JAGIELLONIAN UNIVERSITY
IN KRAKOW

Correlations in Random Graphs

by

Andrzej K. Oleś

Thesis written under the supervision of
dr hab. Piotr Białas
presented to the
Faculty of Physics, Astronomy and Applied Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Physics

Kraków 2012

To Małgosia

Contents

Preface

Networks constitute the backbone of many complex systems of high technological and intellectual importance in such diverse areas as physics, economics, biology, ecology, computer science, and the social sciences. This variety of networks can be described by the same graph framework, in which the nodes represent individual elements or units of the system, connected by links associated with the interactions. The most prominent examples of computer and communication networks are the World Wide Web (WWW), an enormous virtual network of web pages connected by hyperlinks, and the Internet which is a physical network of routers and computers linked by various cable and wireless connections. Of special interest are social networks where individuals act as nodes connected by social relationships; these include phone call networks, acquaintance networks, and social networking services, such as Facebook or the more business oriented LinkedIn. Common examples in biosciences are molecular networks, in specific metabolic pathways and genetic regulatory networks, and ecological networks, like food webs, for instance. Among other typical examples are transportation networks, disease spread networks, and citation networks.

Empirical studies of networks have feedback on theoretical developments aiming to describe and explain the observed features. In particular, recent advance in computer technology has opened new possibilities to the acquisition and processing of data and made it possible to gather information about the topology of diverse real-world networks and to analyze them on a scale far larger than previously possible. It turned out that various, often completely unrelated systems, share the same common properties such as the distribution of connections between nodes or the inter-node correlations. For example, it has been found that many of the real-world networks are *scale-free* and their degree distribution follows a power-law, where the *degree* is the number of connections emerging from a vertex. The most notable characteristic of a scale-free network is that nodes with the number of connections greatly exceeding the average are relatively common within it and the network consists of a few highly connected hubs to which rest of the weaker connected nodes attach. All these observations triggered the interest in various network models, including random graphs—a simple, yet extremely useful network model. Extensive pedagogical reviews of the field

of complex networks can be found in Refs. [4–7] and in the books by Newman [8] and Bornholdt *et al.* [9]. For more information specifically oriented towards scale-free networks consult the book by Caldarelli [10].

One of the most fundamental network characteristics is the pattern of connections. In most kinds of networks a few different types of vertices can be distinguished according to various features, and it has been observed that the probability of connection between vertices often depends on them (cf. Ref. [11]). In particular, vertices may tend to connect to others that are like them in some way, which is called *assortative mixing*, or just the opposite: they may connect to others with completely different characteristics, referred to as *disassortative mixing*. Such correlations are interesting mainly because they uncover additional information about the system which may be inferred based on the knowledge of them.

Particularly important, with regard to network structure, are vertex degree correlations. Interestingly, most of the real-world networks show disassortative mixing by degree, that is, their vertices have the tendency to connect to others that have very different degrees from their own. The exact reasons for this pattern are not yet known for certain, but there is a strong indication that it is a consequence of the fact that these networks are simple graphs, i.e., ones that do not have multiple- and self-connections (see Ref. [12]). Therefore, the number of edges that can connect the high-degree vertex pairs is limited.

On the other hand, most social networks have a clear tendency towards assortativity. This can be put down to the common phenomenon that we tend to associate preferentially with people who are similar to ourselves in some way. This in turn results in social networks being divided into groups of vertices weakly connected with the rest of the network [8]. Roughly speaking, the degrees of vertices within small groups tend to have lower degree than vertices in larger groups, so the small groups gather mostly low-degree vertices connected with other low degree-vertices, and similarly for high-degree ones. A quantitative description of this mechanism can be found in Ref. [13].

The study of degree correlations is of vital importance in gaining insight into the structure and patterns of connections appearing in networks. These are essential for an in-depth understanding of the behavior and function of the various complex systems around us, ranging from technological to biological networks. Obviously, if we could gain such understanding, it would give us fresh insight into a vast array of complex and previously poorly understood phenomena.

Random graphs are a network model of common use which mimics the patterns of connections in real-world networks. They are interesting in their own right for the light they shed on the structural properties of these networks. The simplest random network model are *classic random graphs* proposed by the Hungarian mathematicians Erdős and Rényi in their seminal

cycle of papers [14–16]. The model consists of V vertices connected at random with L edges, and is characterized by the Poisson degree distribution. A generalization of this concept are *random graphs* with arbitrary degree distributions. Intuitively, these are model networks which are maximally random under a given set of constraints. The constraints typically include a prescribed degree distribution; for example, a power-like degree distribution such as $p(q) \propto q^{-\gamma}$ results in a scale-free network. The general random graph model provides insight into the expected structure of networks of almost all kinds and is an excellent tool for investigating all sorts of network phenomena.

Random graphs are an example of the so-called *complex network* models, which have a neither purely regular nor purely random pattern of connections between their elements. They typically have a much more complex structure than these of classic random graphs, and display such features as a heavy tailed degree distribution or internode correlations. In fact, the vast majority of real-world networks is complex. Moreover, these networks usually exhibit a relatively small diameter, referred to as the *small-world effect*. For example, Milgram showed in his famous 1967 experiment that the average separation distance in the acquaintance network is around six [17]. The combination of the compactness and complexity of these systems results in a wide spectrum of nontraditional and intriguing critical effects. These include structural changes in networks, various percolation phenomena, emergence of scale-free topologies, epidemic thresholds, phase transitions in cooperative models defined on networks, and many others. A comprehensive review of recent progress in the field of critical phenomena occurring in complex networks can be found in Ref. [18].

There are two natural strategies of modeling complex networks as random graphs: the diachronic one, in which the growth of the system is simulated, and the synchronic one, in which a statistical ensemble of graphs is constructed. In here we are concerned only with the latter one and investigate the general random graph model in the conventional framework of equilibrium statistical mechanics (see Refs. [4, 19–24]). The crucial concept of this approach is the statistical ensemble of graphs defined by ascribing statistical weight to each particular configuration of vertices and edges. The quantities of interest are then expressed as weighted averages over all the graphs from the ensemble. We conduct our studies by means of both analytic calculations and numerical simulations.

The thesis is organized as follows. **Chapter 1** serves as an introduction to the fundamental concepts and mathematical tools employed across the rest of the dissertation. It also settles the notation and defines basic quantities used in the calculations from the subsequent chapters. The presented material may be logically divided into two parts. The first one, constituted by **Chapters 2** and **3**, is based on our two papers [1, 2] and is primarily concerned with structural correlations in connected topologies. We first

discuss nearest-neighbor correlations in connected random graphs in **Chapter 2**. Then, in **Chapter 3** we specialize to random trees and investigate correlations at larger distances. By definition, trees are connected graphs without cycles, where a cycle is a closed path. The second part deals with critical phenomena occurring in random graphs. Our aim is to construct and investigate a simple random geometry model exhibiting a geometrical phase transition. To this end we first introduce an Ising spin model on top of regular random graphs in **Chapter 4**. By regular random graphs we mean maximally random networks composed of vertices which all have the same degree [3]. In this chapter we are interested in the properties of the spin system rather than the underlying topology. The considered model exhibits a mean-field phase transition, which is a direct consequence of the infinite dimensionality of the system. The actual random graph model with the geometric phase transition is studied in **Chapter 5**. It is constituted by maximally random graphs whose vertex degrees are limited to the values $\{2, 3, 4\}$ only. We show that asymptotically the solutions of this model and the Ising spin system from the previous chapter coincide, and so the thermodynamic behavior of these two models is identical. However, due to their different nature and the broken symmetry of the geometrical model, its correlation functions exhibit some unexpected behavior. A summary and a brief discussion of the results and prospects is given in **Chapter 6**. The dissertation is supplemented by some **Appendices** providing additional technical details to the derivations appearing in the course of the thesis.

This dissertation is intended to be mostly self-contained. To make it easily accessible even for readers from other disciplines, a detailed and comprehensive description of the fundamental ideas, theoretical tools, and numerical methods is provided, mainly in the first introductory chapter. Nevertheless, we are aware that since the presented material spans a wide spectrum of theoretical physics and mathematics, one might find the discussion of some concepts either too general or insufficient. We encourage an interested reader to consult the numerous references listed at the end of this thesis and hope that they prove helpful.

Before we proceed any further, we would like to clarify a few points concerning the notation used. We write $f(n) \sim g(n)$ for asymptotic equality meaning

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1,$$

while proportionality is written as $f(n) \propto g(n)$ implying that $f(n) = c\, g(n)$, where c is some constant. The symbol $\approx$ is reserved for general approximately equal values. Sometimes, if the meaning is clear from the context, we omit the arguments of functions for the sake of clarity. Vectors are denoted by bold lowercase symbols such as $\mathbf{v}$ and $\boldsymbol{\phi}$, while bold capital Latin letters such as $\mathbf{M}$ refer to matrices.

Acknowledgments

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Piotr Białas, for introducing me into the fascinating topic of random graphs, scientific guidance, and insightful discussions. Without his constant help, inspiring comments, and endless patience this dissertation would not be possible.

I would like to heartily thank my dear wife Małgosia for her love and encouragement throughout, particularly during the last hectic stage of writing.

Last but not least, I am deeply indebted to my parents, who nurtured my interest in physics and mathematics from the early school years, and who stood by me through the many trials and decisions along my educational career.

This research was supported in part by the PL-Grid Infrastructure. Monte Carlo simulations were performed on the Shiva computing cluster at the Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, and at the Academic Computer Centre CYFRONET AGH using the Zeus cluster.

Andrzej K. Oleś

Chapter 1

Introduction

The study of random graphs is a highly interdisciplinary field that combines ideas from diverse areas, mostly mathematics, physics, and computer science. This may lead to ambiguity in terminology due to the inconsistent usages among its practitioners from different disciplines. To avoid or at least reduce any confusion we briefly define basic terms and concepts, and present the theoretical foundations for the subsequent calculations.

We begin with Sec. 1.1 introducing the terminology and fundamental ideas of graph theory which are crucial for essentially all later developments. Graph theory is a wide branch of mathematics and readers interested in studying it further might like to look at the books by Bondy [25] and Wilson [26].

Then, in Sec. 1.2 we provide some background material on statistical mechanics. It contains only the very basics of the subject and is primarily intended to recall some of the main ideas that are mentioned later on. For further reading please consult any book on statistical physics. Regarding the problems discussed here we recommend the one by Plischke [27].

Random graphs are introduced in Sec. 1.3. First, we briefly present classic random graphs, also known as Erdős-Rényi graphs, characterized by Poisson degree distribution. They have been widely studied and thorough descriptions of the field can be found in the books by Bollobás [28] and Janson [29]. For a mathematical introduction specific on random trees, see the book by Drmota [30]. We then come to the generalized random graph model, i.e., random graphs with arbitrary degree distribution, which are the actual object of our study. We present them in the traditional framework of statistical mechanics—for details please consult Refs. [4, 19–24].

Next, in Sec. 1.4 we come to the notion of correlations in random geometries. We define specific measures used to quantify vertex degree correlations in random graphs. These include the basic two-point correlation function (Sec. 1.4.1), its connected version (Sec. 1.4.2), and the joint degree distribution (Sec. 1.4.3).

For the models of our interest it has not yet been proved possible to find exact analytic solutions to some of the discussed problems. Our approximate techniques include graphical perturbation expansion and numerical methods. We deal with the first one in Sec. 1.5. In particular, we show how one can relate integrals appearing in field theory to an ensemble of graphs—Feynman diagrams (Sec. 1.5.2), elucidate the role of the symmetry factor in defining statistical weights of these diagrams (Sec. 1.5.3), and recall the useful technique of obtaining connected diagrams only (Sec. 1.5.4).

We end this chapter with a discussion of equilibrium Monte Carlo (MC) simulations of statistical ensembles in Sec. 1.6. We explain the notion of importance sampling (Sec. 1.6.1), briefly introduce the classic Metropolis algorithm (Sec. 1.6.2), and present its application to the generation of the ensemble of random graphs (Sec. 1.6.3). To the end of this section we discuss some practical issues related to MC simulations, namely equilibration and measurement (Sec. 1.6.4). In particular, we devote special attention to the evaluation of statistical errors of MC data when the output is time correlated (Sec. 1.6.5).

