Large-Scale Structure of Multi-Optimised Networks

Morel-Balbi, Sebastian

Award date:
2022

Awarding institution:
University of Bath

Large-Scale Structure of Multi-Optimised Networks

submitted by

Sebastian Morel-Balbi

for the degree of *Doctor of Philosophy*

of the

University of Bath

Department of Mathematical Sciences

May 2022

COPYRIGHT

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

. . . . . .

Sebastian Morel-Balbi

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally in collaboration with my supervisor Tiago P. Peixoto.

. . . . . .

Sebastian Morel-Balbi

Summary

Generative network models (i.e. models that aim to uncover the underlying mechanisms that drive network formation) have played a prominent role in the literature of networked systems, and many hypotheses on how networks grow and evolve have been proposed. Recently, structural optimisation has been suggested as one of the main drivers underlying network formation processes. Indeed, many networks are designed or required to perform one or more specific tasks as efficiently as possible. How optimally a network will perform a set of pre-specified tasks will depend, at least partially, on its large-scale structure. Consequently, these optimality requirements should result in selective pressures driving the network toward particular large-scale topologies. The general prescription behind this family of models is to define a cost function over a network ensemble; by minimising the cost function, one can evaluate whether particular topologies characterise the ensemble and compare these topologies with those observed in real-world networks. However, real-world networks rarely emerge as the result of a single generating mechanism but are more likely to be the end product of several co-existing processes and limitations, such as dynamical rules, exogenous constraints, and optimality requirements. This conflation of generating processes makes it difficult to assess and quantify to what extent optimality might have played a role in the formation of any given network. In this thesis, we develop a framework to construct null models of optimised networks that allow us to isolate the effects of optimisation on network structure from other external artefacts. Furthermore, our framework can accommodate an arbitrary number of optimisation criteria, allowing us to study more realistic scenarios in which a network is simultaneously subject to multiple selective pressures. We apply the proposed framework to study networks driven to optimise for modularity and robustness against random failures. We first analyse the case in which the criteria are imposed separately and uncover the topologies most likely to emerge as a result of optimisation. We then combine the two criteria and study the effects of joint optimisation on the network structure. We uncover a rich phase diagram of optimised networks, characterised by a series of phase transitions at which the optimal topologies change according to the desired degree of optimality. We also identify regions of the parameter space where synergistic and antagonistic effects are present, such that optimising for one criterion can help or hinder optimising for the other.

Acknowledgements

First and foremost, I would like to thank my supervisor, Tiago Peixoto, for his guidance and patience during my PhD, his continued support throughout the pandemic, and for introducing me to the beautiful world of complex networks. I have learned an enormous amount from him, academically and not, and his passion for the field and attention to detail are lessons I will carry with me in the future.

I would also like to thank my examiners, Mathew Nunes and Leto Peel, for their time and effort in reading and correcting my thesis and for their valuable observations and suggestions.

I want to thank all of the Complex Networks and Collective Behaviour group members at the University of Bath for providing a stimulating and friendly environment to work in. A special thank goes to all of the members of 5.8 for making my time there so much fun, particularly Enrico, for all those bus rides to Bristol and for showing me around Bath, and Ben, whom I had the luck to rejoin in Vienna.

I want to thank all of the Department of Network and Data Science at the Central European University for welcoming me so warmly. A special thank goes to Fede for showing me around and to my main climbing partners, Elsa, Sandeep, Ludo, and Martina. I will miss those Wednesday dinners at Riva.

I would also like to thank my friends back in Rome, particularly Sergio, Sofia, and Sammy, for always being there and supporting me.

Finally, I would like to thank my parents and my brother, without whose continued support and encouragement I would have been lost. I cannot truly express with words the love and gratitude I feel for you.

Contents

List of Figures

Chapter 1

Introduction

Many complex systems share the unifying property that they can be modelled as networks. Examples include biological systems [1–3], the Internet [4], social interaction networks [5, 6], and power grids [6, 7]. With the aim of exploiting this universality, a great deal of research has been devoted both to the empirical study of complex networks and to the development of general network models that are valid across different domains.

Of particular interest is the study of the large-scale topological structures that real-world networks exhibit, as they are known to be closely related to the network function [8, 9]. For example, many real-world networks possess nodes of remarkably high degree, also called *hubs*, which have been shown to have profound effects on the resilience of a network to the failure of its components [10], or the spread of a disease across a network [11]. Similarly, the *small-world effect*, i.e. the observation that the average geodesic distance between node pairs in a network is remarkably short, has been linked to efficient communication [12] and has been observed in networks as varied as the collaboration of Hollywood actors [6] and the human brain [13]. Another topological feature of great interest is the presence of *communities*, groups of nodes that tend to be more connected amongst each other than the rest of the network. These clusters of nodes can provide valuable information on the underlying organisation of the network and the interactions between its components.

Given the overwhelming importance that network structure plays in shaping the behaviour of networks and the processes occurring on them, it is no surprise that a wealth of mathematical methods and models aimed at extracting and replicating the large-scale structure of networks has been developed. However, many of these methods fail to address the underlying processes driving the network structure. This has prompted the development of models of network formation, i.e. models which aim to capture the mechanisms behind network formation rather than only reproducing the desired large-scale properties. The earliest and by far the most studied class of network formation models are the so-called *preferential attachment models*. Preferential attachment models are growing network models in which nodes (or edges) are added to a network in subsequent time steps and connect to the existing vertices according to some pre-specified scheme. The main idea is that most real-world networks are not static but rather grow over time. Thus, one hopes to reproduce the structural characteristics commonly observed in real-world networks by capturing the mechanisms with which new nodes are incorporated into the network. The most prominent preferential attachment model is without a doubt the Barabási-Albert

model (BA model) introduced by Laszlo Barábasi and Réka Albert in [14]. The model is centred around two basic assumptions:

Network growth: Networks grow over time. The authors then argue that it is unrealistic to attempt to model them via static models in which the number of nodes N is held fixed.

Preferential attachment: New nodes do not randomly connect to the network, which would be highly unrealistic. Instead, there must be an underlying mechanism guiding the growth of the network. Barabási and Albert propose a scheme by which new nodes preferentially connect to nodes that already possess a high degree.

The BA model can reproduce some of the large-scale properties commonly observed in real-world networks. Most notably, the highly heterogeneous degree distributions that many real-world networks display. However, it fails to account for others, such as the presence of a community structure. An enormous number of extensions to the BA model have subsequently been proposed to circumvent this, and other issues (such as extending the model to directed networks, including node removal, considering the possibility of internal links, and extending the model beyond power-law distributions with exponent $\gamma = 3$) [15–21]. However, the key ingredients of all these models remain the same, network growth and a preferential attachment scheme.

Other models of network formation exist. *Vertex copying models* [22, 23] are a popular example of models which can generate both power-law degree distributions and communities and have been widely employed to model biological networks [24–26]. Recently, significant interest has been devoted to another class of network formation models based on structural optimisation.

Many networks evolve or are designed to accomplish one or more specific tasks. We might, for example, want to optimise the flow of traffic over a transportation or communication network. Similarly, a gene regulatory network needs to be robust against random errors in gene expression. How optimal the performance of these networks is will depend, amongst other things, on their large-scale topological structure. This results in a selective pressure toward particular network structures, depending on the desired fitness.
The general mechanism behind optimisation-based models is then the following. Given a set of pre-specified characteristics we want our networks to satisfy, we can associate a cost function to the networks generated by the model. By minimising the cost function, we can then examine if some particular topology characterises the resulting networks.

Some of the earliest work employing optimal network models as means to understand the observed network structure arose from the study of allometric scaling in biology [27] and the study of river networks [28, 29], where results based on optimisation criteria proved to be in spectacular accord with real-world observations. In the context of brain networks, Cherniak investigated whether neural networks optimise the placement of their components to minimise the length of interconnections between them and found supporting evidence for this hypothesis [30]. These results prompted efforts by researchers to investigate whether some of the large-scale properties commonly observed in real-world networks might arise as the outcome of particular optimisation processes. Mathis and Gopal [31] used a Metropolis-based

optimisation procedure to show that in physical networks (i.e. networks whose nodes are separated by a Euclidean distance), the small-world phenomenon can arise as a tradeoff between maximal connectivity and minimal wiring.
Furthermore, unlike previous models of small-world networks, the authors uncovered that small-worldness arises in optimised networks via the formation of hubs, resulting in networks with seemingly power-law degree distributions, also known as *scale-free* networks. Networks obtained this way were also shown to be more "optimal" than the ones obtained by previous models, as they displayed shorter average path lengths and required less wiring.

Mathis and Gopal's work provided the first evidence of competitive minimisation as a possible formation mechanism of small-world networks. Further corroboration came from Valverde *et al.* [32] who provided empirical evidence of the emergence of both small-worldness and scale-freeness in networks arising from local optimisation principles and whose design makes no explicit assumptions regarding preferential attachment or small-worldness. In a widely celebrated paper, Ferrer-i-Cancho and Solé [33], examined the emerging topologies in a network driven to optimise both the average path length between nodes and the link density. The authors uncovered a series of phase transitions where the network topology can vary from a random tree with exponential degree distribution to a scale-free network characterised by high degree hubs to a quasi star-graph topology, where all nodes in the network are connected to a single central node with few if any, other connections permitted. Similar conclusions were reached by Colizza *et al.* [34] who studied networks driven to minimise the path length between their components while also avoiding highly congested nodes (i.e. large-degree nodes). The authors showed that the interplay between the two objectives could give rise to a wide range of optimal topologies depending on the values of the model parameters.

More broadly, optimal networks have been applied to study a wide variety of problems, such as efficiency in transportation networks [35], minimising the search cost over a network [36], optimal traffic networks [37], and network robustness [38–40].

The emergence of properties commonly observed in real-world networks solely from optimisation principles makes the critical suggestion that, for a large class of networks, growth mechanisms might not be exclusively responsible for the observed structures of real-world networks. Instead, the observed topologies are likely to be the end product of an interplay between growth mechanisms and selective pressures. In [33], the authors provide some evidence that metabolic networks might indeed arise as an outcome of both mechanisms of network formation (optimisation and preferential attachment).

Understanding the evolutionary forces shaping the structure of complex systems can help us better elucidate the interplay between network structure and function and prove crucial for designing networks aimed at optimising one or more predefined tasks or properties. Furthermore, identifying structures that optimise specific external criteria may help us understand how networks evolve when these external conditions are changed, providing possible insights into the processes driving network resilience and failure. However, as mentioned above, selective pressures generally combine with other kinds of dynamical rules, exogenous constraints, and historical artefacts. Furthermore, a given system may be subject to multiple selective pressures simultaneously, e.g. it may need to run efficiently while being simultaneously

robust to errors or damage. Since we seldom observe any given formation process in detail, we are forced to disentangle these different driving forces from each other based only on the structural patterns they produce.

In this thesis, we contribute to this disentangling effort by constructing *null models* of optimised networks. These models correspond to ensembles of networks that possess some pre-specified characteristics but which are otherwise maximally random. By investigating the emerging structural features in these models, we can understand the inherent effect a particular kind of fitness criterion has on the network's structure without the interference of any other kind of constraint. Moreover, we can also combine multiple fitness criteria to determine how they interact in determining the preferred network structure. This gives us a controlled platform to delineate the effects of different kinds of selective pressures on network structure in a principled manner.

In the following, we will employ this approach to investigate two central properties of networked systems, namely the robustness of a network against the random failure of its components [41], and its modularity [42], characterised by the existence of groups of nodes that are more connected among themselves than with the rest of the network. Robustness to failure is believed to play a key role in infrastructure [43] as well as technological networks such as the Internet [44], but also on biological systems [45]. Modularity, on the other hand, has been associated with the adaptability of biological networks [46] and is a necessary ingredient for the scheduling of interdependent processes with minimal amount of communication [47]. By enforcing these two optimisation criteria both in isolation and simultaneously, we analyse which large-scale network structures are most likely to emerge as a result of their interaction. Our main result is the identification of a series of phase transitions at which the optimal structure of the network changes in response to the varying selective pressures. We also identify regions in the parameter space where the interplay between the selective pressures gives rise to synergistic effects, i.e. one kind of fitness pressure contributes to the second, such that it becomes easier to optimise for both at once, as well as antagonistic effects, where both optimisations compete against each other.

The thesis is organised as follows. Chapter 2 briefly presents some basic network theory concepts and introduces the necessary background and tools essential for developing our framework in later chapters. In particular, we focus on models of network structure (i.e. models whose aim is to replicate the observed structure of a network as closely as possible) and present the concepts of *network ensembles* and *random graph models*. In addition, we also introduce the fundamental concept of entropy and show how it can be used as a guideline in the construction of network models. Finally, we conclude the chapter by providing the basic theoretical underpinnings that allow us to place our chosen fitness criteria (modularity and robustness against random failures) on solid quantitative grounds.

Chapter 3 introduces our modelling framework and frames the optimisation of the fitness criteria as an entropy maximisation problem.

In Chapter 4, we apply the formalism developed in Chapter 3 to networks driven to optimise for either modularity or robustness against random edge removals. We consider the two fitness criteria independently and analyse the large-scale structures that are most likely to emerge as a result of the network being placed under varying degrees of selective pressure.

Chapter 5 considers the more realistic scenario in which networks are subject to both selective pressures simultaneously and investigates the effects of the interplay between the two selective pressures on the emerging network topologies.

Finally, in Chapter 6, we draw our conclusions and present possible future research outlooks.

Chapter 2

Background

2.1 Fundamentals of network theory

In its simplest form, a network is composed by a set of elements, called *nodes* or *vertices*, connected amongst each other through *edges*, Fig. 2.1.

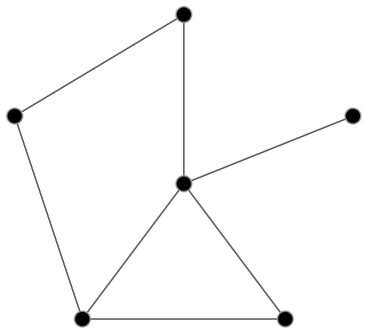


Figure 2.1: An example of a simple network made up of $N = 6$ nodes and $E = 7$ edges.

Mathematically speaking, given a set of nodes V and edges E, a *network* can be defined as an ordered pair $G = (V, E)$. The great advantage of networks is that they can model a wide array of complex systems. For example, in the case of the Internet, the network's nodes represent computers, and the edges the physical links between them. For a power grid, the nodes would represent power stations, and the edges would represent the transmission lines of the network. These are two examples of "physical" networks, but networks can also be used to capture more abstract relationships. For example, in a social network, nodes represent people, and the edges can represent friendships or other types of interactions. If we take some time to look around, we soon realise that networks are ubiquitous, ranging from biological to technological and social networks. As such, it is of great interest to study and analyse networked systems.

Different types of networks exist depending on the nature of the interaction we are attempting to model as a network. For example, a network such as the Internet, where nodes are computers and edges are the physical links between them, can be modelled as a *simple network* such as the one in Fig. 2.1. In a simple network, edges have no directionality; they either exist or do not. However, if we

consider the World Wide Web, where nodes represent web pages and edges are the hyperlinks between them, we can see a directionality in how the network is wired. The fact that some page A has a hyperlink pointing to some other page B does not automatically imply that page B points back to page A. A similar picture holds in friendship networks, where the concept of friendship might not be reciprocated amongst individuals. These types of systems are bettered described by what is known as a *directed network*, i.e. one in which links have a directionality, Fig. 2.2.

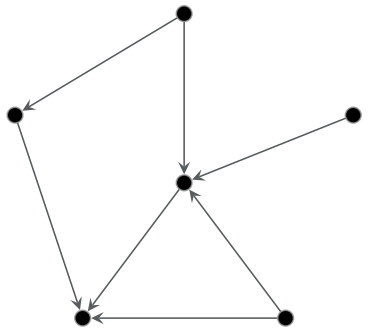


Figure 2.2: An example of a directed network

Other types of networks are also possible, for example, *weighted networks* in which the edges are assigned a weight which typically models some distinctive feature of the interaction, such as the strength of the interaction or the capacity of the tie between two nodes. An example of this could be the amount of traffic that flows along the edges in a transportation network [48]. Other kinds of networks are so-called *multigraphs*, in which multiple edges between nodes are possible, and nodes are also allowed to have self-edges, Fig 2.3.

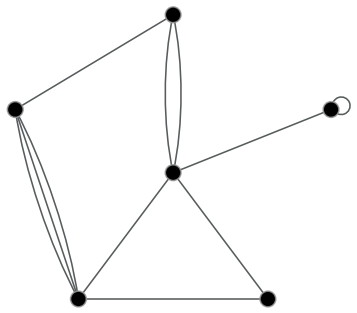


Figure 2.3: An example of a multigraph displaying both multiple edges and self-loops.

If we label the nodes of a network so as to make them distinguishable, then the network can be fully specified by its *adjacency matrix* $\mathbf{A}$. For a simple network, the adjacency matrix is a matrix whose elements A_{ij} are

$$A_{ij} = \begin{cases} 1 & \text{if node } i \text{ and node } j \text{ are connected by an edge,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

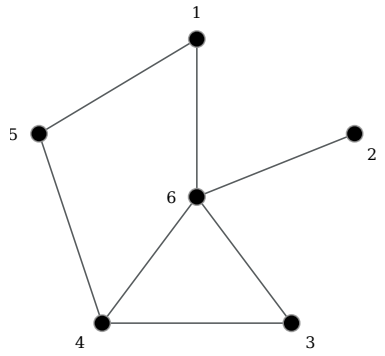


Figure 2.4

For example, if we take the simple graph introduced previously and label it as in Fig. 2.4, the corresponding adjacency matrix would be given by

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} \tag{2.2}$$

Different label orderings will give rise to different adjacency matrices, but they all have some common characteristics. First of all, for a simple network, the absence of self edges means that the diagonal terms of the adjacency matrix will always be equal to zero. Also, as a simple network is undirected, the adjacency matrix will be a symmetric matrix with respect to the diagonal. Finally, it is easy to see that the total number of edges in the network E is equal to

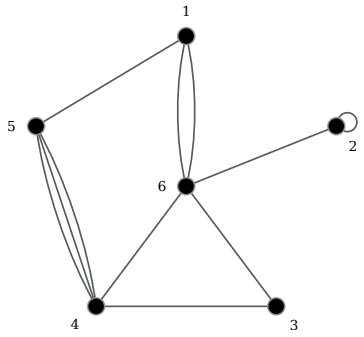
$$E = \frac{1}{2} \sum_{ij} A_{ij}. \tag{2.3}$$

The adjacency matrix can be easily generalised to the case of multigraphs. In this case, the entries of the adjacency matrix correspond to the total number of edges between two nodes, with the small caveat that self-edges have to be counted twice. For example, Fig. 2.5 shows the adjacency matrix of the multigraph in Fig. 2.3 with the same labelling used above.

For directed networks, the convention is to define the adjacency matrix as

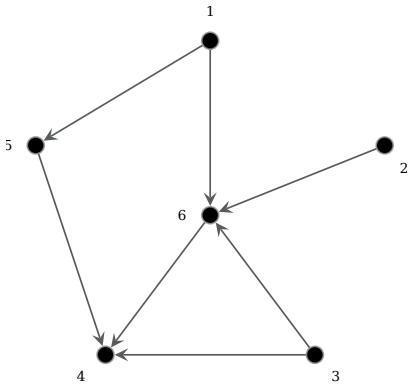
$$A_{ij} = \begin{cases} 1 & \text{if there is an edge running from node } j \text{ to node } i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

Fig. 2.6 shows a directed network with its corresponding adjacency matrix Notice that, for a directed network, the adjacency matrix need not be symmetric, as there is no guarantee that the connections between the elements of the network will be reciprocated. There are many other interesting properties of the adjacency matrix which are beyond the scope of this thesis; we refer the interested reader to [49].



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 3 & 1 \\ 1 & 0 & 0 & 3 & 0 & 0 \\ 2 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$$

Figure 2.5: An example of a multigraph and its corresponding adjacency matrix.



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 2.6: An example of a directed network and its corresponding adjacency matrix. Notice that the adjacency matrix is no longer symmetric.

The degree

We now briefly touch upon another important aspect in graph theory, the concept of *degree*. The degree of a node in the network refers to the number of edges that connect to it. Given a particular node i in a network with N nodes, we can express its degree via the adjacency matrix as

$$k_i = \sum_{j=1}^{N} A_{ij}. \tag{2.5}$$

The *average degree* of the network is then given by

$$\langle k \rangle = \frac{1}{N} \sum_{i} k_i. \tag{2.6}$$

If we look at Eq. (2.3), we see that

$$2E = \sum_{ij} A_{ij} = \sum_{i} k_i. \tag{2.7}$$

By substituting this expression into Eq. (2.6), we see that the average degree of the network is given by the simple relationship

$$\langle k \rangle = \frac{2E}{N}. \tag{2.8}$$

For directed networks, things are slightly different, and one has to define both an *in-degree* and an *out-degree*. However, since directed networks are not the focus of this thesis, we will not delve further into the argument.

The degree distribution

The *degree distribution* $p(k)$ of a network indicates the fraction of nodes in the network that have degree k. More specifically, let

$$N(k) = \sum_{i=1}^{N} \delta_{k,k_i} \tag{2.9}$$

be the number of nodes in the network with degree k, where $\delta_{x,y}$ is the Kronecker delta (defined so that $\delta_{x,y} = 1$ if $x = y$ and 0 otherwise). The degree distribution is then defined as

$$p(k) = \frac{N(k)}{N}, \tag{2.10}$$

and corresponds to the probability that a node chosen uniformly at random from the network has degree k.

Let us consider the network in Fig. 2.4 of $N = 6$ nodes and $E = 7$ edges. We have one node of degree one, three nodes of degree two, one node of degree three, and one node of degree four. The degree distribution is then given by

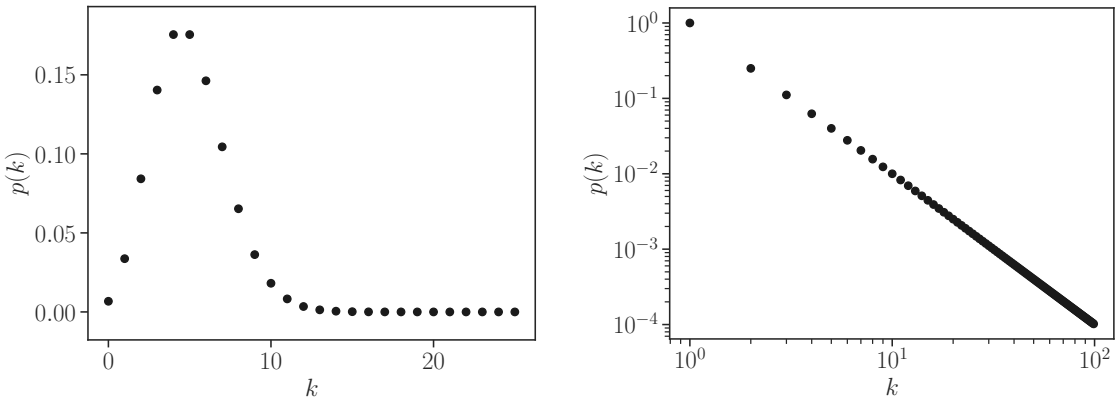
$$p_0 = 0 \quad p_1 = \frac{1}{6} \quad p_2 = \frac{1}{2} \quad p_3 = \frac{1}{6} \quad p_4 = \frac{1}{6} \quad p_5 = 0 \quad . \tag{2.11}$$

Where we have indicated $p(k = c)$ as p_c. The distribution only goes up to $k = 5$ because, in a simple network, the maximum degree a node can achieve is $N - 1$, corresponding to the case in which it connects to every other node in the network.

The degree distribution is a crucial aspect of a network as it gives us precious information on the network's connectivity. One of the most notable discoveries in network science is that many real-world networks possess highly heterogeneous degree distributions. This is the case with the Internet [4], the World Wide Web [10], the network of citations amongst scientific papers [50], and metabolic networks [51] amongst others. Specifically, many of these networks are said to be *scale-free*, meaning that their degree distribution $p(k)$ can be approximated by a power law for large values of k, i.e.

$$p(k) \simeq Ck^{-\alpha} \text{ for } k \gg 1 \tag{2.12}$$

The crucial aspect of power-law distributions is that they decay relatively slowly. While for a Poisson or Gaussian distribution, one expects most of the nodes to have degrees close to the average of the distribution, see Fig. 2.7a, a power-law distribution tells a different story. In a power-law distribution, most nodes possess a low degree, but a small fraction of nodes can have extremely high degrees Fig. 2.7b. These highly connected nodes are commonly referred to as *hubs*. Hubs have profound effects on the way a networked system functions. They have been linked to small-world effects and can facilitate the transmission of information along the network, be it packages of information over the World Wide Web or the spread of disease along a social network. Perhaps one of the most fundamental consequences of hubs is the effect they have on a network's *robustness*.



(a) (b)

Figure 2.7: A Poisson distribution (a) and a power law distribution (b). While the Poisson distribution decays rapidly and most of the density is concentrated around the mean, in a power-law distribution there is a finite probability of observing high degree nodes. The power-law distribution has been plotted in log-log space for clarity.

The robustness of a network represents the network's ability to keep functioning when a fraction of its components malfunctions or are removed. Network robustness is a problem of great interest, especially given past breakdowns that have occurred in real-world networks, such as outages in power-grids [43, 52]. Understanding the resilience of networked systems to random failures or malicious attacks is then of fundamental importance when designing a new network or optimising existing ones [38, 39]. Furthermore, understanding network robustness can give us insights into the resilience observed in many biological [53] and technological networks. We shall give a more extensive summary of network resilience in Section 4.2.

Community structure

Another commonly observed feature in complex networks is the presence of communities. In the context of networks, a *community* can be generally understood as a subset of the network whose nodes are more densely connected amongst each other than with the rest of the network. The general idea being that nodes that fall within the same community are somehow alike or play a similar role in the network, see Fig. 2.8.

Strictly speaking, identifying a network's functional communities as sets of more densely connected nodes is not always correct. For example, in food webs — networks that encode an ecosystem's trophic levels —, it is generally accepted that predators tend to eat prey more than they eat other predators. Similarly, in an economic network, buyers might connect more to sellers than to other buyers. Nevertheless, many real-world networks do exhibit this kind of assortative mixing pattern described above. Examples include biological networks [46, 54] and communities on the World Wide Web [55].

Understanding the community structure of a network can provide us with information on its functional units and can be relevant for processes taking place on the network. Furthermore, different communities can often display varying topologies

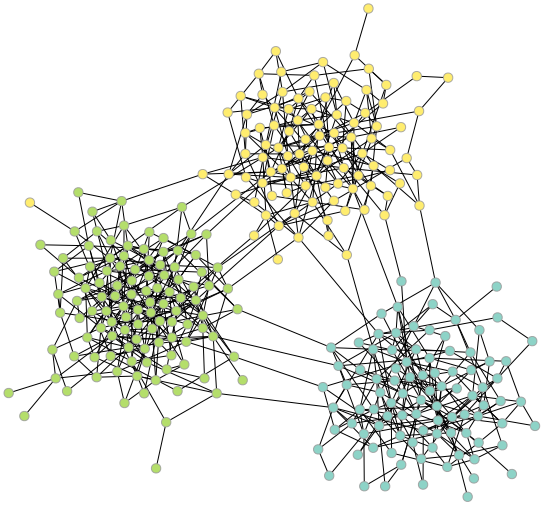


Figure 2.8: Example of a network displaying a division into three communities.

and structural features, allowing us to gain information on the isolated functional modules themselves. Consequently, a significant amount of effort has gone into identifying the communities that naturally divide a network. Sometimes, information on the community structure is available in the form of metadata. Often, however, researchers are faced with the task of identifying communities solely from the pattern of connections within the network, a problem commonly known as *community detection*.

Community detection is one of the most vastly studied problems in the network literature and is beyond the scope of this work. An excellent review can be found in [56]. In Section 4.1, however, we will analyse in greater detail *modularity*, a commonly used metric to assess the goodness of a partition of a network into communities.

2.2 Network models

In the previous section, we have introduced some of the basic properties of networked systems. Many more exist, which we have not covered. Given a network, we can then measure these and other properties to gain information on the system which the network describes. An obvious next step consists in constructing network models. Although many different models exist, we limit ourselves to studying models of network structure. These are models whose goal is to replicate the structure of observed networks as closely as possible. One can then utilise these models to generate a large number of networks with similar properties as the observed one and gain statistical insights into the attributes of specific networks, or to generate synthetic networks that can then be used as a substrate for other simulations, such as processes occurring on the network.

In particular, we will be interested in *random graph models*, i.e. models of networks that display some fixed attributes but which are otherwise maximally random. In the remainder of this section, we introduce three random graph models that will be used later.

2.2.1 The Erdős-Rényi model

The most classic example of a random graph is one in which we consider a network with a fixed number of nodes, N, and edges, L to be randomly distributed amongst the nodes. In mathematics, this is commonly known as the $G(N, L)$ model and was famously studied by Erdős and Rényi in a series of papers [57–61].

Newtork ensembles

It is straightforward to notice that, given N nodes and L edges to be randomly distributed amongst them, many different network realisations that satisfy the constraints of the $G(N, L)$ model are possible. Thus, a more precise definition of the $G(N, L)$ model would be to say that the model network is obtained by sampling uniformly at random from all the networks with exactly N nodes and L edges. Indeed, random graph models are generally not defined in terms of a single network; they instead define an *ensemble* of networks—a probability distribution over the set of all networks that satisfy the constraints imposed by the model. Formally, a *network ensemble* is defined as a tuple $\mathcal{G} = (G, \Omega, P(G))$, where G is any possible network $G = (V, E)$ belonging to the set of networks Ω, and $P(G)$ is the probability associated to the network within the ensemble. Thus, in the $G(N, L)$ model, Ω is the set of all networks with exactly N nodes and L edges, and $P(G) = 1/\Omega$ is the uniform distribution.

Ensemble models have been extensively used in network science [8, 62–65], on the one hand, they lend themselves well to analytic calculations, but, perhaps more importantly, they are apt at answering the questions one is generally interested in. Usually, we are interested in the typical behaviour of a network rather than the behaviour of one particular network. For example, knowing how a disease can spread over a typical social network can prove more valuable than knowing how it spreads over a specific instance of a social network. Ensemble models provide an elegant way to achieve this. Indeed, given a set of graph observables $\{x_i\}$, the typical properties can usually be adequately approximated by the ensemble average of the observables.

$$\langle x_i \rangle = \sum_G x_i(G) P(G). \tag{2.13}$$

The $\mathbf{G(N, p)}$ ensemble

Although the $G(N, L)$ ensemble is straightforward to define, some of its properties are complicated to evaluate analytically. Furthermore, the number of links rarely stays fixed in real-world networks. A much more widely studied model is the $G(N, p)$ model. In the $G(N, p)$ model, rather than fixing the number of edges in the network, we fix the probability that an edge exists between any two nodes in the network. This model was first introduced by Solomonoff and Rapoport [66] and later, independently, by Gilbert [67]. However, an overwhelming contribution to the model came once more from Erdős and Rényi [57–59], to the extent that the $G(N, p)$ model is also commonly referred to as the *Erdős-Rényi random graph* (ER random graph). In what follows, we quickly review some of its properties.

Since in the $G(N,p)$ model we fix the probability p of observing an edge between any vertex pair, the probability of any particular network with L links is

$$P(G) = p^L(1-p)^{\binom{N}{2}-L}. \tag{2.14}$$

And, seeing as there are $\binom{\binom{N}{2}}{L}$ possible ways of distributing L edges amongst N nodes, the total probability of observing a network with L edges is given by a binomial distribution

$$P(L) = \binom{\binom{N}{2}}{L} p^L(1-p)^{\binom{N}{2}-L}. \tag{2.15}$$

The expected number of edges is then given by

$$\langle L \rangle = \sum_{L=0}^{\binom{N}{2}} LP(L) = \binom{N}{2} p. \tag{2.16}$$

Eq. (2.16) then tells us that

$$p = \frac{2\langle L \rangle}{N(N-1)}. \tag{2.17}$$

Recalling that the average degree of a network is given by $\langle k \rangle = 2L/N$, the probability p of an edge existing between two nodes can be written as

$$p = \frac{2\langle L \rangle}{N(N-1)} = \frac{\langle k \rangle}{N-1} \simeq \frac{\langle k \rangle}{N}. \tag{2.18}$$

We arrive at the same result if we directly compute the average degree of the random graph,

$$\begin{aligned} \langle k \rangle &= \sum_{L=0}^{\binom{N}{2}} \frac{2L}{N} P(L) \\ &= \frac{2}{N} \sum_{L=0}^{\binom{N}{2}} L \binom{\binom{N}{2}}{L} p^m (1-p)^{\binom{n}{2}-L} \\ &= \frac{2}{N} \binom{N}{2} p \\ &= \frac{2}{N} \cdot \frac{N(N-1)}{2} p = (N-1)p. \end{aligned} \tag{2.19}$$

And, by inverting the last equation, we re-obtain Eq. (2.18).

As a final note, let us compute the degree distribution of the $G(N,p)$ ensemble. If a node i has degree k, it connects to exactly k out of the remaining $N-1$ nodes in the network. For a given set of k nodes, this happens with probability

$$p^k(1-p)^{N-1-k}. \tag{2.20}$$

And, seeing as there are $\binom{N-1}{k}$ ways of choosing the k vertices, the degree distribution of the $G(N,p)$ ensemble is also given by a binomial distribution

$$p(k_i = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \tag{2.21}$$

As real-world networks can contain millions if not billions of nodes, we are generally interested in studying the properties of networks in the limit of $N \to \infty$, also known as the *thermodynamic limit* in the physics nomenclature. In this limit, the binomial coefficient can be approximated as follows

$$\binom{N-1}{k} = \frac{(N-1)!}{(N-1-k)!k!} \simeq \frac{(N-1)^k + \mathcal{O}\left((N-1)^{k-1}\right)}{k!} \simeq \frac{(N-1)^k}{k!}. \tag{2.22}$$

Another common characteristic of real-world networks is that they tend to be *sparse*, that is, the number of edges in the network is small compared to the maximum possible number of edges that can exist. More formally, we will say that a network is sparse if the number of edges E is of the same order of magnitude as the total number of nodes N in the network

$$E = \mathcal{O}(N) \tag{2.23}$$

It is then easy to see that, for sparse networks,

$$\langle k \rangle = \frac{2E}{N} = \mathcal{O}(1). \tag{2.24}$$

Under these conditions, the probability p given in Eq. (2.18) becomes vanishingly small for $N \to \infty$ and we can write

$$\begin{aligned} \ln(1-p)^{N-1-k} &= (N-1-k)\ln\left(1 - \frac{\langle k \rangle}{N-1}\right) \\ &\simeq -(N-1-k)\frac{\langle k \rangle}{N-1} + \mathcal{O}\left(N^{-2}\right) \simeq -\langle k \rangle \end{aligned} \tag{2.25}$$

where we have expanded the logarithm as a Taylor series. By exponentiating both sides, we are left with

$$(1-p)^{N-1-k} \simeq e^{-\langle k \rangle}. \tag{2.26}$$

Substituting everything back into Eq. (2.21), we have that

$$p(k_i = k) = \frac{(N-1)^k}{k!} p^k e^{-\langle k \rangle} = \frac{(N-1)^k}{k!} \frac{\langle k \rangle^k}{(N-1)^k} e^{-\langle k \rangle} = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}. \tag{2.27}$$

Therefore, at the thermodynamic limit, the degree distribution of the $G(N,p)$ ensemble is given by a Poisson distribution. For this reason, the $G(N,p)$ model is often referred to as the *Poisson random graph*.

The $G(N,p)$ model has numerous other properties which we will not cover here [49] and its relative simplicity and analytical tractability have provided considerable insights into the structure of complex networks. However, the Poisson random graph also has significant drawbacks when used as a model for real-world networks, the most notable of which is highlighted by Eq. (2.27). Eq. (2.27) tells us that, for large networks, the $G(N,p)$ ensemble has a Poisson degree distribution. However, as we have seen previously, most real-world networks do not. Indeed, many of them appear to have scale-free degree distributions. We next introduce a generalised random graph model which allows to model networks with arbitrary degree distributions.

2.2.2 The configuration model

The *configuration model* is a random graph model in which we fix the degree sequence of the nodes. That is, we fix the degree k_i of each node i in the network.[1] A random network can then be generated according to the following procedure:

1. Assign k_i stubs to each node i of the network (the stubs represent the vertex degrees).

2. Select stub pairs uniformly at random and connect them, repeating the procedure until there are no more stubs left to connect, see Fig. 2.9.

