a cluster in two alternative clusterings of a schematic network. (a) The clustering as optimized by the map equation (minimum L). (b) The clustering as optimized by modularity (maximum Q). The thicker links have double the weight of the thinner links. Example from [65].

4.7 Conclusions

In summary, no general purpose algorithm will ever serve all applications or data types [59] because each perspective emphasizes a particular core aspect: a cut-based method provides good separation of balanced groups, a clustering method provides strong cohesiveness of groups with high internal density, stochastic block models provide strong similarity of nodes inside a group in terms of their connectivity profiles, and methods that view communities as dynamical building blocks aim to provide node groups that influence or are influenced by some dynamics in the same way. As more and more diverse types of data are collected, leading to ever more complex network structures, including directed [47], temporal [35, 74], multi-layer or multiplex networks [12], the differences between the perspectives presented here will become even more striking – the same network might have multiple valid partitions depending on the question about the network we are interested in. We might moreover not only be interested in partitioning the nodes, but also in partitioning edges [1], or even motifs [9]. Rather than striving to find a “best” community-detection algorithm for a better understanding of complex networks, we argue for a more careful treatment of what network aspects we seek to understand when applying community detection.

Acknowledgements

We thank Aaron Clauset, Leto Peel, and Daniel Larremore for fruitful discussions. MR was supported by the Swedish Research Council grant 2016-00796. MTS, JCD, and RL acknowledge support from FRS-FNRS, the Belgian Network DYSCO (Dynamical Systems, Control and Optimisation) funded by the Interuniversity Attraction Poles Programme initiated by the Belgian State Science Policy Office, and the ARC (Action de Recherche Concertée) on Mining and Optimization of Big Data Models funded by the Wallonia-Brussels Federation. MTS received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 702410.

References

1. Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, June 2010. doi: 10.1038/nature09182.
2. C. Aicher, A. Z. Jacobs, and A. Clauset. Learning latent block structure in weighted networks. *Journal of Complex Networks*, page cnu026, 2014.
3. C. J. Alpert and A. B. Kahng. Recent directions in netlist partitioning: a survey. *Integration, the VLSI journal*, 19(1):1–81, 1995.
4. R. Andersen, F. Chung, and K. Lang. Local graph partitioning using pagerank vectors. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS’06)*, pages 475–486. IEEE, 2006.
5. C. J. Anderson, S. Wasserman, and K. Faust. Building stochastic blockmodels. *Social Networks*, 14(1):137–161, 1992.
6. P. Arabie, S. A. Boorman, and P. R. Levitt. Constructing blockmodels: How and why. *Journal of Mathematical Psychology*, 17(1):21–63, 1978.
7. A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96(11):114102, Mar. 2006.
8. K. A. Bacik, M. T. Schaub, M. Beguerisse-Díaz, Y. N. Billeh, and M. Barahona. Flow-based network analysis of the Caenorhabditis elegans connectome. *PLoS Comput Biol*, 12(8):1–27, 08 2016.

9. A. R. Benson, D. F. Gleich, and J. Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016. ISSN 0036-8075.
10. P. J. Bickel and P. Sarkar. Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):253–273, 2016.
11. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
12. S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.
13. A. Browet, J. Hendrickx, and A. Sarlette. Incompatibility boundaries for properties of community partitions. *IEEE Transactions on Network Science and Engineering*, 83–99, 2017.
14. M. Chen, K. Kuzmin, and B. K. Szymanski. Community detection via maximization of modularity and its variants. *IEEE Transactions on Computational Social Systems*, 1(1):46–65, 2014.
15. M. Chen, T. Nguyen, and B. K. Szymanski. A new metric for quality of network community structure. *arXiv:1507.04308*, 2015.
16. M. Coscia, F. Giannotti, and D. Pedreschi. A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining*, 4(5):512–546, 2011.
17. M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, 5(1):011027, 2015.
18. A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107:065701, Aug. 2011.
19. J.-C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, 107(29):12755–12760, 2010.
20. J.-C. Delvenne, M. T. Schaub, S. N. Yaliraki, and M. Barahona. The stability of a graph partition: A dynamics-based framework for community detection. In *Dynamics On and Of Complex Networks, Volume 2*, pages 221–242. Springer, 2013.
21. W. E. Donath and A. J. Hoffman. Algorithms for partitioning of graphs and computer logic based on eigenvectors of connection matrices. *IBM Technical Disclosure Bulletin*, 15(3):938–944, 1972.
22. W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development*, 17(5):420–425, 1973.
23. M. G. Everett and S. P. Borgatti. Regular equivalence: General theory. *Journal of Mathematical Sociology*, 19(1):29–52, 1994.
24. M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.
25. M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25(4):619–633, 1975.
26. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3): 75–174, 2010.
27. S. Fortunato and M. Barthélemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007.
28. S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, 2016.
29. B. H. Good, Y.-A. de Montjoye, and A. Clauset. Performance of modularity maximization in practical contexts. *Physical Review E*, 81(4):046106, Apr. 2010.
30. R. Guimera, M. Sales-Pardo, and L. A. N. Amaral. Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2):025101, 2004.
31. I. Guyon, U. Von Luxburg, and R. C. Williamson. Clustering: Science or art. In *NIPS 2009 Workshop on Clustering Theory*, pages 1–11, 2009.
32. L. Hagen and A. B. Kahng. New spectral methods for ratio cut partitioning and clustering. *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems*, 11(9):1074–1085, 1992.
33. R. A. Hanneman and M. Riddle. *Introduction to social network methods*. University of California Riverside, 2005.
34. P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.
35. P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3): 97–125, 2012.
36. D. Hric, R. K. Darst, and S. Fortunato. Community detection in networks: Structural communities versus ground truth. *Physical Review E*, 90(6):062805, 2014.
37. R. Kannan, S. Vempala, and A. Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM*, 51(3):497–515, 2004.