1.1 Fundamentals of graph theory

A *graph* is, in its simplest form, a collection of points joined together by lines, as illustrated in Fig. 1.1. In the jargon of the field the points are referred to as *vertices* or *nodes* and the lines are referred to as *edges* or *links*. Throughout this dissertation we will typically denote the number of vertices in a graph by V and the number of its edges by L. Graphs are sometimes called *networks* and we will use these terms interchangeably. Although a graph can be conveniently described by means of a diagram, its exact form is immaterial—the only thing that matters is connectivity, that is whether a given pair of points is joined by a single or more lines, or not.

It follows that there is no unique way of drawing a graph and the relative positions of points representing vertices and lines representing edges have no significance. In particular, two edges in a diagram may intersect at a point

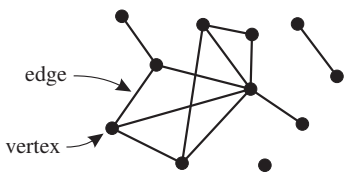

Figure 1.1: A small network composed of eleven vertices and twelve edges. The position of nodes in the picture is irrelevant and the only information that matters is connectivity.

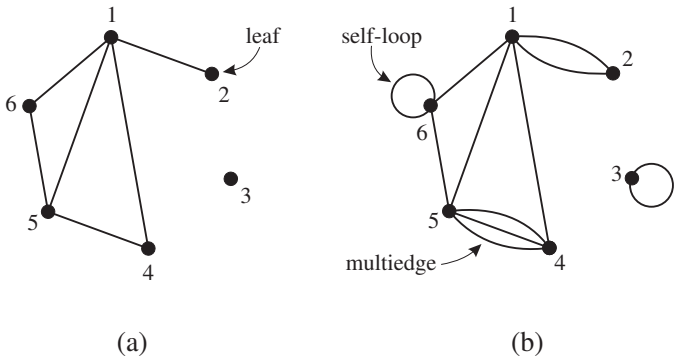

Figure 1.2: Two small networks: (a) a simple graph composed of $V = 6$ vertices and $L = 6$ edges; (b) a pseudograph, i.e., a graph with both multi- and self-edges.

that is not a vertex, which is the case in Fig. 1.1. A graph which can be drawn on a plane in such a way that no lines cross and its edges meet only at their endpoints is called *planar*. The graph in Fig. 1.1 can be rearranged that way, so it is planar.

A graph does not necessarily need to be connected. It may consist of two or more separate *components* disconnected from one another. The exemplary network in Fig. 1.1 consists of three components, one of which is a single isolated vertex not connected to the rest of the network. Further in this section we provide a more formal definition of connected graphs using the concept of path.

Two or more edges connecting the same pair of vertices are called collectively a *multiple edge* or shortly *multiedge*. A *self-edge* or a *loop* is an edge connected with both its ends to one and the same vertex. We shall call graphs which have multiple edges and loops *multigraphs* or *pseudographs* to distinguish them from *simple graphs* which have at most a single edge between any pair of distinct vertices (see Fig. 1.2).

In some applications weighted graphs, in which the connections have given weights, or directed graphs, in which each edge has a direction, are considered. In here we restrict ourselves to undirected and unweighted networks only where the edges form simple on/off connections between vertices. However, undirected networks can be thought of as directed networks in which each undirected edge is represented by a pair of directed ones running in opposite directions. Such representation is especially useful in computer programs and we actually employ it in our MC simulations.

A *degree* of a vertex is the number of edges attached to it. We will typically denote it by q or k. On simple graphs the degree of a vertex equals the number of its neighbors, which in general does not hold when multigraphs are considered. Note that each loop raises the degree of the

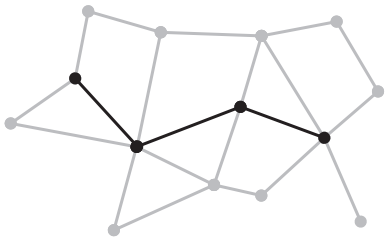

Figure 1.3: A path of length three in a graph. It is a geodesic path, i.e., the shortest path between these two vertices.

vertex to which it connects by two, as both its ends are counted. Vertices of degree equal one, that is connected to only one other vertex, are called *external vertices* or *leaves*, and the remaining ones *internal vertices*. For instance, the graph shown in Fig. 1.2(a) has one external vertex labeled 2, and the other vertices are internal ones.

The sum of the degrees of all the vertices in a graph equals the number of ends of edges, which is twice the total number of edges

$$2L = \sum_{i=1}^{V} k_i. \tag{1.1}$$

On the other hand, the mean degree of a vertex is given by

$$\bar{k} = \frac{1}{V} \sum_{i=1}^{V} k_i, \tag{1.2}$$

which combined with Eq. (1.1) yields

$$\bar{k} = \frac{2L}{V}. \tag{1.3}$$

This relation will come up repeatedly in later developments.

A *path* in a graph is a sequence of vertices such that every consecutive pair of vertices in the sequence is connected by an edge. A path which starts and ends at the same vertex is called a *cycle*. The *length* of a path is the number of edges traversed along the path (see Fig. 1.3). The *geodesic distance* between two vertices on a graph is the shortest path length between this pair of vertices, i.e., the minimum number of edges one would have to traverse in order to get from one vertex to the other. Vertices directly connected, called *nearest neighbors*, have geodesic distance equal one. In most network models the mean geodesic distance is rather short and increases logarithmically with the number of vertices in the network, which is referred to as the small-world effect.

Now we can introduce a more formal definition of a *connected* graph, as mentioned at the beginning of this section. Namely, a graph is said to be connected if every vertex in it is reachable from every other via some path through the graph. Conversely, if there is no such path for at least two vertices, the network is *disconnected* and consists of *components* or *subgraphs*. Technically, these are subsets of all the vertices of the graph such that there exists at least one path between each pair of their members, and such that no other vertex in the graph can be added to the subset while preserving this property.

One possible representation of a graph is its *adjacency* or *connectivity matrix*. Consider a graph containing N labeled nodes $1, \ldots, N$. Its adjacency matrix is a $N \times N$ matrix whose non-diagonal entries A_{ij} count the number of edges between vertices i and j, whereas the diagonal elements A_{ii} equal twice the number of loops attached to the i-th vertex. A loop is counted twice essentially because it is connected with both its ends to vertex i. For example, the adjacency matrix of the graph from Fig. 1.2(b) is

$$\mathbf{A} = \begin{pmatrix} 0 & 2 & 0 & 1 & 1 & 1 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 3 & 0 \\ 1 & 0 & 0 & 3 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 2 \end{pmatrix}. \tag{1.4}$$

The adjacency matrix of undirected graphs is symmetric, and the degree of vertex i is given by the sum of elements in the i-th row or column. For simple graphs all elements are either zero or one, except the diagonal ones, which are all equal zero.

We also distinguish between *labeled* graphs, where the vertices have assigned different numbers, and *unlabeled* graphs, where vertices are indistinguishable. We will often refer to unlabeled graphs as *shapes* or *topologies*. This is particularly important for the counting problems. For example, the topologies A and D in Fig. 1.4 have a 1:1 mapping on their labeled counterparts, while shapes B and C can be labeled in three distinct ways each. In simple graphs each link is uniquely determined by its endpoints, so there is no need to label them additionally. In the case of pseudographs, however, the labeling of links does matter. This is because of the ambiguity of multi-edges joining the same vertices and self-edges connected with both ends to one vertex. These may be freely exchanged unless we explicitly distinguish the endpoints of each edge. We will call graphs with labeled both vertices and edges *fully labeled* graphs.

Depending on some specific characteristics and properties we distinguish a number of special classes of graphs. Here we mention only one, namely *regular pseudographs*, to which we refer in Chapter 4. A regular, simple graph of degree q, or shortly *q-regular graph*, is a graph whose all vertices have the

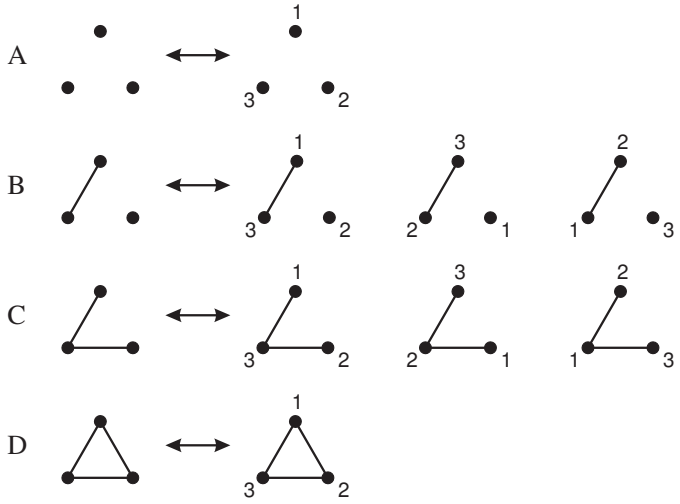

Figure 1.4: Unlabeled simple graphs with $V = 3$ vertices and their labeled counterparts.

same degree q. Regular multigraphs or pseudographs are a generalization of regular graphs in which multiedges and loops are permitted.

1.1.1 Trees

Trees are connected, simple graphs without cycles. A connected graph is a tree, if and only if the number of edges L equals the number of nodes V less one: $L = V - 1$. A tree is a minimal connected graph, so deleting any single edge disconnects it, while adding an extra edge between any two vertices creates a cycle. A disconnected graph in which all individual parts are trees is called a *forest*.

Probably the most important property of trees for our purposes is that, since they have no cycles, there is exactly one path between any pair of vertices, which makes some certain kinds of calculations particularly simple. By definition trees are simple graphs, so the degree of a vertex in a tree equals the number of its neighbors.

We distinguish between rooted and unrooted trees (see Fig. 1.5). A *rooted* tree is a tree with one node marked, which is called the *root*. The presence of the root introduces an ordering in the tree, which may be described in terms of generations or levels: the root constitutes the 0-th generation, its neighbors are the first generation, and so on. In general, a node at distance k from the root belongs to the k-th generation. For a node at level k its neighbors at level $k+1$ are called *children* or *successors*, while the node at level $k-1$ to which it connects is its *parent*.

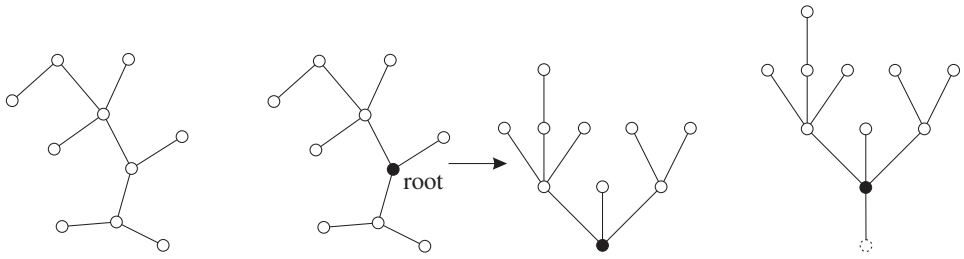

Figure 1.5: Three different types of trees: (a) a non-rooted, (b) a rooted, (c) and a planted tree.

Furthermore, if v is a node in a rooted tree T, then v may be considered as the root of a subtree of T consisting of all iterated successors of v. This means that rooted trees may be constructed recursively, which is an extremely useful property to which we will often refer in the upcoming calculations.

Another important class of trees are *planted* trees. These are rooted trees with the root connected (or planted) to an additional phantom node, that is not taken into account (Fig. 1.5(c)). Usually the additional external node is omitted and the root is drawn with a *stem* attached, i.e., a link which has one loose end. Similarly as for the phantom node, we do not include the stem into the total number of links in a tree. Nevertheless, it contributes to the degree of the root.

Trees play an important role in the study of random graphs. For instance, the local groups of vertices—the so-called small components—form a forest and we exploit this property in Chapter 2 to derive our results.

In Chapter 4 we will be interested in a special class of trees, namely *Cayley trees*. These are regular trees in which each internal vertex has a constant number of branches q, where q is called the coordination number—see Fig. 1.6. In physics an infinite Cayley tree is often referred to as a *Bethe lattice* [31].