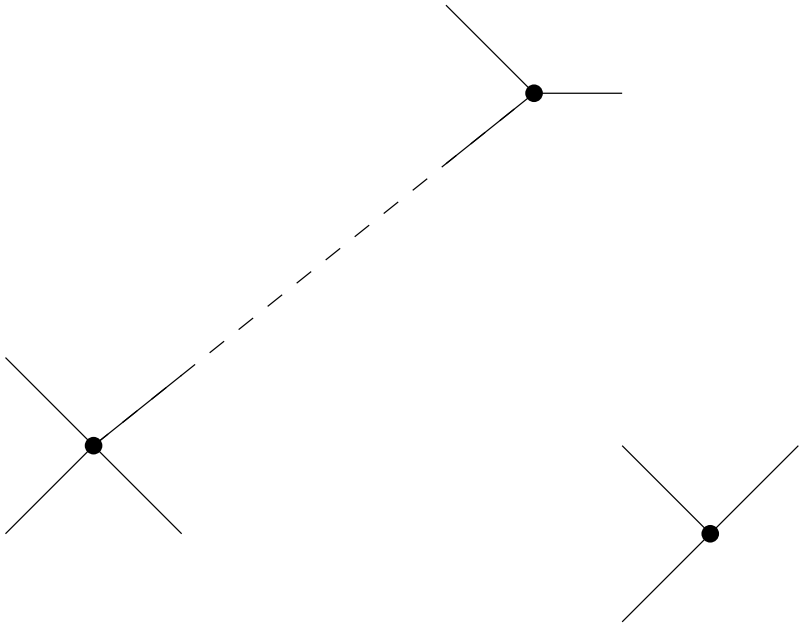


Figure 2.9: Example of the configuration model. Each node is assigned a series of stubs which are then randomly connected amongst each other.

This procedure generates a random network with a pre-specified degree sequence. The configuration model can then be defined as the ensemble of all possible matchings of a specified degree sequence, where each matching has the same probability of appearing. Notice that, since we consider all possible matchings between the stubs, the configuration model generates multigraphs rather than simple graphs, as nothing forbids joining two stubs belonging to nodes that already share an edge. Consequently, both self-edges and multi-edges may be present in the resulting network. Most real-world networks are not multigraphs, so the presence of self and multi-edges might seem like a significant drawback of our model. However, it can be shown that the density of self-edges and multi-edges goes to zero as $N \to \infty$, making the configuration model a suitable model for large networks [49]. Furthermore, the procedure described above is still viable for generating simple graphs; one simply has to discard any realisation in which multiple edges occur and repeat the procedure from the start. As we shall see, this still allows a uniform sampling of all simple graphs, as these are generated by the configuration model with equal probability.

[1] In many respects, the configuration model is similar to the $G(N, L)$ model seen before, as, by fixing the degree sequence, we are de facto fixing the number of edges in the network. Indeed, $\sum_i k_i = 2E$.

It is easy to notice that, for the configuration model to generate a simple network, the degree sequence must be such that $\sum_i k_i$ be even, as we cannot have disconnected stubs left over. More precisely, the degree sequence has to be *graphical*. A degree sequence $\boldsymbol{d} = \{k_1, k_2, ..., k_n\}$ is said to be graphical if there exists a simple graph G for which $\boldsymbol{d}$ is the degree sequence. The Erdős-Gallai theorem [68] gives a quick way to verify if a degree sequence is graphical.

Theorem 2.2.1 (Erdős-Gallai)**.** *A non increasing sequence $k_1 \geq \ldots \geq k_n$ is graphical if and only if*

- $\sum_i k_i$ *is even;*
- *For all $m \in \{1, \ldots, n\}$*

$$\sum_{i=1}^{m} k_i \leq m(m-1) + \sum_{i=m+1}^{n} \min\{m, k_i\} \tag{2.28}$$

Another consequence of the presence of self-loops and multi-edges is that, while all stub pairings appear with equal probability, *networks* do not. The reason behind this is that different permutations of the stubs can correspond to the same network, Fig. 2.10.

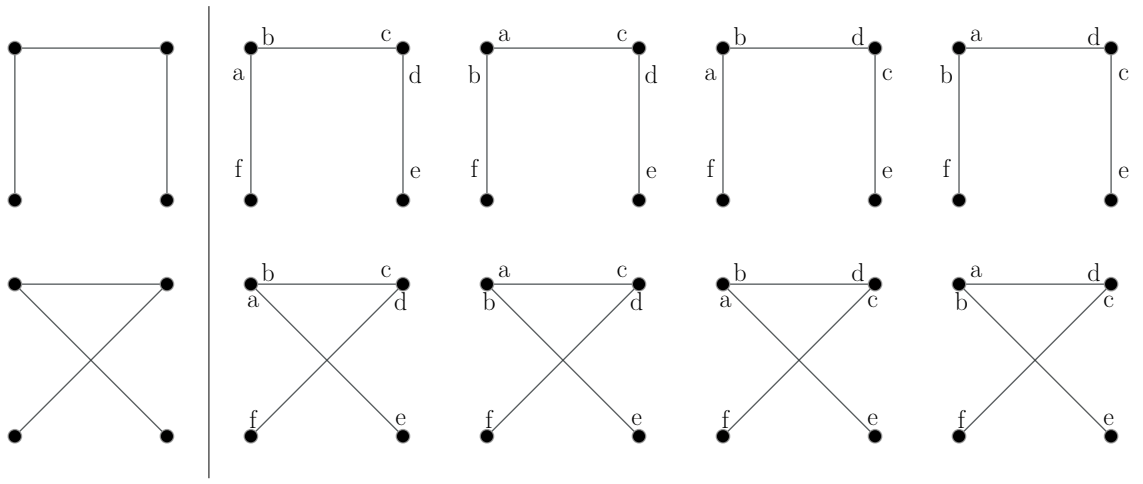


Figure 2.10: (a) Two sample networks with degree sequence $\{2, 2, 1, 1\}$. (b) All the possible stub permutations which give rise to the networks in (a). In this particular case, both the networks shown in (a) can emerge from four different stub permutations.

For a simple network, every possible pairing of the stubs can be obtained by permuting the stubs and, since node i has degree k_i, the number of permutations at node i is $k_i!$. Therefore, the number of matchings corresponding to a given network is $N(\{k_i\}) = \prod_i k_i!$. Since the degrees of the nodes remain unchanged, it follows that all simple graphs have the same probability of appearing in the configuration model. Thus, if we have $\Omega(\{k_i\})$ different matchings, each occurring with the same probability, the probability of observing any given network is given

by N/Ω. However, things change if we include self-loops and multi-edges in the picture. Self-loops and multi-edges can give rise to stub matchings which remain unchanged under permutation of the stubs, Fig. 2.11. In general, for multi-edges,

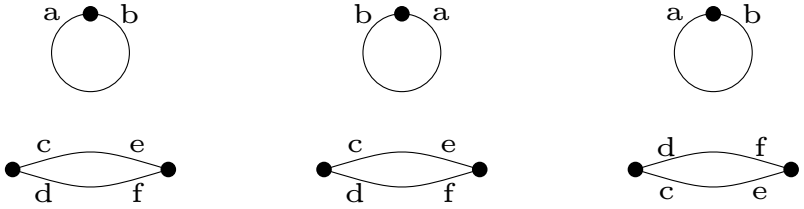


Figure 2.11: Three possible stub permutations which do not change the matching of the stubs. For the graph in (b), stub a is still connected to stub b even if we permute the labels (with respect to (a)). Similarly, in (c), stub c is still paired with stub e and stub d is still paired with stub f. Thus, the matchings are the same for all three networks irrespective of the permutations.

any permutation of the stubs will not produce a new matching as long as we also permute the stubs on the opposing vertex. Recalling that A_{ij} corresponds to the number of edges between two nodes in a multigraph, then the number of possible matchings is reduced by a factor of $A_{ij}!$. There is a further factor of two for self-edges, as permuting the two ends of a self-edge does not generate a new matching. Therefore, the number of matchings corresponding to a particular network in the configuration model is given by

$$N = \frac{\prod_i k_i!}{\prod_{i<j} A_{ij}! \prod_i A_{ii}!!} \tag{2.29}$$

where $n!! = n(n-2)(n-4)\ldots 2$. The total probability of observing a network is once more N/Ω. We observe that the denominator of Eq. (2.29) depends on the structure of the network, so different networks will appear with different probabilities. However, these differences should be small for large networks, as the density of self-loops and multi-edges tends to zero in the large N limit.

Computing the probability of observing an edge between two nodes is pretty straightforward. Consider a stub on node i. It can connect to any other stub in the network with equal probability. Since node j has k_j stubs, the probability that our selected stub will connect to a stub in node j is given by

$$\frac{k_j}{2E-1}, \tag{2.30}$$

and, seeing as there are k_i stubs on node i, the total probability that i connects to j is given by

$$p_{ij} = \frac{k_i k_j}{2E-1}. \tag{2.31}$$

Eq. (2.31) is actually the expression for the average number of edges between nodes i and j (seeing as we are simply summing over the independent probabilities for the

stubs of node i to connect to the stubs of node j). However, as previously mentioned, the density of multi-edges and self-loops goes to zero as $N \to \infty$. In this limit, the number of edges between nodes i and j is either 0 or 1, and the average number of edges coincides with the probability of observing an edge. Further disregarding the -1 at the denominator, we can write

$$p_{ij} = \frac{k_i k_j}{2E}. \tag{2.32}$$

Notice, however, that for Eq. (2.32) to correctly represent a probability measure, the degrees of our network cannot take arbitrarily large values, as we must ensure that $p_{ij} < 1$. Let then k_{max} be the maximum degree in our network. It follows that

$$p_{ij} = \frac{k_i k_j}{2E} \leq \frac{k_{max}^2}{2E} \leq 1 \implies k_{max} \leq K_S = \sqrt{2E}, \tag{2.33}$$

where K_S is commonly known as the *structural cut-off* of the network. It can be shown that if Eq. (2.33) is not satisfied when generating simple graphs via the configuration model, the resulting networks will exhibit some degree of *structural disassortativity*, whereby low degree nodes will preferentially attach to high degree ones, introducing correlations between the degrees of the network. Alternatively, if we do not wish to introduce any correlation between the vertex degrees, we must allow for multi-edges and self-edges in the network, in which case Eq. (2.32) no longer represents the probability that node i is connected to j. Therefore, for Eq. (2.32) to hold, the degrees of the network must be upper bounded by the structural cut-off,

$$k_i \ll K_S = \sqrt{2E}. \tag{2.34}$$

The excess degree distribution

We look at one last interesting property exhibited by the configuration model. Suppose we randomly select a node in the network and follow one of its edges to the opposite vertex; what is the probability that the vertex reached this way has degree k? Following a similar argument as before, we know that, excluding the stub we have randomly selected, there are $2E - 1$ stubs in the network. k of these stubs attach to any specific vertex of degree k. Therefore, our selected stub has a probability $k/(2E-1)$ of leading to a particular vertex of degree k. Since there are Np_k vertices of degree k in the network, the probability that our randomly selected edge ends up on a node of degree k is given by

$$\frac{k}{2E-1} \cdot Np_k \simeq \frac{Nk}{2E} p_k = \frac{kp_k}{\langle k \rangle} \tag{2.35}$$

This fraction depends on kp_k rather than p_k, and this property has intriguing consequences. Let us evaluate the average degree of the neighbours of a randomly selected node. This is nothing else than k multiplied by the probability that the node we reach by following an edge has degree k, summed over all values of k,

$$\sum_k k \frac{kp_k}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} \tag{2.36}$$

So the average degree of a neighbour is different from the average degree of a node in the network, $\langle k \rangle$—generally, it is larger.

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} = \frac{\langle (k - \langle k \rangle)^2 \rangle}{\langle k \rangle} \tag{2.37}$$

Since both the numerator and the denominator of Eq. (2.37) are strictly non-negative (unless all nodes in the network have degree zero), Eq. (2.37) implies that

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle > 0 \Longrightarrow \frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle \tag{2.38}$$

This intriguing phenomenon is colloquially known as "your friends have more friends than you do". It is important to stress that this is a property of the configuration model. Generally, in real-world networks, the degrees of connected vertices are correlated. The probability of reaching a node of degree k is not simply proportional to kp_k but depends on the starting node. Remarkably, however, this basic property appears to also hold in many real-world networks [69].

Eq. (2.35) allows us to introduce a quantity which will be of use later, the *excess degree distirbution*. Many times, we are interested in how many vertices a neighbour node has *other* than the one we followed to get to it. This is known as the *excess degree*, and its distribution can be easily computed via Eq. (2.35) by simply setting k to $k + 1$,

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}. \tag{2.39}$$

Indeed, if a node has an excess degree k, it must necessarily have degree $k + 1$.

2.2.3 Stochastic block models

Stochastic block models are generative models for groups or communities in networks falling under the category of random graph models. Stochastic block models were first introduced in the context of social sciences [70] as a way of combining the properties of deterministic block models (models which map actors in a social network to one of B groups) such as [71] and stochastic models. The rationale is that stochastic models would allow for a framework capable of accounting for variability in the data. The original work of Holland *et al* assumed the partitioning of the nodes to be known *a priori* through additional knowledge of properties of the actors comprising the network. However, in most cases relating to real-world networks, such partitions are not known beforehand. To circumvent this problem, Wasserman and Anderson developed techniques that allow to perform what is known as *a posteriori* blockmodelling, i.e. fitting an SBM to real-world network data to infer an optimal partition [72].

As the field of network science evolved, SBMs found widespread use also outside the context of social sciences [70, 72–74], particularly in community detection problems [56, 75–77], where the aim is to uncover groups of nodes which are densely connected amongst each other. More generally, SBMs can be used to infer generic mesoscale network topologies [65, 78]. This is a significant advantage over most other community detection methods that generally operate under the assumption of assortative mixing patterns or focus on recovering only one kind of topology. Another fundamental characteristic that makes SBMs highly attractive models is their

generative nature. SBMs can recover the underlying partitioning of the nodes and infer the interactions between the elements of the various groups. This information provides us with a putative model of network formation, which we can then use to generate synthetic networks.

In its most simple formulation, a stochastic block model divides N nodes into B groups (typically with $B \ll N$), where each node is assigned a label $b_i \in \{1, \ldots, B\}$ specifying its group membership. Undirected edges are then placed amongst the vertex pairs with probabilities $p_{b_i b_j}$ which depend solely on the group memberships of the nodes. As mentioned above, the number of groups is generally unknown and must be inferred from the data. However, for our purposes, we shall assume this number to be known a priori.

The matrix p_{rs} governing the probabilities of an edge between a node in group r and one in group s is known as the *affinity matrix*, and, by appropriately choosing the values of the p_{rs}, a wide array of different network topologies can be generated, see Fig. 2.12.

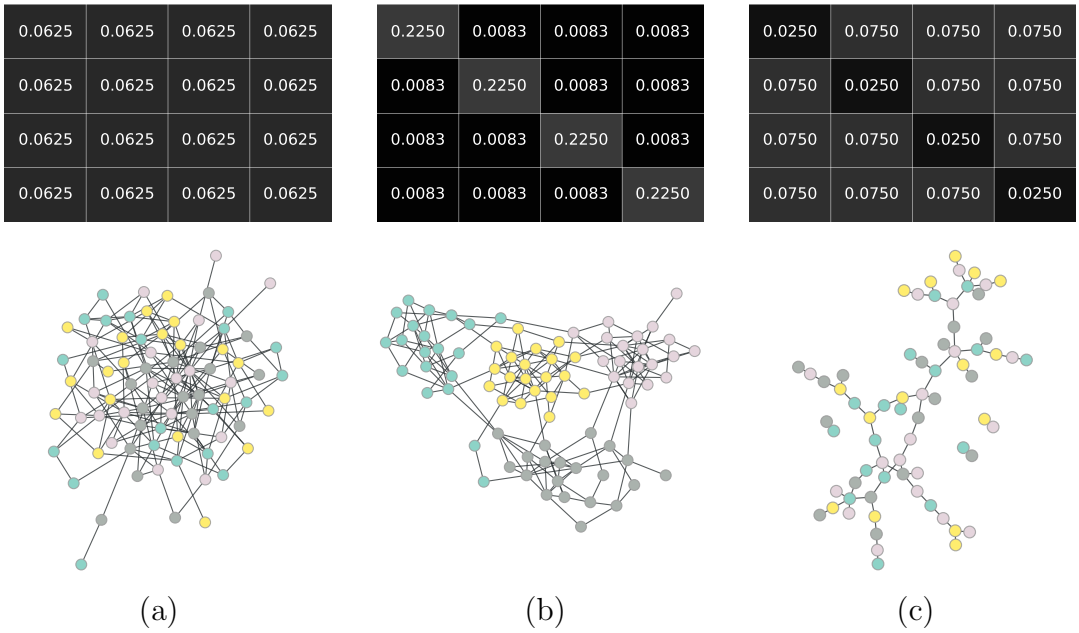


Figure 2.12: Three different structures generated via a SBM. The top panels show the affinity matrix, p_{rs}, whose elements correspond to the probability of observing an edge between blocks r and s. The bottom panels display examples of networks generated using the corresponding affinity matrices. (a) A random graph, characterised by having all identical elements in its affinity matrix (i.e. all groups are equally probable to be connected) (b) A cluster structure, where the corresponding affinity matrix is characterised by having higher values along its diagonal (i.e. a higher probability of connections amongst nodes belonging to the same group). (c) A disassortative structure, in which nodes from one group preferentially attach to nodes of other groups. This structure is characterised by having higher off-diagonal elements.

Since edges are independently placed amongst vertex pairs with probabilities p_{rs}, the probabilities p_{ij} of observing an edge between node i and node j are independent Bernoulli random variables with Bernoulli parameters p_{rs}. The total probability of observing a specific graph $\boldsymbol{A}$ in the ensemble is then given by

$$P(\boldsymbol{A}|\boldsymbol{b}, \boldsymbol{p}) = \prod_{i<j} p_{b_i,b_j}^{A_{ij}} (1 - p_{b_i,b_j})^{1-A_{ij}}, \tag{2.40}$$

where $A_{ij} = 1$ if an edge exists between nodes (i, j) or $A_{ij} = 0$ otherwise. Notice that for the case $B = 1$, where we only have one group, we recover the $G(N, p)$ ensemble studied in Section 2.2.1, and Eq. (2.40) coincides with Eq. (2.14).

The advantage of working with the stochastic block model is that we now have to consider only the degrees of freedom associated with the individual groups, and the adjacency matrix can be summarised by a smaller but not necessarily symmetric $B \times B$ matrix. This grouping of nodes into blocks is a natural form of dimensionality reduction. Since we cannot consider all possible pairwise interactions in the network, we reduce ourselves to analysing the interactions amongst $B \ll N$ objects. Although this might appear as a very crude approximation, we note that it can be made ever more refined by increasing the number of blocks B, allowing us to model increasingly realistic scenarios. In fact, setting $B = N$ means that the probability of each node can be individually controlled.

The SBM, as introduced here, generates simple graphs according to the probabilities encoded in the affinity matrix, but can easily be generalised to account for multi-edges [65, 79]. Extensions to directed networks have also been considered [65, 78].

Despite their usefulness and popularity, SBMs also present some limitations. The statistical indistinguishability of the nodes means that once the group division is known, edges are assigned amongst vertex pairs in an Erdős-Rényi random fashion. Consequently, SBMs cannot reproduce some local features commonly observed in real-world networks, such as cliques and other motifs. In particular, for the classic version of the stochastic block model presented here, this statistical indistinguishability can have disastrous consequences, as all nodes in each group will be assigned on average the same number of edges. This is unrealistic with respect to real-world networks, which generally display highly heterogeneous degree distributions, and can lead to severely incorrect partitions when the SBM is used to infer the community structure. To obviate this problem, Karrer and Newman provide an important generalisation which they name *degree corrected stochastic block model* (DC-SBM) [75], in which an additional parameter is associated with each node, controlling its expected degree. The DC-SBM has the advantage of generating networks with arbitrary degree distributions and performs considerably better than the standard SBM when used as a tool for community detection. Additional variants of the SBM have been proposed to capture further variability of the network structure. Examples include overlapping and mixed membership models [80–82], hierarchical block models [78, 83], weighted networks [84, 85], and networks whose nodes are annotated with metadata [86–88]. More generally, SBMs provide a general framework that can be applied to many problems arising in network science.

2.2.4 Maximum entropy network ensembles

The previous sections introduced examples of random graph models that generate networks with some pre-specified characteristics but which are otherwise random. In all cases, networks were generated by fixing the values of the quantities of interest and then sampling uniformly from the ensemble of networks with the desired values. While the choice of a uniform distribution over the ensemble is intuitive, we did not motivate its choice when introducing the models. When working with these kinds of models, a crucial aspect is characterising the probability distribution $P(G)$ over

the ensemble. Indeed, the number of constraints we impose on the ensemble by fixing the properties of interest is generally far smaller than the number of networks in it. This leads to a heavily underdetermined system in which many different combinations of the degrees of freedom of the probability distribution can result in the same prescribed values of the constraints. We are then faced with having to pick a particular distribution $P(G)$ out of all the ones that satisfy the constraints.

Intuitively, we would like to pick the arrangement we are most likely to find, given no additional information other than that carried by the imposed criteria. Jaynes' *maximum entropy principle* [89], states that this most likely arrangement should coincide with that carrying the largest degeneracy, or *maximum entropy*. Effectively, this amounts to selecting the distribution which is less biased given our knowledge of the constraints.

To give a better understanding of why this is so, we briefly go over an example introduced in [89] and first attributed to Wallis. Suppose we are given N balls that are to be distributed amongst K urns such that n_i balls are placed in the i-th urn and $\sum_i n_i = N$. (To put this into context, we can think of N as the number of networks sampled from a particular ensemble and K as the number of distinct configurations for each network's degrees of freedom. n_k then represents the number of networks that are in configuration k). Given a certain arrangement $\{n_i\}$ of balls in the urns, the number of different ways that this arrangement can be obtained is given by the multinomial coefficient

$$\Omega = \frac{N!}{n_1! n_2! \dots n_K!} \tag{2.41}$$

For $N \to \infty$, we can use Stirling's approximation of the factorial, $m! \simeq m^m e^{-m}$ on each term in Ω,

$$\begin{aligned} \Omega &\simeq \frac{N^N e^{-N}}{n_1^{n_1} \cdots n_K^{n_K} e^{-(n+1+\cdots+n_K)}} = \frac{N^N}{n_1^{n_1} \cdots n_K^{n_K}} \\ &= e^{\ln(N^N) - \ln(n_1^{n_1} \cdots n_K^{n_K})} = e^{N \ln N - \sum_{i=1}^K n_i \ln n_i} \\ &= e^{N\left[-\sum_{i=1}^K \frac{n_i}{N}(\ln n_i - \ln N)\right]} = e^{-N \sum_{i=1}^K \frac{n_i}{N} \ln \frac{n_i}{N}} \\ &= e^{N\Sigma}, \end{aligned} \tag{2.42}$$

where Σ is the *entropy* of the arrangement $\{n_i\}$. Note that since $\sum_{i=1}^K n_i = N$, then $n_i/N \in [0, 1]$. This, in turn, means that the entropy is always greater than or equal to zero, with the equality holding only if $n_i = N$ and $n_j = 0 \quad \forall j \neq i$.
What Eq. (2.42) is telling us is that, for large N, some arrangements can be realised in a vast number of ways (exponential in N) and will have larger entropies, while others have very small degeneracies. For example, the case in which all balls end up in a single urn can be realised in only $\Omega = 1$ way and will have entropy $\Sigma = 0$. The entropy is then really a measure of the microscopic multiplicity underlying a macroscopic state of the system.

The terms n_i/N can be interpreted as the probability that a ball sampled uniformly at random comes from urn i. Under this interpretation, the entropy can be written as

$$\Sigma = -\sum_{i=1}^{K} p_i \ln p_i, \tag{2.43}$$

and the constraint that $\sum_{i=1}^{K} n_i = N$ simply corresponds to the condition that our probabilities be correctly normalised, $\sum_{i=1}^{K} p_i = 1$.

It is intuitive that if we randomly distribute the balls amongst the urns, then one would expect the final arrangement, $\{n_i\}$ (or the most likely distribution of probabilities $\{p_i\}$) to correspond to one of high degeneracy rather than one with a small multiplicity. In fact, the most likely distribution $\{p_i^*\}$ should coincide with the one that maximises the entropy given the constraint that $\sum_i p_i = 1$.
This is the gist of the maximum entropy principle. If one is to infer a distribution $\{p_i\}$ given some constraints, then amongst all possible $\{p_i\}$ compatible with the constraints, one should select the distribution that maximises Eq. (2.43).

Proposition 2.2.1. *Consider a random variable x, which can take M distinct values from a set Ω. If the only constraint considered is the normalisation of probabilities, the uniform distribution maximises the entropy, and the maximum entropy is given by*

$$\Sigma = \ln M. \tag{2.44}$$

Proof. We can maximise the entropy over all normalised distributions by resorting to the method of Lagrange multipliers. Given the functional

$$\mathcal{F} = \Sigma - \gamma \left(\sum_{x \in \Omega} p(x) - 1 \right) = - \sum_{x \in \Omega} p(x) \ln p(x) - \gamma \left(\sum_{x \in \Omega} p(x) - 1 \right), \tag{2.45}$$

the corresponding saddle-point equations are

$$\frac{\partial \mathcal{F}}{\partial p(x)} = - \ln p(x) - 1 - \gamma = 0 \tag{2.46}$$

$$\frac{\partial \mathcal{F}}{\partial \gamma} = \sum_{x \in \Omega} p(x) - 1 = 0, \tag{2.47}$$

where the partial derivative with respect to the Lagrange multiplier γ is simply the normalisation condition. From Eq. (2.46)

$$p(x) = e^{-(1+\gamma)}, \tag{2.48}$$

and, since the cardinality of Ω is $|\Omega| = M$, from Eq. (2.47) we have that

$$\sum_{x \in \Omega} p(x) = M e^{-(1+\gamma)} = 1 \quad \Longrightarrow \quad p(x) = \frac{1}{M}. \tag{2.49}$$

By substituting the expression obtained in Eq. (2.49) in our expression for the entropy, Eq. (2.43), we have that

$$\Sigma = - \sum_{x \in \Omega} p(x) \ln p(x) = - \sum_{x \in \Omega} \frac{1}{M} \ln \frac{1}{M} = \sum_{x \in \Omega} \frac{1}{M} \ln M = \ln M. \tag{2.50}$$

Where we have made use of the fact that $|\Omega| = M$. □

This is the case when the normalisation of probabilities is the only constraint. However, additional constraints can give rise to different maximum entropy distributions reflecting the additional information that the constraints introduce into the inference problem.

An important consideration is that, since maximum entropy distributions correspond to the ones with the highest degeneracy, they are the most unbiased distributions given the constraints. Indeed, any other distribution, having a lower degeneracy, would automatically exclude some valid configurations of the microscopic degrees of freedom. Thus, in essence, the maximum entropy principle (MEP) tells us that the least biased ensemble that satisfies a set of constraints is the one that maximises the entropy given the constraints.

Canonical and Microcanonical ensembles

We now turn to how we can leverage the maximum entropy principle to construct network models, and, following the work developed in [63, 64, 90, 91], we introduce the concept of canonical and microcanonical ensembles.

The general idea is that one would like to fix the probability $P(G)$ of the network ensemble by using the maximum entropy principle with a given set of constraints. So what are the possible constraints we can consider? Let $\{x_i\}$ be a set of network measures whose value we wish to fix. Then the general prescription is to consider two principal classes of constraints [90]. *Hard constraints* are constraints that are strictly satisfied by every network in the ensemble

$$x_i(G) = c_i \text{ for } i = 1, 2, \ldots, M, \tag{2.51}$$

where M is the cardinality of $\{x_i\}$. Examples of hard constraints are fixing the number of links L in the network or fixing the network's degree sequence. *Soft constraints*, on the other hand, are constraints that are satisfied on average over the ensemble of networks.

$$\sum_{G\in\Omega} x_i(G)P(G) = c_i \text{ for } i = 1, 2, \ldots, M. \tag{2.52}$$

For example, we can fix the average number of edges in a network or the mean degree of a node.

Given these two classes of constraints, we can now define the *canonical* and *microcanonical* ensembles as follows:

- The *canonical* ensemble is the maximum entropy ensemble which satisfies a set of soft constraints
$$\sum_{G\in\Omega} x_i(G)P(G) = c_i.$$
- The *microcanonical* ensemble is the maximum entropy ensemble which satisfies a set of hard constraints
$$x_i(G) = c_i.$$

Let us study some of their properties.

The canonical ensemble: For the canonical ensemble, we wish to maximise the entropy given in Eq. (2.43) subject to the constraints

$$\sum_{G\in\Omega} x_i(G)P(G) = c_i, \tag{2.53}$$

and the condition that $P(G)$ be properly normalised

$$\sum_{G\in\Omega} P(G) = 1. \tag{2.54}$$

This can be achieved by resorting once more to the method of Lagrange multipliers. The Lagrangian is now given by

$$\mathcal{L} = -\sum_{G\in\Omega} P(G)\ln P(G) - \sum_{i=1}^{M}\beta_i\left(\sum_{G\in\Omega} x_i(G)P(G) - c_i\right) - \gamma\left(\sum_{G\in\Omega} P(G) - 1\right), \tag{2.55}$$

and the corresponding saddle point equations are

$$\begin{cases} \dfrac{\partial\mathcal{L}}{\partial P(G)} = -\ln P(G) - 1 - \displaystyle\sum_{i=1}^{M}\beta_i x_i(G) - \gamma = 0, & (2.56)\\ \dfrac{\partial\mathcal{L}}{\partial\beta_i} = \displaystyle\sum_{G\in\Omega} x_i(G)P(G) - c_i = 0, & (2.57)\\ \dfrac{\partial\mathcal{L}}{\partial\gamma} = \displaystyle\sum_{G\in\Omega} P(G) - 1 = 0. & (2.58) \end{cases}$$

From Eq. (2.56), we have

$$\ln P(G) = -\sum_{i=1}^{M}\beta_i x_i(G) - 1 - \gamma \tag{2.59}$$

$$\implies P(G) = e^{-1-\gamma}e^{-\sum_{i=1}^{M}\beta_i x_i(G)} = \frac{e^{-H(G)}}{Z}, \tag{2.60}$$

where $Z = e^{-1-\gamma}$ is known as the *partition function* and $H(G) = \sum_{i=1=}^{M}\beta_i x_i(G)$ is the *graph Hamiltonian*. The value of Z can be determined from the normalisation condition

$$\sum_{G\in\Omega} P(G) = \sum_{G\in\Omega}\frac{e^{-H(G)}}{Z} = 1 \implies Z = \sum_{G\in\Omega} e^{-H(G)} \tag{2.61}$$

while the Lagrange multipliers are obtained by substituting Eq. (2.60) into Eq. (2.53)

$$\begin{aligned} c_i = \sum_{G\in\Omega} x_i(G)\frac{e^{-H(G)}}{Z} &= \frac{1}{Z}\sum_{G\in\Omega} x_i(G)e^{-\sum_{i=1}^{M}\beta_i x_i(G)} \\ &= -\frac{1}{Z}\frac{\partial}{\partial\beta_i}\sum_{G\in\Omega} e^{-\sum_{i=1}^{M}\beta_i x_i(G)} = -\frac{1}{Z}\frac{\partial Z}{\partial\beta_i} = -\frac{\partial\ln Z}{\partial\beta_i}. \end{aligned} \tag{2.62}$$

$\ln Z$ is sometimes written as $F \equiv \ln Z$ and called the *free energy* of the system. Because of the exponential form of Eq. (2.60), canonical network ensembles are often referred to as *exponential random graphs*.

Once the distribution over the ensemble, $P(G)$, is known, we can use it to evaluate the expected value of any graph property $y(G)$

$$\langle y\rangle = \sum_{G\in\Omega} y(G)P(G) = \frac{1}{Z}\sum_{G\in\Omega} y(G)e^{H(G)}. \tag{2.63}$$

While very elegant, the exponential random graph model can be tricky to implement from a technical standpoint. In particular, evaluating the sum in the partition function can be difficult and is often impossible to do so analytically. In these cases, one usually has to resort to numerical methods to evaluate Z. In [62], Park and Newman give an excellent overview of a series of techniques that can be employed to obtain exact or approximate analytical solutions in the exponential random graph model.

The microcanonical ensemble: The microcanonical network ensemble is the maximum entropy ensemble given a set of hard constraints $x_i(G) = c_i, \;\; i \in [1, M]$. Therefore, we have that $P(G) > 0$ if and only if $x_i(G) = c_i$. The entropy of the ensemble can then be written as

$$\Sigma = -\sum_{G\in\Omega|\{x_i(G)=c_i\}_{i=1,2,\ldots,M}} P(G)\ln P(G) \tag{2.64}$$

where the sum runs only over those networks that exactly satisfy the constraints. The cardinality of Ω will then be given by

$$Z_{MC} = \sum_{G\in\Omega}\prod_{i=1}^{M}\delta\left(x_i(G), c_i\right). \tag{2.65}$$

We have previously seen that the maximum entropy distribution over a set of possible outcomes with fixed cardinality is the uniform distribution. The probability $P(G)$ of observing a particular network is then

$$P(G) = \frac{1}{Z_{MC}}\prod_{i=1}^{M}\delta\left(x_i(G), c_i\right) \tag{2.66}$$

Substituting this into our previous expression for the entropy we have that

$$\Sigma = -\sum_{G\in\Omega|\{x_i(G)=c_i\}_{i=1,2,\ldots,M}} \frac{1}{Z_{MC}}\ln\left(\frac{1}{Z_{MC}}\right) = \ln Z_{MC}. \tag{2.67}$$

The microcanonical entropy is thus equal to the logarithm of the number of networks in the ensemble.

An example: the $\mathbf{G(N, p)}$ ensemble Maximum entropy ensembles provide a general framework to model networks subject to a wide array of constraints, including the random graph and the configuration model as special cases. Consider, for example, the case in which we want to fix the average number of links in the network, $\langle L\rangle$. Our constraint is then given by

$$\sum_{G\in\Omega}\left(\sum_{i<j}A_{ij}\right)P(G) = \langle L\rangle. \tag{2.68}$$

According to the general theory of exponential random graphs, the maximum entropy distribution will be given by

$$P(\boldsymbol{A}) = \frac{1}{Z}e^{-\beta\sum_{i<j}A_{ij}}, \tag{2.69}$$

where we have made use of the fact that, seeing as a network is fully specified by its adjacency matrix, $P(\boldsymbol{A}) = P(G)$. We can evaluate the partition function as

$$Z = \sum_{\{\boldsymbol{A}\}} e^{-\beta \sum_{i<j} A_{ij}} = \prod_{i<j} \sum_{A_{ij}=\{0,1\}} e^{-\beta A_{ij}} = \prod_{i<j} \left(1 + e^{-\beta}\right) = \left(1 + e^{-\beta}\right)^{\binom{N}{2}}. \quad (2.70)$$

Given Z, we can now evaluate the probability p_{ij} of a link existing between nodes i and j. p_{ij} will be given by the average of the corresponding element A_{ij} in the adjacency matrix,

$$\begin{aligned} p_{ij} = \langle A_{ij} \rangle &= \frac{1}{Z} \sum_{\{\boldsymbol{A}\}} A_{ij} e^{-\beta \sum_{r<s} A_{rs}} \\ &= \frac{1}{Z} e^{-\beta} \prod_{r<s|(r,s)\neq(i,j)} \sum_{A_{rs}=\{0,1\}} e^{-\beta A_{rs}} \\ &= \frac{e^{-\beta}}{Z} \left(1 + e^{-\beta}\right)^{\binom{N}{2}-1} \\ &= \frac{e^{-\beta}}{1 + e^{-\beta}}, \end{aligned} \quad (2.71)$$

which is the same for any pair (i, j). The expected number of edges is then given by

$$\langle L \rangle = \sum_{\{\boldsymbol{A}\}} \left(\sum_{i<j} A_{ij} \right) P(\boldsymbol{A}) = \sum_{i<j} \left(\sum_{\{\boldsymbol{A}\}} A_{ij} P(\boldsymbol{A}) \right) = \sum_{i<j} p_{ij} = \binom{N}{2} p, \quad (2.72)$$

and the probability p of an edge existing between two edges can then be expressed as

$$p = \frac{2\langle L \rangle}{N(N-1)} \simeq \frac{\langle k \rangle}{N}. \quad (2.73)$$

This is the same probability we observed in Eq. (2.18) when studying the $G(N, p)$ model. Indeed, the $G(N, p)$ model is a particular case of the random graph model. Similarly, both the $G(N, L)$ ensemble and the configuration model can be obtained from the microcanonical ensemble by fixing respectively the number total number of edges and the degree sequence respectively.