38. B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.
39. B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49(2):291–307, 1970.
40. J. Kleinberg. An impossibility theorem for clustering. *NIPS'02 Proceedings of the 15th International Conference on Neural Information Processing Systems*, 463–470, 2003.
41. K. Kloster and D. F. Gleich. Heat kernel based community detection. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1386–1395. ACM, 2014.
42. F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.
43. R. Lambiotte, J.-C. Delvenne, and M. Barahona. Random walks, Markov processes and the multiscale modular organization of complex networks. *IEEE Transactions on Network Science and Engineering*, 1(2):76–90, 2014.
44. A. Lancichinetti and S. Fortunato. Limits of modularity maximization in community detection. *Physical Review E*, 84:066122, Dec. 2011.
45. A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4):046110, Oct. 2008.
46. F. Lorrain and H. C. White. Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1(1):49–80, 1971.
47. F. D. Malliaros and M. Vazirgiannis. Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4):95–142, 2013.
48. L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014.
49. E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708.
50. M. Newman. Equivalence between modularity optimization and maximum likelihood methods for community detection. *Physical Review E*, 94(5): 052315, 2016.
51. M. E. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, 2006.
52. M. E. Newman. Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1):25–31, 2012.
53. M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
54. M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, Feb 2004.
55. V. Nicosia, P. E. Vértés, W. R. Schafer, V. Latora, and E. T. Bullmore. Phase transition in the economically modeled growth of a cellular nervous system. *Proceedings of the National Academy of Sciences*, 110 (19):7880–7885, 2013.
56. K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.
57. G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
58. S. Parthasarathy, Y. Ruan, and V. Satuluri. Community discovery in social networks: Applications, methods and emerging trends. In *Social network data analytics*, pages 79–113. Springer, 2011.
59. L. Peel, D. B. Larremore, and A. Clauset. The ground truth about metadata and community detection in networks. *Science Advances*, 3(5): e1602548, 2017.
60. T. P. Peixoto. Parsimonious module inference in large networks. *Physical Review Letters*, 110(14):148701, 2013.
61. T. P. Peixoto. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Physical Review E*, 92(4):042807, 2015.
62. T. P. Peixoto and M. Rosvall. Modelling sequences and temporal networks with dynamic community structures. *Nature Communications*, 8(1):582, 2017.
63. C. Persson, L. Bohlin, D. Edler, and M. Rosvall. Maps of sparse Markov chains efficiently reveal community structure in network flows with memory. *arXiv:1606.08328*, 2016.
64. A. Pothen. Graph partitioning algorithms with applications to scientific computing. In *Parallel Numerical Algorithms*, pages 323–368. Springer, 1997.
65. M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.
66. M. Rosvall and C. T. Bergstrom. Mapping change in large networks. *PloS One*, 5(1):e8694, 2010.

67. M. Rosvall and C. T. Bergstrom. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS One*, 6(4):e18209, 2011.
68. M. Rosvall, D. Axelsson, and C. Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, 2009.
69. M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5, 4630 2014.
70. A. Saade, F. Krzakala, and L. Zdeborová. Spectral clustering of graphs with the bethe hessian. In *Advances in Neural Information Processing Systems 27 (Proceedings of Neural Information Processing Systems)*, Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence and K.Q. Weinberger (eds), pages 406–414, 2014.
71. V. Salnikov, M. T. Schaub, and R. Lambiotte. Using higher-order Markov models to reveal flow-based communities in networks. *Scientific Reports*, 6:23194, 3 2016.
72. S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.
73. M. T. Schaub, J.-C. Delvenne, S. N. Yaliraki, and 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(2):e32210, 2012.
74. V. Sekara, A. Stopczynski, and S. Lehmann. Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, 113(36):9977–9982, 2016.
75. C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
76. J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
77. D. A. Spielman and S.-H. Teng. Spectral partitioning works: Planar graphs and finite element meshes. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pages 96–105. IEEE, 1996.
78. D. A. Spielman and S.-H. Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM Journal on Computing*, 42(1):1–26, 2013.
79. U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.
80. U. Von Luxburg, R. C. Williamson, and I. Guyon. Clustering: Science or art? In *JMLR Workshop and Conference Proceedings: ICML Unsupervised and Transfer Learning*, volume 27, pages 65–80, 2012.
81. H. C. White, S. A. Boorman, and R. L. Breiger. Social structure from multiple networks. I. blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780, 1976.
82. J. Xie, S. Kelley, and B. K. Szymanski. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys*, 45(4):43, 2013.
83. X. Yan. Bayesian model selection of stochastic block models. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 323–328. IEEE, 2016.
84. J. Yang and J. Leskovec. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1): 181–213, 2015.