1.2 Basic equilibrium statistical mechanics

Statistical mechanics is primarily concerned with the calculation of properties of systems composed of very many particles, typically atoms or molecules. Although these individual particles usually obey simple equations of motion and the behavior of the entire system is mathematically well defined and predictable, it is the magnitude of the problem which makes it impossible to solve the mathematics exactly. Instead of solving the equations of motion,

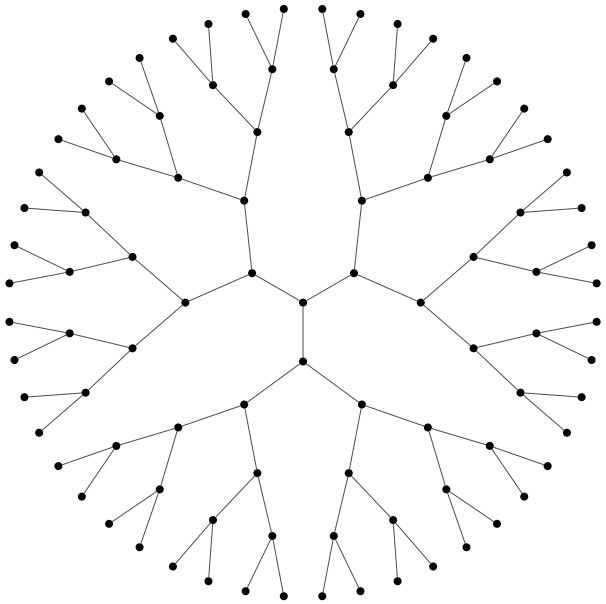

Figure 1.6: Cayley tree with coordination number $q=3$ and five generations.

statistical mechanics attempts to calculate the properties of large systems by treating them in a probabilistic fashion. Such approach turns out to be extremely useful because the reasonably probable behavior of such systems falls into a very narrow range, allowing us to state with very high confidence that the system of interest will display behavior within that range.

Complex systems can be characterized by the Hamiltonian function H, expressing the total energy of the system in any particular state. For a system in thermal equilibrium at temperature T the probability of the system being in state μ is given by the Boltzmann distribution,

$$p_\mu = \frac{1}{Z}\, \mathrm{e}^{-E_\mu/k_B T}, \tag{1.5}$$

where E_μ is the energy of state μ given by the Hamiltonian, and k_B is the Boltzmann's constant. The quantity $1/k_B T$ is traditionally denoted by β and we will follow that convention. The normalizing factor Z is the *partition function* of the system,

$$Z = \sum_\mu \mathrm{e}^{-\beta E_\mu}. \tag{1.6}$$

The average or expectation value of an observable $\mathcal{O}$ is given by

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_\mu \mathcal{O}_\mu\, \mathrm{e}^{-\beta E_\mu}. \tag{1.7}$$

In particular, the internal energy U, which is the expectation value of the energy itself, is

$$U \equiv \langle E \rangle = \frac{1}{Z} \sum_{\mu} E_{\mu} \, \mathrm{e}^{-\beta E_{\mu}}. \tag{1.8}$$

Apart from the internal energy U we define the *free energy* F, which can be directly calculated from the partition function Z,

$$F = -\frac{1}{\beta} \ln Z. \tag{1.9}$$

The free energy is especially useful, because it relates the system's conjugate variables. These are parameters coupled in such a way that the change of one of them is followed by the response of the counterpart. Suppose that X is the conjugate variable to Y. They correspond to a term in the Hamiltonian of the form $-XY$, thus the expectation value of X is given by

$$\langle X \rangle = \frac{1}{Z} \sum_{\mu} X_{\mu} \, \mathrm{e}^{-\beta E_{\mu}} = \frac{1}{\beta Z} \frac{\partial}{\partial Y} \sum_{\mu} \mathrm{e}^{-\beta E_{\mu}}, \tag{1.10}$$

because of the $-X_{\mu} Y$ term on which the derivative acts on. In terms of the free energy the above average reads

$$\langle X \rangle = \frac{1}{\beta} \frac{\partial \ln Z}{\partial Y} = -\frac{\partial F}{\partial Y}. \tag{1.11}$$

This introduces a useful technique of calculating thermal averages, even if there is no appropriate field coupling to the quantity of interest. We can simply introduce a fictitious field coupled to that quantity, which we set to zero after performing the derivative.

The expectation value of some quantity alone does not provide any information about its fluctuations. These are quantified by the variance

$$\mathrm{var}(X) \equiv \langle X^2 \rangle - \langle X \rangle^2 , \tag{1.12}$$

which can be calculated from the second derivative of the free energy

$$\begin{aligned} \frac{\partial^2 F}{\partial Y^2} &= -\frac{1}{\beta} \frac{\partial}{\partial Y} \left(\frac{1}{Z} \frac{\partial Z}{\partial Y} \right) = -\frac{1}{\beta} \left[\frac{1}{Z} \frac{\partial^2 Z}{\partial Y^2} - \left(\frac{1}{Z} \frac{\partial Z}{\partial Y} \right)^2 \right] \\ &= -\beta \left(\langle X^2 \rangle - \langle X \rangle^2 \right). \end{aligned} \tag{1.13}$$

Combining Eqs. (1.12) and (1.13) we finally get

$$\mathrm{var}(X) = -\frac{1}{\beta} \frac{\partial^2 F}{\partial Y^2} = \frac{1}{\beta} \frac{\partial \langle X \rangle}{\partial Y}. \tag{1.14}$$

The bare derivative $\partial \langle X \rangle / \partial Y$ is called the *susceptibility* of X to Y,

$$\chi \equiv \frac{\partial \langle X \rangle}{\partial Y}. \tag{1.15}$$

We see that the fluctuations in a variable are proportional to the variable's susceptibility, which is known as the *linear response theorem*. It provides a practical way of calculating the susceptibility in MC simulations by measuring the fluctuations of a variable.

1.3 Random graphs

In general, a *random graph* is a model network in which some specific set of parameters take fixed values, but the graph is random in all other respects (cf. Ref. [8]). Random graphs constitute a useful tool in exploring various features of the networks covered by models possessing particular properties of interest, such as a specified degree distribution, but which are otherwise random. We may, for instance, specialize to connected random graphs, as in Chapter 2, or to random trees, that is to acyclic connected random graphs with $L = V - 1$ edges, discussed in Chapter 3. Chapters 4 and 5 deal with some even more specific models: regular random graphs which have a constant degree across all vertices, but are completely random otherwise, and a random graph model with restricted vertex degrees intended to exhibit a geometrical phase transition.

1.3.1 Erdős-Rényi graphs

The simplest example of a random graph is the classic model proposed by Erdős and Rényi in 1959 [14], in which we fix only the number of vertices V and the number of edges L. That is, we take V vertices and connect them with L edges at random. Alternatively, one can think of creating the network by choosing uniformly at random from the set of all possible graphs with exactly V vertices and L edges. The original approach is restricted to simple graphs only, however, a variant with multi- and self-edges is also possible.

The Erdős-Rényi (ER) model is closely related to the so-called binomial model, in which we fix not the number but the probability of edges between vertices (see Refs. [32, 33]). Again we have V vertices, but now we place an edge between each distinct pair with independent probability. Thus, the number of edges is not fixed anymore. In the large-size limit these two models are equivalent (see, e.g., Refs. [21] or [24]), so we may use the latter one, which is considerably easier to handle mathematically, to infer the properties of the ER model. For example, one can show (see, e.g., Refs. [34] or [8]) that the degree distribution in ER random graphs is given by the Poisson distribution

$$p_k = \mathrm{e}^{-z}\frac{z^k}{k!}, \tag{1.16}$$

where z is the mean degree given by Eq. (1.3): $z = 2L/V$. In fact, ER graphs are maximally random simple graphs under a single constraint of fixed mean degree.

In mathematical literature the term random graph means by definition a graph with Poisson distribution of connections. Here, we prefer to refer to it as *classic* or *Poisson random graph* and reserve the general term *random graph* for the maximally random graph with arbitrary degree distribution.

Although classic random graphs are one of the most widely studied model networks, they have one serious shortcoming: in many ways they are completely unlike real-world networks. For instance, there is no correlation between the degrees of vertices, necessarily due to their completely random formation process. In actual networks, by contrast, the degrees are usually correlated, as mentioned in the Preface. Another severe divergence is the shape of the degree distribution. In typical networks encountered in reality most of vertices have low degrees, yet there is a small fraction of high-degree hubs in the tail of the distribution. The Poisson random graphs have, on the other hand, a degree distribution which rapidly decreases at large degrees and is not heavy-tailed. These differences have a profound effect on various properties of the networks and make classic random graphs inadequate when it comes to explain most of the interesting phenomena of real-world systems. Therefore, in our studies we concentrate on the most general random graph model allowing for arbitrary degree distribution.

1.3.2 Statistical ensemble of random graphs

The random graph model is defined in terms of the statistical ensemble, which is constructed by ascribing a statistical weight to each graph from the given set, proportional to its occurrence probability in random sampling (see Refs. [19, 22, 24, 35]). When we talk about the properties of random graphs we mean the expectation values of the observables defined on the statistical ensemble. These are calculated as weighted averages over all graphs in the ensemble. For example, the average of an observable $\mathcal{O}$ over the ensemble of random graphs $\mathcal{G}$ characterized by the partition function Z is given by

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_{G \in \mathcal{G}} \mathcal{O}(G)\, P(G), \tag{1.17}$$

where $P(G)$ is the statistical weight associated with graph G, and the partition function equals

$$Z = \sum_{G \in \mathcal{G}} P(G). \tag{1.18}$$

In the above definition of $\langle \mathcal{O} \rangle$ we have assumed that the observable in question is well defined on all graphs from the ensemble $\mathcal{G}$. This is, in general, not always the case. In such situations we restrict the sum in Eq. (1.17) to the subset of $\mathcal{G}$ containing only those graphs for which $\mathcal{O}(G)$ can be measured.

G						
$n(G)$	72	36	72	36	18	9
$\pi(G)$	1/2	1/4	1/2	1/4	1/8	1/16

Table 1.1: Canonical ensemble of pseudographs with $V = 3$ vertices and $L = 2$ edges. $n(G)$ indicates the number of all distinct labelings of each of the topologies, and $\pi(G)$ is the corresponding configuration space weight.

The statistical weight of a graph $P(G)$ is often distinguished into two parts: the *configuration space weight* $\pi(G)$ and the *functional weight* $w(G)$. The configurational weight is proportional to the uniform probability measure defined on the configuration space, i.e., a collection of graphs. It tells us how to choose graphs from a given set with equal probability. One has to be careful which graphs are meant as equiprobable. For a given number of vertices there are namely two common choices, depending on whether the vertices are distinguishable or not. If we consider all nodes as equal we may define all topologies or shapes to be equiprobable. Alternatively, one may treat the nodes as distinct by attaching labels to them and consider such labeled graphs to have the same probability (see Fig. 1.4). Because the number of possible labelings of a graph depends on its topology, these two approaches result in two different probability measures. When it comes to numerical simulations the latter definition is more natural, so we will stick to the labeled graphs rather than to the bare topologies.

There are $V!$ ways to permute the indices of a graph with V vertices, so it is convenient to choose the configurational weight of each labeled graph equal to $1/V!$. The weight of a topology G is then given by the ratio of the number of its distinct labelings $n(G)$ and the $V!$ factor: $\pi(G) = n(G)/V!$.

In the case of pseudographs the definition of the configuration space weight becomes even more complex. In the presence of multiple connections and self-links the edges are no longer unambiguously specified by its endpoints. Thus, we may label them as well and only treat such fully labeled graphs as equiprobable. Their configurational weight becomes then $1/V!(2L)!$, as we have to account for the permutations of edges. However, when multiple edges and loops are present or when the graph possesses some special symmetries, the number of distinct labelings is smaller than that, which is illustrated on the example of graphs with three vertices and two edges in Tab. 1.1. One can show (see, e.g., Ref. [24]) that the configuration space weight of the topologies equals the inverse of the symmetry factor of Feynman diagrams discussed in Sec. 1.5.2.