2.3 Modelling the optimisation criteria

The previous section has introduced some of the most fundamental network models that can be employed to generate synthetic networks. The resulting networks can then be analysed and their properties and characteristics better studied and compared to what is observed in real-world networks.

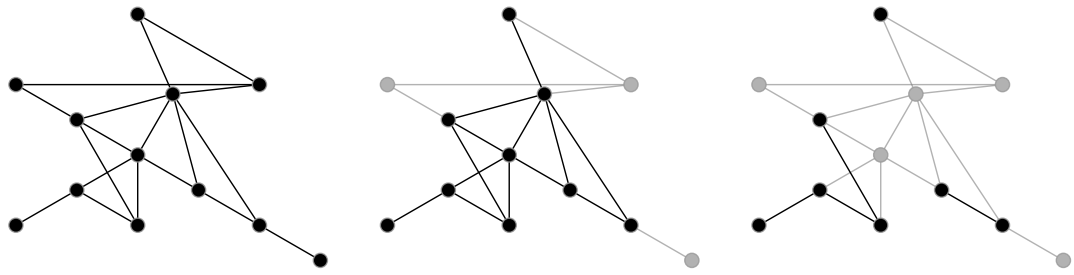
As mentioned in the Introduction, our principal goal is to develop null models for optimal networks. That is, networks subject to one or more selective pressures driving them to optimise for some pre-specified characteristics. As such, a crucial step in our formulation is the choice of the network properties and features we wish to select for. While a vast collection of diverse properties characterising complex networks exists, we focus our attention on two key attributes commonly observed

in real-world networks: their robustness against the random failures of their components and the emergence of assortative mixing patterns amongst the network elements. In the following section, we introduce the main concepts and modelling techniques used to analyse these properties.

2.3.1 Percolation and robustness against random failures

The *robustness* (or *resilience*) of a network generally refers to the network's ability to keep functioning when a certain fraction of its components fails or malfunctions and is a characteristic of fundamental importance in the context of network design. Indeed, as failures are bound to occur, one would like to have an accurate picture of how this affects the network's ability to operate correctly and what fraction of the network is still functioning. Furthermore, network resilience is a commonly observed property in numerous real-world networks across a wide array of diverse fields ranging from technological networks to biological ones. As a consequence, network robustness has been an extensively studied topic over the years. Although various methods for quantifying a network's robustness exist, in this section, we review how this problem can be elegantly framed as a *percolation* process.

By *percolation* we denote the process of removing a certain fraction of nodes or edges from the network. The former case is known in the literature as *site percolation*, the latter as *bond percolation*. Intuitively, this process provides us with a qualitative basis to understand a network's robustness. Indeed, the functionality of a network depends on its connectivity. If a network is completely disconnected, it cannot achieve its desired function, whatever that may be. For example, consider the network in Fig. 2.13. Panel (a) displays the original network, in which no nodes have been removed. As we begin to remove nodes, panel (b), the network shrinks, but the remaining nodes are still connected. If we continue to remove nodes from the network, we eventually reach a point at which the network splits into smaller

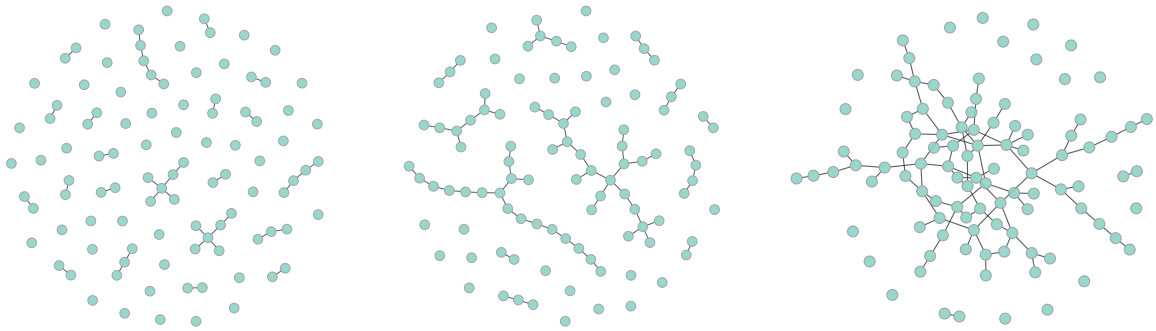


(a) (b) (c)

Figure 2.13: Toy network depicting the topology of the network as an increasing number of nodes is removed from the network. Gray nodes represent those nodes which have been removed from the network, along with their corresponding edges. Panel (a) represents the initial network with no nodes removed. In panel (b) we have removed a certain number of nodes, but the remaining nodes are still connected. Panel (c) represents the case in which so many nodes have been removed that the network splits into smaller disconnected components.

components, and its functionality is compromised, panel (c). This connection between a network's ability to function and its long-range connectivity is elegantly captured by percolation theory.

Percolation processes are typically parametrised by an *occupation probability* ϕ, representing the probability that a node (edge) has *not* been removed from the network. By varying ϕ, one can observe an abrupt transition in which the network topology shifts from being characterised by small disconnected components to forming a *giant connected component* (GCC) spanning most of the network. See Fig. 2.14. The critical value $\phi = \phi_c$ at which this transition takes place is known as the *percolation threshold*.



(a) (b) (c)

Figure 2.14: Emergence of a giant connected component in a random graph subject to a bond percolation process.

A common measure of a network's robustness is then given by the fraction of nodes that are part of the giant connected component. While the presence of a giant component is an indicator that the network is at least partially performing its intended function, the size of the giant component indicates the fraction of the network that is still functional.

Following [49], we review the theory of percolation on a configuration model network in which the nodes are randomly removed.

Consider a particular vertex v that has not been removed from the network. For it to be part of the giant connected component, it must necessarily connect to it via one of its neighbours. Let then u be the average probability that a node in the network is *not* connected to the giant component via one of its neighbours. If the node in question has degree k, then the total probability that it is not part of the giant component is given by u^k. Averaging over the degree distribution p_k, we obtain the average probability that a node is not connected to the giant component via one of its neighbours

$$\sum_{k=0}^{\infty} p_k u^k \coloneqq f_0(u), \tag{2.74}$$

where $f_0(z) = \sum_k p_k z^k$ is the *generating function* of the degree distribution. The average probability that a node *is* part of the GCC is then simply given by

$$P(v \in \text{GCC}) = 1 - f_0(u) \tag{2.75}$$

Eq. (2.75), however, assumes that the node in question has not been removed from the network, which happens with probability ϕ. Therefore, the fraction of nodes in the giant connected component is then given by

$$S = \phi[1 - f_0(u)]. \tag{2.76}$$

We must now compute the average probability u that a node is not connected to the giant component via one of its neighbours. There are two possible way for this to occur:

1. The neighbour in question has been removed from the network, which happens with probability $1 - \phi$.

2. The neighbour in question is present in the network, but it is not connected to the GCC via one of its neighbours.

If the selected neighbour has degree k, then the probability that it is not connected to the GCC is given by u^k. Adding these terms together, we have that the probability u that a node is not connected to the giant component via a specific neighbour is given by

$$1 - \phi + \phi u^k. \tag{2.77}$$

When averaging Eq. (2.77) on the degree distribution of the selected neighbour, we must consider the fact that, since the neighbour is reached following an edge, k will be distributed according to the excess degree distribution introduced previously.

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle} \tag{2.78}$$

The average probability for a node not to connect to the giant component is then given by

$$\begin{aligned} u &= \sum_{k=0}^{\infty} q_k \left(1 - \phi + \phi u^k\right) \\ &= 1 - \phi + \phi \sum_{k=0}^{\infty} q_k u^k \\ &= 1 - \phi + \phi f_1(u), \end{aligned} \tag{2.79}$$

where $f_1(z) \coloneqq \sum_k q_k z^k$ is the generating function of the excess degree distribution, which can also be obtained from the degree generating function $f_0(z)$ as

$$f_1(z) = \frac{f_0'(z)}{f_0'(1)}. \tag{2.80}$$

Solving for Eq. (2.76) and Eq. (2.79) allows us to fully characterise the giant connected component. Unfortunately, Eq. (2.79) is generally not solvable in closed form. However, using graphical arguments, it is possible to show that it has only two solutions [49]: a trivial solution at $u = 1$ for which $S = 0$, and a non-trivial solution, whose existence depends on the value of the occupation probability ϕ. For small values of ϕ, the trivial solution remains the only solution, and there is no giant connected component. However, as ϕ grows, we reach a critical point $\phi = \phi_c$

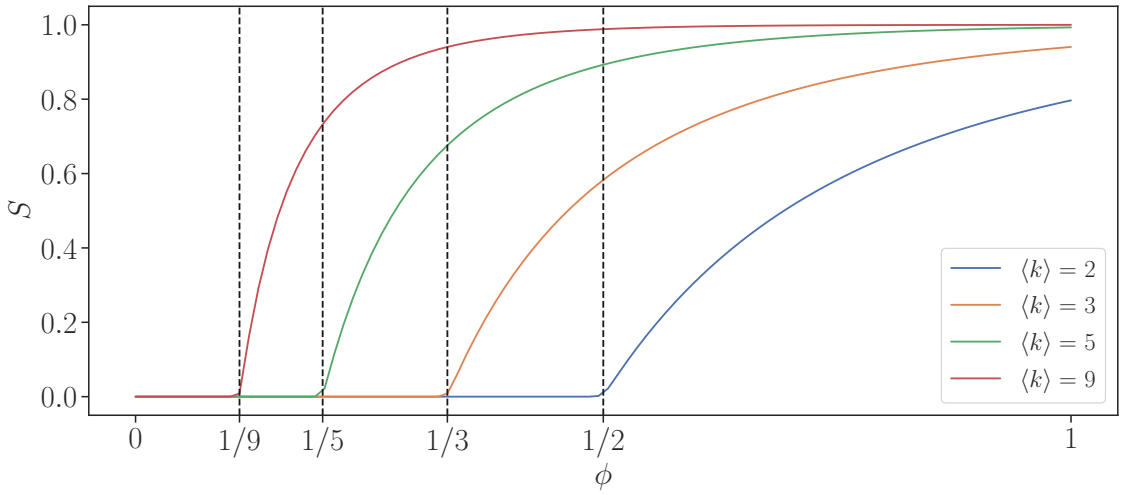


Figure 2.15: Plots of S as a function of the dilution probability ϕ for different values of $\langle k \rangle$. The vertical dashed lines indicate the positions of the percolation transitions.

at which the non-trivial solution appears, and we observe the formation of a giant connected component. This critical value ϕ_c is precisely the percolation transition.

Specifically, it can be shown that the condition at which the non-trivial solution first appears corresponds to the point at which Eq. (2.79) is tangent to the line $y = u$.

$$\left[\frac{d}{du}\left(1 - \phi + \phi f_1(u)\right)\right]_{u=1} = 1 \implies \phi_c = \frac{1}{f_1'(1)}. \tag{2.81}$$

Performing the derivative of $f_1(u)$, we have that

$$f_1'(1) = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}, \tag{2.82}$$

and the critical value of ϕ corresponding to the percolation transition is given by

$$\phi_c = \frac{1}{f_1'(1)} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{2.83}$$

Let us consider the case in which a Poisson degree distribution characterises our network,

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}. \tag{2.84}$$

In this case,

$$\langle k \rangle = \lambda \quad ; \quad \langle k^2 \rangle = \lambda(\lambda - 1), \tag{2.85}$$

and ϕ_c is given by

$$\phi_c = \frac{\lambda}{\lambda^2 + \lambda - \lambda} = \frac{1}{\lambda} = \frac{1}{\langle k \rangle}. \tag{2.86}$$

Fig. 2.15 depicts the onset of the percolation transition for Poisson random graphs with different average degrees, $\langle k \rangle$.

We conclude this section with a few remarks regarding bond percolation. In the case of bond percolation, the occupation probability ϕ corresponds to the fractions of edges rather than nodes that have been removed from the network. By repeating

the same procedure as for the site percolation case, one can show that the average probability for a node not to be connected to the giant component and the percolation threshold are still given by Eq. (2.79) and Eq. (2.83) respectively. However, the relation characterising the fraction of nodes belonging to the giant connected component is slightly different and is given by

$$S = 1 - f_0(u). \tag{2.87}$$

Eq. (2.87) differs from Eq. (2.76) by a factor of ϕ. This is because, in a bond percolation process, nodes are not being removed from the network.

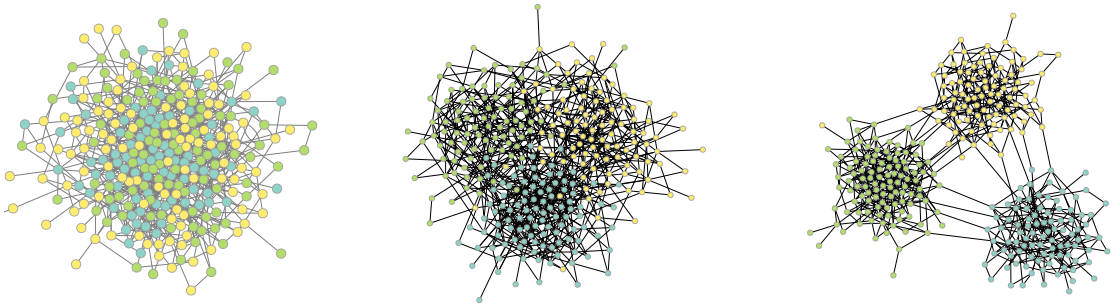
2.3.2 Assortative mixing and modularity

A commonly observed characteristic across many complex networks is the tendency for a network's elements to form ties according to some notion of similarity amongst the components. A notable example of this is the formation of social interactions in a social network. People belonging to the same age group, for example, will tend to interact more amongst each other than with people of other age groups. Similarly, people are generally more prone to engage with people who share a common belief or political view rather than an opposing one. More generally, humans often tend to form bonds based on a wide range of characteristics such as race, age, political views, education level and many more. This kind of connectivity pattern, where nodes perceived as similar in some sense tend to be more tightly knit together, is known as *assortative mixing* or *homophily*. Homophily is in no way limited to human interactions, and instances of assortative mixing have also been observed in citation networks (where papers tend to preferentially cite papers belonging to the same research area) and food webs [42], to name a few.

Assortative mixing patterns can play an important role in the network function. For example, they can lead to the emergence of *communities* in networks [92] (i.e. groups of nodes that are more densely connected amongst each other than with the rest of the network), which can have profound implications on the dynamics of the network. A disease or rumour is more likely to spread amongst a group of nodes if these are tightly knit together. Groups or communities have also been associated with occupying functional roles in a network, thus providing a source of specialisation, where separated tasks performed by different communities can then be integrated at a system-wide level. Furthermore, this type of mixing pattern has proven to be beneficial for the stability [93] and adaptability [46] in biological networks.

Given the prevalence and importance that homophily can have on networked systems, one would like to develop tools that allow us to capture and quantify assortative mixing patterns in complex networks. In what follows, we introduce one of the most popular instruments for this purpose: the *modularity function*.

Suppose we are given a network whose nodes have been classified into distinct groups according to some labels $b_i \in [1, B]$, where B is the total number of groups. The main idea behind modularity is then to compare the number of edges that fall between nodes belonging to the same group with what would be expected by chance given a pre-specified null model. The choice of the null model is, in principle, arbitrary. However, not all choices are equally good. A $G(N, p)$ ensemble, for



(a) (b) (c)

Figure 2.16: Examples of networks displaying different degrees of modularity. (a) A random graph, with no cluster structure present. (b) A network with a moderate amount of modularity. An emerging cluster structure is visible. (c) A network with a high value of modularity and clear cluster structure.

example, would generate networks with Poisson degree distributions, which are very different from the degree distributions commonly observed in real-world networks, making our comparison questionable. A more reasonable choice for the null model is to use the configuration model.

Given the labelling of the nodes, the number of edges within a group can then be calculated as

$$\frac{1}{2}\sum_{ij} A_{ij}\delta_{b_i b_j}, \tag{2.88}$$

where δ_{xy} is the Kronecker delta.

To evaluate the number of intra-group edges expected by chance, consider a node i of degree k_i. As seen previously, the probability that one of its stubs connects to some node j will be given by $k_j/2E$ (as exactly k_j stubs out of the total $2E$ connect to node j). Seeing as node i has k_i stubs attached to it, the expected number of edges between nodes i and j will be given by $k_i k_j/2E$, and the expected number of edges between nodes belonging to the same community is

$$\frac{1}{2}\sum_{ij} \frac{k_i k_j}{2E}\delta_{b_i b_j}. \tag{2.89}$$

The difference between the number of edges within a group and what we expect to observe by chance is then

$$\frac{1}{2}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2E}\right)\delta_{b_i b_j}. \tag{2.90}$$

If instead of considering the exact number of edges, we consider the fraction of edges, we obtain the modularity function:

$$Q = \frac{1}{2E}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2E}\right)\delta_{b_i b_j}. \tag{2.91}$$

High values of Q indicate an abundance of intra-group edges with respect to what is expected by chance and correspond to network topologies that are highly clustered, see Fig. 2.16. On the other hand, negative modularity values correspond to a

scarcity of intra-group edges with respect to what is expected by chance and indicate disassortative topologies. We also note that Q is always less than 1.

Modularity can then be interpreted as an elegant measure of the degree of homophily in a network.

Chapter 3

Null models of optimised modular networks

This chapter describes the framework, first introduced in [40], which we adopt to develop null models of optimised modular networks.

3.1 The model

As mentioned in Chapter 1, we wish to construct null models of optimised networks, corresponding to network ensembles that possess some pre-specified fitness level but which are otherwise maximally random. We approach this problem via generative models. This means that instead of describing individual networks, we are interested in formulating network ensembles, such that the probability of observing a given network is associated with its particular fitness value, given a predefined fitness criterion.
Attempting to consider the ensemble of all possible graphs which satisfy a set of given criteria is typically a gargantuan task, as the degrees of freedom needed to describe the network is generally much larger than the number of criteria we impose. In order to reduce the number of free parameters to a tractable amount, we constrain ourselves to networks that exhibit modular structure, i.e. the nodes are divided into groups, which share a similar role in the network structure. More specifically, we consider networks that are generated from the previously introduced stochastic block model [70, 72, 75].

In the stochastic block model, N nodes are divided into B groups, such that each node i is given a group membership label $b_i \in \{1, \ldots, B\}$, and an edge between a node in group r and another in group s exists with probability p_{rs}. This yields a network ensemble where a network $\boldsymbol{A}$ occurs with probability

$$P(\boldsymbol{A}|\boldsymbol{b}, \boldsymbol{p}) = \prod_{i<j} p_{b_i,b_j}^{A_{ij}} (1 - p_{b_i,b_j})^{1-A_{ij}}, \tag{3.1}$$

where $A_{ij} = 1$ if an edge exists between nodes (i, j), or $A_{ij} = 0$ otherwise. Although this is just one of a large set of possible network ensembles, as we saw in Section 2.2, the SBM is capable of capturing arbitrary mixing patterns between groups by appropriate choices of the matrix $\boldsymbol{p}$, and if the number of groups is increased, it can account for arbitrarily elaborate network structures [94]. Although this does not

give us the full breadth of all possible network structures — in particular, we cannot describe the details of the network structure at a local level, e.g. by stipulating desired propensities of observing triangles or other small subgraphs — as we will see, this is a sufficiently flexible framework to express the kind of null models we have in mind.

For a given arbitrary fitness function $R(\boldsymbol{A})$, which maps a network to a scalar fitness value, the average fitness over the SBM ensemble is then given by

$$R(\boldsymbol{b}, \boldsymbol{p}) = \sum_{\boldsymbol{A}} R(\boldsymbol{A}) P(\boldsymbol{A}|\boldsymbol{b}, \boldsymbol{p}). \tag{3.2}$$

Based on such a function, we could, in principle, proceed by finding the SBM parameters $\boldsymbol{b}$ and $\boldsymbol{p}$ such that the mean fitness $R(\boldsymbol{b}, \boldsymbol{p})$ is maximised, and in this way, uncover how a fitness criterion favours specific patterns of network structures. However, this optimisation problem is ill-defined in the general case, as many parameter choices yield the same optimal fitness value. Therefore, we formulate our question differently. Instead of optimising the mean fitness $R(\boldsymbol{b}, \boldsymbol{p})$, we impose its value as a pre-determined parameter, and we select the SBM parameters that yield the most random network ensemble and therefore is the most agnostic about the unimportant properties of the network structure. More formally, this means we employ the maximum entropy principle [95] introduced in Section 2.2, such that for any imposed fitness value $R(\boldsymbol{b}, \boldsymbol{p}) = R^*$, the choice of the model parameters $\boldsymbol{b}$ and $\boldsymbol{p}$ from all those that fulfil this constraint is the one that maximises the ensemble entropy [90],

$$\Sigma(\boldsymbol{b}, \boldsymbol{p}) = -\sum_{\boldsymbol{A}} P(\boldsymbol{A}|\boldsymbol{b}, \boldsymbol{p}) \ln P(\boldsymbol{A}|\boldsymbol{b}, \boldsymbol{p}). \tag{3.3}$$

In this way, if we specify a set of fitness functions $\{R_i(\boldsymbol{b}, \boldsymbol{p})\}$ and their imposed set of values $\{R_i^*\}$, we are interested in the following constrained optimisation problem

$$\hat{\boldsymbol{b}}, \hat{\boldsymbol{p}} = \underset{\boldsymbol{b},\boldsymbol{p}}{\operatorname{argmax}}\, \Sigma(\boldsymbol{b}, \boldsymbol{p}), \quad \text{subject to } R_i(\boldsymbol{b}, \boldsymbol{p}) = R_i^* \ \ \forall i. \tag{3.4}$$

The SBM parameters obtained in this way can be interpreted as *null models* of networks, which contain only the most essential ingredients to achieve the prespecified values of fitness, and otherwise are maximally random. The imposed fitness values themselves can be increased arbitrarily to achieve any level of optimised structures, as we will show.

We can compute the entropy of the SBM ensemble by substituting Eq. (3.1) into Eq. (3.3), which yields [65]

$$\Sigma(\boldsymbol{b}, \boldsymbol{p}) = \sum_{r<s} n_r n_s H_{\mathrm{b}}(p_{rs}) + \sum_r \frac{n_r(n_r-1)}{2} H_{\mathrm{b}}(p_{rr}), \tag{3.5}$$

where $n_r = \sum_i \delta_{b_i,r}$ is the number of nodes in group r, and $H_{\mathrm{b}}(x) = -x\ln x - (1-x)\ln(1-x)$ is the binary entropy function. This can be further simplified if we take into account that most networks in the real world are sparse with $p_{rs} = O(1/N)$, so that using $H_{\mathrm{b}}(x) = -x\ln x + x + O(x^2)$, and taking the limit $N \gg 1$ we obtain

$$\Sigma(\boldsymbol{b}, \boldsymbol{p}) = -\frac{1}{2}\sum_{rs} n_r n_s \left(p_{rs}\ln p_{rs} - p_{rs}\right). \tag{3.6}$$

For some choices of fitness functions, arbitrarily high fitness values can be obtained simply by increasing the network density. In order to differentiate between the effect of increased density and favored mixing patterns, we will take the average degree $\langle k \rangle = \sum_{rs} n_r n_s p_{rs}/N$ as an external parameter not subject to optimization. With this in mind, it will be useful for our calculations to use the following re-parametrisation over intensive variables,

$$\omega_r = \frac{n_r}{N}, \qquad m_{rs} = \frac{n_r n_s p_{rs}}{N\langle k \rangle}. \tag{3.7}$$

Note that the above implies the normalisation $\sum_r \omega_r = 1$ and $\sum_{rs} m_{rs} = 1$. Given this choice, the ensemble entropy can be written as

$$\Sigma(\boldsymbol{\omega}, \boldsymbol{m}) = -\frac{\langle k \rangle N}{2} \sum_{rs} m_{rs} \ln \frac{m_{rs}}{\omega_r \omega_s} + \frac{\langle k \rangle N}{2} \tag{3.8}$$

Note that we no longer reference the actual partition $\boldsymbol{b}$ itself, but rather the fraction of nodes ω_r that belong to a given group r, since these are the relevant macroscopic quantities as $N \gg 1$.

Based on the above model parametrisation, we can perform the constrained optimisation of Eq. (3.4) by employing the method of Lagrange multipliers, which involves finding the saddle points of the Lagrangian function

$$\Lambda(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{\beta}) = \Sigma(\boldsymbol{\omega}, \boldsymbol{m}) + \sum_i \beta_i \left[R_i(\boldsymbol{\omega}, \boldsymbol{m}) - R_i^* \right], \tag{3.9}$$

where β_i are the Lagrange multipliers that enforce each constraint. This means we need to find $\boldsymbol{\omega}$, $\boldsymbol{m}$, and $\boldsymbol{\beta}$ such that the gradient of Λ is zero, i.e. $\partial\Lambda(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{\beta})/\partial\omega_r = \partial\Lambda(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{\beta})/\partial m_{rs} = \partial\Lambda(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{\beta})/\partial\beta_i = 0$. Note that the last derivative yields simply the equation $R_i(\boldsymbol{\omega}, \boldsymbol{m}) = R_i^*$, which means that the problem of fixing R_i^* and finding $\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{\beta}$ is equivalent to first taking $\boldsymbol{\beta}$ as fixed parameters and minimising the function

$$\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m}) = -\sum_i \beta_i R_i(\boldsymbol{\omega}, \boldsymbol{m}) - \Sigma(\boldsymbol{\omega}, \boldsymbol{m}), \tag{3.10}$$

with respect to $\boldsymbol{\omega}$ and $\boldsymbol{m}$ alone and then varying $\boldsymbol{\beta}$ until we obtain $R_i(\boldsymbol{\omega}, \boldsymbol{m}) = R_i^*$.

The above formulation puts us in a standard setting in equilibrium statistical physics, as the function $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m})$ can be interpreted as the *free energy* of the network ensemble where the sum $-\sum_i \beta_i R_i(\boldsymbol{\omega}, \boldsymbol{m})$ plays the role of the mean energy. Following this analogy, the values of β_i play the role of inverse temperatures, or perhaps more appropriately to our setting, *selective pressures*, which if increased cause the corresponding energy functions to decrease (and thus the fitness values to increase), and thus settling on a particular balance between energy and entropy.

To summarise, our protocol to generate null network models is as follows:

1. We establish a set of fitness functions $\{R_i(\boldsymbol{\omega}, \boldsymbol{m})\}$.

2. Given a choice of selective pressures $\{\beta_i\}$ we find the parameters $\boldsymbol{\omega}, \boldsymbol{m}$ which minimize the free energy $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m})$ of Eq. (3.10).

3. We vary the values $\{\beta_i\}$ to investigate the trade-off between competing fitness functions as well as entropy.

The constrained optimisation of step 2 is the most central part of our approach. Although it is straightforward to compute the gradient of the entropy $\Sigma(\boldsymbol{\omega}, \boldsymbol{m})$ analytically, in the general case this will not be possible for arbitrary fitness functions $R_i(\boldsymbol{\omega}, \boldsymbol{m})$, and even when it is, setting the gradient of $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m})$ to zero usually yields an implicit system of nonlinear equations that cannot be solved in closed form. Therefore, in the following, we will proceed by performing the minimisation numerically, via the L-BFGS-B conjugate gradient descent algorithm [96], using automatic differentiation [97] whenever the gradient cannot be obtained in closed form. As a final implementation note, the used algorithms require us to convert step 2 into an unbounded optimisation problem, which we do via a simple exchange of variables given by

$$\omega_r = \frac{\mathrm{e}^{\mu_r}}{\sum_s \mathrm{e}^{\mu_s}}, \qquad m_{rs} = \frac{\mathrm{e}^{\nu_{rs}}}{\sum_{tu} \mathrm{e}^{\nu_{tu}}}, \tag{3.11}$$

with $\mu_r \in [-\infty, \infty]$ and $\nu_{rs} \in [-\infty, \infty]$, which keep both ω_r an m_{rs} bounded in the range $[0, 1]$, and enforces normalisation.

3.2 Discussion

In this chapter, we have introduced the general framework we will employ throughout this thesis to generate null models of optimised networks. Although we have focussed on the standard stochastic block model, as it will be the principal focus of this thesis, it is important to stress that the presented framework is very general and, in principle, any network ensemble or additional constraint can be incorporated.

In what follows, we will apply this framework to analyse the topologies most likely to emerge in networks driven to optimise modularity and robustness against random failures. In the following chapter, we will consider the two criteria separately; then, in Chapter 5, we will study networks subject to both selective pressures simultaneously.

Chapter 4

Optimisation of criteria in isolation

This chapter applies our methodology to the fitness criteria introduced previously —namely, modularity and the robustness of a network to random failures of its components. These characteristics are properties of great interest in real-world networks across many different domains. For example, network resilience plays a key role in understanding the possible collapse of ecosystems in response to external perturbations [98] or in developing strategies to stem the spread of disease [49]. At the same time, modularity is often a desirable design feature, favouring flexibility and re-usability of components in technological and industrial settings and is also a commonly observed property in many naturally arising systems. As such, analysing what topological features are more likely to emerge when networked systems are driven to optimise for these characteristics is a problem of great interest.

In what follows, we consider the case in which networks are subject to a single selective pressure at a time, as this enables us to obtain a blueprint of the most fundamental structural characteristics selected by each fitness criterion. By analysing these network structures, we can then determine the topological features central in driving an increase (or decrease) of the chosen criterion.

Using the modularity function and the fraction of nodes in the giant connected component as our model's fitness functions and minimising the corresponding free energy, we uncover a series of optimal network topologies and structural transitions that emerge in response to the varying selective pressures.

4.1 Modularity

As mentioned previously, some networks tend to cluster into groups of nodes that are more connected amongst themselves than the rest of the network. This feature is of great importance in a wide array of settings. For example, in technological systems, modules are often associated with tasks that can be executed in parallel. One such example is parallel processing, which is of fundamental importance in scientific computing applications. However, for these tasks to be executed properly, it is essential to minimise the communication between modules so as to reduce the overhead introduced by the communication time [47]. Communication networks are another example where the design of highly modular structures is common practice, as this allows for a clearer understanding of a network's functional, logical, and

physical components, thus facilitating the design of the overall network and optimising its performance [99]. Furthermore, modularity (and its optimisation) is not limited to technological or artificial networks. In the context of biological networks, for example, genes and proteins involved in similar biological functions have been shown to cluster in the same network neighbourhood [100], which has recently led to applications in network medicine [101]. Moreover, modularity has shown to be beneficial for the adaptability [46, 54] and stability [93] of biological systems, leading some authors to argue that natural selection itself drives the origin of modularity either by direct or indirect effects [46, 102]. In view of this, it is highly interesting to analyse what large-scale structures are more likely to emerge in networks driven to optimise their modular structure, as these could prove helpful in the design of optimal networks, the development of efficient algorithms, and gaining insights into the processes driving the emergence of modular structures.

The most typical way to quantify this kind of assortativity pattern is via the modularity function introduced in Section 4.1. We then begin this section by expressing the modularity function in the context of stochastic block models, which we shall later employ as a fitness criterion in our formalism. By minimising the corresponding free energy, we are able to identify a smooth transition from a completely random graph topology to a highly modular structure composed of symmetric clusters whose nodes prevalently connect to other nodes within the same cluster.

4.1.1 Modularity in the stochastic block model formalism

Let us recall the definition of modularity,

$$Q = \frac{1}{2E} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2E} \right) \delta_{b_i b_j}. \tag{4.1}$$

Since the only contributions to the sum come from vertices belonging to the same cluster, we can group these terms and rewrite the summation as a sum over the groups.

$$Q = \sum_{r=1}^{B} \left[\frac{e_{rr}}{2E} - \left(\frac{e_r}{2E} \right)^2 \right], \tag{4.2}$$

where B is the total number of partitions, e_{rr} corresponds to twice the number of edges in group r, and e_r is the total number of half-edges incident on group r (which coincides with the sum of the degrees of the nodes in r). Considering the parametrisation introduced in Eq. (3.7), the modularity function can then be expressed as

$$Q = \sum_{r=1}^{B} m_{rr} - m_r^2, \tag{4.3}$$

where $m_r = \sum_s m_{rs}$ is the total fraction of edges incident on group r. Note that for completely assortative SBMs with $m_{rs} = \delta_{rs}/B$, we have $Q(\boldsymbol{m}) = 1 - 1/B$, so we achieve maximal modularity $Q(\boldsymbol{m}) \to 1$ for an infinite number of perfectly isolated groups.

Before continuing with our maximisation of Q, it is important to stress that here, we are interested in maximising the expected modularity conditioned on a known

partition, i.e.,

$$\max_{\boldsymbol{b},\boldsymbol{p}} \sum_{\boldsymbol{A}} Q(\boldsymbol{A},\boldsymbol{b}) P(\boldsymbol{A}|\boldsymbol{p},\boldsymbol{b}) \tag{4.4}$$

This is different from maximising the modularity conditioned on a specific network with respect to an unknown partition, i.e.,

$$\max_{\boldsymbol{b},\boldsymbol{p}} \sum_{\boldsymbol{A}} P(\boldsymbol{A}|\boldsymbol{p},\boldsymbol{b}) \max_{\boldsymbol{b}'} Q(\boldsymbol{A},\boldsymbol{b}'), \tag{4.5}$$

as is usually the case in the context of community detection [56]. This is because here, we consider the node types responsible for the value of Q to be an intrinsic property of the system, based on which the modularity is being optimised. For example, these node types could correspond to the assembly of distinct items in a production line or different types of metabolites in a metabolic network, and the system's overall fitness would be improved if there were fewer interdependencies between these subsets of nodes. This means that alternative partitions of the generated networks with a higher modularity value but unrelated to these intrinsic types would be irrelevant for the system's fitness.

We can include the modularity as a fitness criterion into our framework by making $R(\boldsymbol{\omega},\boldsymbol{m}) = Q(\boldsymbol{m})$ and coupling with its selective pressure β_Q, and proceeding to minimize the free energy

$$\mathcal{F}(\boldsymbol{\omega},\boldsymbol{m}) = -\beta_Q Q(\boldsymbol{m}) - \Sigma(\boldsymbol{\omega},\boldsymbol{m}). \tag{4.6}$$

4.1.2 Results

The $B = 2$ case

We begin our analysis by considering the simple case in which only two groups are allowed ($B = 2$).

The minimisation of the free energy function was carried out via the `SciPy` implementation of the L-BFGS-B conjugate gradient descent method [96, 103, 104], where the gradients have been evaluated via automatic differentiation[1]. As optimisers rarely work "out of the box", both a calibration of the L-BFGS-B parameters and the introduction of a rescaled free energy were necessary to ensure the consistent and correct identification of the free energy minima.

The L-BFGS-B algorithm pertains to a general class of optimisation methods known as *quasi-Newton methods*. These approaches aim to circumvent the generally expensive computation and inversion of the Hessian matrix used in the classical Newton's method by directly developing approximations of the Hessian inverse. In broad terms, the L-BFGS-B algorithm works by starting from an initial estimate of the optimum and then iteratively improving this estimate using the approximation of the Hessian inverse to guide the trajectory through the variable space. The norm of the projected gradient is typically used as a termination criterion, and the algorithm stops if the norm of the projected gradient is sufficiently small. However, another crucial stopping criterion often used in determining the algorithm's termination

[1] Although the gradient for the modularity function (and thus the free energy) can be computed analytically, this will generally not be the case when considering other fitness functions. In the sake of generality, we proceed to use automatic differentiation even when considering modularity.

is the relative difference of function values at successive iteration steps, typically controlled by an `ftol` parameter such that the algorithm terminates whenever the following condition is met.

$$\frac{f^t - f^{t+1}}{\max\{f^t, f^{t+1}, 1\}} < \texttt{ftol}, \tag{4.7}$$

Eq. (4.7) prevents the algorithm from drastically slowing down when the dynamics remains trapped in valleys with small gradients (but not small enough to guarantee termination). In our case, running the optimisation procedure with `ftol` values that are too large can lead to the optimisation procedure being terminated in states which still exhibit large gradient values, see Fig. 4.1. These high gradient states

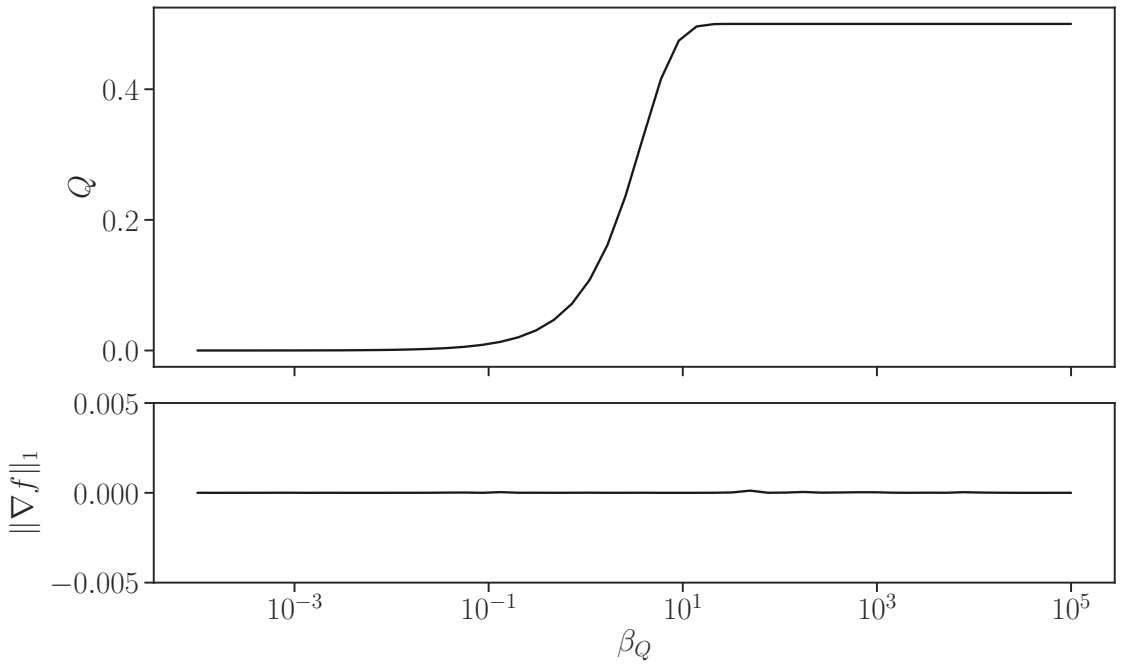


Figure 4.1: The modularity, $Q(\boldsymbol{m})$, (top) and the L_1 norm of the corresponding gradient (bottom) as a function of the selective pressure β_Q. The figure has been obtained by randomly initialising a network with $B = 2$ groups and average degree $\langle k \rangle = 5$ at each value of the selective pressure β_Q, and subsequently minimising the free energy via the L-BFGS-B conjugate gradient algorithm. As can be seen, results obtained this way can result in the optimiser finding high gradient states which do not correspond to the free energy minima.

occur predominantly for high values of the selective pressure, where the modularity function plateaus around its theoretical maximum, and even significant parameter changes lead to minor improvements of the objective function. However, simply reducing the value of `ftol` is still not sufficient to ensure correct minimisation, as large β_Q values can introduce considerable numerical instabilities. To circumvent this problem, we also introduce a rescaled free energy defined as

$$f \to \frac{f}{\max\{\beta_Q, 1\}}, \tag{4.8}$$

and use it as our objective function to be fed to the L-BFGS-B optimiser.

To determine an optimal choice of the `ftol` parameter for the rescaled free energy, we can repeat the minimisation procedure leading to Fig. 4.1 for different choices of the `ftol` parameters and keep track of the maximum values attained by the L_1 norm of the gradient. Fig. 4.2 displays the behaviour of this maximum as a function of `ftol`. A simple elbow method suggests that a value of `ftol` of 1×10^{-16} is sufficient to guarantee a good convergence of the L-BFGS-B algorithm. With

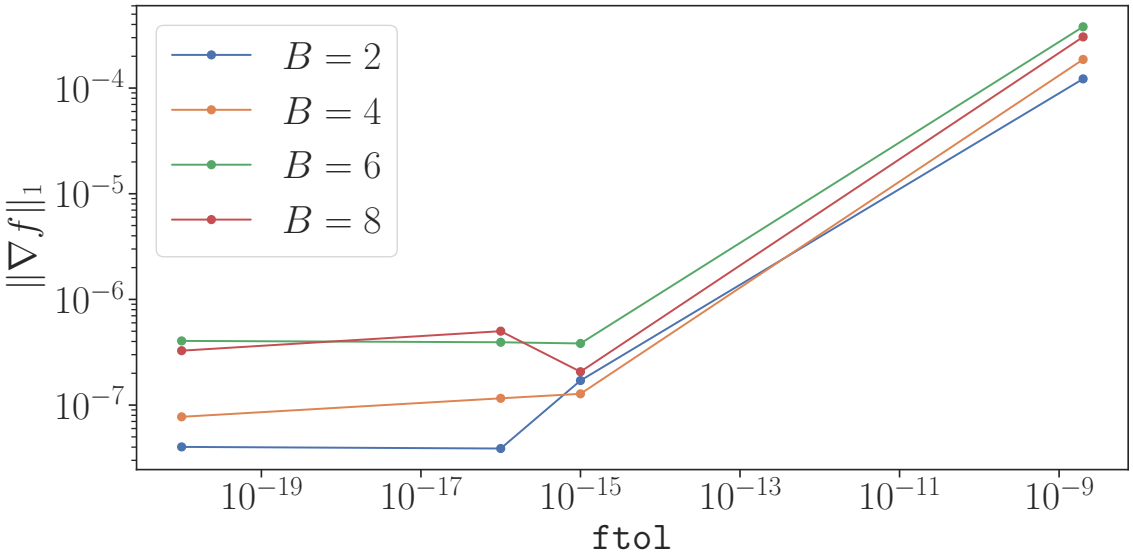


Figure 4.2: Maximum of the L_1 norm of the gradient of the free energy as a function of the `ftol` parameter discussed in the text. The norm decreases as the `ftol` parameter is lowered until, at a values between 10^{-15} and 10^{-16} (roughly coinciding with the machine precision limit), the curve shows a sharp elbow.

these precautions, the optimisation procedure correctly identifies the free energy minima for the B = 2 case, see Fig. 4.3. As the selective pressure increases, the network structure transitions from a random graph topology at low beta values (corresponding to zero modularity) to ever-increasing modular structures until, for large enough selective pressures, the modularity attains its maximum theoretical value and the network topology is described by two symmetric and disconnected modular structures.

Increasing the number of groups

As we have seen, the maximum value that the modularity function can attain is directly linked to the number of groups we use to model the network. Increasing the number of groups the network has access to allows for the possibility that a larger number of them will be populated as the selective pressure is increased. Populating a large number of groups has a high entropic cost. Therefore, for low values of the selective pressure, one would expect to recover the two-block structures observed in Fig. 4.3. As the selective pressure increases, the drive to maximise modularity should lead to the network gradually occupying a larger number of groups. Fig. 4.4 appears to confirm this hypothesis. However, the minimisation procedure can lead to the network getting stuck in states corresponding to local minima of the free energy around values of the selective pressure where the optimal number of groups

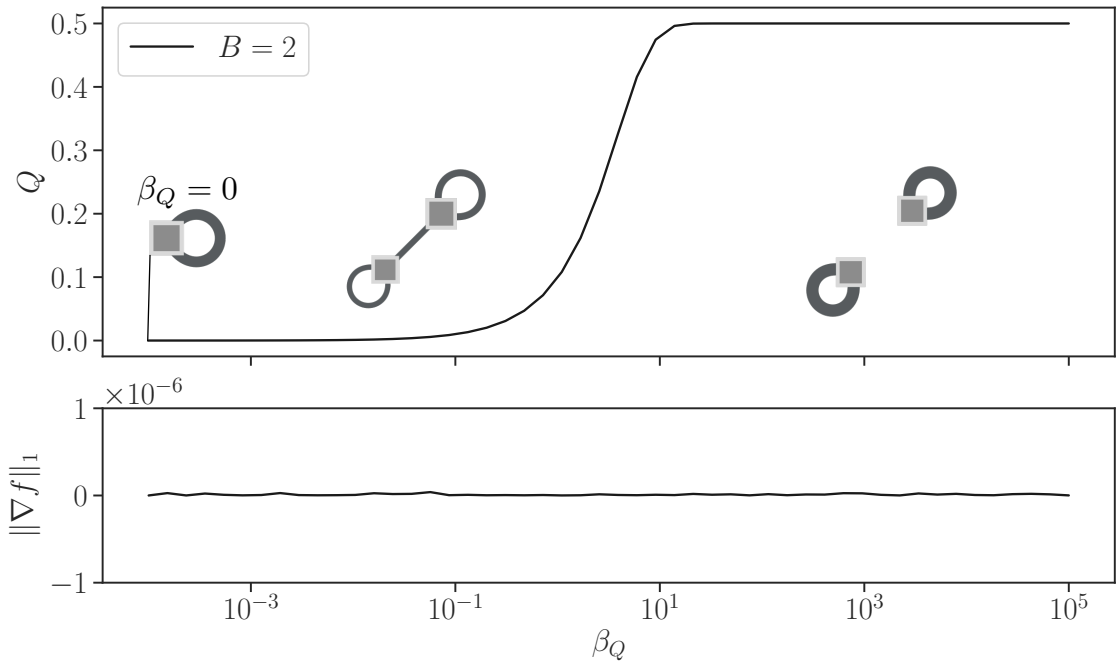


Figure 4.3: The modularity, $Q(\boldsymbol{m})$, (top) and the L_1 norm of the corresponding gradient (bottom) as a function of the selective pressure β_Q for a network with $B = 2$ groups. The optimised network structures are shown schematically in the insets, where each square corresponds to one of the groups of our model, with size proportional to ω_r and edge thickness between them proportional to m_{rs}. As the selective pressure increases, the networks smoothly splits into two identical cluster structures. We note that differently from Fig. 4.1, the L_1 norm of the gradient is always $\mathcal{O}(10^{-8})$.

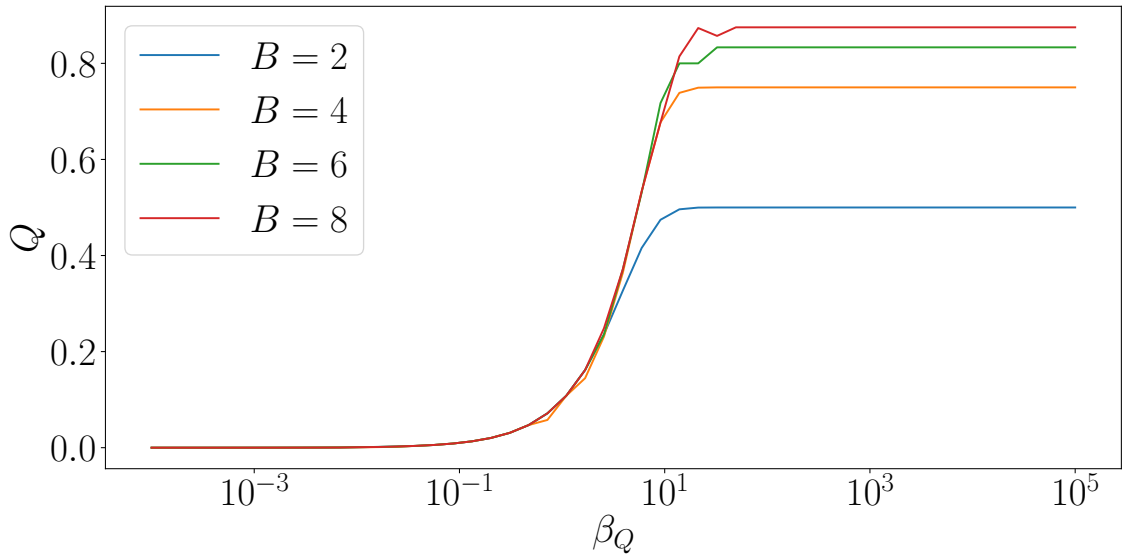


Figure 4.4: Modularity Q as a function of the selective pressure β_Q, for different choices of the allowed number of groups B. The kinks observed along some of the curves correspond to the local minima states of the system discussed in the text.

required to describe the network changes from B to $B+1$. These local minima states appear as kinks in our modularity curves, as the occupation of a lower number of groups leads to a drop in modularity, see Fig. 4.5 This problem can be readily

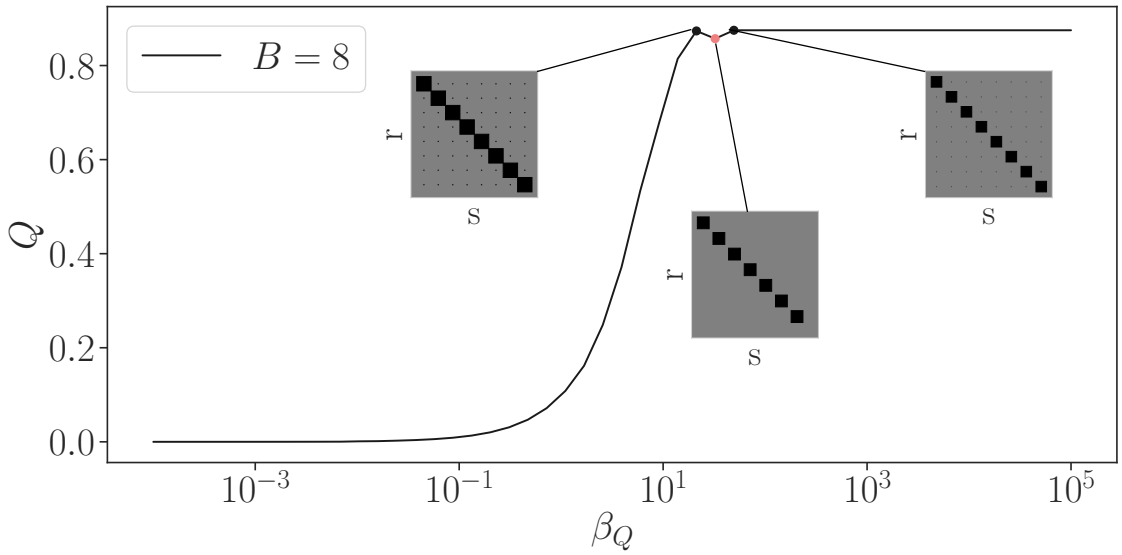
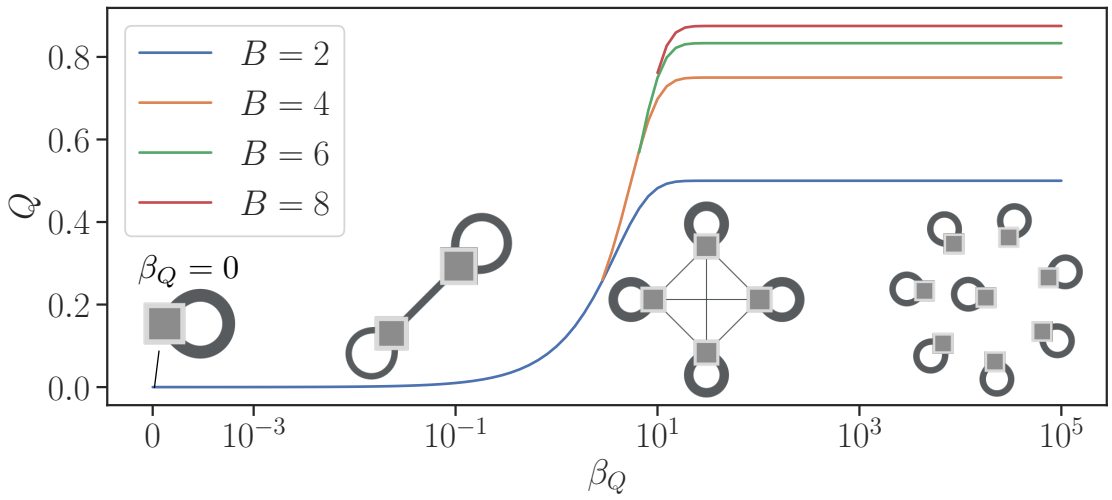
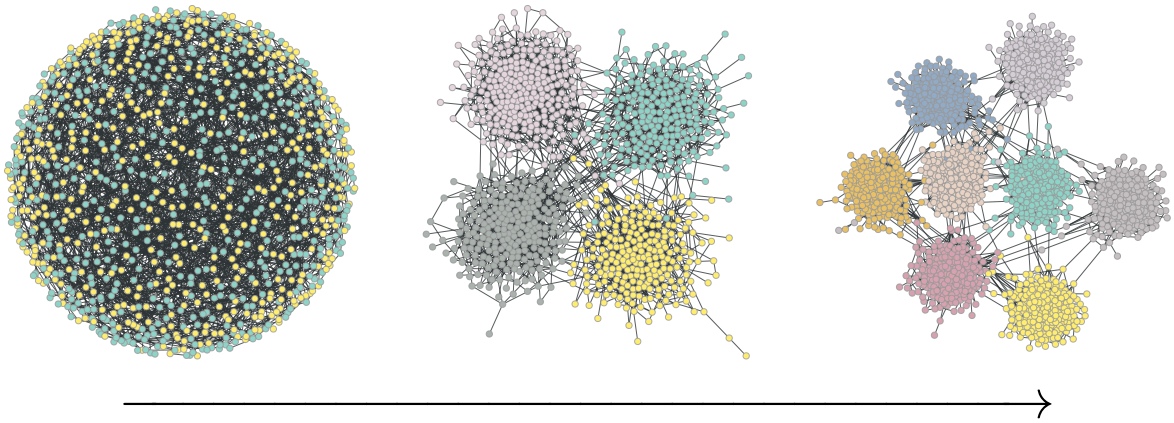


Figure 4.5: Modularity Q as a function of the selective pressure β_Q for a network with $B = 8$ groups and average degree $\langle k \rangle = 5$. The insets show the affinity matrices m_{rs} for three specific states. As can be seen, the intermediate state (marked in red) only presents seven occupied groups and thus has a lower modularity value with respect to the neighbouring states which have all eight available groups occupied.

resolved by changing the procedure used to minimise the free energy. Specifically, we resort to an *adiabatic* pass of our algorithm. We begin by minimising the free energy at some low value of the selective pressure, where we know our algorithm to consistently return the correct result. This optimised state is then used as the initial condition from which to perform the minimisation at the subsequent value of beta. The newly minimised state is then used as the next starting point and the procedure is repeated along the entire range of β_Q values. Proceeding this way, we can correctly follow the states to high values of the selective pressure. Fig. 4.6 displays the behaviour of Q as a function of the selective pressures for different choices of the allowed number of groups B. As the selective pressure increases, the network splits smoothly and progressively into fully symmetric groups of equal size with a larger number of connections inside each group. For low values of β_Q, the results obtained with different numbers of groups coincide, and then they diverge for higher values. This is because, as expected, the actual number of groups populated starts off small and progressively increases for larger values of β_Q.



(a)



(b)

Figure 4.6: (a) Modularity Q as a function of the selective pressure β_Q for different choices of the allowed number of groups B. (b) Network samples from the ensemble with average degree $\langle k \rangle = 5$ at different values of the β_Q. As the selective pressure increases, the network splits into a growing number of groups of increasing modularity.

4.2 Robustness against random failures

In this section, we consider robustness against random failure as a fitness criterion. As mentioned in Section 2.3.1, robustness is a characteristic of fundamental importance in networked systems. Indeed, if a networked system is to function in the first place, then being connected is bound to be one of its fundamental characteristics. As we increasingly live and rely on an interconnected society, when networks fail, this can have catastrophic consequences. The disruption of the airline network caused by the 2010 eruptions of the Eyjafjallajökull volcano in Iceland [105] had an estimated financial cost of approximately £1.1 billion according to The International Air Transport Association (IATA). Similarly, power outages like the one in Italy in 2003 [106] can significantly impact both the economy and the well-being of individuals. More generally, failures of real-world networks can have widespread consequences on human health [107], the economy [108], and the environment [109], amongst others. Understanding how the network topology affects the robustness of networked systems (and how to optimise their robustness) is then a problem of utmost importance, with possible applications ranging from designing robust infrastructures to controlling the spread of epidemics or avoiding mass extinctions in ecological networks.

As introduced in Section 2.3.1, we model the robustness against random failures as a percolation process. Specifically, we examine a bond percolation process and consider the case in which a fraction $1 - \phi$ of edges are randomly removed from the network and measure the fraction S of nodes that remain connected, forming a giant connected component [110]. In many respects, this is similar to the work carried out by Peixoto and Bornholdt in [40], which considered site percolation. There are, however, a few key differences. In [40], the authors consider a slightly different ensemble in which the degree distrbutions of the groups are fixed to be modified Poisson distributions such that nodes of degree zero are forbidden. The rationale behind this choice is that since nodes of degree zero are never part of the giant connected component, their contribution to the robustness properties of the network can be neglected. However, this could potentially lead to artificial constraints in the emerging topologies, as all nodes are forced to have at least one connection. Moreover, this might also not always be the desired behaviour. In the case of an epidemic, for example, disconnecting the network might be the desired outcome. In what follows, we consider the standard stochastic block model and allow for nodes of degree zero.

The fitness criterion considered by the authors in [40] is also slightly different from the one used here. Specifically, in [40], the authors consider a measure of robustness, previously introduced in [39], which corresponds to the fraction of nodes in the giant component, averaged over all possible ϕ values.

$$R = 2 \int_0^1 S(\phi) d\phi, \tag{4.9}$$

where the factor 2 ensures that $R \in [0, 1]$. With respect to other common measures of robustness, such as the critical value ϕ_c at which the network collapses, both S and R have the distinct advantage of also considering the size of the connected component, yielding information not only on whether the network is functioning

but also what fraction of it is functioning. However, The two measures are slightly different, as R collapses all the information from ϕ into a single value.

While R provides a comprehensive measure of robustness which considers all possible degrees of failure, it might lack specificity for more distinct scenarios. For example, it is a well-known fact that about 3% of routers on the Internet are non-functional at any given time [49]. It might then be preferable to study the resilience properties of the network for a small range of ϕ values rather than over all possible types of failure. In general, it would appear improbable for a sudden massive random failure of an entire network to take place, and these critical system failures are typically the end product of dynamic processes such as cascading failures. In this respect, R appears to be more suited for designing optimal networks against malicious attacks in which the extent of the damage is unknown a priori. On the other hand, S appears to be better suited to analyse the resilience properties of networks at specific levels of degradation. Moreover, utilising S allows us to explicitly model the effects that the dilution probability has on the robustness properties of the network rather than considering a summary statistic (for example, topologies which are more robust on average might not be so for specific values of ϕ). This allows us to have a more nuanced view of how robustness develops at different levels of damage.

We then use the fraction of nodes in the giant component as the fitness function of our model and analyse the topologies that are most likely to emerge when the network is subject to a varying range of selective pressures. We consider these effects for a wide range of occupation probabilities ϕ, obtaining a phase diagram of the robustness against random failures in the (β_S, ϕ) plane (where β_S is the selective pressure driving to optimise for robustness), which allows us to gain a more comprehensive view of the combined effects that selective pressures and occupation probabilities have on a network's robustness. Our main result is the identification of three principal topologies describing the network structure whose emergence depends on the values of ϕ and β_S. Furthermore, we distinguish the region around the percolation threshold, ϕ_c, as the one where more significant gains in robustness can be made with respect to a random graph topology.

4.2.1 Bond percolation in the stochastic block model

Closely following [40], we begin by extending the formalism of bond percolation to stochastic block model ensembles.

Let u_r be the average probability that a node in group r is not connected to the giant component via a specific one of its edges, and let ϕ_k^r be the occupation probability of an edge in group r, i.e. the fraction of edges emanating from a node of degree k in group r which have not been removed from the network. In what follows, we shall consider the case in which all edges are removed uniformly at random irrespective of the group assignment or the degrees of the nodes they connect. We therefore have

$$\phi_k^r = \phi, \quad \forall\ k,\ r. \tag{4.10}$$

Consider now a node in group r, and let us evaluate the probability that our selected node is not connected to the giant component via a specific one of its edges leading to some group s. This can occur in one of two ways.

1. The selected edge is unoccupied (i.e. it has been removed from the network), which happens with probability $1 - \phi$.

2. The edge is occupied (i.e. it has not been removed from the network), but the vertex at the other end is not connected to the giant component via one of its edges. If the vertex at the other end has k extra neighbours, then the probability of it not belonging to the giant component will be given by u_s^k.

The total probability that a node in group r does not connect to the giant component via a specific edge leading to some group s is then given by

$$u_r = 1 - \phi + \phi u_s^k. \tag{4.11}$$

Since the vertex in group s is reached by following an edge, k will be distributed according to the excess degree distribution of group s,

$$q_k^s = \frac{p_{k+1}^s (k+1)}{\kappa_s} \tag{4.12}$$

where p_k^s is the degree distribution of group s and

$$\kappa_r = \sum_\ell \frac{e_{r\ell}}{n_r} = \langle k \rangle \frac{m_r}{\omega_r} \tag{4.13}$$

is the average degree of the nodes in group s. Averaging over k, we have the following expression,

$$1 - \phi + \phi \sum_{k=0}^{\infty} q_k^s u_s^k = 1 - \phi + \phi f_1^s(u_s), \tag{4.14}$$

where $f_1^s(z)$ is the generating function for the excess degree distribution of group s. Eq. (4.14) is valid for one specific edge leading to some group s. Since the fraction of edges running from group r to group s is given by m_{rs}, the probability of randomly selecting one such edge will be given by m_{rs}/m_r. Summing over all possible values of s, we arrive at the following self-consistent equation for u_r:

$$u_r = 1 - \phi + \phi \sum_s \frac{m_{rs}}{m_r} f_1^s(u_s). \tag{4.15}$$

To evaluate the fraction of nodes S which belong to the giant connected component, we note that the probability that a node of degree k in group r does not belong to the giant component is simply u_r^k, and the average of such probability over group r will be

$$\sum_k p_k^r u_r^k = f_0^r(u_r). \tag{4.16}$$

Therefore, the fraction of nodes in group r which are connected to the giant component is given by

$$S_r = 1 - f_0^r(u_r) \tag{4.17}$$

and the total fraction of nodes in the giant component can be obtained by averaging over the group sizes,

$$S = \sum_r \omega_r S_r = 1 - \sum_r \omega_r f_0^r(u_r), \tag{4.18}$$

where $f_0^r(z) = \sum_k p_k^r z^k$ is the generating function for the degree distribution of group r.

The case of site percolation has been studied in [40], and it is straightforward to show that the equations governing site percolation are similar to the ones observed in Section 2.3.1.

$$u_r = 1 - \phi + \phi \sum_s \frac{m_{rs}}{m_r} f_1^s(u_s), \tag{4.19}$$

$$S = \phi \left[1 - \sum_r \omega_r f_0^r(u_r) \right], \tag{4.20}$$

where the self-consistent equation for u_r is the same for both bond and site percolation, but Eq. (4.20) is now multiplied by the occupation probability ϕ.

In what follows, we limit ourselves to considering the classical stochastic block model. Although generally not representative of most real-world networks, the standard SBM allows us to study the simplest scenario and gain insight into the most fundamental consequences that optimising for random failures can have on the network structure without the contribution of additional details.

For the standard SBM, the degree distributions p_k^r are Poisson distributions with mean $\kappa_r = \langle k \rangle m_r / \omega_r$. Substituting into the expression for $f_0^r(z)$, we have that

$$f_0^r(z) = \sum_{k=0}^{\infty} \frac{e^{-\kappa_r} \kappa_r^k}{k!} z^k = e^{-\kappa_r} \sum_{k=0}^{\infty} \frac{(\kappa_r z)^k}{k!} = e^{\kappa_r (z-1)}. \tag{4.21}$$

Recalling that $f_1^r(z)$ can be obtained in terms of $f_0^r(z)$ as $f_1^r(z) = f_0^{r'}(z)/f_0^{r'}(1)$, we have that, in our case

$$f_0^r(z) = f_1^r(z) = e^{\kappa_r (z-1)}. \tag{4.22}$$

Even though we possess exact analytical expressions for $f_0^r(z)$ and $f_1^r(z)$, Eq. (4.15) cannot be solved in closed form, and we must resort to solving it numerically. We do so by repeated iteration, starting at some $u_r < 1$ and iterating Eq. (4.15) until some desired convergence criterion is met. Once the u_r are known, the fraction S of nodes belonging to the giant component can be obtained via Eq. (4.18).

For any given SBM, the behaviour of S as a function of the fraction ϕ of edges that are not removed is that we have $S = 0$ for $\phi \in [0, \phi^*]$, where ϕ^* is a critical value, so that for $\phi > \phi^*$ we have a positive fraction of connected nodes $S > 0$ that increases continuously [40].

If we now consider the fitness function $R(\boldsymbol{\omega}, \boldsymbol{m}) = S(\boldsymbol{\omega}, \boldsymbol{m})$, our resulting free energy becomes

$$\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m}) = -\beta_S S(\boldsymbol{\omega}, \boldsymbol{m}) - \Sigma(\boldsymbol{\omega}, \boldsymbol{m}). \tag{4.23}$$

By minimising the above function, we find null network models that are robust against random failures, with the robustness increasing for higher β_S values.

4.2.2 Implementation details and parameter setting

We once more work with a rescaled free energy, which is now defined as

$$f^* = \frac{f}{\max\{\beta_S, 1\}}, \tag{4.24}$$

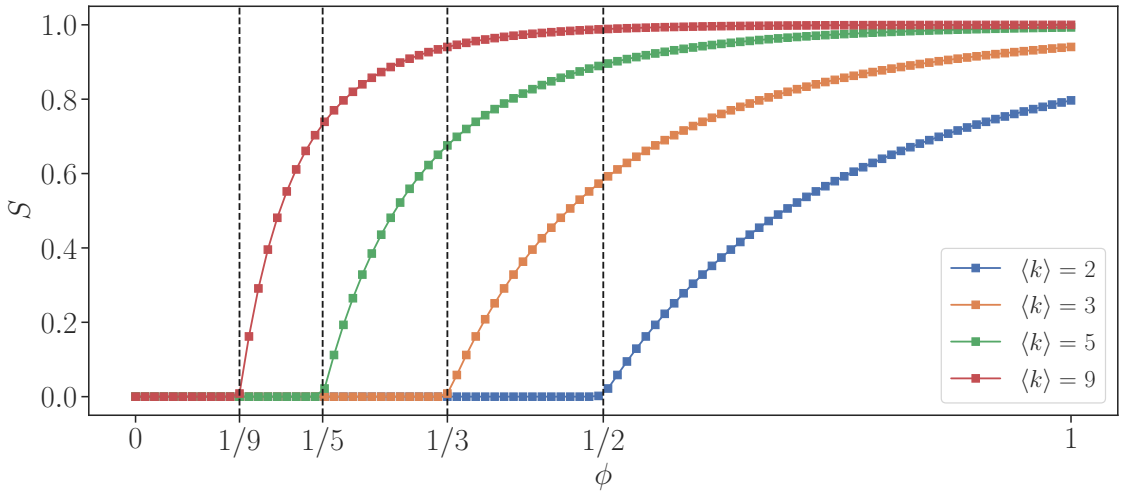


Figure 4.7: Plots of S as a function of the dilution probability ϕ for different values of $\langle k \rangle$, $\beta_S = 0$, and $B = 4$. The dashed black lines mark the positions of the percolation transitions, ϕ_c, for ER random graphs with corresponding values of $\langle k \rangle$.

and the gradient computations are performed again via the L-BFGS-B algorithm, using automatic differentiation to compute the gradient of S and with the `ftol` parameter set to 1×10^{-16} as in the modularity case seen previously.

The computation of the self-consistent equation Eq. (4.15) is potentially problematic, as repeated iterations can introduce numerical artefacts in the gradient computation. However, preliminary tests conducted by fixing the number of iterations to some n_{iter} and then gradually increasing n_{iter} show that the self-consistent equations can be safely run until convergence, where we consider the equations converged when $u_r^{(t+1)} - u_r^{(t)} < \epsilon$, where $u_r^{(t)}$ is the value of the probability of not being connected to the giant component at time-step t, and ϵ is an external parameter fixed to $\epsilon = 1 \times 10^{-8}$.

Finally, the optimisation procedure has been performed by taking subsequent adiabatic passes to avoid the system getting stuck in local minima. A first pass from low to high values of the selective pressures, followed by an additional pass from high to low values of β_S.

4.2.3 Results

We begin by studying the case in which there is no selective pressure driving to optimise robustness, $\beta_S = 0$, as we expect this to yield random graph topologies. Fig. 4.7 shows that our framework is indeed capable of correctly capturing the percolation transition observed in random graphs in the absence of any selective pressure driving to optimise robustness.

In Fig. 4.8 we show the properties of the obtained models for $\langle k \rangle = 5$ and $B = 2$ groups. As β_S increases, the network ensemble undergoes two abrupt transitions, where the structure first changes from fully random (I) to a core-periphery (CP) structure (II), and finally to an asymmetric bipartite structure (III). The core-periphery structure corresponds to a smaller and denser set of "core" nodes connected among themselves and a larger and sparser set of "periphery" nodes that connect mostly to the core nodes and not among themselves. The asymmetric bi-

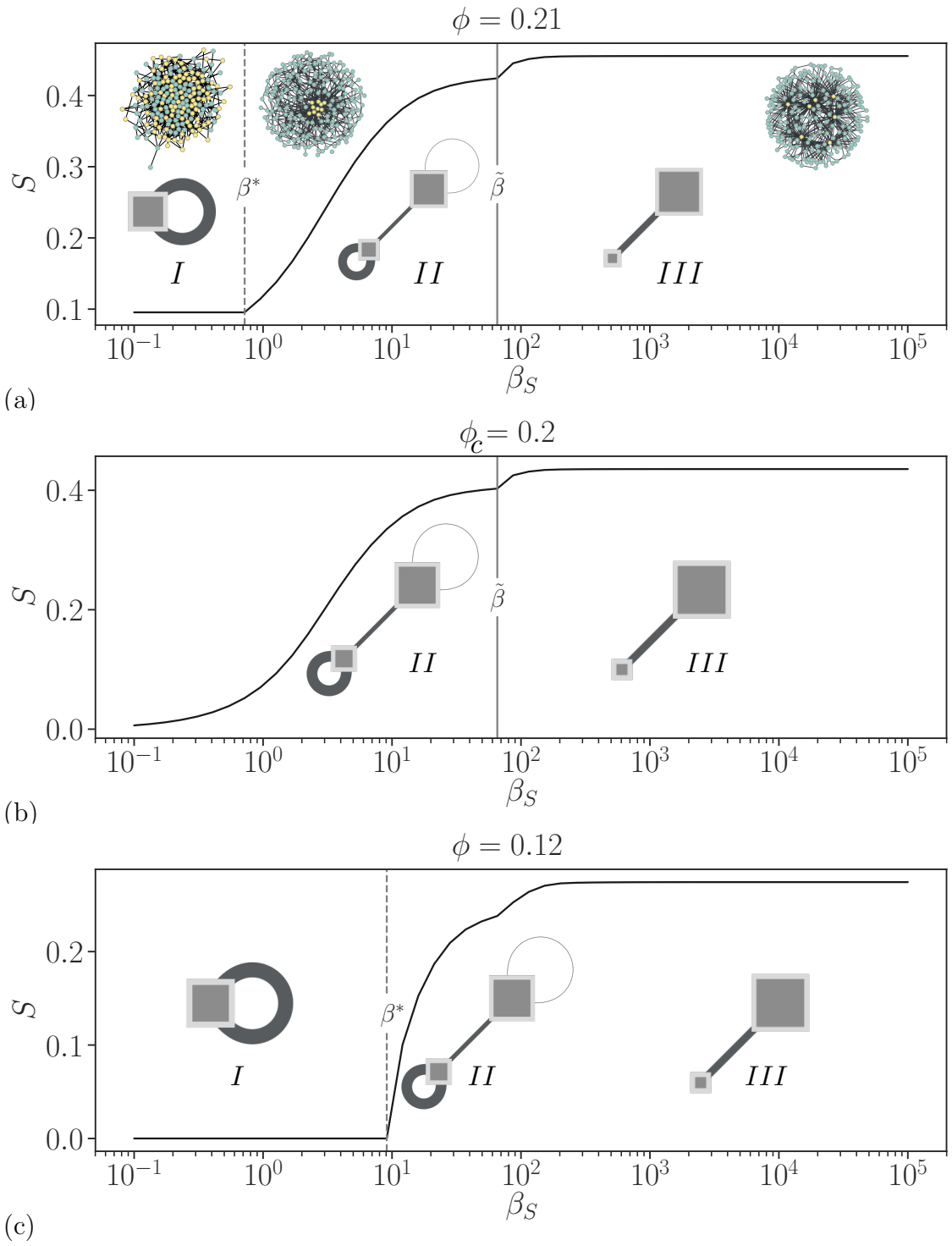


Figure 4.8: Relative size of the giant component S as a function of the selective pressure for robustness to damage β_S for different values of the edge dilution probability ϕ. The dashed vertical lines indicate the value $\beta_S = \beta^*$ at which we observe a transition from a random structure to a core-periphery one. The solid vertical lines indicate the value $\beta_S = \tilde{\beta}$ at which the network structure transitions from a core-periphery to a bipartite pattern. The optimized network structures are shown schematically in the insets, where each group corresponds to one of the groups of our model. Samples from the ensemble depicting the three structures are also shown in Figure (a).