Let us now come back to the functional weight $w(G)$. It depends ex-

plicitly on the graph's topology and is typically assumed to factorize into one-point weights w_{q_i} which depend solely on the i-th node degree q_i,

$$w(G) = \prod_{i=1}^{V} w_{q_i}. \tag{1.19}$$

Allowing for both the configuration space weight and the functional weight, the partition function of the ensemble of graphs given by Eq. (1.18) takes the form

$$Z = \sum_{G \in \mathcal{G}_{fl}} \frac{1}{V!(2L)!}\, w(G) = \sum_{G \in \mathcal{G}} \frac{n(G)}{V!(2L)!}\, w(G) = \sum_{G \in \mathcal{G}} \frac{1}{s(G)}\, w(G). \tag{1.20}$$

By $\mathcal{G}_{fl}$ and $\mathcal{G}$ we mean the ensembles of fully labeled and unlabeled graphs, respectively, and $s(G)$ is the mentioned symmetry factor. For quantities which depend on the topology of the graph but not on the labeling of nodes, the average (1.17) becomes

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \sum_{G \in \mathcal{G}} \mathcal{O}(G)\, \frac{1}{s(G)} \prod_{i=1}^{V} w_{q_i}, \tag{1.21}$$

where we have used the explicit expression (1.19) for the functional weight.

1.4 Correlations

Correlation is the measure of dependence between observables in a system. Its knowledge is interesting and useful because from the practical point of view correlation means additional information. If two quantities are correlated, the knowledge of one of them implies certain information about the other one. In a physical system correlation usually indicates the presence of interactions between its parts. A model example is the Ising spin system in which local interactions induce long-range correlations leading to a phase transition.

The situation in random graphs is somewhat different. It is known that for fluctuating geometries even in the absence of any explicit terms inducing interactions between vertices their degree may be correlated. These correlations are generated by model constraints rather than by direct interactions.

The observables of our particular interest are vertex degrees of random graphs and spins of an Ising model defined on these graphs. We investigate if there is any correlation between the adjoining vertices and vertices at some distance apart, and how this relation changes with the distance.

1.4.1 Two-point correlation functions

Correlations can be captured and quantified in a number of ways. Typically, for a system with fixed coordinates they are expressed by the two-point correlation function (see, e.g., Ref. [36]). It is defined for some observables of interest A and B by the thermal average

$$G^{AB}(\mathbf{i},\mathbf{j}) \equiv \langle A_i \cdot B_j \rangle \,, \tag{1.22}$$

where A_i and B_j are the observables' values at sites i and j, respectively. In general G^{AB} depends on the position vectors $\mathbf{i}$ and $\mathbf{j}$ of the corresponding sites. However, for a transitionally invariant and isotropic system it becomes a function of the distance $|\mathbf{i}-\mathbf{j}| = r$ between the observables in question,

$$G^{AB}(r) = \langle A(0) \cdot B(r) \rangle \,. \tag{1.23}$$

This is the case of a system on a regular lattice, although a finite lattice cannot be strictly isotropic. Nevertheless, it appears to be so if its directions are all equivalent, and the system is probed at a much lower resolution than that required to resolve individual sites.

Compared to systems with a fixed metric and coordinates, the problem of defining a correlation function on random geometries is much more challenging (see Ref. [37]). This is because in fluctuating geometries the distance between any two points is constantly changing and it is not possible to consider fixed points at some distance apart, as in Eq. (1.23). Instead, a global correlation measure of the system is introduced, defined by the sum over all pairs of points at a given distance. For the observables A and B associated with each vertex it reads

$$G^{AB}(r) \equiv \frac{1}{n}\left\langle \sum_{i,j} A_i B_j \delta_{d(i,j),r} \right\rangle, \tag{1.24}$$

where the subscripts label vertices, A_i and B_j are corresponding observables' values, n is the size of the system, and $d(i,j)$ is the geodesic distance defined as the length of the shortest path connecting vertices i and j. The above average is taken over all instances of the geometry (configurations).

It should be emphasized here that the distance dependent correlator defined above is not just a pure two-point function, as it is the case of fixed lattice models (cf. Refs. [37, 38]). This is due to the fact that the distance $d(i,j)$ does depend on the whole configuration rather than on the endpoints i and j only. Therefore, $G^{AB}(r)$ is highly non-local, which may lead to some interesting and non-intuitive behavior.

The average number of all pairs of points at distance r can be calculated from Eq. (1.24) by setting $A = B = 1$,

$$G^{11}(r) = \frac{1}{n}\left\langle \sum_{ij} \delta_{d(i,j),r} \right\rangle. \tag{1.25}$$

It serves to define the average distance-dependent product

$$\langle AB(r)\rangle \equiv \frac{G^{AB}(r)}{G^{11}(r)}, \tag{1.26}$$

which on random graphs corresponds to the two-point degree probability distribution described in the following section.

1.4.2 Connected correlation functions

The ordinary two-point correlation function $G^{AB}(r)$ defined in Eq. (1.23) measures the degree of correlation of two quantities, but does not give any information about their fluctuations. If the observables in question tend to be in a specific state the correlator becomes biased by the contribution from the overall ordering of the system. The remedy to this is to use the connected correlation function in which the one-point contribution from the observables separately is subtracted off (see, e.g., Ref. [36]).

The concept of a connected correlation function on random geometries is, due to the lack of fixed metric, highly non-trivial and there is no one standard definition corresponding to the fixed lattice counterpart. Instead, there are different ways to proceed, depending on application (cf. Ref. [37]). One possibility is to use the definition proposed in Ref. [38],

$$\begin{aligned} G_c^{AB}(r) &\equiv \frac{1}{G^{11}(r)}\left\langle \sum_{i,j} \left[A_i - \langle A(r)\rangle\right]\left[B_j - \langle B(r)\rangle\right]\delta_{d(i-j),r}\right\rangle \\ &= \langle AB(r)\rangle - \langle A(r)\rangle\langle B(r)\rangle\,, \end{aligned} \tag{1.27}$$

where

$$\langle A(r)\rangle \equiv \frac{G^{1A}(r)}{G^{11}(r)} \tag{1.28}$$

is the origin-independent average of the observable A over a sphere of radius r. On a fixed lattice this would never depend on the distance, but here it does. The reason for it is that we associate the functions of the geometry with the distance, which is itself a function of the geometry. As apparent from the first line of Eq. (1.27), the connected correlation function measures the correlation between the fluctuations of the observables.

Although $G_c^{AB}(r)$ vanishes at large distances as expected, it does not integrate to susceptibility. This is a desired property, especially for the models discussed in Chapters 4 and 5. As shown in Ref. [39], this can be fixed by calculating fluctuations relative to the global averages $\langle A\rangle$ and $\langle B\rangle$ instead of to their distance-dependent counterparts $\langle A(r)\rangle$ and $\langle B(r)\rangle$,

$$\begin{aligned} G_{c'}^{AB}(r) &\equiv \frac{1}{G^{11}(r)}\left\langle \sum_{i,j} \left(A_i - \langle A\rangle\right)\left(B_j - \langle B\rangle\right)\delta_{d(i-j),r}\right\rangle \\ &= \langle AB(r)\rangle - \langle A\rangle\langle B(r)\rangle - \langle B\rangle\langle A(r)\rangle + \langle A\rangle\langle B\rangle\,. \end{aligned} \tag{1.29}$$

The average $\langle A \rangle$ is the usual mean value of A in a system. Of course, for fixed geometries $\langle A(r) \rangle = \langle A \rangle$, and the two connected correlation functions defined above are identical.

1.4.3 The two-point degree probability distribution

An intuitive correlation measure on an ensemble of graphs is the joint distribution $p_{q,r}(l)$ describing the probability that a pair of vertices at distance l apart has degrees q and r,

$$p_{q,r}(l) \equiv \left\langle \frac{n_{q,r}(l)}{n(l)} \right\rangle, \tag{1.30}$$

where $n_{q,r}(l)$ is the number of pairs of points in a graph which have degrees (q, r) and are separated by the distance l,

$$n_{q,r}(l) \equiv \sum_{i,j \in G} \delta_{k_i,q} \delta_{k_j,r} \delta_{d(i-j),l}, \tag{1.31}$$

and $n(l)$ is the total number of all pairs of vertices at distance l,

$$n(l) \equiv \sum_{i,j \in G} \delta_{d(i-j),l} = \sum_{q,r} n_{q,r}(l). \tag{1.32}$$

The two-point degree distribution $p_{q,r}(l)$ is in fact a generalization of the joint distribution of directly connected vertices proposed in Refs. [40–42] to distances greater than one. We discuss the special case of $p_{q,r}(1)$ in the following section in the context of nearest-neighbor correlations.

The average $\langle \cdot \rangle$ in the definition of the joint probability (1.30) is taken over the ensemble including only graphs with diameter equal l or more for which $n(l)$ is non-zero. However, this is not the only possibility of defining $p_{q,r}(l)$. Another one is to take pairs of vertices from the collection of all configurations and first calculate the averages $\langle n_{q,r}(l) \rangle$ and $\langle n(l) \rangle$ separately, and only then take their ratio,

$$\tilde{p}_{q,r}(l) \equiv \frac{\langle n_{q,r}(l) \rangle}{\langle n(l) \rangle}. \tag{1.33}$$

This definition corresponds to the two-point correlation function (1.26) from the previous section with $A = \delta_{k_i,q}$ and $B = \delta_{k_j,r}$, i.e.,

$$\tilde{p}_{q,r}(l) = \left\langle \delta_{k_i,q} \delta_{k_j,r} \right\rangle (l). \tag{1.34}$$

Although the first definition (1.30) seems to be more natural in the context of random graphs, it is much more difficult to work with. Fortunately, in most of the typical cases relative fluctuations disappear in the large-volume limit and the properties of the whole ensemble can be inferred from just

one large graph (see Refs. [1, 2] and Ref. [43]). For such systems, called *self-averaging*, $p_{q,r}(l)$ and $\tilde{p}_{q,r}(l)$ are asymptotically equal

$$\left\langle \frac{n_{q,r}(l)}{n(l)} \right\rangle \sim \frac{\langle n_{q,r}(l) \rangle}{\langle n(l) \rangle}. \tag{1.35}$$

For independent vertex degrees, the probability $p_{q,r}(l)$ should factorize,

$$p_{q,r}(l) = p_q(l)\, p_r(l), \tag{1.36}$$

where

$$p_q(l) = \sum_r p_{q,r}(l) = \sum_r \left\langle \frac{n_{q,r}(l)}{n(l)} \right\rangle = \left\langle \frac{n_q(l)}{n(l)} \right\rangle \tag{1.37}$$

is the probability that the degree of either of the two vertices at distance l is q, and $n_q(l)$ is to be interpreted as a number of pairs of points at distance l such that one of them has degree q. Using the above relation we can also write Eq. (1.36) as

$$\left\langle \frac{n_{q,r}(l)}{n(l)} \right\rangle = \left\langle \frac{n_q(l)}{n(l)} \right\rangle \left\langle \frac{n_r(l)}{n(l)} \right\rangle. \tag{1.38}$$

One should, however, keep in mind that this defines the lack of correlations in the *ensemble* of graphs. A more appropriate question might be: are the vertices of *individual* graphs uncorrelated? The condition for absence of correlations between vertices in each individual graph G is

$$\frac{n_{q,r}(l)_G}{n(l)_G} = \frac{n_q(l)_G\, n_r(l)_G}{[n(l)_G]^2}, \tag{1.39}$$

or, after averaging,

$$\left\langle \frac{n_{q,r}(l)}{n(l)} \right\rangle = \left\langle \frac{n_q(l)\, n_r(l)}{[n(l)]^2} \right\rangle. \tag{1.40}$$

As already pointed out, for a large class of ensembles conditions (1.38) and (1.40) are equivalent in the large-volume limit.

In practice, checking the condition (1.36) is tricky as it entails measuring a two-dimensional distribution with good accuracy. Therefore, we will use another quantity to express dependencies between nodes in a graph: the average degree of the vertices at distance l from a vertex of a given degree q,

$$\bar{k}_l(q) \equiv \sum_r \left\langle \frac{r\, n_{q,r}(l)}{n_q(l)} \right\rangle. \tag{1.41}$$

It is defined only for those degrees q and at distances l for which the denominator $n_q(l)$ is non-zero. This approach is a generalization of the idea proposed in Ref. [40] where only correlations at distance $l = 1$, i.e., between

degrees at the ends of individual edges, were considered. With reference to the original formulation $\bar{k}_l(q)$ can be also interpreted as the first moment of the conditional probability

$$p(r|q;l) = \frac{p_{q,r}(l)}{p_q(l)}. \tag{1.42}$$

Namely, using the above definition we may write

$$\sum_r r\, p(r|q;l) = \sum_r \frac{r\, p_{q,r}(l)}{p_q(l)} = \sum_r \left\langle \frac{r\, n_{q,r}(l)}{n(l)} \right\rangle \left\langle \frac{n(l)}{n_q(l)} \right\rangle, \tag{1.43}$$

which, assuming self-averaging, yields

$$\sum_r r\, p(r|q;l) \sim \sum_r \left\langle \frac{r\, n_{q,r}(l)}{n_q(l)} \right\rangle = \bar{k}_l(q). \tag{1.44}$$

If vertex degrees are independent, $\bar{k}_l(q)$ does not depend on the degree q and takes a constant value for a given distance. From the relation (1.40) we have

$$\bar{k}_l(q) = \sum_r r \left\langle \frac{n_r(l)}{n(l)} \right\rangle. \tag{1.45}$$

Any deviation from this regularity signals the presence of two-point correlations. Note that for $l = 0$ Eq. (1.45) simply reduces to the average vertex degree z,

$$\bar{k}_0(q) = \sum_r r p_r = z, \tag{1.46}$$

since $n_r(0) = n_r$ is the number of vertices with degree r, and $n(0) = n$ is the total number of the graph's nodes.

Degree-degree correlations may be also conveniently expressed using the notion of connected two-point probability

$$\tilde{p}^c_{q,r}(l) \equiv \tilde{p}_{q,r}(l) - \tilde{p}_q(l)\, \tilde{p}_r(l), \tag{1.47}$$

used to define the connected degree correlation function (cf. Refs. [37, 38, 44]),

$$\tilde{p}^c_{\bar{q},\bar{r}}(l) \equiv \sum_{q,r} q\, r\, \tilde{p}^c_{q,r}(l). \tag{1.48}$$

1.4.4 Nearest-neighbor correlations

Let us now focus on correlations between nearest neighbors, i.e., vertices directly connected. We devote them special attention in Chapter 2, thus, it will be convenient to introduce a shorthand notation for the quantities defined in the previous section. Namely, we will omit the distance argument if equal one, i.e.,

$$p_{q,r} \equiv p_{q,r}(1), \quad \tilde{p}_{q,r} \equiv \tilde{p}_{q,r}(1). \tag{1.49}$$

The joint probability $p_{q,r}$ can be used to express the probability that a randomly chosen edge connects two vertices of degrees q and r, which is given by the symmetric function $(2-\delta_{q,r})p_{q,r}$. If we assume self-averaging, we have in the asymptotic limit

$$p_{q,r} = \left\langle \frac{n_{q,r}}{2L} \right\rangle \sim \frac{\langle n_{q,r} \rangle}{\langle 2L \rangle} = \tilde{p}_{q,r}, \tag{1.50}$$

where $n_{q,r} \equiv n_{q,r}(1)$ is simply the number of links whose ends have degrees q and r. We use the representation in which each of the L undirected edges is counted twice as a pair of two opposite directed ones. Therefore,

$$n_{q,r} = n_{r,q}, \quad \sum_{q,r} n_{q,r} = 2L, \quad \text{and} \quad \sum_{r} n_{q,r} = q n_q. \tag{1.51}$$

Please note that for the canonical ensemble of graphs, in which the number of links is fixed, Eq. (1.50) turns into an exact equality.

We will express correlations by means of the nearest-neighbor average degree. From Eq. (1.41) one finds

$$\bar{k}(q) \equiv \sum_{r} r \left\langle \frac{n_{q,r}}{q\, n_q} \right\rangle, \tag{1.52}$$

because $n_q(1) = q\, n_q$ is the total number of edges emerging from vertices with degree q. As we already know, $\bar{k}(q)$ can be alternatively expressed using the conditional probability (1.42),

$$\bar{k}(q) \sim \sum_{r} r\; p(r|q) = \frac{\sum_r r\, p_{q,r}}{\sum_r p_{q,r}}. \tag{1.53}$$

In the absence of correlations between nearest neighbors, Eq. (1.40) takes the form

$$\left\langle \frac{n_{q,r}}{2L} \right\rangle = qr \left\langle \frac{n_q\; n_r}{(2L)^2} \right\rangle. \tag{1.54}$$

This implies that $\bar{k}(q)$ should then equal

$$\bar{k}(q) = \sum_{k} k^2 \left\langle \frac{n_k}{2L} \right\rangle \sim \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{1.55}$$

It is interesting to note that the average degree of a neighbor $\bar{k}(q)$ is thus larger than the average vertex degree $\langle k \rangle$. This can be seen from their difference,

$$\frac{\langle k^2 \rangle}{\langle k \rangle} - \langle k \rangle = \frac{1}{\langle k \rangle} \big(\langle k^2 \rangle - \langle k \rangle^2 \big) = \frac{\sigma_k^2}{\langle k \rangle} \geq 0, \tag{1.56}$$

because both σ_k^2, which is the variance of the degree distribution, and $\langle k \rangle$ are non-negative. This result is very counter-intuitive: in a social network,

for instance, it translates into your friends having on average more friends than you have!

Any non-trivial dependence of $\bar{k}(q)$ on degree is a signature of correlations between nearest-neighbors. A classification of complex networks according to this property was first proposed by Newman in Ref. [45]. When $\bar{k}(q)$ is an increasing function of q the graph is said to experience *assortative* mixing by degree, while a decreasing function $\bar{k}(q)$ is typical of *disassortative* mixing. In assortative networks highly connected vertices tend to connect to other vertices with many connections, and those with few connections to ones weakly connected. Conversely, disassortative mixing means that high-degree vertices are more probably connected to low-degree ones and vice versa.

1.5 Diagrammatic perturbation expansion

In the analytical studies of the properties of the ensembles of random graphs in Chapters 3–5 we will use methods borrowed from field theory. We formulate the problem as a toy field theory in zero dimensions, referred to by the authors of Refs. [19, 35, 46] as the *minifield theory*. The chief idea is to relate the Feynman diagrams appearing in the graphical perturbative expansion of this theory to the elements of the random graphs' ensemble [19, 35, 46–50].

We start by introducing the generating function of a zero-dimensional field theory toy model. It involves integrals that are computed in power series, each term of which can be put in correspondence with a set of graphs [51], namely the Feynman diagrams. In the following section we illustrate this concept by some examples. Section 1.5.3 addresses the issue of symmetry factors associated with the generated Feynman diagrams. In the last section we introduce the generating function of connected random graphs.

1.5.1 The zero-dimensional field theory

In the general case of a k-component real scalar field $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)$ with sources $J_1, \ldots, J_k$, the normalized vacuum generating function defining the zero-dimensional field theory is given by

$$Z(J) = (2\pi)^{-\frac{k}{2}} \int \mathrm{d}^k \boldsymbol{\phi}\, \mathrm{e}^{-\mathcal{S}}, \tag{1.57}$$

where the integration variables $\phi_1, \ldots, \phi_k$ are just real numbers. The action $\mathcal{S} = \mathcal{S}_0 + \mathcal{S}_I$ is composed of the free part

$$\mathcal{S}_0 = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \phi_i \Delta_{ij}^{-1} \phi_j - \sum_{i=1}^{k} J_i \phi_i \tag{1.58}$$

and the interaction which is typically a polynomial in ϕ,

$$\mathcal{S}_I = -\sum_{i=1}^{k} g_i \sum_n \lambda_{i,n} \frac{\phi_i^n}{n!}, \quad n > 1. \tag{1.59}$$

By assumption g and the real coupling constants $\lambda_{i,n}$ are non-negative. Thus, strictly speaking, the integral in Eq. (1.57) does not exist. Nevertheless, we shall show that after expansion of the exponential $\mathrm{e}^{-\mathcal{S}}$, $Z(J)$ can be treated as a generating function defining the perturbative series. The factorials appearing in Eq. (1.59) are a convention which will prove useful in the Feynman diagram expansion described in the following section.

1.5.2 Feynman rules

The generating function of the ensemble of random graphs is given by the single-component partition function

$$Z(J) = \frac{1}{\sqrt{2\pi\kappa}} \int_{-\infty}^{+\infty} \mathrm{d}\phi \, \exp\left[\frac{1}{\kappa}\left(-\frac{1}{2}\phi^2 + J\phi + \mathrm{e}^{-\mu}\sum_{n>1} w_n \frac{\phi^n}{n!}\right)\right]. \tag{1.60}$$

The auxiliary constant κ introduced in the above definition controls the dependence of Z on the number of loops in the diagrams, which is identical to the power of κ. By expanding the exponential in Eq. (1.60) in powers of $\mathrm{e}^{-\mu}$,

$$\begin{aligned} Z(J) = \frac{1}{\sqrt{2\pi\kappa}} \int_{-\infty}^{+\infty} \mathrm{d}\phi \, \exp\left[\frac{1}{\kappa}\left(-\frac{1}{2}\phi^2 + J\phi\right)\right] \\ \times \left[1 + \frac{\mathrm{e}^{-\mu}}{\kappa}\sum_n w_n \frac{\phi^n}{n!} + \frac{1}{2!}\left(\frac{\mathrm{e}^{-\mu}}{\kappa}\right)^2 \sum_{n,m} w_n w_m \frac{\phi^{n+m}}{n!\,m!} + \cdots\right], \end{aligned} \tag{1.61}$$

we get a well defined series whose terms are consecutive moments of a Gaussian. Each such integral is equal to a sum of contributions which have a graphical representation by Feynman diagrams. These are constructed according to a set of rules mentioned below. A diagram corresponding to a term of order $\mathrm{e}^{-\mu V}$ in Eq. (1.61) consists of V labeled nodes connected pairwise in all possible ways, including self-connections and multiple connections between nodes. Each edge introduces a factor κ and each vertex of degree q introduces a factor $\mathrm{e}^{-\mu} w_q/\kappa$. Note that the factorials in the interaction part of the definition (1.60) are cancelled by all possible $q!$ rearrangements of the labels attached to the edges emerging from the vertex. The total weight of a Feynman diagram D is then

$$w(D) = \frac{1}{s(D)} \kappa^{L-V} \frac{\mathrm{e}^{-\mu V}}{V!} \prod_{i=1}^{V} w_{q_i}, \tag{1.62}$$

where E and V are the numbers of edges and vertices in the graph, respectively. When self-connections and multiple connections are present or the diagram exhibits some symmetries, it has to be multiplied by the so-called symmetry factor $s(D)$. Its origin and the rules of calculating it are described in detail in the following section. In the case of non-degenerate graphs without any symmetries $s(D) = 1$.

According to the above rules the series representation of Eq. (1.61) reads

$$Z(J) = Z_0(J) \sum_{V,L} \frac{\mathrm{e}^{-\mu V}}{V!} \kappa^{L-V} \sum_{D} \frac{1}{s(D)} \prod_{i=1}^{V} w_{q_i}, \tag{1.63}$$

where $Z_0(J)$ stands for the free field generating function

$$Z_0(J) = \frac{1}{\sqrt{2\pi\kappa}} \int_{-\infty}^{+\infty} \mathrm{d}\phi \, \exp\left[\frac{1}{\kappa}\left(-\frac{1}{2}\phi^2 + J\phi\right)\right]. \tag{1.64}$$

Individual terms of the expansion (1.63) can be actually written as appropriate order derivatives of $Z_0(J)$ with respect to J, which leads to an easy correspondence with the graphical representation.