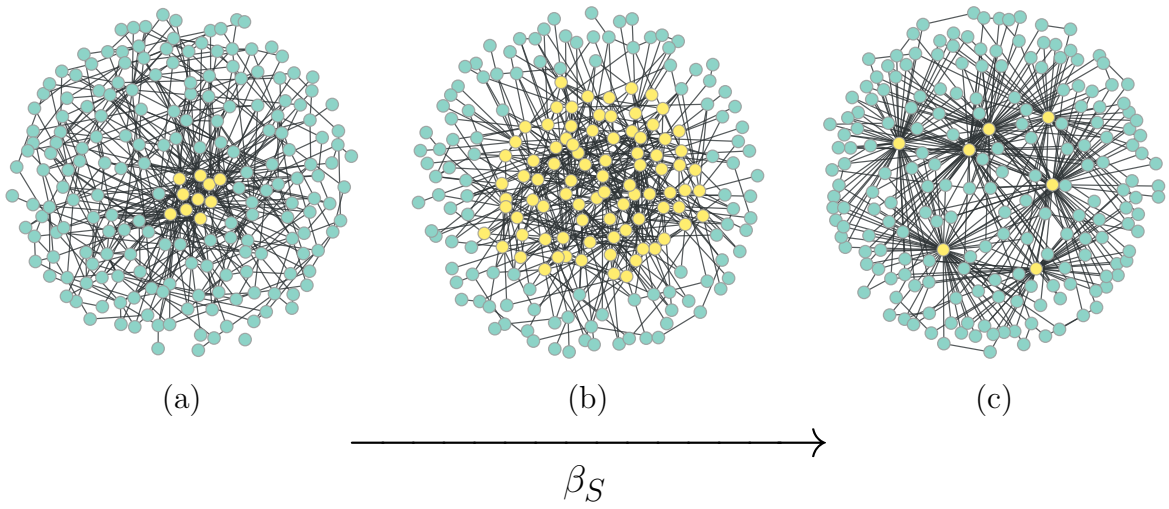


Figure 4.9: Ensemble samples depicting the typical evolution of the core-periphery structure as a function of the selective pressure β_S. (a) When the core-periphery structure first appears, it is composed of a small high-degree core. (b) As β_S increases, the size of the core group becomes larger, (c) before eventually transitioning to a bipartite structure.

partite structure is similar to the core-periphery pattern, but the "core" nodes no longer preferentially connect to themselves. Instead, they predominantly connect to the periphery nodes, although they remain a smaller and denser set. An illustration of these structures can be seen in Fig. 4.9, where we show network samples from the obtained ensembles. In Figs. 4.10 and 4.11, we also show the size and density of the two groups as a function of the selective pressure, for different values of the edge dilution probability ϕ.

It is easy to understand why a core-periphery structure increases the robustness to random edge removal: the core group corresponds to a denser subgraph, which remains connected with a large probability after the removal of a given fraction of edges, and the peripheral nodes benefit directly from this by connecting directly to the core, rather than among themselves. What is perhaps more surprising is the eventual onset of the bipartite structure, at which point the core group becomes so dense that its nodes tend to remain in the giant component even if they are not connected preferentially among themselves, which would incur a high entropic cost for no significant additional benefit, but instead, connect mostly to periphery nodes. The latter group tends to remain connected since its nodes tend to receive multiple connections to the denser core nodes. (Similar structures to the core-periphery one encountered here were also seen in similar setups where the robustness was integrated over all possible dilution values ϕ [40, 111] as well different ones based on dynamical robustness against noise [112], but the onset of the bipartite structures were not seen in these other cases.)

In most cases, the results tend to change predictably with different values of the edge dilution probability ϕ. However, a qualitative change in behaviour is seen when we cross the $\phi = \phi_c$ value, where $\phi_c = 1/\langle k \rangle$ is the critical percolation value for a fully random graph. For $\phi > \phi_c$, a fully random graph has a nonzero giant component even for $\beta_S = 0$, and thus the progression to core-periphery and bipartite structures proceeds as discussed above. However, for $\phi < \phi_c$, a fully random graph

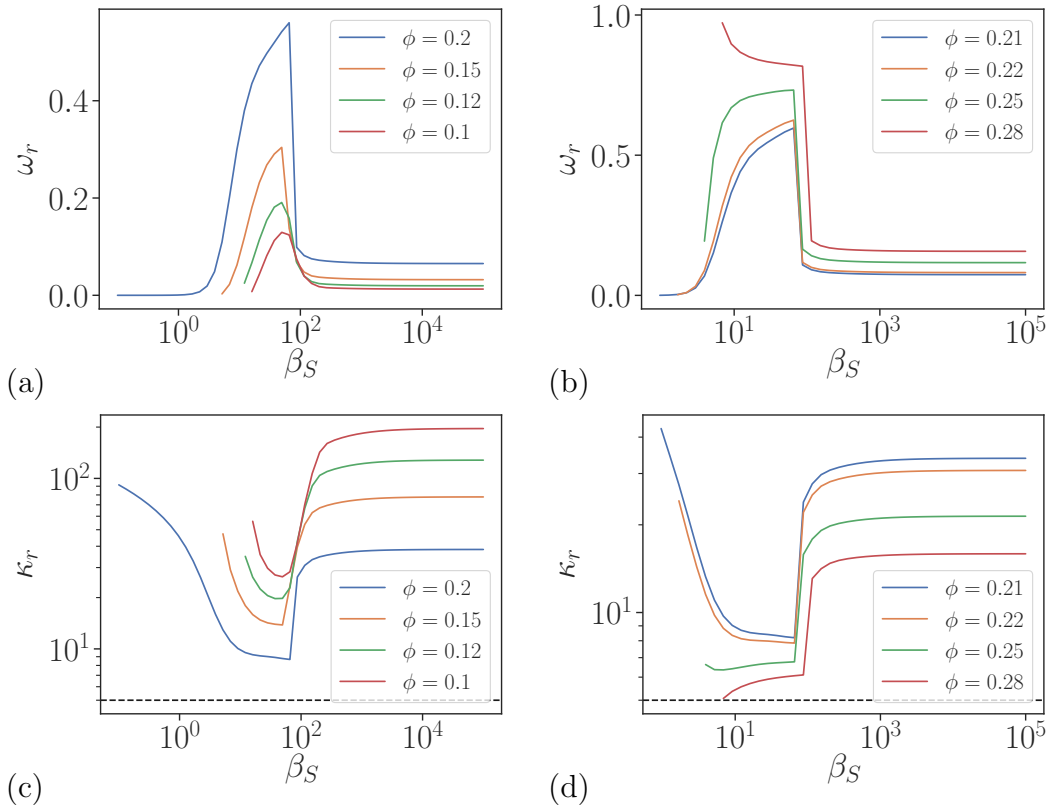


Figure 4.10: Fraction of nodes and average degree of the core groups as a function of the selective pressure β_S. Panels on the left display curves for values of $\phi \leq \phi_c$. Panels on the right display curves for values of $\phi > \phi_c$. The black dashed line in the plots for κ_r indicates the average degree of the network, which has been externally fixed to $\langle k \rangle = 5$.

gets completely disconnected. Therefore the response of the structural changes to increasing β_S is not continuous but happens more abruptly, with the onset of a core-group that is typically much denser. We also observe an interesting behaviour for sufficiently large values of ϕ. The core group spans almost the entire network at its onset, with an average degree coinciding with the whole network. The mechanism driving the network structure as β_S increases appears to be slightly different in this case, as it is the smaller set of "periphery" nodes that end up forming the smaller group of the eventual bipartite structure.

For $\phi = \phi_c$, we also observe a different behaviour, where the onset of the core-periphery structure ceases to be abrupt, and the change happens continuously. This seems to indicate that an infinitesimal optimisation of networks that lie on the critical percolation threshold has an infinitesimal entropic cost (a similar behaviour had been observed previously in the context of Boolean networks optimised against stochastic fluctuations [112]).

A more detailed overview of the combined effect of β_S and ϕ can be seen in Fig. 4.12, which shows both the value of $S(\beta_S, \phi)$, but also the relative improvement $\Delta S(\beta_S, \phi) = S(\beta_S, \phi) - S(0, \phi)$ with respect to a fully random graph. Indeed most of the improvement happens around the critical value $\phi = \phi_c$.

Changing the value of the imposed averaged degree $\langle k \rangle$ only shifts the position of the transitions, which remain qualitatively the same.

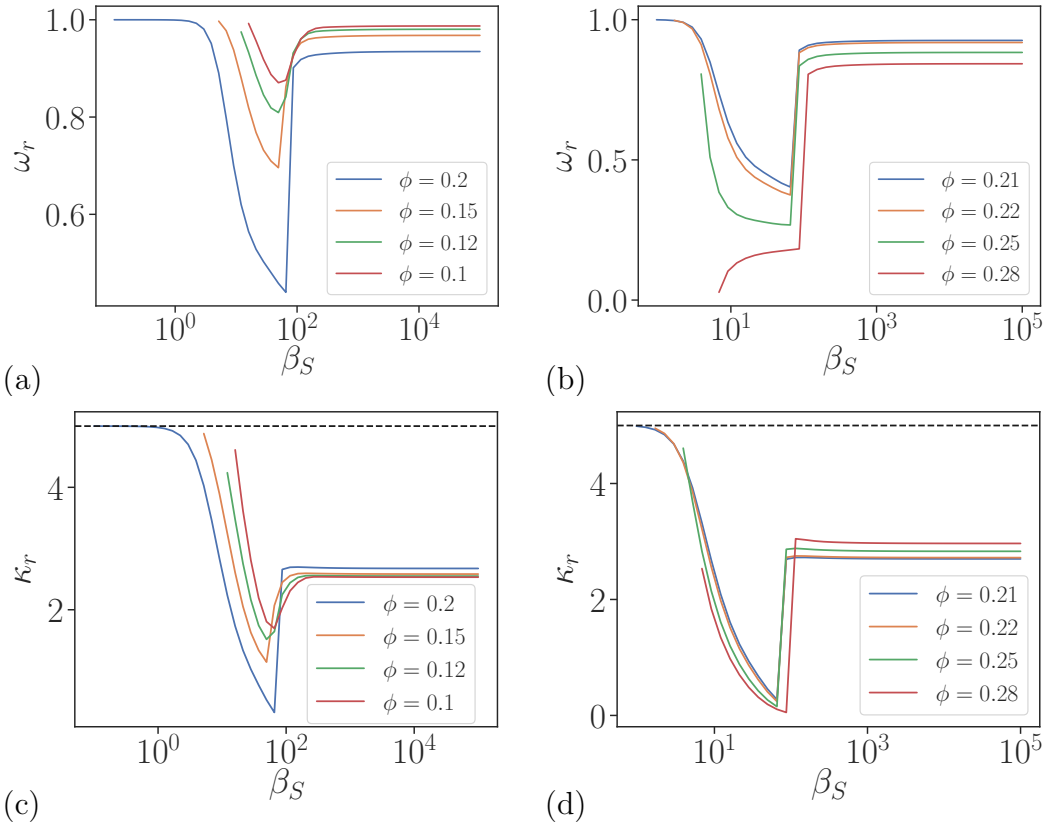


Figure 4.11: Fraction of nodes and average degree of the periphery groups as a function of the selective pressure β_S. Panels on the left display curves for values of $\phi \leq \phi_c$. Panels on the right display curves for values of $\phi > \phi_c$. The black dashed line in the plots for κ_r indicates the average degree of the network, which has been externally fixed to $\langle k \rangle = 5$.

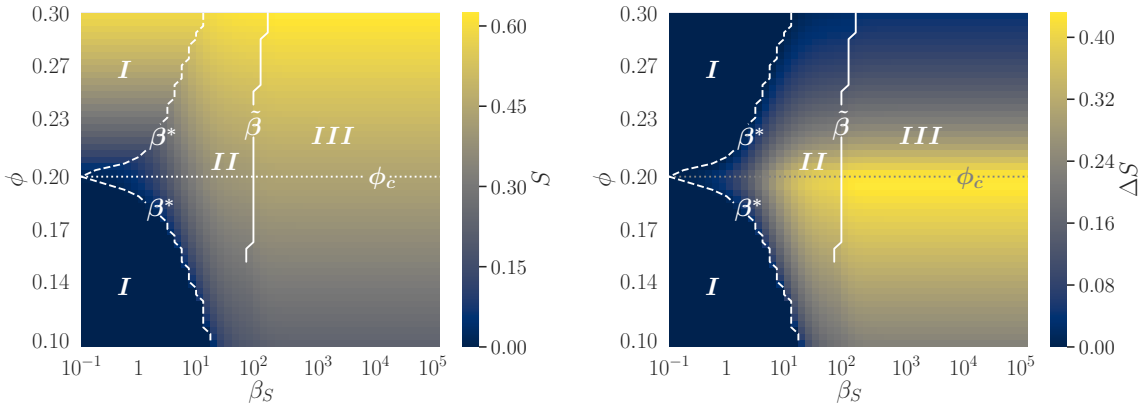
Increasing the number of groups

As a final note, we analyse the effects of increasing the number of groups on the system. Differently from modularity, whose value directly depends on the chosen number of groups, percolation properties are a network characteristic that we are modelling via a stochastic block model. As such, we would like to make no stipulations a priori regarding the number of partitions in the network. Therefore, when increasing the number of groups, we must account for the possibility that some of the generated partitions be topologically indistinguishable, i.e. with edges distributed randomly over two or more groups. A simple strategy to recover the correct number of groups required to describe the system is then the following

1. We propose "merge moves" amongst the various groups, where we attempt to merge group pairs.

2. We accept the proposed merging if it leaves the free energy unchanged.

In the case of robustness against random failures, we observe that increasing the number of groups B has no effect on the results obtained. Indeed, for any value $B > 2$, we find it possible to merge two or more groups without changing the ensemble properties until only two groups remain. The structures identified above are



(a) (b)

Figure 4.12: (a) Value of the fraction of nodes S which are part of the giant connected component as a function of the selective pressure β_S and dilution probability ϕ for networks with average degree $\langle k \rangle = 5$. (b) Variation in S with respect to the case where no selective pressure is applied as a function of the selective pressure β_S and dilution probability ϕ.

then to be considered the only ones to emerge when the selective pressure against random removal of edges is the only driving mechanism.

4.3 Discussion

In this chapter, we have applied our methodology to analyse the network topologies most likely to emerge when networks are driven to optimise either their modularity or their robustness against random deletions of their components. For modularity, a proliferation of symmetric clusters with an increasing number of internal edges has been identified as the most robust. On the other hand, a study of the resilience properties of networks driven to maximise their robustness against random failures of their connections has revealed three optimal structures depending on the value of the selective pressure. For low selective pressures, where there is little incentive to optimise robustness, a random or quasi-random topology describes the system. As the selective pressure increases, we have observed a series of transitions first to a core-periphery structure, described by a small and densely connected "core" group of nodes and a "periphery" group whose nodes predominantly attach to the core nodes, and subsequently to an asymmetric bipartite structure, similar to a CP topology, but in which the core nodes no longer connect to each other. In the last two scenarios, increased robustness against random failures is predominantly guaranteed by the presence of a small group of nodes vehiculating most of the network's connectivity. As deletions are random, the probability of removing one of these highly connected nodes is small, and the network's structure remains intact. Furthermore, we have shown that the observed structures are the only ones that can emerge if a selective pressure driving for increased robustness against random failure is the only driving mechanism of network formation —effectively providing a blueprint of robust topologies against random deletions.

While the results provided in this chapter help us to better understand how fitness towards a particular objective shapes network topology, they fail to address the more realistic scenario in which networks are subject to multiple selective pressures simultaneously. Modularity and robustness against random failures present themselves as ideal candidates for consideration in a combined scenario as they select for seemingly contrasting network topologies. Therefore, while they are individually characterised by well-defined structures, the interplay between the two selective pressures could potentially give rise to unforeseen effects. In the next chapter, we apply our framework to study networks subject to concurrent selective pressures driving to optimise both modularity and robustness against random failures.

Chapter 5

Multiple optimisation criteria

5.1 Introduction

The previous chapters have delineated the effects that selecting for specific characteristics can have on the emerging network topology. However, we limited ourselves to considering the effects of one single selective pressure at a time. While this allows us to gain valuable insight into the fundamental characteristics driving an increase in modularity or robustness against random failures, it overlooks the possible effects that can arise due to the interplay between selective pressures.

This chapter analyses the effects on network topology induced by different fitness criteria when these are applied simultaneously. Specifically, we once again focus on modularity and robustness against random failures. Modularity and robustness against edge removal appear to select for contrasting network topologies (a cluster structure in the case of modularity, core-periphery/bipartite structures for robustness against edge removal), and thus present themselves as attractive candidates for a study of the effects of competing selective pressures on network structure. However, far from being merely a question of convenience, analysing networks subject to optimise both their modularity and their robustness to random failure could potentially have interesting applications. For example, in the case of power grids, splitting the network into self-sufficient modules has been investigated as a possible mitigation strategy to prevent the propagation of cascading failures [113], linking the concepts of modularity and robustness. It is important to note that this is a different concept of robustness from the one considered in this Thesis. Indeed, cascading failures are a *dynamic* process where the failure of a node (or edge) can change the balance of flows on the network, leading to critical overloads, which result in a cascade of failures. By contrast, our approach is *static*, meaning that we remove nodes (or edges) without needing to redistribute any quantity transported by the network. Generally speaking, static robustness is more apt for describing the resilience of networks where no quantity is transported on the network, such as the ties in a social network. That being said, this form of static robustness is still essential for networks such as power grids, as networks that can be easily disconnected might be more robust to cascading failures yet still be unable to accomplish their intended function after the failure of even a small fraction of nodes. It is then interesting to analyse what structures emerge as a result of attempting to optimise for both criteria, as these structures could be beneficial in constructing robust networks against both forms of failures.

We begin by providing a brief introduction and some notes regarding the implementation details of our methodology and proceed to present our results. Our main finding is the identification of a series of phase transitions where the network topology is characterised by different entanglements of the previously encountered structures. Furthermore, we observe regions of the phase space where the interplay between the selective pressures can give rise to synergistic or antagonistic effects, i.e. selecting for one kind of fitness criterion can aid or hinder optimising for the other. Finally, we conclude with some considerations on the effects that increasing the number of available groups has on the emerging topologies.

5.2 Implementation details

We consider the situation where we seek to optimise modularity and robustness against random edge removal. In principle, this would amount to a free energy given by

$$\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m}) = -\beta_S S(\boldsymbol{\omega}, \boldsymbol{m}) - \beta_Q Q(\boldsymbol{m}) - \Sigma(\boldsymbol{\omega}, \boldsymbol{m}). \tag{5.1}$$

However, this would mean that the same division of the network used to compute modularity would also be used to obtain the robustness to edge removal. However, in general, there is no reason to impose that these quantities are related, i.e., the network structure that is responsible for increased robustness to edge removal may be unrelated to the patterns that cause increased modularity. Because of this, we want to be more general and allow the modularity of the network to refer to a division that is not necessarily related to the one used to obtain the robustness to damage. We do so by assuming that the partition used for the computation of robustness is a subdivision of the one used to obtain modularity, such that each of its B_Q groups can be further divided into one, two, or more groups, totalling $B_S \geq B_Q$ groups. This assumption is made without loss of generality since any two independent partitions into B_1 and B_2 groups can always be equivalently decomposed into one with at most $B_1 \times B_2$ groups, which is itself a subdivision of a smaller one with $\min(B_1, B_2)$ groups[1]. Based on this, we have the free energy given by

$$\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{c}) = -\beta_S S(\boldsymbol{\omega}, \boldsymbol{m}) - \beta_Q Q[\boldsymbol{m}'(\boldsymbol{m}, \boldsymbol{c})] - \Sigma(\boldsymbol{\omega}, \boldsymbol{m}), \tag{5.2}$$

where $\boldsymbol{c} = (c_1, \ldots, c_{B_S})$ is a hierarchical grouping of the B_S groups, with $c_r \in [1, B_Q]$ being the group membership of the group r used to compute the giant component S. The modularity is therefore computed with the condensed matrix

$$m'_{tu}(\boldsymbol{m}, \boldsymbol{c}) = \sum_{rs} m_{rs} \delta_{r,c_t} \delta_{s,c_u}. \tag{5.3}$$

We stress that for our calculations, the identity of the group memberships are irrelevant, as we concern ourselves only with the resulting network structures. Therefore,

[1]Note that we consider the case $B_S \geq B_Q$, but not the opposite, $B_Q \geq B_S$. Indeed, considering B_Q to be a subdivision of B_S could potentially introduce artificial constraints into our model. As mentioned in Section 4.1, when introducing modularity, we are looking to maximise Q conditioned on a known partition. S, on the other hand, is not; i.e. in principle, we do not care how the network is divided as long as the overall robustness is achieved. Allowing B_Q to be a subdivision of B_S would then result in a constraint on B_S, as group structures with less than the number of occupied B_Q partitions would not be possible. We would essentially be conditioning both Q and S on the partition used to model modularity.

we select $B_Q = q$ and $B_S = ql$, where each of the q groups used to compute Q are subdivided into exactly l groups. Again, this comes without a loss of generality, as we do not make any provisions about how large each group is or even if they are occupied at all. Therefore this scheme is purely conventional and does not impose any inherent symmetry or network structure on its own. By choosing q and l sufficiently large, we can obtain any kind of modular structure used to compute either S or Q independently. For our calculations, we have used mostly $q = l = 2$, which have proved sufficient to capture most of the structures seen, but we have investigated higher values as well, as we discuss later.

5.3 Results

We minimized Eq. (5.2) for an ensemble of networks with $\langle k \rangle = 5$, and edge dilution probability $\phi = 0.21$. Fig. 5.1 shows the relative changes of the optimization criteria as a function of the selective pressures β_S and β_Q, where $\Delta S(\beta_S, \beta_Q)$ and $\Delta Q(\beta_S, \beta_Q)$ are defined as

$$\Delta S(\beta_S, \beta_Q) = S(\beta_S, \beta_Q) - S(\beta_S, 0), \tag{5.4}$$
$$\Delta Q(\beta_S, \beta_Q) = Q(\beta_S, \beta_Q) - Q(0, \beta_Q), \tag{5.5}$$

and represent the relative variations in S and Q induced by the interplay between the selective pressures with respect to the case in which we optimised for each constraint in isolation. As the selective pressures are changed, we observe various structural phases, representing diverse combinations of the modular, core-periphery and bipartite structures encountered previously. The transitions between the various structures can be either smooth or abrupt. In the latter case, we can distinguish three types of transitions. The first type of transition is linked to abrupt changes in the network structure and can be identified by sudden jumps in the group parameters. The second kind of transition occurs when the number of groups required to describe the system changes, but no significant jumps in the group parameters are observed. Finally, the third type of transition is a mixed transition, where a change in the number of groups required to describe the system is accompanied by an abrupt change of the group parameters. Furthermore, we also observe synergistic and antagonistic effects, whereby selecting for one fitness criterion can help (or hinder) optimising for the other. We will discuss these effects in more detail depending on the region where they occur in the phase diagram.

5.3.1 Regions in the phase diagram

The low β_S and low β_Q regimes: For low values of β_S, we can recover the behaviour observed when selecting for modularity in isolation by varying β_Q, and the network structure varies from a random graph, [see Fig. 5.1(a)], to increasingly separated and modular structures [5.1(c) and 5.1(d)]. Conversely, we note that the behaviour observed when selecting for robustness against random failures in isolation is not recovered for low β_Q. By increasing β_S at some fixed low β_Q, the network initially follows the expected behaviour and transitions from a random graph [Fig. 5.1(a)] to a core-periphery structure [Fig. 5.1(e)]. However, for high β_S, the network structure is now described by a four-group structure composed of two

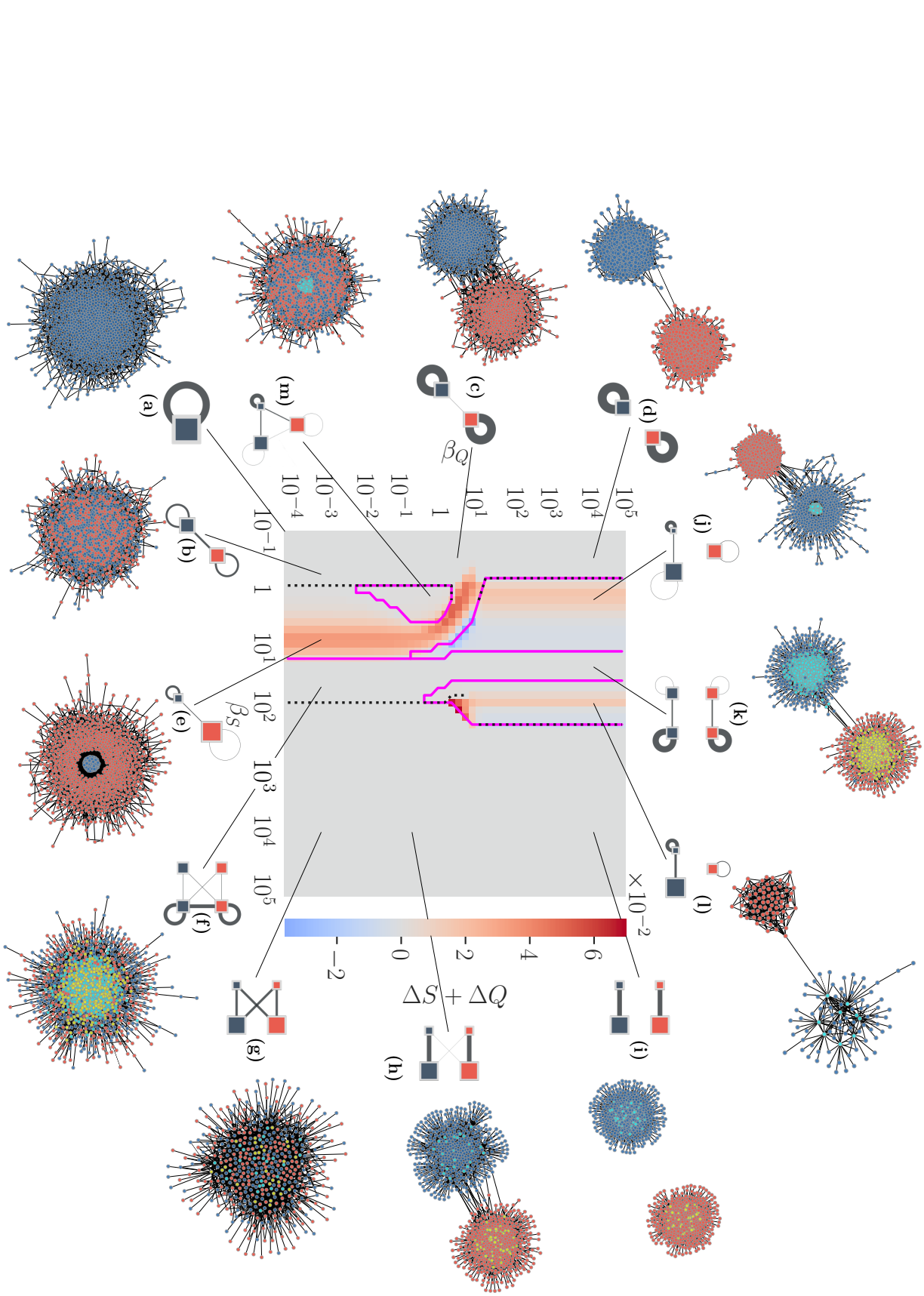


Figure 5.1: Relative change $\Delta S + \Delta Q$ of the fitness values as a function of the selective pressures β_S and β_Q for an ensemble of networks with $\langle k \rangle = 5$ and dilution probability $\phi = 0.21$. The black dashed lines correspond to transitions linked to abrupt changes in the network parameters, the solid magenta lines correspond to transitions in which the number of groups required to describe the system changes. Schematics of the optimized network structures for each region are shown in the margins, with each group corresponding to one of the B_S groups of our model and the colour of each group indicating its B_Q membership. Network samples drawn from the ensemble are also shown.

identical and interconnected core-periphery or bipartite structures [Figs. 5.1(f) and 5.1(g)]. This symmetric effect can be understood in terms of modularity. As β_S increases, the selective pressure against random edge removal pushes the network towards increasingly stronger bipartite structures. Since those structures have edges running predominantly between different groups, they would yield negative modularity values. Therefore, by splitting both "core" and "periphery" groups into two random subgroups used for the computation of modularity, the network can escape the negative values with negligible entropic cost. Note that, in principle, one could recover a modularity of zero *and* keep a two-group structure by simply keeping one of the two B_Q groups empty. However, as we can see from Fig. 4.6, modularity is a monotonically increasing function of β_Q, meaning it will only be zero exactly at $\beta_Q = 0$. Maintaining a four-group structure where both B_Q groups are populated allows the network to attain infinitesimally positive modularity values for $\beta_Q > 0$.

The high β_S and high β_Q regimes: If we increase β_Q at some fixed high value of β_S, we once again observe that the optimisation of modularity causes the symmetric structures observed above to become less interconnected until two separate and identical structures coexist [i.e. Figs. 5.1(g), 5.1(h), and 5.1(i)]. This symmetric pattern effect can be understood as a direct consequence of both optimisation criteria competing with each other: since forming a single mixed core-periphery/bipartite structure would yield low modularity, the overall structure is mirrored to preserve high fitness values according to both criteria.

More interesting effects occur if we consider the impact that increasing β_S has at some fixed high value of β_Q. In this scenario, we once again observe symmetric structures [see Figs. 5.1(k) and 5.1(i)]. However, we also see the presence of regions where an asymmetric three-group pattern describes the network structure [see Figs. 5.1(j) and 5.1(l)]. In these regions, we again observe either a core-periphery or bipartite structure as a result of the selective pressure towards robustness against random edge removal. The requirement to have a high fitness for modularity is instead reflected by the presence of an accompanying and distinct modular structure. This accompanying modular structure is always denser than a fully random graph. It becomes increasingly dense as β_S is increased, suggesting that the effects of the selective pressure against random edge removal are not limited to the core-periphery or bipartite structures.

Intermediate regimes: For intermediate values of β_S and β_Q, the network transitions smoothly and abruptly between the same structures described above. The only difference is the presence of an "island" where a three-group pattern again describes the network structure [see Fig. 5.1(m)]. In this region, the structure is that of a core-periphery pattern in which we now have two peripheries preferentially connecting to a dense set of core nodes. This structure remains substantially unchanged if we vary β_S. However, by increasing β_Q, one of the two peripheries becomes progressively smaller and less connected to the core, and the overall network structure closely resembles the one observed in Fig. 5.1(j).

5.3.2 Synergistic and antagonistic effects

To better understand the synergistic and antagonistic effects seen in Fig. 5.1, it is convenient to consider the relative variations over Q and S individually, as shown in Fig. 5.2 and Fig. 5.3 respectively. Based on this, we consider each effect in isolation as follows.

Modularity: Inspecting the diagram for ΔQ in Fig. 5.2, we can see that for low values of β_Q and β_S, the network structure is essentially that of a fully random graph. By increasing β_S, we eventually encounter a synergistic region just above the β^* transition line that exists when $\beta_Q = 0$ (see Fig. 4.8). This indicates that merely transitioning to a core-periphery structure is enough to guarantee some degree of improvement in modularity with respect to a random graph. This synergistic region extends until moderate values of β_Q, corresponding to the region in Fig. 4.6 in which modularity shows a rapid increase. For high values of β_Q the synergistic effects vanish, as we now find ourselves in the region of Fig. 4.6 where the modularity reaches its plateau value, and no structural transition can provide an additional benefit with respect to the case in which we optimise for modularity in isolation.

What is perhaps more interesting is the small synergistic region in ΔQ around the $\tilde{\beta}$ transition line. In this region of the phase space, the network structure is described by a bipartite pattern and a separate modular division. It would appear that the emergence of a bipartite structure — driven by the selective pressure towards robustness against edge removal — forces more edges to be distributed within their own groups than would be the case had we selected for modularity alone, thus providing an increased fitness.

Robustness against random failures: In the ΔS phase space, we observe two principal regions in which synergistic (antagonistic) effects are present, labelled A and B in Fig. 5.3. In region A, the network structure is described by a core-periphery pattern accompanied by an isolated cluster which is always denser than a fully random graph. This structure is initially able to provide greater robustness against random failures than the corresponding two-group core-periphery structures we observed in Fig. 4.10. However, it also has a higher entropic cost, which is accounted by the selective pressure for modularity, and we observe a synergistic interplay between the two selective pressures. This three-group structure displays no significant changes as β_S increases, and, eventually, the evolution of the core-periphery structures observed in Fig. 4.10 can provide greater robustness. At this point, the selective pressure for modularity reverses its role by pinning the less optimal three-group structure in place, and we observe an antagonistic interplay between the two selective pressures. Increasing β_S even further, we eventually reach the point where it is more beneficial for the network to pay a further cost in entropy and split into two symmetric structures in exchange for larger mutual fitness.

A similar picture occurs in region B, where the network structure is characterised by a bipartite pattern and an accompanying cluster which is always denser than a fully random graph. The onset of region B happens for values of $\beta_S \leq \tilde{\beta}$, and the added bipartitness initially provides an increased fitness against random edge removal. However, the role of the selective pressure for modularity once again reverses as soon as $\beta_S > \tilde{\beta}$ and we cross the bipartite transition line observed when

optimising for robustness against random edge removal in isolation.

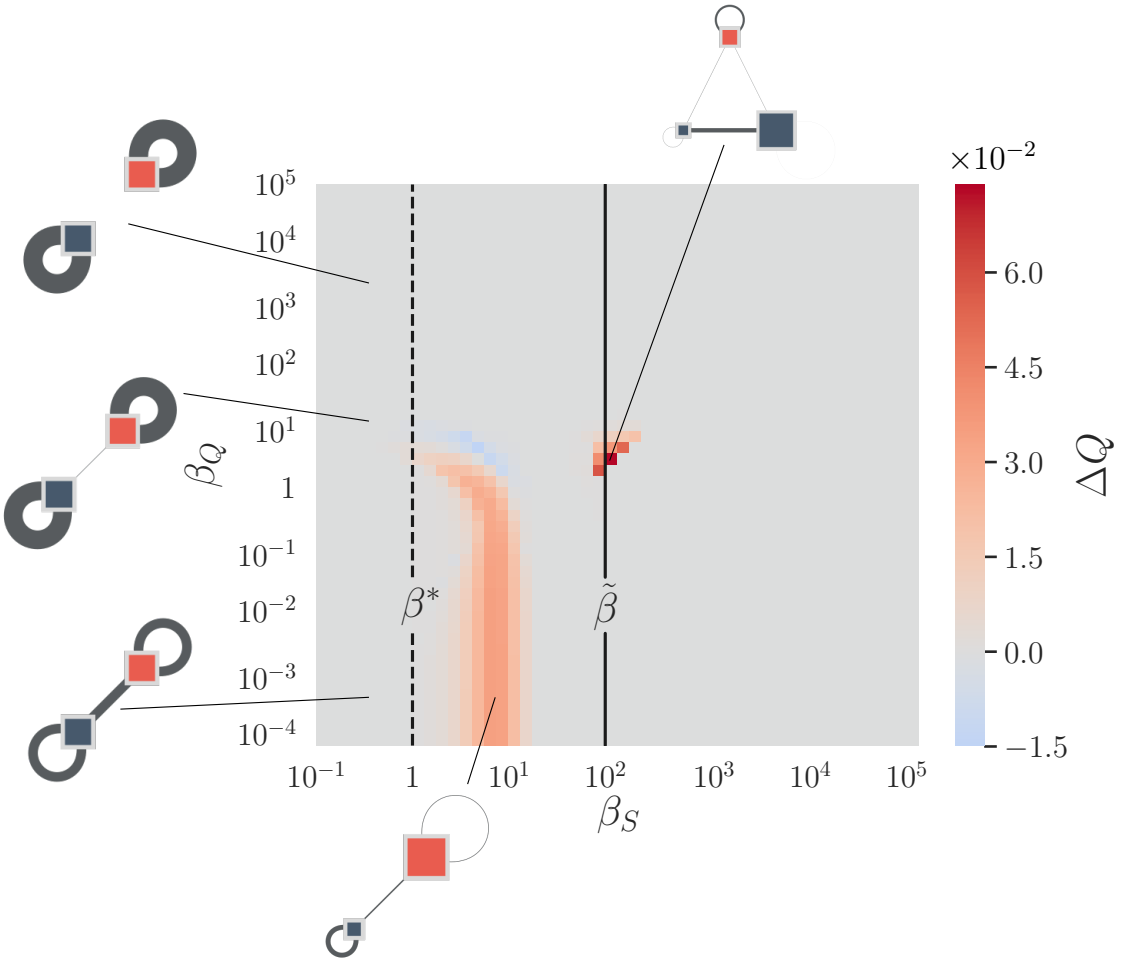


Figure 5.2: Change in modularity, Q, with respect to the case $\beta_S = 0$ as a function of the selective pressures β_S and β_Q. The dashed and solid black lines indicate respectively the values of β_S at which abrupt transitions to core-periphery and bipartite structures are observed when optimizing for robustness against random edge removal in isolation. Schematics of the optimized structures are shown around the margins, where each group corresponds to one of the B_S groups in our model and the color of each group indicates its B_Q membership. The results are obtained for an ensemble of networks with $\langle k \rangle = 5$ and edge dilution probability $\phi = 0.21$.