Let the example of 3-regular random graphs, i.e., graphs build only of vertices with degree equal three, serve as an illustration of the method. In this case the partition function of the ensemble is defined by the ϕ^3 field generating function

$$Z(J) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\phi \, \exp\left(-\frac{1}{2}\phi^2 + J\phi + \frac{1}{3!}\phi^3\right), \tag{1.65}$$

where for simplicity we have set κ and w_3 equal one, and $\mu = 0$. The n-th term of the expansion in ϕ^3,

$$Z(J) = \frac{1}{\sqrt{2\pi}} \int \mathrm{d}\phi \, \exp\left(-\frac{1}{2}\phi^2 + J\phi\right) \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{1}{3!}\phi^3\right)^n, \tag{1.66}$$

equals the $3n$-th derivative of $Z_0(J)$, so eventually

$$\begin{aligned} Z(J) &= \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{(3!)^n} \left(\frac{\partial^3}{\partial J^3}\right)^n \int \mathrm{d}\phi \, \exp\left(-\frac{1}{2}\phi^2 + J\phi\right) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{(3!)^n} \left(\frac{\partial^3}{\partial J^3}\right)^n \mathrm{e}^{\frac{1}{2}J^2} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{1}{(3!)^n} \left(\frac{\partial^3}{\partial J^3}\right)^n \sum_{l=0}^{\infty} \frac{1}{l!} \left(\frac{J^2}{2}\right)^l, \end{aligned} \tag{1.67}$$

where we have calculated the Gaussian integral in $Z_0(J)$ and performed series expansion of the resulting exponent.

The above form of $Z(J)$ is especially useful, because it can be directly translated into the diagrammatic notation using Feynman diagrams [36, 52,

Figure 1.7: Graphical notation of $\partial^3/\partial J^3$ and J^2.

53]. Let us represent each 3rd derivative by a vertex with 3 line segments attached, and each J^2 as a line segment between two sources J (see Fig. 1.7). A single derivative $\partial^3/\partial J^3$ acting on a source J annihilates it and connects the now free line end to the segment emerging from the vertex. The diagrams are constructed by matching the derivatives with the sources in all possible ways, which corresponds to connecting the vertices to the J^2 segments. Suppose now that we distinguish J's and their derivatives by attaching distinct labels to them. Then the number of combinations the $3n$ derivatives can act on the $2l$ sources is $(2l)!/(2l-3n)!$. However, many of such generated expressions are algebraically identical and are represented by the same diagram. In the end the total resulting counting factor is

$$n!\,(3!)^n\,l!\,2^l, \tag{1.68}$$

and we would expect it to cancel with the factors from both of the expansions in Eq. (1.67). However, generally this is not the case. That is because this procedure leads to an overcounting of the possibilities. This may happen when some rearrangement of the derivatives leads to the same match-up as a rearrangement of the sources. This is covered by the symmetry factor.

For vacuum bubbles $J = 0$ and in the expansion (1.67) only the term with $2l = 3n$ survives. This corresponds to diagrams constructed from n vertices and $l = 3n/2$ edges by pairing all the free links emerging from the vertices with the ends of the J^2 links. Diagrams representing the first term of the expansion with $n = 2$ are depicted in Fig. 1.8.

1.5.3 The symmetry factor

The counting procedure described in the previous section may lead to an overcounting of the number of diagrams that give identical results, especially in the presence of loops or multiple connections. This happens when some

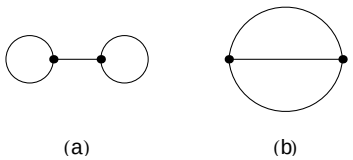

Figure 1.8: Vacuum diagrams of the ϕ^3 theory with $n = 2$.

rearrangement of the derivatives results in the same matching to sources as some rearrangement of the sources. As this is related to some symmetry property of the diagram, the overcounting factor is called the *symmetry factor*.

A general Feynman diagram does not need to be connected and may consist of a product of several connected parts. Let us first focus on determining the symmetry factor of a connected diagram. Each self-connection accounts for a factor of 2, because the exchange of the derivatives at the vertex can be duplicated by the swapping of the edge ends. Similarly, each connection with multiplicity k introduces a factor $k!$, since the $k!$ rearrangements of the links are reproduced by exchanging the derivatives at the vertices. Additionally, if the diagram exhibits some symmetries, its symmetry factor is even larger—see the following examples. These symmetries, however, are not always evident from the picture itself as there are typically many possibilities to draw a diagram on a plane, so its appearance may vary.

In the general case of a diagram build of more than one connected part we also have to take into account the possible exchange of vertices and edges among different connected subdiagrams. However, these will leave the total diagram unchanged only if the exchanges affect different but identical connected parts, and they involve all of their vertices and edges. In other words, there are $n_i!$ ways of rearranging n_i instances of a subdiagram C_i. Thus, the resulting total symmetry factor of the diagram D is

$$s(D) = \prod_i n_i! s(C_i)^{n_i}, \tag{1.69}$$

where the product runs over all connected diagrams.

As an example consider the two diagrams in Fig. 1.8. In the first diagram (Fig. 1.8(a)) each of the two loops introduces a factor of 2. Furthermore, the diagram has a left-right symmetry: the swapping of the vertices along with the loops attached to them can be duplicated by reversing the edge connecting them. Thus, $s = 2^2 \times 2 = 8$. The second diagram (Fig. 1.8(b)) consists of two vertices connected by three edges, which accounts for a factor of 3!. Similarly as in the first example the diagram has a 2-fold symmetry. The endpoints of each edge can be simultaneously swapped, and the effect duplicated by exchanging the vertices, which introduces an additional factor of 2. Altogether, we end up with $s = 3! \times 2 = 12$.

Let us illustrate the present approach by some more complicated cases. Consider for example the diagram in Fig. 1.9(a). It contains two double-links contributing a factor of 2^2. But this diagram exhibits additionally a left-right and a top-bottom symmetry. Namely, if we exchange the two left vertices along with the links connecting them for their right counterparts, and reverse simultaneously the two horizontal links connecting these two parts, we will end up with a diagram identical to the original one. Similarly, the exchange of the upper part of the diagram with the bottom one can

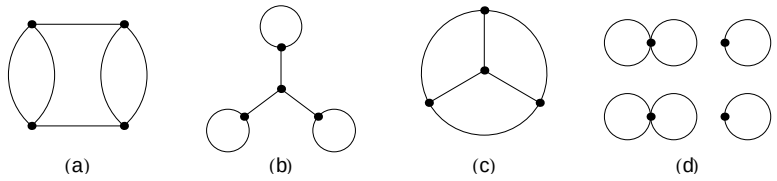

Figure 1.9: Sample Feynman diagrams with $V = 4$ vertices and $L = 6$ links.

be compensated by swapping the vertical edges. Both these symmetries account for a total factor of 2^2 each, resulting in $s = 2^2 \times 2^2 = 16$. Now look at the diagram in Fig. 1.9(b). It has a permutation symmetry accounting for a factor of 3!: the rearrangements of the edges connected to the middle vertex can be duplicated by the exchanges of the outer vertices together with their corresponding loops. This together with the three loops results in $s = 2^3 \times 3! = 48$.

An especially interesting case is the diagram in Fig. 1.9(c). It is a connected 3-regular graph, i.e., a cubic graph. All its four vertices have the same order and are connected to the three remaining vertices, and so they can be rearranged in 4! ways. Since this effect can be duplicated by swapping the edges and there are no other symmetries, the symmetry factor of this diagram is $s = 4! = 24$.

As our last example consider the diagram shown in Fig. 1.9(d). Contrary to all previous diagrams this one comes in more than one connected component. It consists of two different pairs of identical single-vertex diagrams, four separate components in total. Let us first determine the symmetry factors associated with the subdiagrams. The symmetry factor of the one-loop diagram is simply 2. The double-loop of the second diagram introduces a total factor of 2^3, because the exchange of the two loops can be compensated by the rearrangement of the derivatives at the vertex. Recall that each pair of identical diagrams accounts for an additional factor of 2, so finally the total symmetry factor of the diagram is $s = 2 \times 2^2 \times 2 \times (2^3)^2 = 1024$.

1.5.4 Connected diagrams

The generating function $Z(J)$ given by Eq. (1.60) defines an ensemble of all possible graphs, which in general are not connected and consist of a product of several connected components. Following the arguments of Refs. [36] and [52] we can show, however, that the subset of connected graphs only is defined by the logarithm of $Z(J)$.

Let i label the members of the set of all connected diagrams and let C_i stand for the expression corresponding to the i-th diagram, including its symmetry factor. Any disconnected diagram is determined by the number of times n_i each of the connected diagrams appears in it. Individual copies of

identical diagrams can be freely rearranged, so the expression D associated with the general diagram written in terms of the connected ones is

$$D(\{n_i\}) = \prod_i \frac{(C_i)^{n_i}}{n_i!}. \tag{1.70}$$

As the partition function is the sum over all possible diagrams,

$$Z = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots D(\{n_i\}) = \prod_{i=1}^{\infty} \left(\sum_{n_i=0}^{\infty} \frac{(C_i)^{n_i}}{n_i!} \right) = \exp \sum_i C_i, \tag{1.71}$$

we have that $\ln Z$ contains connected diagrams only.

1.6 Equilibrium Monte Carlo simulations

Monte Carlo (MC) simulations are the most important and widely used numerical methods in statistical physics [54]. They have evolved from the idea of statistical sampling, which has a much longer history than the computer and stretches back as far as into the nineteenth century. The name "Monte Carlo" is relatively recent and was coined by John von Neumann, Nicolas Metropolis, and Stanisław Ulam in the 1940s while working on neutron diffusion at the Los Alamos National Laboratory. Being a part of the Manhattan Project their work was secret and so required a code name. This was chosen after the Monte Carlo Casino in Monaco, and first used in the 1949 paper by Metropolis and Ulam [55]. Since then the advance in MC techniques is closely connected with the rapid development of modern computers and the exponential growth of their computational power.

All MC techniques share the same general concept: given some probability distribution π on some configuration space we wish to generate many random samples from π. These methods can be classified as static or dynamic. In the dynamic methods the sequence of generated samples is, unlike in the static methods, not statistically independent but is an outcome of some stochastic process having π as its unique equilibrium distribution. This process simulates the thermal fluctuations of the system from state to state over some period of time. The expectation values of the quantities of interest can be then calculated as time averages over the states the model system passes through, provided that the probabilities of these states equal the real system's weights. This can be achieved by choosing an appropriate dynamics of the simulation, i.e., a rule that governs how to change from one state to another. The main advantage of this method is that we need to sample only a small subset of the configuration space in order to get decent estimates of the quantities of interest. This, unfortunately, necessarily introduces statistical errors.

1.6.1 Importance sampling

The direct calculation of the expectation value $\langle Q \rangle$ given by Eq. (1.7) as a sum over all states of the system is only tractable for small systems. In larger ones we rather average over a subset of states, typically generated by a MC procedure which randomly samples from some given probability distribution p_μ. The estimator of the quantity of interest Q is then the average over the M states the system passes through during the course of a simulation,

$$Q_M = \frac{\sum_{i=1}^{M} Q_{\mu_i} p_{\mu_i}^{-1} \mathrm{e}^{-\beta E_{\mu_i}}}{\sum_{i=1}^{M} p_{\mu_i}^{-1} \mathrm{e}^{-\beta E_{\mu_i}}}. \tag{1.72}$$

It has the property that it becomes more and more accurate as the number of sampled states increases,

$$\lim_{M \to \infty} Q_M = \langle Q \rangle . \tag{1.73}$$

The quality of the estimator Q_M depends not only on the number of sampled states M, but also on the probability distribution p_μ. By picking the states which contribute the most to the sum (1.7) and ignoring the rest we could improve the estimate of $\langle Q \rangle$ significantly, even with only a small number of terms. The idea of picking out only the important states, which is the essence of MC methods, is called *importance sampling*.

As it was put forward in Sec. 1.2, typical systems sample only a very narrow range of configuration space according to the Boltzmann probability distribution given by Eq. (1.5). If the sample states were picked so that each one would appear with its correct Boltzmann probability, then the estimator (1.72) would take a particularly simple form,

$$Q_M = \frac{1}{M} \sum_{i=1}^{M} Q_{\mu_i}. \tag{1.74}$$

This can be achieved by applying the *Markov process* which, given a system in state μ, randomly generates a new system's state ν. The transition probability $P(\mu \to \nu)$ does not vary over time and does not take into account the history of the system, i.e., it depends only on the properties of the current states μ and ν, and not on any previous states the system has passed through. Furthermore, it has to obey the conditions of ergodicity and detailed balance.