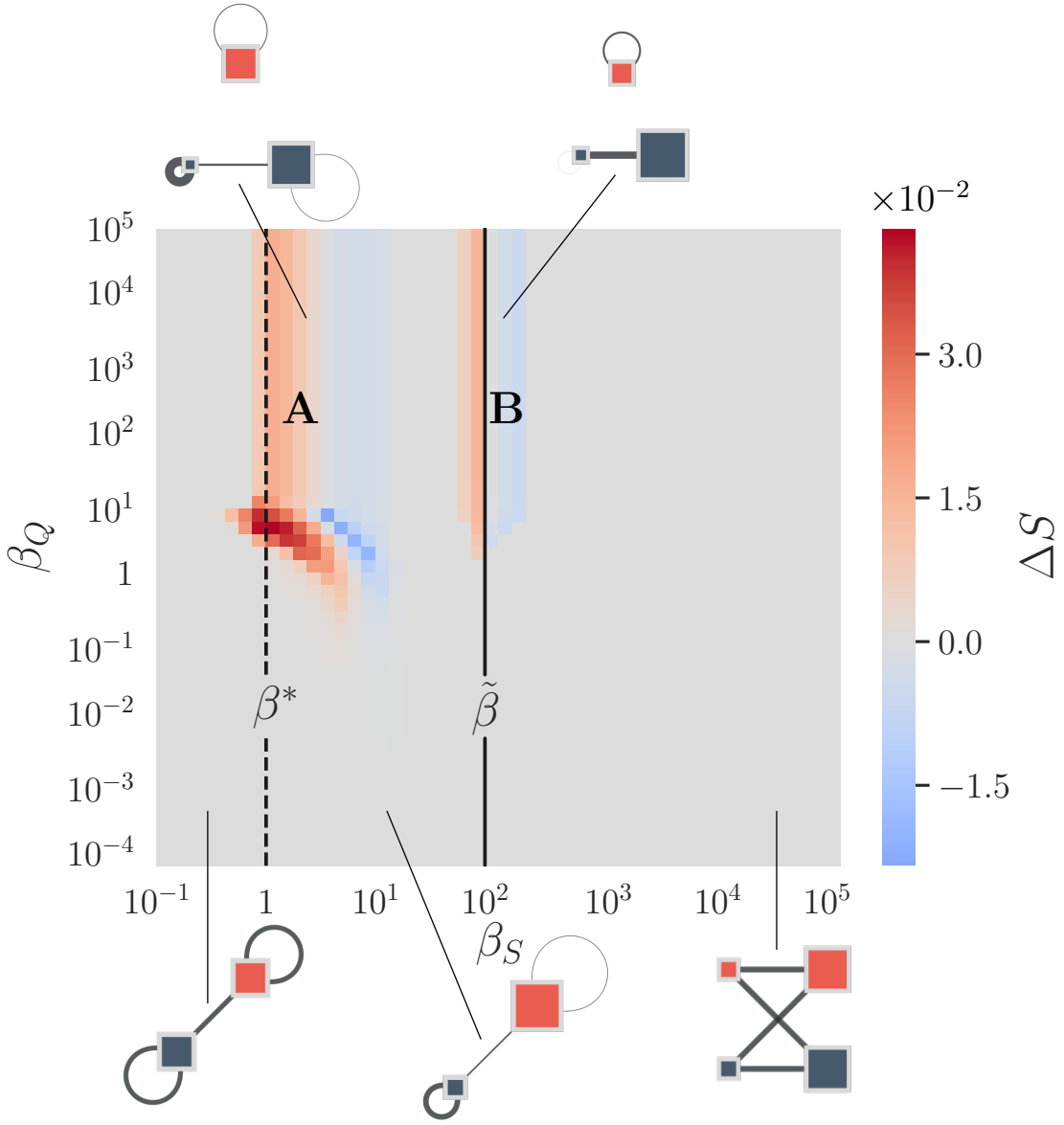


Figure 5.3: Change in the size of the largest component, S, with respect to the case $\beta_Q = 0$ as a function of the selective pressures β_S and β_Q. The dashed and solid black lines indicate respectively the values of β_S at which abrupt transitions to core-periphery and bipartite structures are observed when optimizing for robustness against random edge removal in isolation. Schematics of the optimized structures are shown around the margins, where each group corresponds to one of the B_S groups in our model and the color of each group indicates its B_Q membership. The results are obtained for an ensemble of networks with $\langle k \rangle = 5$ and edge dilution probability $\phi = 0.21$.

5.4 Increasing the number of groups

As mentioned at the beginning of the chapter, it would, in principle, be possible to model any kind of modular structure by choosing high enough values of q and l. However, the free energy computation grows quadratically with B_S, making it computationally expensive to increase the number of groups used to model the network. Nevertheless, we have investigated regions of the phase diagram, allowing us to probe in more detail how the allowed number of groups affects the results. Our findings indicate that increasing the number of groups available can exacerbate the synergistic and antagonistic effects observed previously but does not alter the regions in which these occur. However, increasing the number of groups can potentially give rise to different entanglements of the core-periphery, bipartite, and modular structures observed above. In what follows, we present our results for network ensembles with $\langle k \rangle = 5$ and edge dilution probability $\phi = 0.21$.

The total number of groups used to model the networks can be increased by increasing either l or q, where changing the former corresponds to varying the number of groups B_S used to model robustness, while the latter controls the number of groups used to model modularity.[2] Increasing l without varying q appears not to affect the emerging topologies. The additional groups can always be merged back into the original $q = l = 2$ structures, and the overall observables of the system remain unvaried, see Fig. 5.4. This behaviour is to be expected. As discussed in Section 4.2, a total of two groups is sufficient to capture the three topologies that characterise the resilience properties of the network. The three and four group structures observed in Fig. 5.1, where the core-periphery and bipartite topologies are either mirrored or accompanied by an additional modular structure, arise as a response to the need to guarantee a pre-determined degree of modularity. For whatever choice of q we impose, we would therefore expect $l = 2$ to be sufficient to capture the structural properties responsible for robustness against random failures.

On the other hand, increasing q will directly impact the number of groups used to model the modularity. As the value of modularity directly depends on the number of occupied groups, increasing q allows the system to potentially increase its modularity (thereby decreasing its free energy), by occupying a larger number of groups. However, occupying a large number of groups also has a high entropic cost, and one would therefore expect a balance between fitness and entropy to be reached based on the imposed values of the selective pressures.

Let us consider again the network shown in Fig. 5.4 and this time fix $l = 2$ and vary the value of q. Fig. 5.5 shows the behaviour of the observables of the system as a function of q. As expected, we observe an increase in modularity alongside a decrease in entropy as q is increased. However, we also observe a sharp drop in S followed by a gradual increase to a plateau value.

To gain a clearer understanding of this behaviour, we can analyse the topology of the networks at different values of q. Fig. 5.6 shows schematics of the network topologies alongside some network samples for the cases $q = 2, 3, 4$, and 5. The sharp drop in the fraction of nodes belonging to the giant connected component can be explained by observing the transition in topology from Fig. 5.6(a) to Fig. 5.6(b). While both structures are characterised by having three groups, in Fig. 5.6(b) the

[2]Note that varying q will also influence the partitions used to model the percolation properties of the network in view of the relationship $B_S = ql$.

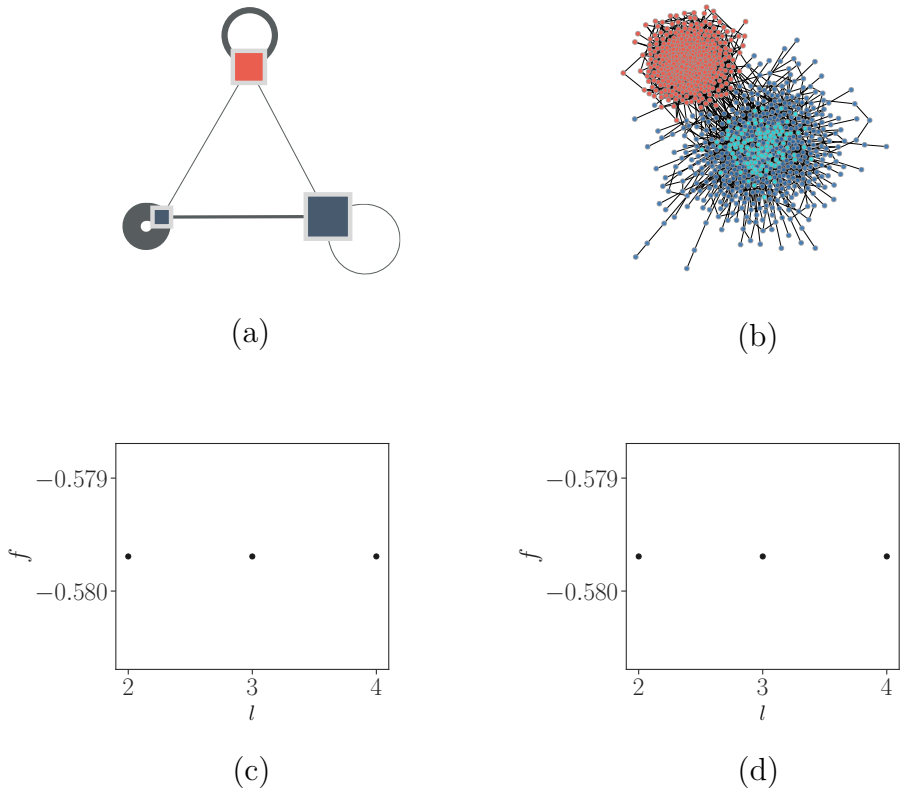


Figure 5.4: Schematic depicting the network structure at $\beta_S = \beta_Q = 9$ in the $q = 2$, $l = 2$ case (a). Sample from the ensemble (b). Free energy (c) and fraction of nodes that belong to the giant connected component (d) as a function of l. As the number of B_S groups increases, there is no observed effect on the network properties and the additional groups can always be merged back into the three block structure shown in (a). Further tests, conducted on other points of the phase space, also display the same behaviour.

network loses its core-periphery component in favour of an additional modularity group, and three modular structures now describe the topology, resulting in a loss of robustness. Notice, however, that the modular structures are no longer symmetric, with one of the groups typically much larger than the others and characterised by a lower average degree. As q increases, the number of occupied groups increases, but the network topology remains characterised by a modular structure. The free energy presents a minimum at $q = 4$, indicating that a total of four groups is required to adequately describe the network. Above $q = 5$, it becomes virtually impossible for the optimiser to select states in which all groups are occupied as the entropic cost is too large. Consequently, we observe the presence of numerous empty groups, and the resulting structures are equivalent to those observed in the $q = 4$ or the $q = 5$ cases. Finally, we note that, although in principle higher values of S can be obtained by mirroring the core-periphery structure observed in Fig. 5.6(a), the entropic cost of this is too high, and the network prefers to split into separate modules.

In conclusion, while increasing the number of groups used to model the percolation properties of the system does appear not affect the emerging topologies, increasing the number of groups used to model modularity can lead to different entanglements of the previously observed core-periphery, bipartite, and modular structures.

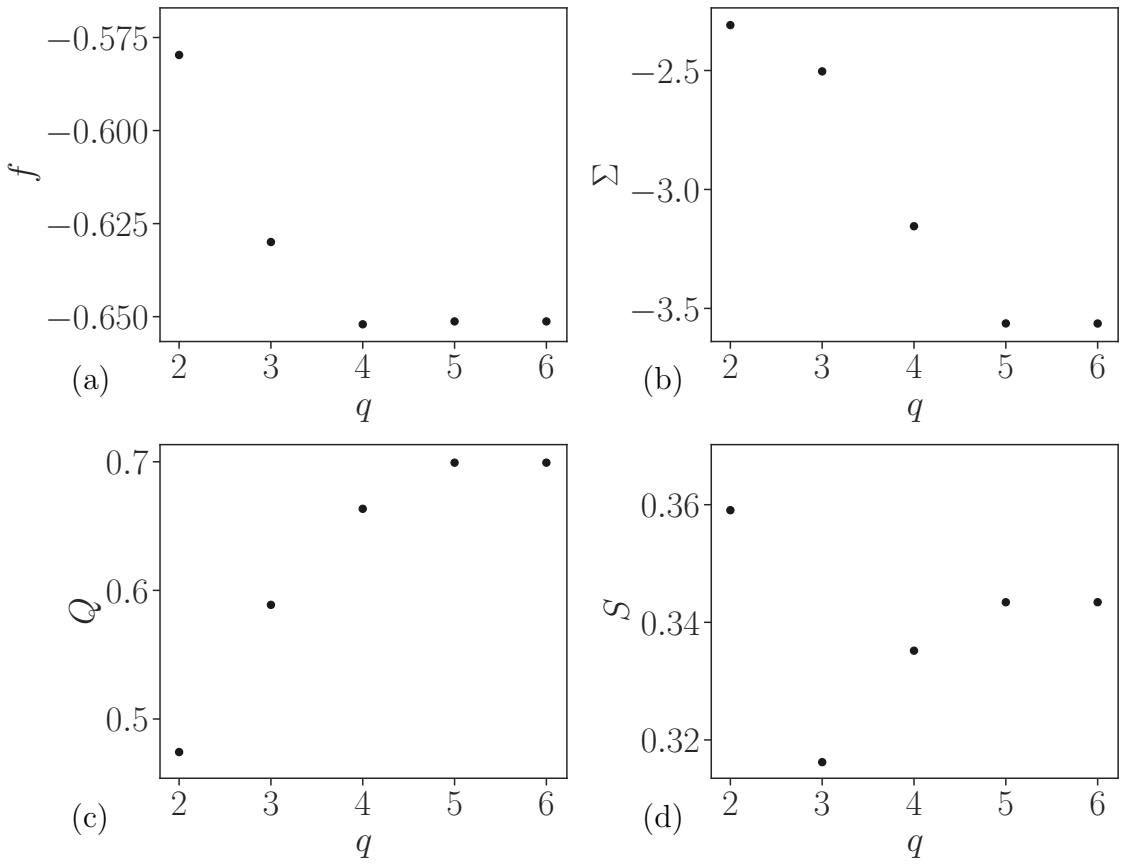


Figure 5.5: Free energy (a), entropy (b), modularity (c), and the fraction of nodes belonging to the giant connected component (d) as a function of q for a network sample at $\beta_S = \beta_Q = 9$ and $l = 2$.

While computational constraints prevent us from systematically exploring the entire phase space, we can nevertheless gain a broader perspective of the effects that increasing the allowed number of groups has on the network topologies by studying the emerging structures along some pre-specified slices of the phase space. As an example, we consider the two slices at fixed β_S shown in Fig. 5.7, and for each of them, we fix $q = 8$ and $l = 2$. Fig. 5.8 and Fig. 5.9 show a comparison of the modularity as a function of β_Q for both the $q = l = 2$ case studied above, and this new case with $q = 8$ and $l = 2$.

For slice A, we can see that the two curves coincide for low to moderate values of β_Q, with the network structure transitioning from a two-group core-periphery structure to a three-group core-periphery, with two peripheries connecting to a dense core group [see Fig. 5.1(m)]. For higher values of β_Q, the curves diverge, as the higher value of q in the $q = 8$, $l = 2$ case allows the network to populate more groups, thus increasing its modularity. The number of populated groups increases with β_Q, and the network topology is described by interconnected modular structures that become progressively disconnected from each other as the selective pressure is raised. Again, in contrast to what we observed when we optimised for modularity in isolation, we note that these new modular structures are not symmetric, with some groups being denser than a random graph and others less so.

For slice B, we find ourselves in a region of the parameter space where the

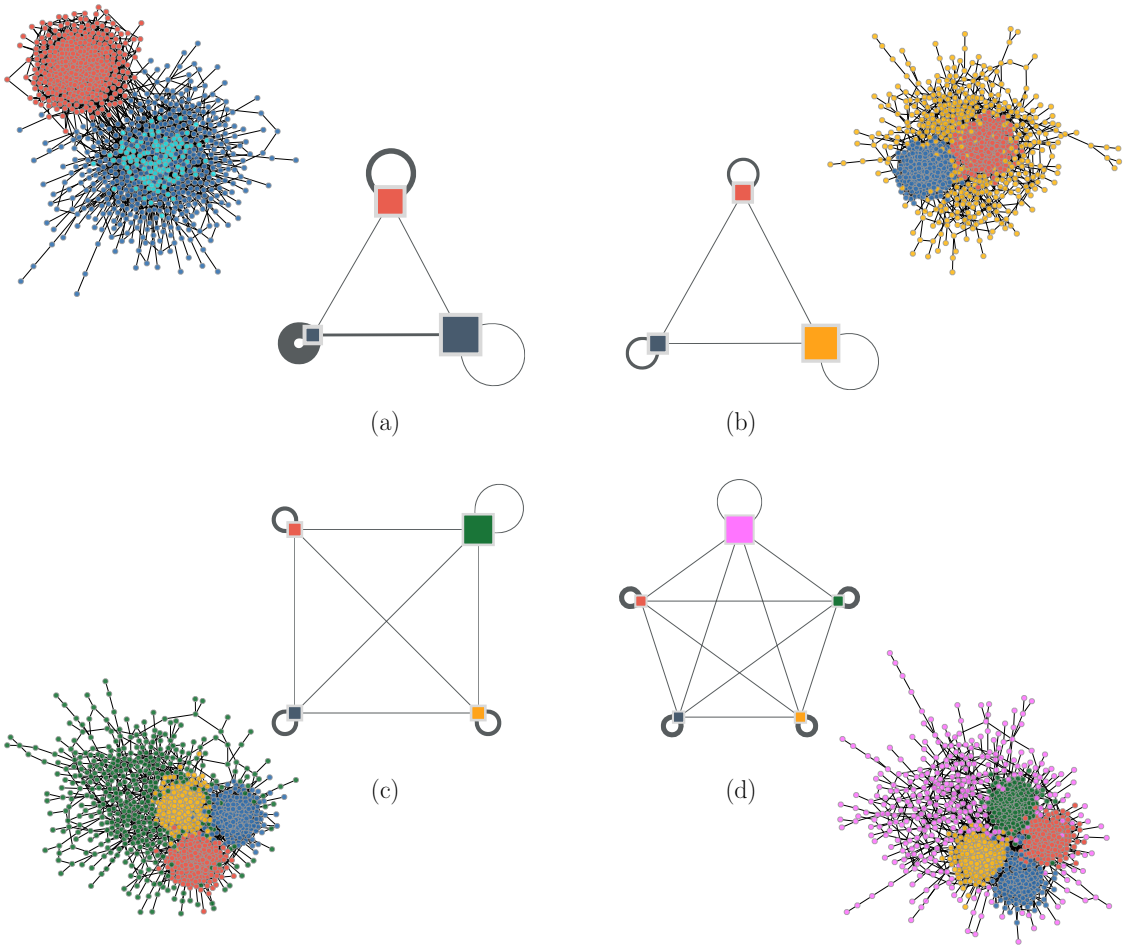


Figure 5.6: Network topologies at $\beta_S = \beta_Q = 9$ for $l = 2$ and varying values of q. A core-periphery accompanied by a modular structure corresponding to the case $q = 2$ (a), and modular structures corresponding to $q = 3$ (b), $q = 4$ (c), and $q = 5$ (d) respectively.

network topology is described by two symmetric core-periphery structures, which get progressively disconnected as β_Q is raised. Once again, the two curves coincide for low to moderate values of β_Q, but, as β_Q increases, the access to a higher number of B_Q groups in the $q = 8, l = 2$ case allows for more groups to be populated. We observe different entanglements of core-periphery structures accompanied by isolated clusters, and, for high enough β_Q, we once again observe a mirroring effect in which eight symmetric core-periphery structures describe the network topology.

Synergistic effects The variations in Q and S observed in Fig. 5.5 suggest that increasing the number of groups might alter the synergistic/antagonistic effects encountered previously. However, preliminary results suggest that, while increasing the number of groups used to model the network can exacerbate the synergistic/antagonistic effects, it does not alter the regions in which these are observed.

As an example, consider slice A in Fig. 5.7. In the $q = l = 2$ case, slice A selects a region of the phase space displaying synergistic effects on S for high β_Q values and negligible effects on modularity (see Fig. 5.10). The variation in modularity, ΔQ, and robustness against random edge removal, ΔS, along the slice for both the

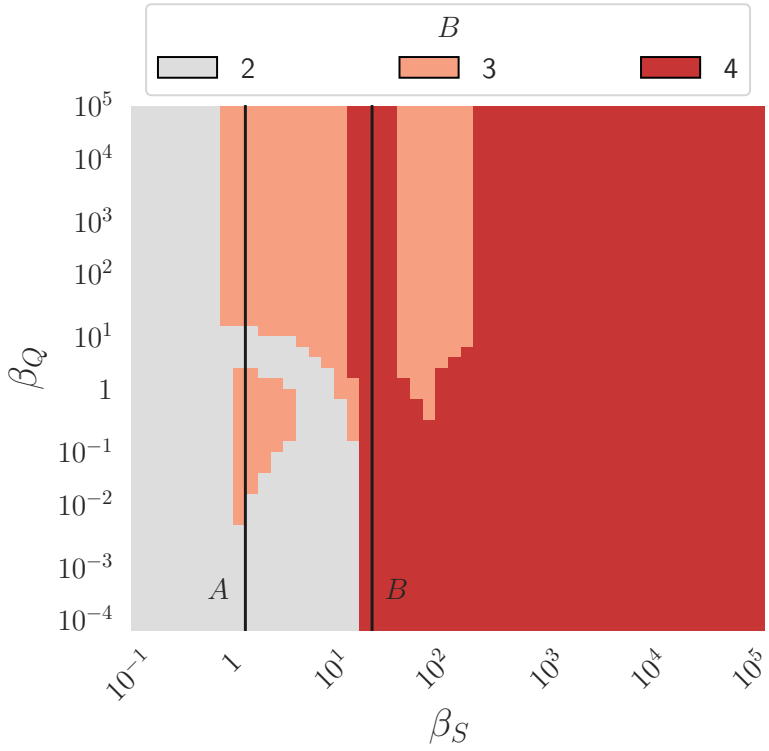


Figure 5.7: Total number of groups required to describe the system as a function of the selective pressures β_S and β_Q for the $q = l = 2$ case. The solid black lines indicate the slices A (at $\beta_S \simeq 1.2$) and B (at $\beta_S \simeq 17.5$) discussed in the text.

original $q = l = 2$ case and the case in which $q = 8$ and $l = 2$ are shown in Fig. 5.11 and Fig. 5.12 respectively.

For modularity, we observe that for large β_Q values, where Q has reached its plateau value and the drive to optimise modularity is dominant, the selective pressure driving for robustness plays no effect and $\Delta Q = 0$. A similar picture holds for low β_Q values, where a two-group core-periphery structure characterises the network topology in both the $q = l = 2$ and the $q = 8$, $l = 2$ cases. As the selective pressure driving for modularity increases, we observe slight deviations between the two curves, with the $q = 8$, $l = 2$ curve providing a marginally increased synergistic effect. In this region, the network topologies are characterised by a three-group structure in both the $q = l = 2$ and the $q = 8$, $l = 2$ cases. However, access to a third modularity group in the $q = 8$, $l = 2$ case slightly increases the synergistic effects. Overall, the two curves coincide throughout most of the range of β_Q values considered, the only notable exception being the presence of rapid oscillations in the $q = 8$, $l = 2$ case around the values of β_Q where both the modularity and the number of groups required to describe the network present a sharp increase. We suspect these oscillations to be the result of numerical artefacts arising during the free energy minimisation procedure.

The variation in the fraction of nodes S that belong to the giant connected component follows a similar pattern, where the two curves coincide in the region where a two or three-group structure is sufficient to adequately describe the network, but diverge as β_Q is increased. In particular, the access to a larger number of groups exacerbates the synergistic behaviour observed for high betas, and the interplay

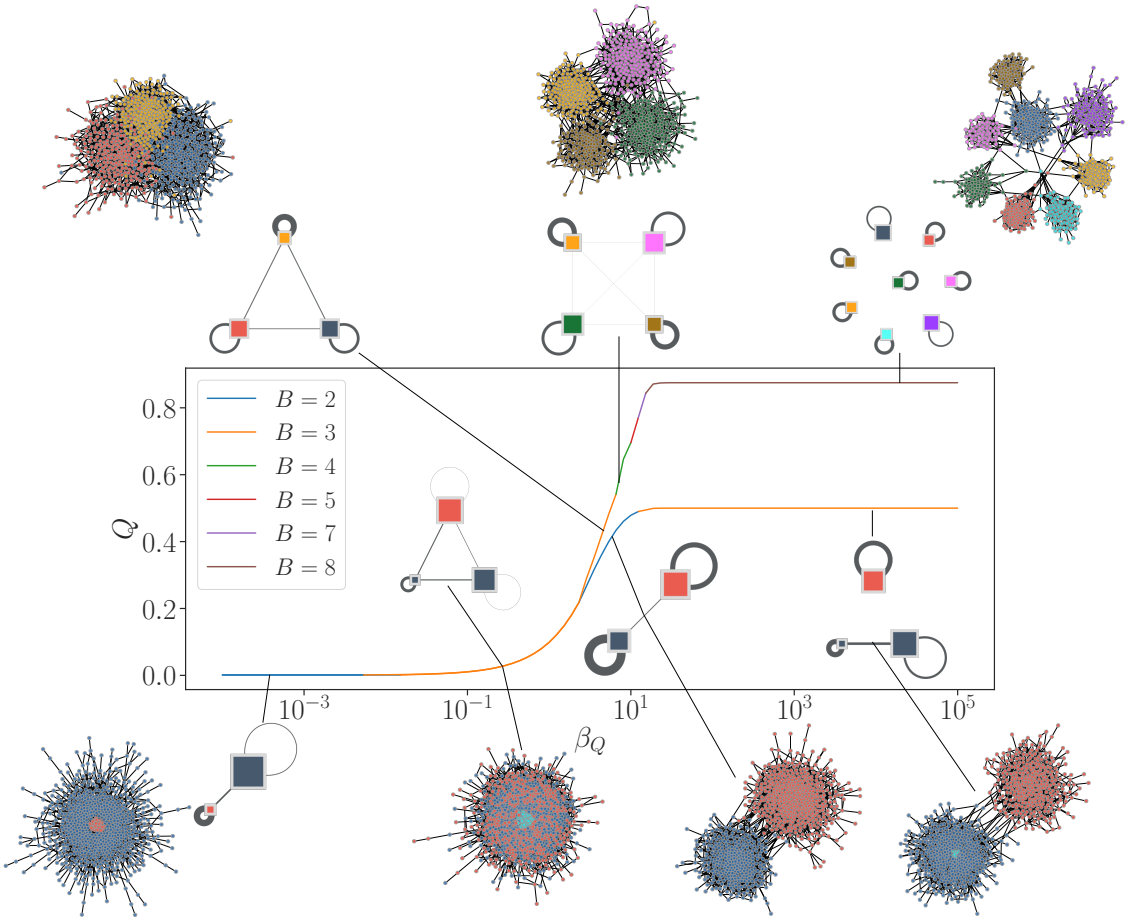


Figure 5.8: Modularity as a function of the selective pressure β_Q for slice A. The bottom curve displays the behaviour observed in the $q = l = 2$ case, while the top curve represents the $q = 8, l = 2$ case. Changes in colour indicate a change in the number of groups required to describe the system. Schematics of the optimized structures are shown in the insets, where each group corresponds to one of the B_S groups in our model and the colour of each group indicates its B_Q membership. Samples from the ensemble with average degree $\langle k \rangle = 5$ and edge dilution probability $\phi = 0.21$ are also shown.

between the two selective pressures proves even more beneficial.

From the definition of ΔS given in Eq. (5.4), we observe that ΔS measures the difference in the fraction of nodes that belong to the giant connected component between any state of the slice and the "first" state of the slice (where by first we intend the state at $\beta_Q = 0$), which we shall term the *reference* state. This leads to a series of interesting considerations. While structures such as the one in Fig. 5.12(c) (which we observe at high β_Q in the $q = l = 2$ case) do provide increased robustness when compared to the two-group core-periphery of our reference state [see Fig. 5.12(a)], the peak of the synergistic effects occurs at $\beta_Q \simeq 5$, where the network topology is characterised by an asymmetric modular structure comprised of two interconnected modules of different sizes, see Fig. 5.12(b). The reason that this modular structure can provide greater robustness than the core-periphery structure of the reference state can be understood by looking at Fig. 4.10. When $\beta_Q = 0$ and we are optimising solely for the robustness against random edge removal, the size

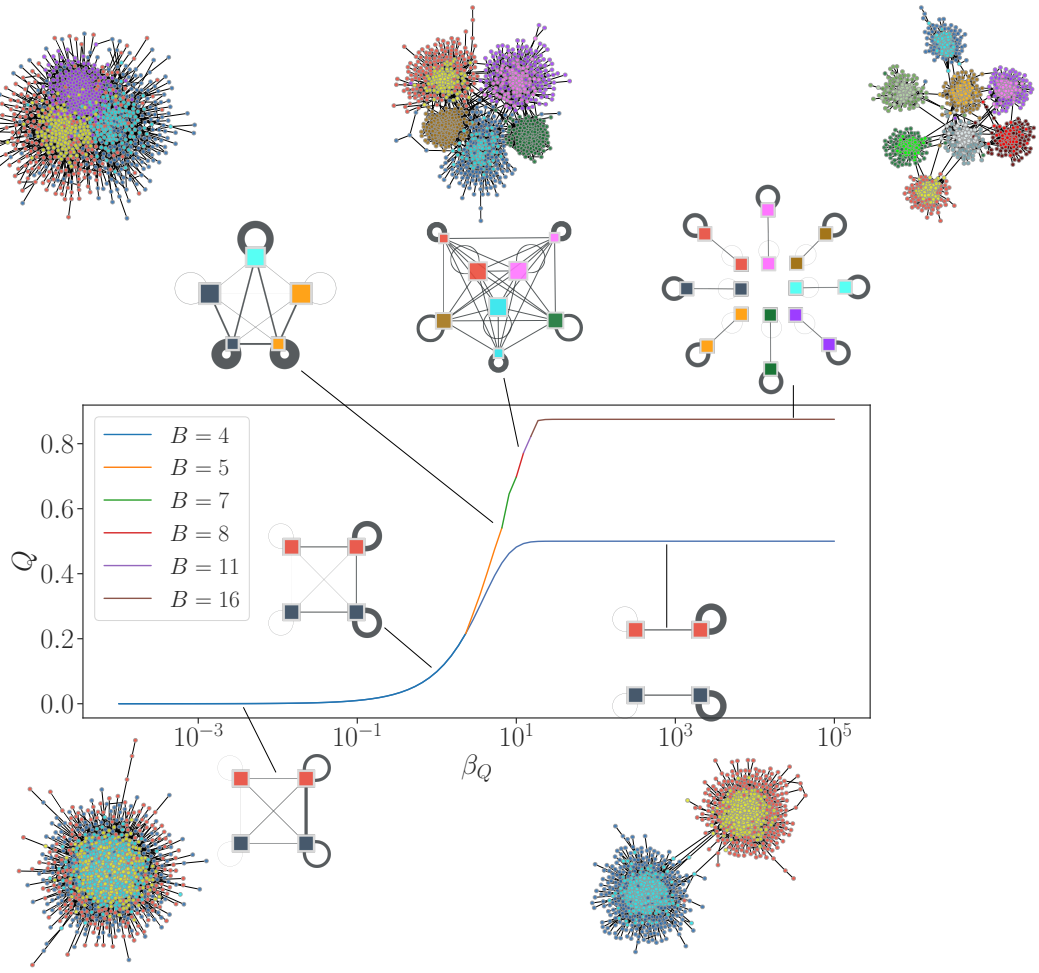
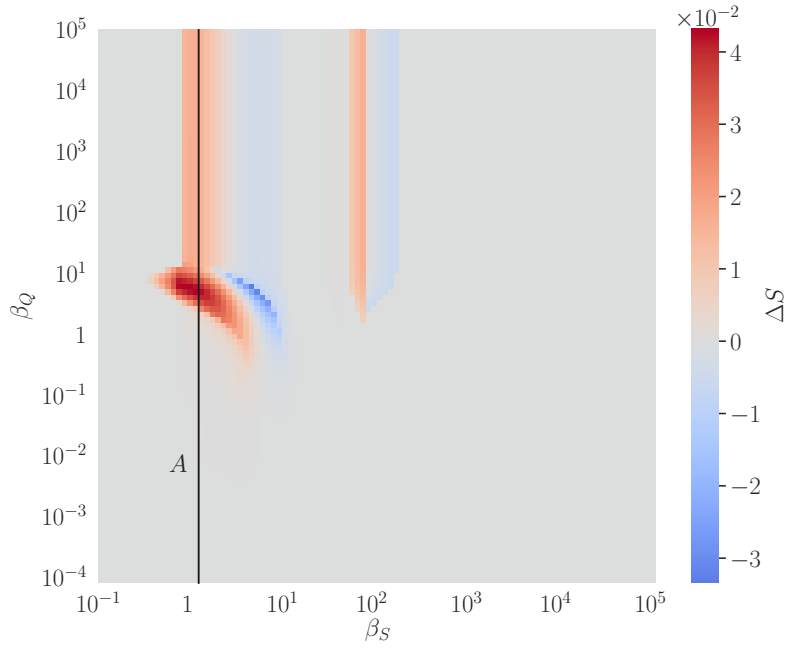
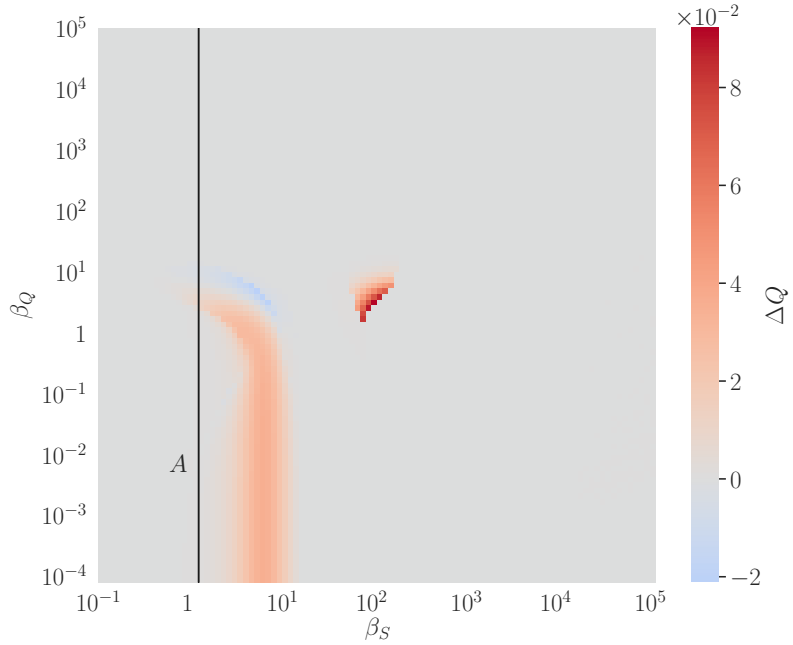


Figure 5.9: Modularity as a function of the selective pressure β_Q for slice B. The bottom curve displays the behaviour observed in the $q = l = 2$ case, while the top curve represents the $q = 8, l = 2$ case. Changes in colour indicate a change in the number of groups required to describe the system. Schematics of the optimized structures are shown in the insets, where each group corresponds to one of the B_S groups in our model and the colour of each group indicates its B_Q membership. Samples from the ensemble with average degree $\langle k \rangle = 5$ and edge dilution probability $\phi = 0.21$ are also shown.

of the core group at $\beta_S \simeq 1.2$ is infinitesimal. While this infinitesimal core does improve robustness over a completely random graph, the improvement is small, and two asymmetric interconnected modular structures can guarantee greater resilience. In the $q = l = 2$ case, as β_Q increases, the network structure becomes increasingly modular with fewer and fewer inter-group connections, and the robustness against random edge removal decreases until, eventually, we reach a point at which it is more beneficial for the network to split into three groups and replicate the core-periphery structure observed at $\beta_Q = 0$ [see Fig. 5.12(c)]. This behaviour helps shed some light on the asymmetric nature of the modular structures that arise as the result of the interplay between the two selective pressures. If we compare the fraction of nodes belonging to the giant connected component in our reference state with that of



(a)



(b)

Figure 5.10: Position of slice A in the ΔS (a) and ΔQ phase space (b). Slice A crosses a strongly synergistic region in the ΔS phase space for high β_Q while presenting only minimal effects in ΔQ.

perfectly symmetric modular structures such as the ones observed in Section 4.1, we can see that $S_{ref} > S_{\beta_S=0}$ for any β_Q value, as shown in Fig. 5.13. The asymmetric nature of the groups then plays a fundamental role in guaranteeing a certain degree of robustness against random edge removal. By breaking the symmetry of the modules, the network is able to mimic, to a certain degree, a core-periphery structure, with the smaller and more densely connected groups playing the role of cores. In turn,

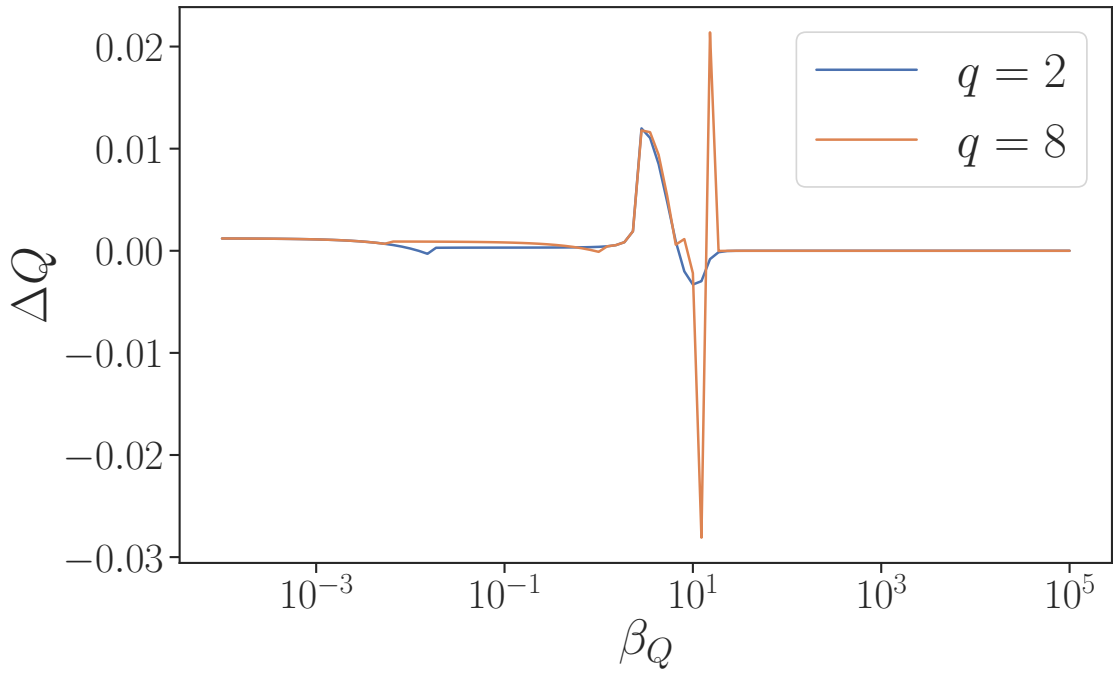


Figure 5.11: Variation in modularity ΔQ as a function of the selective pressure β_Q along slice A shown in Fig. 5.7 for both the case $q = l = 2$ and the case $q = 8$, $l = 2$.