The condition of *ergodicity* requires that it should be possible to reach any state of the system from any other state in some finite number of steps. This requirement follows directly from the fact that the Boltzmann probability of each state is non-zero. In practice this means that although we are allowed to set some of the transition probabilities of our Markov process to

zero, we must be careful to leave at least one path of non-zero transition probabilities between any two states.

The actual mechanism that guarantees that the equilibrated system will be governed by the Boltzmann probability distribution, rather than any other distribution, is the condition of *detailed balance*

$$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu). \tag{1.75}$$

It ensures that the overall rate for the transition from state μ into ν is the same as for the reverse transition. Its key feature of the condition of detailed balance is that it forbids dynamics leading to the so-called limit cycles, in which the equilibrium probability distribution rotates around a number of different values changing the occupation probabilities of some or all of the states in a cyclic fashion.

If we now choose the values of p_μ to be the Boltzmann probabilities, we get from the detailed balance condition (1.75) the following constraint

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{p_\nu}{p_\mu} = \mathrm{e}^{-\beta(E_\nu - E_\mu)}, \tag{1.76}$$

which, as well as the ergodicity and the fact that the Markov process must always generate some state ν,

$$\sum_\nu P(\mu \to \nu) = 1, \tag{1.77}$$

restricts the possible choices of transition probabilities. All these constraints together guarantee that the equilibrium distribution in the Markov process will be the desired Boltzmann distribution.

The actual choice of the transition probabilities depends on the particular algorithm used. In the following section we present the Metropolis algorithm, which was used to perform the simulations from the following chapters.

1.6.2 The Metropolis algorithm

The Metropolis algorithm, introduced in a 1953 paper by Nicolas Metropolis and his co-workers [56], is without doubt the most famous and widely used MC method in statistical physics. It is based on the idea of separating the probability of proposing a new configuration of the system from the actual probability of accepting that change. The algorithm works by repeatedly generating a new state, say ν, and then accepting or rejecting it at random with the chosen acceptance ratio. If the state is accepted, the system changes to ν, and if not, it just stays as it is. The whole procedure is repeated again and again. It is the particular choice of the acceptance ratio which characterizes the Metropolis algorithm.

Let us start by splitting the transition probability into two parts,

$$P(\mu \to \nu) = g(\mu \to \nu)A(\mu \to \nu). \tag{1.78}$$

The first one, the selection probability $g(\mu \to \nu)$, is the probability that provided an initial state μ the algorithm will generate a new target state ν, whereas the acceptance ratio $A(\mu \to \nu)$ is the probability that the change of the system to this new state ν will be accepted. We have much of freedom in choosing the selection probabilities $g(\mu \to \nu)$ since the constraint of detailed balance, Eq. (1.76) only fixes their ratio,

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{g(\mu \to \nu)A(\mu \to \nu)}{g(\nu \to \mu)A(\nu \to \mu)}. \tag{1.79}$$

In the Metropolis algorithm we set all the selection probabilities $g(\mu \to \nu)$ for each of the possible states ν equal. When $g(\mu \to \nu) = g(\nu \to \mu)$ the selection probabilities in the above equation cancel and we are left with an appropriate choice of acceptance ratios satisfying Eq. (1.76),

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{A(\mu \to \nu)}{A(\nu \to \mu)} = \mathrm{e}^{-\beta(E_\nu - E_\mu)}. \tag{1.80}$$

The algorithm will be efficient only if the acceptance ratios are the highest possible. Otherwise it will stay most of the time in the state that it is in, instead of sampling a wide range of different states. Since Eq. (1.80) fixes only the ratio $A(\mu \to \nu)/A(\nu \to \mu)$, both acceptance ratios can be multiplied by the same factor. The way to maximize them is to set the larger of the two to one, and have the other one take whatever value is necessary to keep the constraint (1.80) satisfied.

Suppose that from the two considered states μ has higher energy than ν: $E_\mu > E_\nu$. Following the above argumentation we set $A(\nu \to \mu)$, which is the larger of the two acceptance ratios, to one. Then, according to Eq. (1.80), $A(\mu \to \nu) = \mathrm{e}^{-\beta(E_\nu - E_\mu)}$. The acceptance probabilities of the Metropolis algorithm are characterized by

$$A(\mu \to \nu) = \begin{cases} \mathrm{e}^{-\beta(E_\nu - E_\mu)} & \text{if } E_\nu - E_\mu > 0, \\ 1 & \text{otherwise.} \end{cases} \tag{1.81}$$

This means that moves to states with energy lower than or equal to the present one are always accepted, and transitions to higher energy states occur randomly with probability given by the above equation.

1.6.3 Simulating random graphs

Random graphs are generated in MC simulations using a standard thermodynamic approach: the graphs are sampled according to their statistical

weights using the Metropolis algorithm [56]. The statistical weighs depend on the degrees of individual nodes which makes it easy to construct graphs with desired degree distributions. We are exclusively concerned with graphs belonging to the canonical ensemble, that is, graphs with a fixed number of vertices and edges.

Generally speaking, the program works by recursively generating new graphs via modifications of the current one. In the jargon of thermodynamics, the rearrangement of the network can be thought of as a sort of relaxational dynamics in the system, and the final outcome may be treated as an equilibrated random graph. The constraints of fixed numbers of vertices and edges are kept using rewiring of edges. This procedure does not introduce a priori any correlations or bias, provided that the rewiring probability depends only on the degree of the vertices involved, and the rewirings are independent.

1.6.4 Equilibration and measurement

The initial state of a MC simulation is often taken to be some arbitrary configuration, which might be far away from representative. For a system of spins on a lattice these are typically the $T = 0$ or the $T = \infty$ states corresponding to having all spins aligned or completely randomly oriented, respectively. These two choices are popular because they are easy to generate and have a well defined temperature. Although it is not particularly important from which state we start off, a careful choice may reduce the time needed to equilibrate the system.

To get a reliable estimate of any quantity of interest at some non-zero temperature, we must run the simulation for a suitably long time in order to allow our system to come to equilibrium at the temperature we are interested in before we begin with the measurements. This may take some time, but once the system has reached equilibrium, we expect it from the construction of the Metropolis algorithm to stay within a representative subset of states, in which its internal energy and other quantities take a relatively narrow range of values.

Unfortunately, there is no systematic procedure to determine the exact value of the *equilibration time* τ_{eq}. Its rough estimate can be gained by examining the change of some quantities of interest over time from the beginning of the simulation. They will change until the system comes to equilibrium, at which they only fluctuate around some average value.

A potential pitfall of this method is that we may confuse some metastable region in which the system becomes trapped with the true equilibrium. This may happen if the system gets temporarily stuck in a local energy minimum rather than reaching the global minimum. To avoid such situation, we base the estimate of τ_{eq} on more simulations than one. There are two typical ways to proceed: one is to start the simulations from different configurations, and

the other one is to have the same configuration, but run the simulation with different random generator seeds. This technique avoids the problem mentioned above, since it will be apparent from the graph if not all of the systems reached the equilibrium.

Apart from the equilibration time τ_{eq} there is another characteristic time related to MC simulations, namely the *autocorrelation*, or simply *correlation time* τ. It is a consequence of the fact that successive states of a Markov chain are correlated and only after a certain period of time τ they can be regarded as statistically independent. Typically, the correlation time is considerably shorter than the equilibration time, $\tau < \tau_{eq}$.

The time correlation of some observable $\mathcal{O}$ within a Markov chain is characterized by the normalized autocorrelation function $\rho(t)$. It returns the correlation coefficient evaluated between the measurements of $\mathcal{O}$ performed on the equilibrated Markov chain $\{\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_n\}$ and its time-shifted copy,

$$\rho(t) = \frac{\sum_{i=1}^{n-t} \left(\mathcal{O}_i - \bar{\mathcal{O}}\right)\left(\mathcal{O}_{i+t} - \bar{\mathcal{O}}\right)}{\sum_{i=1}^{n} \left(\mathcal{O}_i - \bar{\mathcal{O}}\right)^2}, \tag{1.82}$$

where $\bar{\mathcal{O}}$ is the sample mean,

$$\bar{\mathcal{O}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{O}_i. \tag{1.83}$$

The autocorrelation function takes values $|\rho(t)| \leq 1$, and for uncorrelated data $\rho(t) \approx 0$ for all $t > 0$ (for a comprehensive discussion see, e.g., Ref. [57]). Since successive MC states are very similar, they will have a large positive autocorrelation. On the other hand, for times a long way apart, the states will probably be completely unrelated, and their autocorrelation will be close to zero. In fact, the typical time-scale on which the correlation drops off defines the correlation time. At long distances the autocorrelation is expected to decay exponentially,

$$\rho(t) = \mathrm{e}^{-t/\tau_{exp}}, \tag{1.84}$$

which defines the exponential autocorrelation time

$$\tau_{exp} = \int_0^\infty \mathrm{d}t \; \mathrm{e}^{-t/\tau_{exp}}. \tag{1.85}$$

As it is evident from the above definition, there is still a significant correlation between two states taken time τ_{exp} apart: it is only a factor of $1/\mathrm{e}$ lower than the maximum value at $t = 0$. In order to get independent samples then, we must draw them at greater intervals, preferably $2\tau_{exp}$. This is related to the statistical error and discussed further in the following section.

The time needed to produce a statistically independent sample corresponds to the *integrated correlation time* introduced in Refs. [58–61],

$$\tau_{int} = 1 + 2\sum_{t=1}^{\lambda} \rho(t), \tag{1.86}$$

where the sum is cut off by the parameter λ such that $\tau_{int} < \lambda \ll n$. The reason why the numerical estimator of τ_{int} is defined inside a window of width λ rather than over the whole range of t up to n is that, roughly speaking, for $t \gg \tau_{int}$ each $\rho(t)$ adds a constant amount of noise and little signal (which dies out exponentially). The choice of λ is a tradeoff between the bias and the variance of τ_{int}. The bias can be minimized by taking λ large enough so that $\rho(t)$ is negligible for $t > \lambda$, while at the same time λ should be no larger than necessary, in order to keep the variance small (for the discussion of mathematical details consult Refs. [58, 59]). In practice, the appropriate value of λ can be found using the "automatic windowing" algorithm suggested by the authors of Ref. [58]: choose λ to be the smallest integer satisfying $\lambda \geq c\,\tau_{int}(\lambda)$. Typically, it is reasonable to take c around 2 or 3 (cf. Ref. [60]). This procedure works well provided that a sufficient amount of data is available, at least $n = 1000\,\tau_{int}$.

Usually $\tau_{int} \approx 2\tau_{exp}$, so we will use the integrated autocorrelation time to describe the correlation time τ needed to produce a statistically independent sample. Roughly speaking, the effective number of independent samples in a run of length n is around n/τ.

1.6.5 Error estimation

The accuracy of the outcomes of MC simulations depends on the computational budget involved and it improves with the increase of the simulation length. This is so because the statistical error of virtually all MC methods scales with the number of samples n as $1/\sqrt{n}$, which is an intrinsic consequence of the central limit theorem.

The principal source of statistical error in MC calculations are the fluctuations of the measured quantity from one time step to the other, just like in experiments. This suggests the use of similar tools as we would apply to analyze the results of an experiment performed in laboratory. We must, however, bear in mind that successive samples in the outcome of a Markov process are not necessarily independent, and so the formulas used to evaluate errors of statistically independent data are no longer valid. These correlations are quantified by the integrated correlation time τ defined in Eq. (1.86). One can show that once the equilibrium has been attained, the variance of the sample mean, $\mathrm{var}(\bar{x})$, is by a factor τ larger than it would be for statistically independent data (see, e.g., Refs. [54, 58]).

However, in the case of more complex measures, such as correlation functions of the simulated system, this approach turns out to be insufficient.