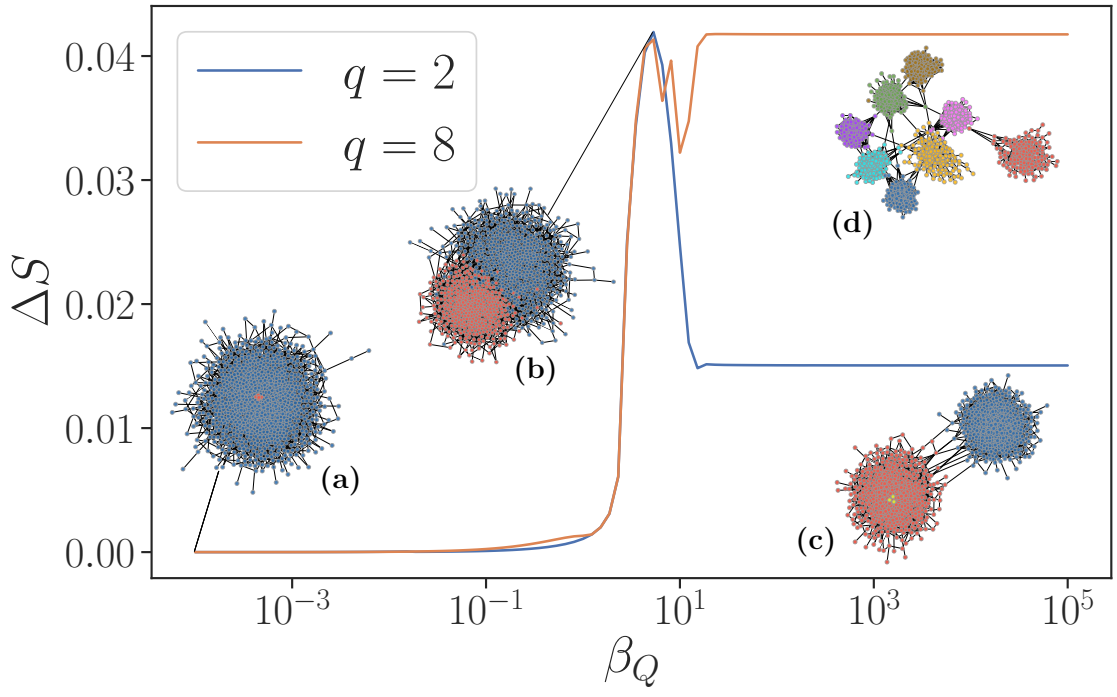


Figure 5.12: Variation in the fraction of nodes that belong to the giant connected component ΔS as a function of the selective pressure β_Q along slice A shown in Fig. 5.7 for both the case $q = l = 2$ and the case $q = 8$, $l = 2$. Samples from the network ensemble at different values of β_Q are shown in the insets.

this ensures a higher degree of robustness against random edge removal.

For the $q = 8$, $l = 2$ case, things are slightly different. While the initial peak in

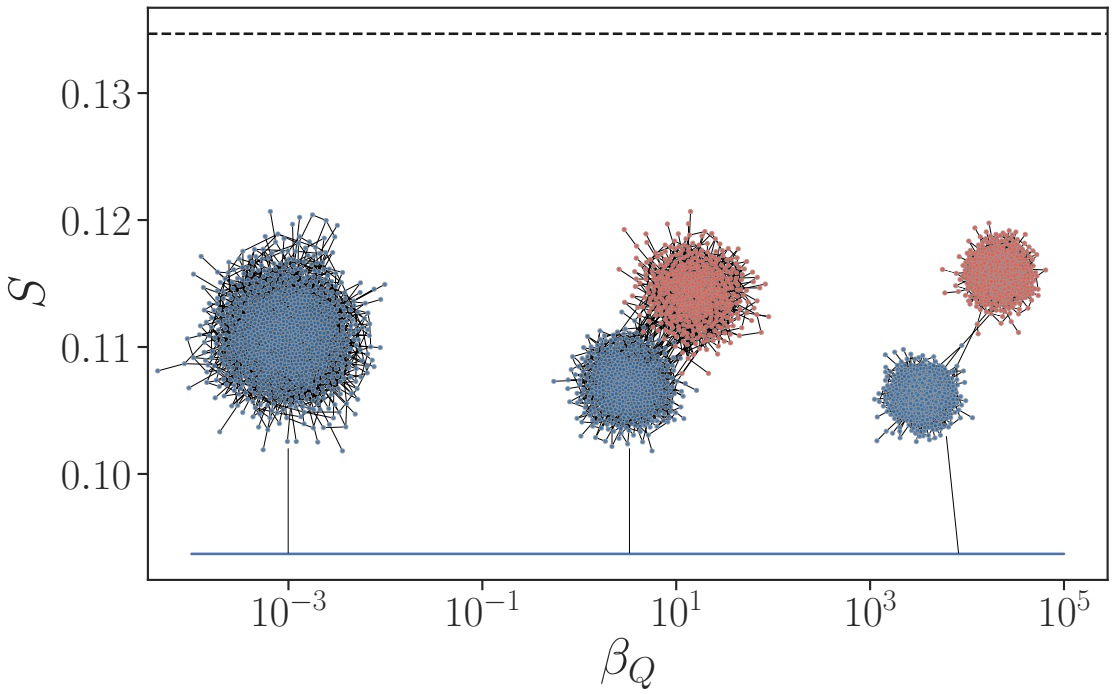


Figure 5.13: Fraction of nodes that belong to the giant connected component as a function of β_Q for states which are under no selective pressure to optimise against random edge removal (i.e. $\beta_S = 0$). The black dashed line indicates the fraction of nodes that belong to giant connected component for the reference state ($\beta_S \simeq 1.2, \beta_Q = 0$) discussed in the text. Notice that, as the nodes within each group are randomly distributed, there is no difference in robustness between a random graph and two weakly connected modular structures each of which randomly distributed.

ΔS is also reached at $\beta_Q \simeq 5$, the ability to access more modularity groups in the $q = 8$, $l = 2$ case allows the network to retain an asymmetric modular structure as the selective pressure increases by simply splitting into a larger number of groups. While the groups do become increasingly disconnected as β_Q increases, the presence of multiple weakly connected asymmetric modules is sufficient to guarantee a higher degree of robustness than our reference state. Indeed, even if the edges connecting two arbitrary groups are completely removed, these can still remain connected via paths traversing the other modules of the network.

Once more, we observe oscillations around the region in which the number of groups required to adequately describe the network rapidly increases.

Although we could not systematically probe the entire phase space due to computational constraints, further tests on single states display the same behaviour. Both the synergistic and antagonistic effects are amplified, but the regions in which these effects occur remain unchanged.

5.5 Discussion

This chapter has implemented the formalism introduced in Chapter 3 to study the more realistic scenario in which networks are driven to optimise for multiple fitness criteria simultaneously. Specifically, we have analysed the effects of interacting

selective pressures on the emerging network topologies.

We have focused on network ensembles characterised by pre-specified values of both modularity and robustness against random edge removal. By minimising the corresponding free energy, we have uncovered a rich phase space in which the network topology can transition between different entanglements of the previously encountered core-periphery, bipartite, and random graph structures depending on the values of the selective pressures. In particular, we have observed both symmetric structures characterised by a "mirroring" phenomenon where two separate core-periphery or bipartite structures are present, as well as asymmetric network topologies in which a core-periphery or bipartite structure is accompanied by an additional modular structure which accounts for the need to optimise modularity. Furthermore, we have identified regions of the phase space where synergistic or antagonistic effects can occur, such that optimising for one criterion aids or hinders optimising for the other.

As a final note, we have shown that if the maximum number of groups that can be potentially used to partition the network is increased, then different entanglements of the previous network topologies can arise. Namely, we observe the proliferation of asymmetric modular structures characterised by smaller and denser modules that are weakly connected to larger and sparser ones. This particular kind of topology guarantees high modularity values while also allowing for improved robustness with respect to the case in which no drive to optimise for modularity is present.

Chapter 6

Conclusions and future outlooks

6.1 Conclusions

In order to efficiently achieve their intended function, many real-world networks are often required to exhibit a certain degree of optimality with respect to one or more specific criteria. As the degree of efficiency of a particular network will depend, amongst other things, on its large-scale structure, this need for optimality naturally results in selective pressures driving the network towards particular network structures. The focus of this thesis has been to develop tools that allow for a principled analysis of the effects that various selective pressures can have on the emerging topologies. Specifically, we have introduced a framework that allows us to generate null models of optimal networks, which can incorporate the effects that selective pressures towards some predefined set of criteria can have on the structural properties of the network. This framework allows us to study these effects in an isolated manner, without additional interferences that might arise from other exogenous constraints or dynamical rules that contribute to the network formation process. Crucially, our framework can accommodate an arbitrary number of criteria, which allows us to analyse more realistic scenarios in which networked systems are subject to multiple interacting selective pressures.

We have applied this framework to analyse the emerging structures in systems subject to the joint optimisation for modularity and robustness against random removal of edges, which we analysed both in isolation and in combination. In the case of modularity alone, we showed that by increasing the selective pressure, we observe network structures that progressively split into an increasing number of symmetric groups whose nodes predominantly connect amongst themselves. In the case of robustness against random failures, we instead identify two phase transitions in which the network structure transitions first to a core-periphery pattern and then into an asymmetric bipartite one. The core-periphery structure is characterised by a smaller and denser set of "core" nodes that connect preferentially amongst themselves and a larger "periphery" whose nodes mostly connect to the core nodes. This structure allows for higher robustness as the random removal of any edge is unlikely to disconnect the core, and peripheral nodes remain connected via the core itself. Increasing the selective pressure further, the core group eventually becomes so dense that its nodes no longer require to preferentially connect amongst each other to ensure a high level of robustness. They instead connect predominantly to the periphery, and we observe an asymmetric bipartite structure.

By combining both fitness criteria, we observed different combinations of the above structures, where the core-periphery and bipartite structures can either appear in duplicate (i.e. we observe two symmetric core-periphery or bipartite structures) or accompanied by an additional cluster which ensures high modularity values. Notably, we observed regions of the parameter space where the interplay between the selective pressures can have synergistic or antagonistic effects, and optimising for a specific characteristic can facilitate or hinder optimising for the other.

Our results show how the interaction between different selective pressures can be combined in simple network models, offering a platform to investigate the effects that different fitness criteria can have on the emerging network structures. Understanding these effects can help us better elucidate the interplay between network structure and function and prove crucial for designing networks aimed at optimising one or more predefined tasks or properties. Furthermore, identifying structures that optimise specific external criteria may help us understand how networks evolve when these external conditions are changed, providing possible insights into network resilience and failure processes.

6.2 Future outlooks

The introduced framework is flexible enough to allow for a wide range of extensions and modifications. In what follows, we present some of the possible extensions that can be explored.

6.2.1 Different optimisation criteria

The consideration of different sets of optimisation criteria naturally emerges as the most immediate modification. Although the choice of the criteria is arbitrary, the most direct extensions related to the work carried out throughout this thesis are

Node percolation:

As we have seen in Section 2.3.1, the formalism describing edge percolation can easily be extended to account for the removal of nodes instead (*node percolation*). Comparing the structures that emerge as a result of edge and node percolation is interesting in its own right, as it is not assured that the observed behaviour should remain the same.

Targeted attacks:

Throughout this thesis, we have analysed the most likely large-scale structures that emerge when a network is driven to optimise its robustness against random edge removals, but different "attack" protocols can be considered. One example is to consider a *non-uniform* removal of nodes (or edges), i.e. a percolation process in which not all nodes (or edges) have the same probability of being removed from the network. Effectively, this amounts to considering the case of targeted (as opposed to random) attacks on a network. Generally, this is achieved by considering occupation probabilities for each node that depend on some pre-specified characteristic. The degree of a node, for example, is a commonly studied choice, in which the occupation

probability ϕ_k depends on the degree, k, of the node, but other choices are possible. One could consider different occupation probabilities for nodes belonging to different groups in the network or a combination of both node degree and group membership. By varying our choice of the occupation probability, a host of complex removal patterns may be examined.

The optimal structures that emerge when considering a non-uniform removal of edges are known to be different than what is obtained by considering a uniform removal [40]. However, real-world networks are often under pressure to optimise (or at the very least look for the best compromise in robustness) both against random failures and targeted attacks. Our framework can readily capture the interplay between these two criteria, providing a basis to further study the emerging topologies.

More generally, it would be interesting to apply the formalism developed in this report to different sets of evolutionary constraints to determine what structures dominate the Pareto front (thus providing the best trade-off between the various selective pressures) and evaluate whether any recurring structures occur across different criteria.

6.2.2 Arbitrary degree distributions and other SBM variants

The framework presented in this thesis is centred around the stochastic block model. While the SBM allows us to gain valuable insights into how optimisation processes can drive network formation, it also presents some limitations. As mentioned in Section 2.2.3, one of the most significant drawbacks of the traditional stochastic block model lies in its inability to capture heterogeneous degree distributions such as those commonly observed in real-world networks. To circumvent this problem, Karrer and Newman introduced the *degree-corrected stochastic block model* (DC-SBM) [75], which can, in principle, account for any arbitrary degree distribution of the nodes. In what follows, we briefly outline how this extension can be incorporated into our modelling framework and analyse some of its advantages and limitations.

Poisson stochastic block models

Section 2.2.3 introduced the standard stochastic block model as a generative network model for simple graphs. We have seen that networks are generated so that each entry of the adjacency matrix, A_{ij}, follows a Bernoulli distribution. For ease of calculation, it is often convenient to introduce a multi-graph version of the stochastic block model in which, rather than placing a single edge between any pair of nodes, we place a Poisson-distributed number of edges with mean λ_{rs} depending only on the group membership of the nodes. The probability of observing a network in the ensemble is then given by

$$P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{b}) = \prod_{i<j} \frac{\lambda_{b_i b_j}^{A_{ij}} e^{-\lambda_{b_i b_j}}}{A_{ij}!} \times \prod_i \frac{\left(\frac{1}{2}\lambda_{b_i b_j}\right)^{A_{ii}/2} e^{-\frac{1}{2}\lambda_{b_i b_i}}}{(A_{ii}/2)!}, \tag{6.1}$$

where, as before, the diagonal entries of the adjacency matrix, A_{ii}, are equal to twice the number of self-edges from node i to itself. In principle, the Poisson SBM allows

both self and multiple edges to exist between nodes, which is unrealistic with respect to most real-world networks. However, in the sparse regime we are interested in, this model can be shown to differ only negligibly from the original Bernoulli model [79, 114]. Intuitively, this is clear, as in the sparse regime $p_{rs} = \lambda_{rs} = \mathcal{O}(N^{-1})$ with $N \gg 1$, the probability of having a multi-edge is small. Consequently, multi-edges can be neglected (a similar picture holds for self-edges). There is then little difference in the networks generated by the two models in the sparse limit.

The degree-corrected stochastic block model

The Poisson SBM introduced above suffers the same problematics as its Bernoulli counterpart. Namely, all nodes within the same group will have the same degree distribution. Moreover, as the degrees of the nodes correspond to sums of independent Poisson variables, the degree distributions of the groups will be Poisson. Consequently, the SBM will tend to avoid placing nodes of very different degrees within the same group.

The DC-SBM circumvents this limitation by incorporating degree heterogeneity within the generating mechanism. To achieve this, each node is assigned an additional parameter, θ_i, which controls its expected degree independently of the node's group membership. The DC-SBM then generates networks with probability

$$P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{\theta},\boldsymbol{b}) = \prod_{i<j} \frac{(\theta_i\theta_j\lambda_{b_ib_j})^{A_{ij}} e^{-\theta_i\theta_j\lambda_{b_ib_j}}}{A_{ij}!} \times \prod_i \frac{\left(\frac{1}{2}\theta_i^2\lambda_{b_ib_i}\right)^{A_{ii}/2} e^{-\frac{1}{2}\theta_i^2\lambda_{b_ib_i}}}{(A_{ii}/2)!}. \tag{6.2}$$

By varying the parameters theta, we can achieve (in expectation) any desired degree sequence. Note that by setting $\theta_i = 1,\ \forall\ i$, we recover the Poisson SBM introduced above.

θ_i and $\lambda_{b_ib_j}$ appear multiplied in Eq. (6.2); this means that they can be arbitrarily rescaled provided their product remains unchanged. Correct identification of the parameters then requires a constraint. A common choice is to require that $\sum_i \theta_i\delta_{b_i,r} = 1$, where $\delta_{x,y}$ is the Kronecker delta. With this choice, the parameters acquire a straightforward interpretation, where λ_{rs} corresponds to the expected number of edges between groups r and s, and θ_i is the probability that an edge connecting to the group to which node i belongs lands on node i.

Given Eq. (6.2), we can evaluate the entropy of the DC-SBM, which, in the sparse regime and with our previously introduced parametrisation, is given by

$$\Sigma(\boldsymbol{\kappa},\boldsymbol{m},\boldsymbol{b}) = \sum_\kappa N_\kappa\kappa\ln\kappa - \frac{N\langle k\rangle}{2}\sum_{rs} m_{rs}\ln\left(\frac{m_{rs}}{m_rm_s}\right), \tag{6.3}$$

where N_κ is the number of nodes with expected degree κ and we have disregarded constant terms. By incorporating this entropy into our framework, we can capture the role that connectivity plays in optimising pre-specified characteristics.

Note that the model described above for the degree-corrected SBM imposes the degree sequence in expectation. This is standard in many settings in network science. As we mentioned in Section 2.2.1, we are often interested in the typical behaviour of networks rather than the behaviour of one particular instance. Fixing the expected value of the degree sequence then allows us to generate networks where the actual degrees of nodes can vary, only their mean value being set. However, in our context,

where we aim to characterise the role that network connectivity plays in optimising specific characteristics, it might be more beneficial to consider the imposed degree sequence as fixed rather than defined only on expectation. This can be achieved by fixing both the degree k_i of each node as well the number e_{rs} of edges between groups r and s. Then, for each edge required to fall between r and s, one picks two stubs at random (one from group r and one from group s) and connects them. This model defines a network ensemble consisting of all possible configurations satisfying a given degree sequence where the number of edges between groups has been fixed and is a direct extension of the configuration model. This ensemble has been extensively studied by Peixoto [65, 78, 83, 115] and is known as the *microcanonical* stochastic block model.

The parameters of the microcanonical SBM are the partition $\boldsymbol{b} = \{b_i\}$ of nodes into B groups, the degree sequence $\mathbf{k} = \{k_i\}$, and the matrix of edge counts between groups $\mathbf{e} = \{e_{rs}\}$, where e_{rs} is the number of edges between groups r and s. Given these parameters, the probability of observing a particular network can be calculated as follows. Let $\Omega(\mathbf{e})$ be the number of configurations (pairings) that respect the edge constraints e_{rs}. As with the configuration model, many of these pairings will correspond to the same network. Let then $\Xi(\boldsymbol{A})$ be the number of configurations corresponding to a particular graph $\boldsymbol{A}$. The probability of observing $\boldsymbol{A}$ given the model parameters is then simply the ratio between $\Xi(\boldsymbol{A})$ and $\Omega(\mathbf{e})$,

$$P(\boldsymbol{A}|\boldsymbol{k}, \boldsymbol{e}, \boldsymbol{b}) = \frac{\Xi(\boldsymbol{A})}{\Omega(\boldsymbol{e})}. \tag{6.4}$$

Let us compute $\Xi(\boldsymbol{A})$ and $\Omega(\boldsymbol{e})$. Since our model allows for multiple and self-edges, then for a given adjacency matrix $\boldsymbol{A}$, the number of matchings $\Xi(\boldsymbol{A})$ compatible with it is the same as in the configuration model introduced previously,

$$\Xi(\boldsymbol{A}) = \frac{\prod_i k_i!}{\prod_{i<j} A_{ij}! \prod_i A_{ii}!!}. \tag{6.5}$$

For $\Omega(\boldsymbol{e})$, the computation is similar to the one for $\Xi(\boldsymbol{A})$. Indeed, the number of configurations that respect the constraints e_{rs} will be the same as $\Xi(\boldsymbol{A})$, but with the matrix of edge counts taking the place of the adjacency matrix and the groups playing the role of nodes,

$$\Omega(\boldsymbol{e}) = \frac{\prod_r e_r!}{\prod_{r<s} e_{rs}! \prod_r e_{rr}!!}. \tag{6.6}$$

We then have,

$$P(\boldsymbol{A}|\boldsymbol{k}, \boldsymbol{e}, \boldsymbol{b}) = \frac{\prod_{r<s} e_{rs}! \prod_r e_{rr}!! \prod_i k_i!}{\prod_r e_r! \prod_{i<j} A_{ij}! \prod_i A_{ii}!!}. \tag{6.7}$$

Taking the logarithm of Eq. (6.7), we can compute the entropy of the model [65] as

$$\Sigma(\boldsymbol{k}, \boldsymbol{e}, \boldsymbol{b}) = -E - \frac{1}{2}\sum_{rs} e_{rs} \ln \frac{e_{rs}}{e_r e_s} - \sum_k N_k \ln k!. \tag{6.8}$$

If we allow each group in the network to have its own degree distribution p_k^r, representing the probability of observing a node of degree k in group r, then Eq. (6.8) can be further re-written as

$$\Sigma(\boldsymbol{e}, \boldsymbol{p}, \boldsymbol{b}) = -E - \frac{1}{2}\sum_{rs} e_{rs} \ln \frac{e_{rs}}{e_r e_s} - \sum_r n_r \sum_k p_k^r \ln k! - \sum_r n_r \sum_k p_k^r \ln p_k^r, \tag{6.9}$$

where n_r is the number of nodes in group r, and we have now included the entropy of the degree distributions of the individual groups. With our parametrisation and disregarding constant terms the entropy becomes

$$\Sigma(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{p}) = -\frac{\langle k \rangle N}{2} \sum_{rs} m_{rs} \ln \frac{m_{rs}}{m_r m_s} - \sum_r \omega_r \sum_k p_k^r \ln k! - \sum_r \omega_r \sum_k p_k^r \ln p_k^r. \tag{6.10}$$

Using Eq. (6.10) as our ensemble entropy, the free energy can be written as

$$\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{p}) = -\sum_i \beta_i R_i(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{p}) - \Sigma(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{p}). \tag{6.11}$$

which we can then proceed to minimise as before.[1]

Incorporating the degree distribution of the network directly as a model parameter allows us to explore the role that network connectivity plays in optimising specific criteria. However, while including the degree distributions of the groups into our modelling framework is straightforward, doing so introduces additional complications. Namely, we must ensure that the constraint $\sum_k k p_k^r = (m_r/\omega_r)\langle k \rangle$ holds at all times. This requirement cannot be easily converted to an unconstrained optimisation problem, and constrained optimisation techniques would probably need to be employed. Another issue that must be addressed is how to compute the infinite summations over the degrees k, as in most cases, it will not be possible to solve them analytically. A possible solution could be to impose a cut-off value k_r^{max} representing the maximum allowed degree of nodes in group r. The underlying idea is that, for large enough k_r^{max}, our framework should prove flexible enough to capture a rich collection of degree distributions. Note, however, that the inclusion of these kB additional parameters used to model the group degree distributions would, invariably, slow down the optimisation procedure. As we have pointed out before and shall see in the following section, speed is one of the key limitations of our framework, and further work to improve its efficiency would be advisable.

Further extensions centred around other variants of the SBM are also possible and would allow us to capture further variability in the network structure. Examples include mixed-membership models, weighted and directed networks, triadic closure, and networks annotated with metadata. However, it is important to note that while all of these extensions could potentially allow us to model increasingly complex scenarios and analyse the role played by different network characteristics in optimising a set of pre-specified features, there is a large class of descriptors for which the model presented in this Thesis is sufficient. Modularity is a prime example. As Q is a function solely of the block parameters (specifically the affinity matrix m), no other network characteristic such as degree distribution or triadic closure will alter the results. Robustness (as measured by the fraction of nodes, S, in the giant connected component) is in some ways similar. Although not as straightforward as modularity, and although it is true that the degree distributions

[1] The use of the microcanonical SBM over the canonical one is an important modelling choice as, in general, the two ensembles are known not to be equivalent, meaning that they will generate different networks with different probabilities even in the sparse limit with $N \gg 1$. We refer the reader to [65, 78] for a more in-depth discussion on the equivalence of these two ensembles.

of the groups could play a role in determining the overall robustness of the network, there are not many other extensions that could significantly come into play. For example, considering triadic closure would be pointless, as any edge used to close a triangle would simply be connecting nodes which are already connected. Furthermore, even in the case of the degree corrected SBM, there could still be vast swaths of the phase space in which the degree distribution remains a Poisson. To see why this is so, we can explicitly write down the minimisation procedure. Our problem amounts to minimising Eq. (4.23), subject to the following constraints,

$$\begin{cases} \sum_r \omega_r = 1, & \text{(6.12a)} \\ \sum_{rs} m_{rs} = 1, & \text{(6.12b)} \\ \sum_k p_r(k) = 1, & \text{(6.12c)} \\ \sum_k k p_r(k) = \dfrac{m_r}{\omega_r}\langle k\rangle, & \text{(6.12d)} \end{cases}$$

where Eqs. (6.12a) - (6.12c) enforce a correct normalisation of the parameters, and Eq. (6.12d) ensures that the average degree of the single groups is well defined. We can then construct the following Lagrangian

$$\begin{aligned} \Lambda(\boldsymbol{\omega}, \boldsymbol{m}, \boldsymbol{p}) = &-\beta R + \frac{\langle k\rangle}{2}\sum_{rs} m_{rs}\ln\frac{m_{rs}}{m_r m_s} + \sum_r \omega_r \sum_k p_k^r \ln k! \\ &+ \sum_r \omega_r \sum_k p_k^r \ln p_k^r - \sum_r \xi_r \left(\sum_k p_k^r - 1\right) \\ &- \sum_r \mu_r \left(\sum_k k p_k^r - \frac{m_r}{\omega_r}\langle k\rangle\right) - \gamma\left(\sum_r \omega_r - 1\right) - \eta\left(\sum_{rs} m_{rs} - 1\right). \end{aligned} \tag{6.13}$$

The corresponding saddle point equations with respect to the degree distributions and group sizes are given by

$$\begin{cases} \nabla_{p_k^r}\Lambda = -\beta\nabla_{p_k^r}R + \omega_r \ln k! + \omega_r\left[\ln p_k^r + 1\right] - \xi_r - \mu_r k = 0, & \text{(6.14)} \\ \nabla_{\omega_r}\Lambda = -\beta\nabla_{\omega_r}R + \sum_k p_k^r \ln k! + \sum_k p_k^r \ln p_k^r - \dfrac{\mu_r m_r}{\omega_r^2}\langle k\rangle - \gamma = 0, & \text{(6.15)} \end{cases}$$

while for the edge parameters, we obtain two equations, one for the diagonal terms of the affinity matrix and one for the off-diagonal entries.

$$\begin{cases} \nabla_{m_{rr}}\Lambda = -\beta\nabla_{m_{rr}}R + \dfrac{\langle k\rangle}{2}\left(\ln\dfrac{m_{rr}}{m_r^2} - 1\right) + \langle k\rangle\dfrac{\mu_r}{\omega_r} - \eta = 0 & \text{(6.16)} \\ \nabla_{m_{rs}}\Lambda = -\beta\nabla_{m_{rs}}R + \langle k\rangle\left(\ln\dfrac{m_{rs}}{m_r m_s} - 1\right) + \langle k\rangle\left(\dfrac{\mu_r}{\omega_r} + \dfrac{\mu_s}{\omega_s}\right) - 2\eta = 0 & \text{(6.17)} \end{cases}$$

Eqs. (6.14) - (6.17) give us a set of coupled non-linear differential equations that we can attempt to solve for the model's parameters. While no general closed-form solution to these equations exists, they can be solved analytically for $\beta = 0$, resulting,

as expected, in a random graph structure. Specifically, for any partitioning of the network into B groups of arbitrary sizes $\omega_i,\ i \in [1, B]$, the free energy minimisation leads to an affinity matrix such that the connections between groups are given by $m_{rs} = \omega_r \omega_s$. This leads to groups with Poisson degree distributions and average degree $\kappa_r = \langle k \rangle,\ \forall r$.

$$\begin{cases} \omega_r = \dfrac{1}{B}, \\ m_{rs} = \dfrac{1}{B^2}, \\ p_r(k) = \dfrac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}. \end{cases} \tag{6.18}$$

This result suggests that, at least for small to moderate values of the selective pressure, the degree distributions of the emerging topologies might still be Poisson, as deviating significantly from a Poisson distribution might incur in high entropic costs.

Although this is merely a hypothesis, which would have to be verified, it does appear to be in line with previous findings. For example, in [40], Peixoto and Bornholdt, analysing the resilience of interdependent networks to random failures and intentional attacks, found a simple core-periphery topology, similar to the ones observed in this Thesis, to be optimal against random failures. No other features often observed in real-world systems, such as scale-free distributions, emerged as a result of needing to optimise robustness. Similarly, in [39], Schneider et al. found that a highly assortative "onionlike" topology, in which nodes of similar degree connect preferentially amongst each other, emerged as the most robust large-scale structure against targeted attacks. This "onionlike" topology is qualitatively equivalent to a stochastic block model, as nodes with the same degree are statistically indistinguishable.

Finally, it is worth mentioning that even this simple model based around the traditional SBM is already surprisingly complicated and worth understanding better.

6.2.3 Speed and computational complexity

One of our proposed framework's main limitations is its speed, as seen in Section 5.4 when attempting to increase the number of groups. In what follows, we briefly overview what makes our approach slow and give a few ideas on how this issue might be tackled to improve computational efficiency.

In general, the standard way of evaluating the computational complexity of an algorithm is via asymptotic or Big-O notation [116]. While useful in compiled languages such as C, it might be somewhat misleading in our setting. Our code makes heavy use of the NumPy library [117] and, as a result, is a mix of interpreted Python and compiled code. This can raise issues, as interpreted code can be orders of magnitude slower than its compiled counterpart, meaning that two operations with the same Big-O complexity can have drastically different running times. Nevertheless, even disregarding these implementation differences, Big-O complexity is sufficient to understand what makes our code slow.

The principal issue slowing down the code is that the free energy computation is $O(B^2)$ (or $O(B_S^2)$ in the case of multiple fitness criteria), making it highly inefficient

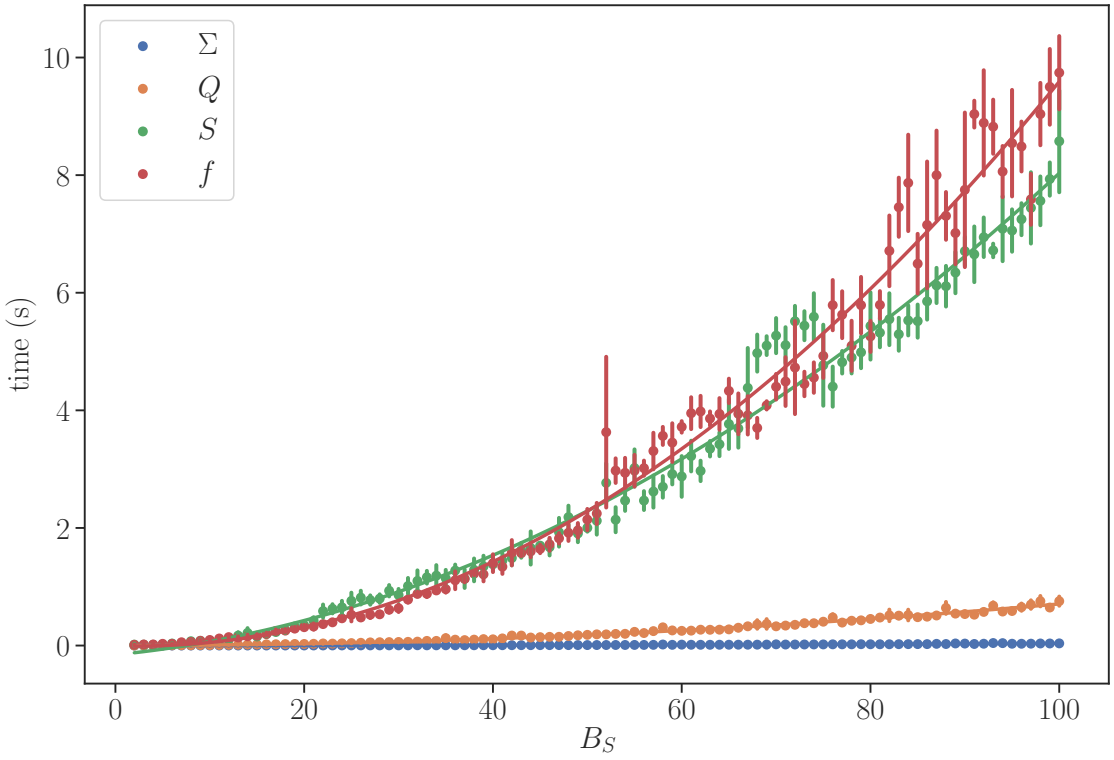


Figure 6.1: Average running times as a function of the system size for the entropy, Σ, modularity, Q, fraction of nodes in the GCC, S, and the free energy, f. Averages and standard deviations are computed over five different network realisations. Regression lines are drawn as a visual guide.

as the number of groups is increased. However, this is only part of the picture, as Big-O notation discards constant terms, which can have significant effects in practice when running code. Moreover, in our particular case, these effects are compounded by the fact that interpreted code is generally slower than compiled code. To better understand what is slowing down our code, we can consider separately the times required to evaluate the modularity, Q, the fraction of nodes in the giant component, S, and the entropy, Σ, as a function of the number of groups B_S[2]. Although all three quantities can be shown to be $O(B^2)$, they can have significantly different running times, as can be seen in Fig. 6.1.
It is straightforward to understand why S is dominating the times required to compute the free energy. As we saw in Section 4.2.2, in order to compute S, we must solve Eq. (4.15) by repeated iterations. Eq. (4.15) requires us to compute a sum over all groups s and to do so once for each group r. It follows that each iteration has a computational cost of $O(B_S^2)$, meaning that the total complexity involved in solving Eq. (4.15) is $O(n_{iter} \cdot B_S^2)$, where n_{iter} is the number of iterations required for Eq. (4.15) to converge. Formally, n_{iter} is a constant and, as such, can be neglected from our Big-O calculations. However, from a practical standpoint, n_{iter} can grow to be quite large (around 250 iterations for some values of β_S) and significantly impact the running times.

[2]In keeping with the main text, we increase B_S by keeping l fixed to two and increasing q so that, as B_S grows, so does B_Q

To make matters worse, we cannot, in general, compute the gradient of $\mathcal{F}$ analytically and must resort to numerical optimisation techniques to perform the minimisation. In our particular case, we have used the L-BFGS-B algorithm, whose complexity can be shown to scale as $O(mn)$ per iteration [118], where n is the input size and m is a user-defined constant [3]. Note that this is the cost per iteration, not per function evaluation (which can be evaluated multiple times per iteration). It is clear then that as the number of groups used to model the network increases, the evaluation and minimisation of the free energy function can become desperately slow.

Despite these difficulties, some steps can be taken to improve the situation. Utilising automatic differentiation is already a good step, as it allows us to avoid potentially costly computations to estimate the gradients during the minimisation process (forward finite differences, for example, would require $O(n)$ function evaluations per gradient estimation). Another potentially computationally efficient step to take is the implementation of just-in-time compilation, which allows to pre-compile part of the interpreted Python code at runtime and can speed up code almost to the levels of C or FORTRAN. However, the biggest hurdle in computational cost still boils down to evaluating and minimising the free energy function. Improving the evaluation of the free energy is, to a certain extent, ill-defined, as the free energy will depend on the evaluation of the chosen fitness criteria, which are arbitrary. The procedure would still benefit from an algorithm capable of evaluating the entropy faster than $O(B^2)$. However, as seen in Fig. 6.1, the entropy computation is negligible with respect to other quantities of interest, at least along a wide range of B values.

Exploring alternative minimisation procedures is another option. While the L-BFGS-B is a state-of-the-art optimisation technique with a time complexity per iteration which scales linearly with the input size, it has recently been argued that it might not be best suited for black-box optimisation problems in which information about the gradients is absent. As gradient information is unavailable, it would have to be estimated by potentially costly methods such as finite differences. Furthermore, L-BFGS computes an approximation of the Hessian inverse starting from m previously-stored gradients. In the case of a large-scale black-box function, the information provided by only m gradients might be small, and the gradient estimations imprecise. In turn, this can scale up the runtimes in the number of function evaluations by a factor of n [119]. As such, it might be worth exploring whether derivative-free algorithms might perform better. Other approaches to reduce the computational complexity of the optimisation process are also possible. For example, O-LBFGS [120, 121] presents an online generalisation of L-BFGS similar to stochastic gradient descent, which reduces the computational complexity by evaluating, at each iteration, the error and gradients on a random subset of the dataset.

[3] As mentioned in Section 4.1.2, L-BFGS-B is a quasi-Newton method that works by approximating the Hessian inverse to guide the search through the phase space. However, computing and storing an $n \times n$ matrix (as in the original BFGS algorithm) can be computationally expensive. The L-BFGS and L-BFGS-B algorithms circumvent this limitation by only storing the last m computed gradients and positions and then constructing the Hessian inverse approximation from these quantities. The constant m determines how many gradients to store in memory. See [118] for more information on the L-BFGS and L-BFGS-B algorithms.

6.2.4 Network inference

As mentioned briefly in Section 2.2.3, stochastic block models have found widespread use as means to infer modular structures in empirical networks [79]. One of the fundamental reasons for their use in this context is their generative nature, which allows for principled inferential approaches based around Bayesian techniques. The framework presented in this Thesis shares the same parametrisation as these models. As such, we expect that the two approaches can eventually be combined, allowing us to identify the dominant driving mechanisms of network formation directly from network data. Furthermore, framing the presented model in an inference setting would open up opportunities in areas as diverse as making predictions that generalise from past observations, identify errors and omissions in data, and provide possibilities for architectural improvements. However, extending the framework to an inferential setting can present some challenges. In what follows, we briefly introduce the problem of inferring modular structures in networks and highlight the main challenges that must be addressed if one is to extend our modelling framework to an inferential setting.

Network inference

Network inference, or a posteriori blockmodelling, is the task of fitting block models to network data in an attempt to uncover the large-scale structure characterising the network. More specifically, the inferential task consists of finding the most likely partition, $\boldsymbol{b}$, that gave rise to an observed network $\boldsymbol{A}$. This can be achieved by resorting to Bayes' rule,

$$P(\boldsymbol{b}|\boldsymbol{A}) = \frac{P(\boldsymbol{A}|\boldsymbol{b})P(\boldsymbol{b})}{P(\boldsymbol{A})}, \tag{6.19}$$

where $P(\boldsymbol{A}|\boldsymbol{b}) = \int d\boldsymbol{\xi} P(\boldsymbol{A}|\boldsymbol{\xi}, \boldsymbol{b})P(\boldsymbol{\xi}|\boldsymbol{b})$ is the marginal likelihood integrated over the remaining model parameters, $P(\boldsymbol{A})$ is known as the evidence and acts as a normalisation constant, and $P(\boldsymbol{\xi}|\boldsymbol{b})$ and $P(\boldsymbol{b})$ are known as prior distributions and encode our beliefs about the block partition before we have observed any data. By maximising this posterior distribution, we can find the most likely partition given the observed network.

Prior selection is particularly important in Bayesian inference, as they affect the shape of the posterior and, therefore, the results of the inferential procedure. Ideally, one would choose the priors based on previous observations. However, this is typically not possible for networks as they are often singletons (i.e. they are unique objects rather than coming from a population). A common approach is to be maximally agnostic about the possible partitions and take the prior $P(\boldsymbol{b})$ as a constant, meaning that all possible partitionings are, a priori, equally likely. With this choice of priors, we have that the posterior is proportional to the likelihood, so that maximising the posterior is equivalent to maximising the likelihood with respect to the model parameters,

$$\hat{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\operatorname{argmax}}\, P(\boldsymbol{A}|\boldsymbol{b}). \tag{6.20}$$

This approach is known as *maximum likelihood estimation* (MLE). In most practical situations, it is easier to maximise the logarithm of the likelihood, $\log P(\boldsymbol{A}|\boldsymbol{b})$, known as the *log-likelihood*. Note that, since the logarithm is a monotonically increasing

function of its argument, the log-likelihood leaves the position of the maximum unchanged.

The first challenge we are faced with when attempting to extend our framework to an inferential setting is that we do not have a clear expression $P(\boldsymbol{A}|\boldsymbol{p},\boldsymbol{b},\boldsymbol{\beta})$ for our likelihood. Indeed, our framework does not specify a particular generating mechanism (throughout this work we have used the stochastic block model, but, in principle, any generating mechanism can be used so long as we can compute its entropy), rather, it employs the maximum entropy principle to generate null models for optimised network structures. It is then not immediately clear how to formulate a likelihood. However, if we consider the log-likelihood of the stochastic block model, we can observe that it has an interesting information-theoretic interpretation. For simplicity, let us work with the Poisson SBM introduced in Section 6.2.2, which generates networks with probability

$$P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{b}) = \prod_{i<j}\frac{\lambda_{b_ib_j}^{A_{ij}}e^{-\lambda_{b_ib_j}}}{A_{ij}!}\times\prod_i\frac{\left(\frac{1}{2}\lambda_{b_ib_j}\right)^{A_{ii}/2}e^{-\frac{1}{2}\lambda_{b_ib_i}}}{(A_{ii}/2)!}. \tag{6.21}$$

We can write its log-likelihood as

$$\begin{aligned}\log P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{b}) = &\sum_{i<j}A_{ij}\log\lambda_{b_ib_j}-\lambda_{b_ib_j}-\log A_{ij}!\\ &+\sum_i\frac{A_{ii}}{2}\log\frac{\lambda_{b_ib_j}}{2}-\frac{\lambda_{b_ib_j}}{2}-\log\left(\frac{A_{ii}}{2}\right)!\end{aligned} \tag{6.22}$$

Ignoring terms which do not depend on the parameters of the model, we have

$$\begin{aligned}\log P(\boldsymbol{A}|\boldsymbol{\lambda},\boldsymbol{b}) &= \sum_{i<j}\left(A_{ij}\log\lambda_{b_ib_j}-\lambda_{b_ib_j}\right)+\sum_i\left(\frac{A_{ii}}{2}\log\frac{\lambda_{b_ib_j}}{2}-\frac{\lambda_{b_ib_j}}{2}\right)\\ &= \frac{1}{2}\sum_{ij}\left(A_{ij}\log\lambda_{b_ib_j}-\lambda_{b_ib_j}\right)\\ &= \frac{1}{2}\sum_{ijrs}A_{ij}\log\lambda_{rs}\delta_{b_ir}\delta_{b_js}-\frac{1}{2}\sum_{ijrs}\lambda_{rs}\delta_{b_ir}\delta_{b_js}\\ &= \frac{1}{2}\sum_{rs}\left(e_{rs}\log\lambda_{rs}-n_rn_s\lambda_{rs}\right),\end{aligned} \tag{6.23}$$

where e_{rs} is the number of edges between groups r and s, n_r and n_s correspond to the number of nodes in groups r and s respectively, and we have once more made use of the Kronecker delta.
Maximising Eq. (6.23) with respect to the parameters λ_{rs} is straightforward, and gives us our maximum likelihood estimate of the parameters λ_{rs},

$$\hat{\lambda}_{rs} = \frac{e_{rs}}{n_rn_s}. \tag{6.24}$$

Substituting Eq. (6.24) back into Eq. (6.23) and ignoring constant terms, we have that, at this maximum

$$\log P(\boldsymbol{A}|\hat{\boldsymbol{\lambda}},\boldsymbol{b}) = \frac{1}{2}\sum_{rs}e_{rs}\log\frac{e_{rs}}{n_rn_s}. \tag{6.25}$$

Eq. (6.25) cannot be maximised analytically with respect to the group memberships of the nodes, and one must resort to numerical techniques. However, we can notice that it has an interesting property. By adding and dividing by constant terms, Eq. (6.25) can be re-written as

$$\log P(\boldsymbol{A}|\boldsymbol{b}) = \sum_{rs} \frac{e_{rs}}{2E} \log \left(\frac{e_{rs}/2E}{n_r n_s/N^2} \right), \tag{6.26}$$

where we have discarded irrelevant constant terms.

Suppose that we are given some fixed assignment of the nodes into groups. Then $e_{rs}/2E$ is the probability, $p_K(r,s)$, that a randomly chosen edge in the network connects two nodes in groups r and s respectively. $n_r n_s/N^2$, on the other hand, corresponds to the same probability in a network with the same group membership of the nodes, but with the edges now being placed entirely at random ($p_{ER}(r,s)$). The log-likelihood can then be written as

$$\log P(\boldsymbol{A}|\boldsymbol{b}) = \sum_{rs} p_K(r,s) \log \frac{p_K(r,s)}{p_{ER}(r,s)}. \tag{6.27}$$

Eq. (6.27) is equal to the Kullback-Leibler divergence [122] (KL divergence) between the probability distributions p_K and p_{ER}, and measures the expected excess surprise from using p_{ER} as a model when the actual distribution is p_K. [4] Therefore, the most likely group assignments for the SBM are those which are most surprising compared to the Erdős–Rényi random graph.

Similarly, the log-likelihood for the DC-SBM can be written as

$$\log P(\boldsymbol{A}|\boldsymbol{b}) = \sum_{rs} \frac{e_{rs}}{2E} \log \frac{e_{rs}/2E}{(e_r/2E)(e_s/2E)} = \sum_{rs} p_K(r,s) \log \frac{p_K(r,s)}{p_{deg}(r,s)} \tag{6.28}$$

where e_r is the total number of half-edges incident on group r and $p_{deg} = (e_r/2E)(e_s/2E)$ corresponds to the probability that a randomly chosen edge connects two nodes in groups r and s respectively in a network with the same group assignment as the original one, but in which all stubs have been randomly rewired according to the configuration model. Therefore, the most likely group assignments for the DC-SBM are those which are most surprising with respect to the given expected degree sequence.

In a similar spirit, we might wish to construct a quantity that measures the distance of the distribution of edge placements being inferred from that of our null model obtained via the framework presented in this Thesis. Therefore, given a network and a set of corresponding graph observables $\{R_i\}$ to be used as our fitness criteria, which take on values $\{R_i^*\}$, one possible way of proceeding is the following:

1. Using our framework, determine the values of the parameters $\boldsymbol{\lambda}^*$ and $\boldsymbol{b}^*$ which maximise the entropy subject to the constraints $R_i(\boldsymbol{\lambda}, \boldsymbol{b}) = R_i^*$.

2. Define a probability distribution

$$p_{evo} = \frac{e_{rs}^*}{2E} \tag{6.29}$$

[4] In information theory, surprise, or surprisal, is a measure of the information content contained in a particular outcome. In this sense, the KL divergence can be thought as measuring the extra number of bits required to encode samples from p_K when p_{ER} is erroneously used as a distribution.

representing the probability that a randomly chosen edge in our null model connects a node in group r to one in group s.

3. Maximise the KL divergence

$$D_{KL}(p_K \parallel p_{evo}) = \sum_{rs} p_K(r,s) \log \frac{p_K(r,s)}{p_{evo}(r,s)}. \tag{6.30}$$

Where, as before, $p_K(r,s) = e_{rs}/2E$. This corresponds to selecting the most likely group assignments as those most surprising with respect to our "evolutionary" null model.

However, the approach outlined above has a number of serious drawbacks. To begin with, the procedure does not correspond to a maximum likelihood estimation. We still have not written down an actual likelihood, and Eq. (6.30) does not correspond to the log-likelihood of the SBM. This is problematic, as we lose the principled probabilistic approach rooted in statistical inference and are merely proposing an ad-hoc metric to infer the partitions. Secondly, this approach assumes that the number of groups in the network is known a priori, which is generally never the case in community detection problems. It is important to note that this second point is common to all community detection methods based on MLE. Indeed, it is a well-known fact that, if the number of groups is not known beforehand, the optimal value of the likelihood function is a strictly increasing function of B [115]. Therefore, maximising the log-likelihood will lead to the trivial partition $B = N$. This limitation is generally overcome by resorting to *minimum description length* methods (MDL) [79, 115, 123], where the description length is defined as the logarithm (generally taken base 2) of the joint distribution of the network and the model.

$$\begin{aligned} \mathrm{L}(\boldsymbol{A}, \boldsymbol{b}) &= -\log_2 P(\boldsymbol{A}, \boldsymbol{b}) \\ &= -\log_2 P(\boldsymbol{A}|\boldsymbol{b}) - \log_2 P(\boldsymbol{b}). \end{aligned} \tag{6.31}$$

The description length describes the number of bits needed to encode the network along with the model parameters. Given $\mathrm{L}(\boldsymbol{A}, \boldsymbol{b})$, Eq. (6.19) can then be re-written as

$$P(\boldsymbol{b}|\boldsymbol{A}) = \frac{2^{-\mathrm{L}(\boldsymbol{A},\boldsymbol{b})}}{P(\boldsymbol{A})}. \tag{6.32}$$

Since the evidence does not depend on the group partition, maximising the posterior is equivalent to minimising the description length. Therefore, the network partition that maximises the posterior corresponds to the choice of model parameters that most compresses the data.

It is easy to see why Eq (6.31) prevents the overfitting behaviour exhibited by the likelihood. The first term in Eq. (6.31) encodes the number of bits required to describe the network if the model parameters are known. The second term encodes the number of bits required to describe the model (via its parameters). As the complexity of the model increases (by increasing the number of groups), it will constrain itself better to the data, and $-\log_2 P(\boldsymbol{A}|\boldsymbol{b})$ will decrease. However, as the complexity of the model increases, more bits will be required to describe it, and $-\log_2 P(\boldsymbol{b})$ will increase. The second term then acts as a penalty, preventing models from becoming overly complex, and the optimal choice corresponds to a balance between the two terms.

A fundamental aspect for this framework to function correctly lies in the choice of the priors. A common choice is to use non-informative priors, which are maximum entropy distributions that are maximally agnostic about the model before observing any data and do not bias the posterior. However, a naive implementation of uninformative priors can lead to erroneous results [79]. A more principled approach is to take a non-parametric standpoint, where priors are themselves sampled from hyperprior distributions [79, 124]. Then, by considering a hierarchy of priors and hyperpriors of the kind $P(\boldsymbol{b}|\boldsymbol{\xi}, B)P(\boldsymbol{\xi}|B)P(B)$, where $\boldsymbol{\xi}$ are additional model parameters, it is possible to infer the number of groups B along with the partitioning $\boldsymbol{b}$ of the network.

However, this procedure could be further complicated in our case, as to determine the parameters $\boldsymbol{\lambda}^*$ and $\boldsymbol{b}^*$ of our evolutionary null model, the number of groups needs to be known beforehand. This is not ideal, as we would like to infer both quantities (the model parameters and the number of groups) directly from the data. Therefore, even if we can construct a generative model with a log-likelihood similar to Eq. (6.30), additional work would be required to formulate appropriate priors for the model. Furthermore, as our framework requires numerically minimising the free energy to determine the null model parameters, the task of formulating the priors could potentially prove to be a challenging one. However, as mentioned at the beginning of the chapter, extending our framework to an inferential setting could potentially open the doors to numerous applications and further work in this direction is desirable.

Bibliography

[1]H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks", Nature **407**, 651 (2000).

[2]D. A. Fell and A. Wagner, "The small world of metabolism", Nature biotechnology **18**, 1121 (2000).

[3]L. Apeltsin, J. H. Morris, P. C. Babbitt, and T. E. Ferrin, "Improving the quality of protein similarity network clustering algorithms using the network edge weight distribution", Bioinformatics **27**, 326–333 (2010).

[4]M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology", Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, 251–262 (1999).

[5]S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge Univ.Press, Cambridge, U.K., 1994).

[6]D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks.", Nature **393**, 440–442 (1998).

[7]L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks", **97**, 11149–11152 (2000).

[8]M. E. Newman, "The structure and function of complex networks", SIAM review **45**, 167–256 (2003).

[9]M. E. Newman, "The structure and function of networks", Computer Physics Communications **147**, 40–45 (2002).

[10]R. Albert, H. Jeong, and A. L. Barabási, "Error and attack tolerance of complex networks", Nature **406**, 378–382 (2000).

[11]R. Pastor-Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks", Physical Review Letters **86**, 3200–3203 (2001).

[12]E. Bullmore and O. Sporns, "The economy of brain network organization", Nature Reviews Neuroscience **13**, 336–349 (2012).

[13]D. S. Bassett and E. T. Bullmore, "Small-world brain networks revisited", The Neuroscientist **23**, 499–516 (2017).

[14]A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", science **286**, 509–512 (1999).

[15]P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of Growing Random Networks", Physical Review Letters **85**, 4629–4632 (2000).

[16]S. N. Dorogovtsev and J. F. F. Mendes, "Scaling behaviour of developing and decaying networks", EPL (Europhysics Letters) **52**, 33 (2000).

[17]A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations", Physica A: Statistical mechanics and its applications **311**, 590–614 (2002).

[18]C. Moore, G. Ghoshal, and M. E. J. Newman, "Exact solutions for models of evolving networks with addition and deletion of nodes", Physical Review E **74**, 036121 (2006).

[19]S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks with aging of sites", Physical Review E **62**, 1842 (2000).

[20]S. N. Dorogovtsev and J. F. F. Mendes, "Effect of the accelerating growth of communications networks on their structure", Physical Review E **63**, 025101 (2001).

[21]G. Bianconi and A.-L. Barabási, "Competition and multiscaling m evolving networks", in *The Structure and Dynamics of Networks* (Princeton University Press, 2011), pp. 361–367.

[22]J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins, "The web as a graph: Measurements, models, and methods", in International Computing and Combinatorics Conference (Springer, 1999), pp. 1–17.

[23]R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the web graph", in Proceedings 41st Annual Symposium on Foundations of Computer Science (IEEE, 2000), pp. 57–65.

[24]F. Chung, L. Lu, T. G. Dewey, and D. J. Galas, "Duplication Models for Biological Networks", Journal of Computational Biology **10**, 677–687 (2003).

[25]R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler, "A model of large-scale proteome evolution", Advances in Complex Systems **05**, Publisher: World Scientific Publishing Co., 43–54 (2002).

[26]A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, "Modeling of protein interaction networks", Complexus **1**, 38–44 (2003).

[27]J. R. Banavar, A. Maritan, and A. Rinaldo, "Size and form in efficient transportation networks", Nature **399**, 130–132 (1999).

[28]A. Rinaldo, A. Maritan, F. Colaiori, A. Flammini, R. Rigon, I. Rodriguez-Iturbe, and J. R. Banavar, "Thermodynamics of fractal networks", Physical review letters **76**, 3364 (1996).

[29]G. Caldarelli, *Scale-free networks: complex webs in nature and technology* (Oxford University Press, 2007).

[30]C. Cherniak, "Neural component placement", Trends in neurosciences **18**, 522–527 (1995).

[31]N. Mathias and V. Gopal, "Small worlds: how and why", Physical Review E **63**, 021117 (2001).

[32]S. Valverde, R. F. Cancho, and R. V. Solé, "Scale-free networks from optimal design", EPL (Europhysics Letters) **60**, Publisher: IOP Publishing, 512 (2002).

[33]R. F. i. Cancho and R. V. Solé, "Optimization in Complex Networks", in *Statistical Mechanics of Complex Networks*, edited by R. Pastor-Satorras, M. Rubi, and A. Diaz-Guilera, Lecture Notes in Physics (Springer, Berlin, Heidelberg, 2003), pp. 114–126.

[34]V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, "Network structures from selection principles", Physical review letters **92**, 198701 (2004).

[35]M. T. Gastner and M. E. J. Newman, "Optimal design of spatial distribution networks", Physical Review E **74**, Publisher: American Physical Society, 016117 (2006).

[36]R. Guimerà, A. Dìaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, "Optimal network topologies for local search with congestion", Physical review letters **89**, 248701 (2002).

[37]M. Barthélemy and A. Flammini, "Optimal traffic networks", Journal of Statistical Mechanics: Theory and Experiment **2006**, Publisher: IOP Publishing, L07002–L07002 (2006).

[38]G. Paul, T. Tanizawa, S. Havlin, and H. Stanley, "Optimization of robustness of complex networks", The European Physical Journal B - Condensed Matter and Complex Systems **38**, 187–191 (2004).

[39]C. M. Schneider, A. A. Moreira, J. S. Andrade, S. Havlin, and H. J. Herrmann, "Mitigation of malicious attacks on networks", Proceedings of the National Academy of Sciences **108**, 3838–3841 (2011).

[40]T. P. Peixoto and S. Bornholdt, "Evolution of robust network topologies: Emergence of central backbones", Physical review letters **109**, 118703 (2012).

[41]D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs", Physical Review Letters **85**, 5468–5471 (2000).

[42]M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the National Academy of Sciences **99**, 7821–7826 (2002).

[43]S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks", Nature **464**, 1025–1028 (2010).

[44]R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Resilience of the Internet to Random Breakdowns", Physical Review Letters **85**, 4626 (2000).

[45]D. Deutscher, I. Meilijson, M. Kupiec, and E. Ruppin, "Multiple knockout analysis of genetic robustness in the yeast metabolic network", Nature Genetics **38**, Number: 9 Publisher: Nature Publishing Group, 993–998 (2006).

[46]G. P. Wagner, M. Pavlicev, and J. M. Cheverud, "The road to modularity", Nature Reviews Genetics **8**, Number: 12 Publisher: Nature Publishing Group, 921–931 (2007).

[47]A. Buluç, H. Meyerhenke, I. Safro, P. Sanders, and C. Schulz, "Recent Advances in Graph Partitioning", in *Algorithm Engineering: Selected Results and Surveys*, edited by L. Kliemann and P. Sanders, Lecture Notes in Computer Science (Springer International Publishing, Cham, 2016), pp. 117–158.

[48]T. Opsahl, V. Colizza, P. Panzarasa, and J. J. Ramasco, "Prominence and control: the weighted rich-club effect", Physical review letters **101**, 168702 (2008).

[49]M. Newman, *Networks* (Oxford university press, 2018).

[50]D. J. Price et al., *Little science, big science... and beyond*, Vol. 480 (Columbia University Press New York, 1986).

[51]A.-L. Barabasi and Z. N. Oltvai, "Network biology: understanding the cell's functional organization", Nat Rev Genet **5**, 101–113 (2004).

[52]M. Zhou and J. Liu, "A memetic algorithm for enhancing the robustness of scale-free networks against malicious attacks", Physica A: Statistical Mechanics and its Applications **410**, 131–143 (2014).

[53]L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology", Nature **402**, C47–C52 (1999).

[54]N. Kashtan and U. Alon, "Spontaneous Evolution of Modularity and Network Motifs", Proceedings of the National Academy of Sciences of the United States of America **102**, 13773–13778 (2005).

[55]G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, "Self-organization and identification of web communities", Computer **35**, 66–70 (2002).

[56]S. Fortunato, "Community detection in graphs", Physics reports **486**, 75–174 (2010).

[57]P. Erdős and A. Rényi, "On random graphs", Publ. Math. Debrecen **6** (1959).

[58]P. Erdős, A. Rényi, et al., "On the evolution of random graphs", Publ. Math. Inst. Hung. Acad. Sci **5**, 17–60 (1960).

[59]P. Erdős and A. Rényi, "On the strength of connectedness of a random graph", Acta Mathematica Hungarica **12**, 261–267 (1961).

[60]P. Erdős and A. Rényi, "Asymmetric graphs", Acta Math. Acad. Sci. Hungar **14**, 15 (1963).

[61]P. Erdős and A. Rényi, "On the existence of a factor of degree one of a connected random graph", Acta Math. Acad. Sci. Hungar **17**, 359–368 (1966).

[62]J. Park and M. E. J. Newman, "Statistical mechanics of networks", Physical Review E **70**, 066117 (2004).

[63]G. Bianconi, "The entropy of randomized network ensembles", EPL (Europhysics Letters) **81**, 28005 (2008).

[64]K. Anand and G. Bianconi, "Entropy measures for networks: Toward an information theory of complex topologies", Physical Review E **80**, 045102 (2009).

[65]T. P. Peixoto, "Entropy of stochastic blockmodel ensembles", Physical Review E **85**, 056122 (2012).

[66]R. Solomonoff and A. Rapoport, "Connectivity of random nets", The bulletin of mathematical biophysics **13**, 107–117 (1951).

[67]E. N. Gilbert, "Random Graphs", The Annals of Mathematical Statistics **30**, 1141–1144 (1959).

[68]P. Erdős and T. Gallai, "Graphs with Prescribed Degrees of Vertices", Mat. Lapok. **11**, 264–274 (1960).

[69]S. L. Feld, "Why your friends have more friends than you do", American journal of sociology **96**, 1464–1477 (1991).

[70]P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps", Social Networks **5**, 109–137 (1983).

[71]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**, 730–780 (1976).

[72]S. Wasserman and C. Anderson, "Stochastic a posteriori blockmodels: Construction and assessment", Social Networks **9**, 1–36 (1987).

[73]C. J. Anderson, S. Wasserman, and K. Faust, "Building stochastic blockmodels", Social Networks **14**, 137–161 (1992).

[74]K. Nowicki and T. A. B. Snijders, "Estimation and Prediction for Stochastic Blockstructures", Journal of the American Statistical Association **96**, 1077–1087 (2001).

[75]B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks", Physical Review E **83**, 016107 (2011).

[76]P. J. Bickel and A. Chen, "A nonparametric view of network models and Newman–Girvan and other modularities", Proceedings of the National Academy of Sciences **106**, 21068–21073 (2009).

[77]A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications", Physical Review E **84**, 066106 (2011).

[78]T. P. Peixoto, "Nonparametric Bayesian inference of the microcanonical stochastic block model", Physical Review E **95**, 012317 (2017).

[79]T. P. Peixoto, "Bayesian Stochastic Blockmodeling", in *Advances in Network Clustering and Blockmodeling* (John Wiley & Sons, Ltd, 2019), pp. 289–332.

[80]E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed Membership Stochastic Blockmodels", J. Mach. Learn. Res. **9**, 1981–2014 (2008).

[81]T. P. Peixoto, "Model selection and hypothesis testing for large-scale network models with overlapping groups", Physical Review X **5**, 011033 (2015).

[82]C. De Bacco, E. A. Power, D. B. Larremore, and C. Moore, "Community detection, link prediction, and layer interdependence in multilayer networks", Physical Review E **95**, 042317 (2017).

[83]T. P. Peixoto, "Hierarchical block structures and high-resolution model selection in large networks", Physical Review X **4**, 011047 (2014).

[84]C. Aicher, A. Z. Jacobs, and A. Clauset, "Learning latent block structure in weighted networks", Journal of Complex Networks, cnu026 (2014).

[85]T. P. Peixoto, "Nonparametric weighted stochastic block models", Physical Review E **97**, 012306 (2018).

[86]M. E. J. Newman and A. Clauset, "Structure and inference in annotated networks", Nature Communications **7**, 11863 (2016).

[87]D. Hric, T. P. Peixoto, and S. Fortunato, "Network Structure, Metadata, and the Prediction of Missing Nodes and Annotations", Physical Review X **6**, 031038 (2016).

[88]L. Peel, "Supervised Blockmodelling", arXiv:1209.5561 [cs, stat], arXiv: 1209.5561 (2012).

[89]E. T. Jaynes, "Information Theory and Statistical Mechanics", Physical Review **106**, 620 (1957).

[90]G. Bianconi, "Entropy of network ensembles", Physical Review E **79**, 036114 (2009).

[91]K. Anand and G. Bianconi, "Gibbs entropy of network ensembles by cavity methods", Physical Review E **82**, 011116 (2010).

[92]M. E. Newman and M. Girvan, "Mixing patterns and community structure in networks", in *Statistical mechanics of complex networks* (Springer, 2003), pp. 66–87.

[93]R. Guimerà, D. Stouffer, M. Sales-Pardo, E. Leicht, M. Newman, and L. A. Amaral, "Origin of compartmentalization in food webs", Ecology **91**, 2941–2951 (2010).

[94]S. C. Olhede and P. J. Wolfe, "Network histograms and universality of blockmodel approximation", Proceedings of the National Academy of Sciences **111**, 14722–14727 (2014).

[95]E. T. Jaynes, *Probability Theory: The Logic of Science*, edited by G. L. Bretthorst (Cambridge University Press, Cambridge, UK ; New York, NY, June 2003).

[96]R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A Limited Memory Algorithm for Bound Constrained Optimization", SIAM Journal on Scientific Computing **16**, 1190 (1995).

[97]A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey", Journal of Marchine Learning Research **18**, 1–43 (2018).

[98]R. V. Sole and M. Montoya, "Complexity and fragility in ecological networks", Proceedings of the Royal Society of London. Series B: Biological Sciences **268**, 2039–2045 (2001).

[99]S. Zhang, J.-l. Sun, Y. Shen, Y. Chen, and A. J. Kavs, "Optimizing performance of composite services in multiple networks enterprise environment", Information Technology Journal **10**, 807–815 (2011).

[100]E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, "Hierarchical Organization of Modularity in Metabolic Networks", Science **297**, 1551–1555 (2002).

[101]A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease", Nature reviews genetics **12**, 56–68 (2011).

[102]D. A., C. R., and P. D., in 6th Neural Computation and Psychology Workshop: Evolution, Learning, and Development (2000).

[103]C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization", ACM Trans. Math. Softw. **23**, ACM ID: 279236, 550–560 (1997).

[104]P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", Nature **17**, 261–272 (2020).

[105]Eurocontrol, *Ash-cloud of April and May 2010: Impact on Air Traffic*, 2010.

[106]V. Rosato, L. Issacharoff, F. Tiriticco, S. Meloni, S. Porcellinis, and R. Setola, "Modelling interdependent infrastructures using interacting dynamical models", International Journal of Critical Infrastructures **4**, 63–79 (2008).

[107]J. G. Venegas, T. Winkler, G. Musch, M. F. Vidal Melo, D. Layfield, N. Tgavalekos, A. J. Fischman, R. J. Callahan, G. Bellani, and R. Scott Harris, "Self-organized patchiness in asthma as a prelude to catastrophic shifts", Nature **434**, 777–782 (2005).

[108]C. Perrings, "Resilience in the dynamics of economy-environment systems", Environmental and Resource Economics **11**, 503–520 (1998).

[109]R. M. May, "Thresholds and breakpoints in ecosystems with a multiplicity of stable states", Nature **269**, 471–477 (1977).

[110]M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications", Physical Review E **64**, 026118 (2001).

[111]C. Priester, S. Schmitt, and T. P. Peixoto, "Limits and Trade-Offs of Topological Network Robustness", PLoS ONE **9**, e108215 (2014).

[112]T. P. Peixoto, "Emergence of robustness against noise: A structural phase transition in evolved models of gene regulatory networks", Physical Review E **85**, 041908 (2012).

[113]H. Li, G. W. Rosenwald, J. Jung, and C.-C. Liu, "Strategic power infrastructure defense", Proceedings of the IEEE **93**, 918–933 (2005).

[114]X. Yan, C. Shalizi, J. E. Jensen, F. Krzakala, C. Moore, L. Zdeborová, P. Zhang, and Y. Zhu, "Model selection for degree-corrected block models", Journal of Statistical Mechanics: Theory and Experiment **2014**, P05007 (2014).

[115]T. P. Peixoto, "Parsimonious Module Inference in Large Networks", Physical Review Letters **110**, 148701 (2013).

[116]S. Dasgupta, C. H. Papadimitriou, and U. V. Vazirani, *Algorithms* (McGraw-Hill Higher Education New York, 2008).

[117]C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy", Nature **585**, 357–362 (2020).

[118]J. Nocedal and S. J. Wright, *Numerical optimization* (Springer, 1999).

[119]I. Loshchilov, "LM-CMA: An alternative to L-BFGS for large-scale black box optimization", Evolutionary computation **25**, 143–171 (2017).

[120]N. N. Schraudolph, J. Yu, and S. Günter, "A stochastic quasi-Newton method for online convex optimization", in Artificial intelligence and statistics (PMLR, 2007), pp. 436–443.

[121]P. Moritz, R. Nishihara, and M. Jordan, "A linearly-convergent stochastic L-BFGS algorithm", in Artificial Intelligence and Statistics (PMLR, 2016), pp. 249–258.

[122]T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 99th (Wiley-Interscience, Aug. 1991).

[123]M. Rosvall and C. T. Bergstrom, "An information-theoretic framework for resolving community structure in complex networks", Proceedings of the National Academy of Sciences **104**, 7327–7331 (2007).

[124]M. Schmidt and M. Morup, "Nonparametric Bayesian Modeling of Complex Networks: An Introduction", IEEE Signal Processing Magazine **30**, 110–128 (2013).