A much more versatile and robust method of estimating statistical errors is the *bootstrap method* introduced by Efron in 1979 [62–64]. It is a computer-based resampling procedure used to estimate the properties of a sample estimator, which enjoys the advantage of being completely automatic. It does not require any theoretical calculation and is therefore especially useful for complicated estimators. We briefly sketch the main idea below.

The basic concept behind bootstrapping is to simulate repeated observations from an unknown probability distribution F using the obtained random sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ as a basis. This is done by drawing randomly with replacement from the original sample, called bootstrapping. Suppose our parameter of interest is $\theta = t(F)$ estimated by $\hat{\theta} = s(\mathbf{x})$, and we would like to estimate the standard error of $\hat{\theta}$. This can be done as follows. First, create B bootstrap samples $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$, which are random samples of length n drawn from the original dataset $\mathbf{x}$ (typically, already $B = 100$ is a sufficient number). Next, for each of the bootstrap samples evaluate the corresponding bootstrap replication of $\hat{\theta}$,

$$\hat{\theta}^*(b) = s(\mathbf{x}^{*b}), \quad b = 1, 2, \ldots, B. \tag{1.87}$$

The bootstrap estimate of the standard error of $\hat{\theta}$ is then given by the sample standard deviation of the B replications,

$$\widehat{\mathrm{se}}_B = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left[\hat{\theta}^*(b) - \frac{1}{B} \sum_{b'=1}^{B} \hat{\theta}^*(b') \right]^2 \right\}^{1/2}. \tag{1.88}$$

For a comprehensive introduction to the bootstrap we refer to the book by Efron [65]. Bootstrapping in the case of dependent data, in particular the block bootstrap method which proved useful for our purposes, is described in Ref. [66]. The key idea of block bootstrap is to account for correlations by resampling the original dataset blockwise.

Chapter 2

Connected random graphs

This chapter is based on the paper [1] *Correlations in connected random graphs* by P. Bialas and A. K. Oleś, Phys. Rev. E **77**, 036124 (2008).

In general, random graphs are model networks in which some parameters are fixed, but which are otherwise completely random. A classic example is the ER ensemble in which the only constraint is the fixed number of vertices and links, with the links distributed randomly with uniform probability (but such that no multiple- and self-connections arise). As it was put forward in Sec. 1.3, vertex degrees of ER graphs follow the Poisson distribution. A generalization of this concept are graphs with arbitrary degree distributions (cf. Refs. [19, 20, 35, 67–69]). Random graph models are extremely useful because they serve as the null hypothesis. For example, it was the deviation of data collected on the WWW graph from the predictions of the ER model that triggered the interest in random networks [70]. It implied that the WWW was not created just by linking web pages at random, but required the existence of a preferential attachment mechanism [71].

An important characteristic of random graphs is the absence of correlations between neighboring nodes' degrees, at least for degree distributions without heavy tails (cf. Ref. [72]). Typically, these graphs are not connected and consist of a number of separate components. However, some real-world structures, like the Internet or other communication networks, exist as a single connected graph. It is therefore interesting to study the properties of connected random graphs, i.e., random graphs with the additional constraint of being connected. A simple argument indicates that correlations will appear in such graphs. Namely, a neighbor of a node from which only a single edge emerges must have its degree greater than one, otherwise both these vertices would form a separate connected component. Similarly, all neighbors of a node cannot have their degree equal one, as such a *hedgehog*

(also referred to as a *star graph*) would form a separate cluster as well (see Ref. [44]). This obviously leads to correlations. We have studied these correlations numerically in Ref. [34] and derived analytic formulas describing them in our paper [1]. We repeat all these calculations here and present them in a slightly more comprehensive way than in the original article.

We begin with Sec. 2.1 discussing the emergence of the giant connected component in random graphs. Then, in Sec. 2.2 we introduce the generating function's method which is used throughout the calculations in this chapter. We apply it in Sec. 2.3 to derive formulas for the average degree and degree distribution in the giant and non-giant parts of random graphs. We come to the calculation of correlations in the giant connected component in Sec. 2.4. The results of the preceding sections are illustrated in Sec. 2.5 by the examples of ER graphs, graphs with exponential degree distribution, and scale-free networks. In Sec. 2.6 we show how to deal with maximally random connected graphs by relating them to the giant connected component and present a useful and efficient technique of generating such graphs in MC simulations based on this approach. The chapter ends with Sec. 2.7 addressing the special case of uncorrelated connected random graphs lacking the vertices with degree equal one, i.e., leaves.

2.1 Connected components

The most important property of random graphs with regard to our purposes is the emergence of a single connected component gathering an extensive fraction of all vertices. Consider a random graph with V vertices. When there are no edges at all the graph is completely disconnected and composed of V separate components of exactly one vertex each. In the opposite limit, when there is an edge between each pair of vertices such that every vertex is directly connected to every other, there is only one single component which spans the entire graph. Let us now focus at the size of the largest component in both cases. Apart from the huge size disproportion, there is an important qualitative difference between them: in the first case its size is constant and does not depend on the size of the network, while in the second it is an extensive quantity, i.e., it is proportional to V. When we increase the number of edges gradually from one extreme to the other an interesting thing happens: instead of a smooth transition between these two regimes the largest component undergoes a sudden change akin to a phase transition. Namely, at some particular network density it switches from constant size to extensive size.

It has been shown by the authors of Refs. [67] and [68] that above the percolation threshold given by

$$\langle k^2 \rangle - 2 \langle k \rangle > 0, \tag{2.1}$$

the largest connected component, called the *giant connected component*, becomes extensive and gathers a finite fraction of the whole network. In the case of ER graphs this condition translates into $\langle k \rangle > 1$. This is directly related to the Poisson distribution, which has the characteristic that its variance is equal to its mean, $\langle k^2 \rangle - \langle k \rangle^2 = \langle k \rangle$.

It is now interesting to take a closer look at the structure of the network in the presence of the giant component. One can show using the arguments from Ref. [8] that there must be at most one component whose size increases with the size of the network. The rest of the vertices which do not belong to the giant component are spread among many small components whose average size is constant and does not change with the size of the network. The crucial insight on which the calculations from this chapter are based is that the small components are trees. In other words, they form a forest. This can be seen by the following argument. Consider a small tree of size s. As explained above, any additional edge between the vertices of such a component would create a loop. But the probability of a link emerging from one of the components' vertices connecting back to a vertex from this component is proportional to s/V. For finite s this tends to zero in the large-V limit.

2.2 Generating functions

Analytical results in this chapter are derived using the generating functions method described in Refs. [8] and [68]. The most basic generating function with respect to random graphs is the generating function G_0 for the probability distribution of vertex degrees k. It is defined by the power series

$$G_0(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \ldots = \sum_{k=0}^{\infty} p_k x^k, \tag{2.2}$$

where p_k is the probability that a randomly chosen vertex has degree k. The generating function G_0 encapsulates all information contained in the discrete probability distribution p_k. Individual probabilities p_k can be easily recovered by differentiation:

$$p_k = \frac{1}{k!} \frac{\mathrm{d}^k G_0}{\mathrm{d}x^k}\bigg|_{x=0}. \tag{2.3}$$

Thus $G_0(x)$ is just a different representation of the probability distribution p_k. As it will become apparent from the upcoming calculations, it is much easier to deal with generating functions than with probability distributions.

There are a number of useful properties of generating functions. The average of the distribution can be calculated by differentiating its generating

function and taking $x = 1$. For example, the average degree $\langle k \rangle$ is given by

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k = G_0'(1). \tag{2.4}$$

This generalizes to higher moments which can be calculated from

$$\langle k^n \rangle = \sum_{k=0}^{\infty} k^n p_k = \left[\left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^n G_0(x) \right]_{x=1}. \tag{2.5}$$

Another important property of generating functions which will prove useful is the following: if the distribution of some observable is generated by a given generating function then the distribution a sum of m independent realizations of that observable is generated by the m-th power of that generating function.

Apart from the vertex degree distribution p_k we will be interested in the probability distributions and their generating functions of some other quantities, which shed more light on the structure, especially the connectivity, of random graphs. Suppose that instead of a vertex we pick some edge at random and follow it to the vertex on one of its ends. The number of edges attached to that vertex other than the one we arrived along, i.e., the degree of that vertex less one, is called the excess degree of that vertex. An interesting question now is: what is the probability distribution q_k of the excess degrees? For a graph containing m edges k of them is connected to any particular vertex with degree k, so the probability that a randomly chosen edge leads to such a vertex is $k/2m$ (the factor 2 comes from the fact that we can choose either of the two ends of each edge). The total number of vertices with degree k is given by np_k, so the probability of an edge being connected to any vertex of degree k is

$$np_k \frac{k}{2m} = \frac{kp_k}{\langle k \rangle} = q_{k-1}, \tag{2.6}$$

since the excess degree of a vertex with degree k is $k-1$. We define $G_1(x)$ to be the generating function for the probability distribution of excess degrees,

$$G_1(x) = \sum_{k=0}^{\infty} q_k x^k = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k+1) p_{k+1} x^k = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} k p_k x^{k-1} = \frac{G_0'(x)}{G_0'(1)}, \tag{2.7}$$

where the last equality follows from Eq. (2.4). We will often refer to the generating function $G_1(x)$ in the considerations below.

The generating functions for the small components are of our particular interest. As already discussed in Sec. 2.1, in a random graph there is at most one giant component, and the rest of the vertices is spread over many finite size components. In the large-size limit these small components are

trees. Let $P_0(s)$ be the probability that a randomly chosen vertex belongs to a small component of size s, and let H_0 be its generating function,

$$H_0(x) = \sum_{s=1}^{\infty} P_0(s) x^s. \tag{2.8}$$

Note that since every vertex belongs to a component of size at least one, the above sum starts from $s = 1$. Let us also define the generating function $H_1(s)$ for the probability $P_1(s)$ that the vertex at the end of a random edge belongs to a small component of size s after that edge is removed,

$$H_1(x) = \sum_{s=1}^{\infty} P_1(s) x^s. \tag{2.9}$$

In other words, $H_1(x)$ generates the size distribution of trees attached to a random edge. These consist of a single vertex reached by following the initial edge and may branch into any number of subtrees which share the same size distribution $P_1(s)$. This can be schematically depicted as in Fig. 2.1. The number of branches is given by the excess degree distribution q_k. Using the aforementioned property of generating functions, the total size distribution of the subtrees attached is generated by the powers of the generating function $H_1(x)$, leading to the following identity:

$$H_1(x) = xq_0 + xq_1 H_1(x) + xq_2 \left[H_1(x)\right]^2 + \cdots . \tag{2.10}$$

It can be expressed in a more compact form using the definition of $G_1(x)$ from Eq. (2.7),

$$H_1(x) = x\, G_1\big[H_1(x)\big]. \tag{2.11}$$

If we now take as our starting point some random vertex instead of an edge then we will have one such component at the end of each outgoing edge, resulting in

$$H_0(x) = xp_0 + xp_1 H_1(x) + xp_2 \left[H_1(x)\right]^2 + \cdots , \tag{2.12}$$

which can be also written using $G_0(x)$ defined in Eq. (2.2),

$$H_0(x) = x\, G_0\big[H_1(x)\big]. \tag{2.13}$$

The total probability that a vertex reached by following an edge belongs to a small component of any size is

$$H_1(1) = \sum_{s=1}^{\infty} P_1(s). \tag{2.14}$$

This quantity will occur frequently in subsequent calculations, so it will be convenient to use the shorthand notation

$$u \equiv H_1(1). \tag{2.15}$$

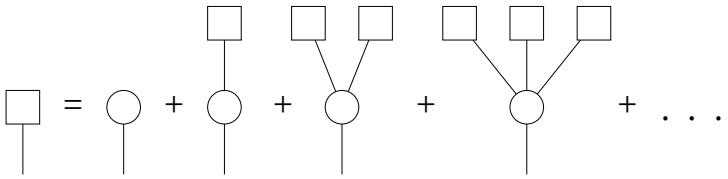

Figure 2.1: Graphical representation of the self-consistency condition (2.10) satisfied by the $H_1(x)$ generating function. The probability of a connected component reached by following a random edge (left-hand side) equals the sum over probabilities of reaching a single vertex, or a vertex connected to one, two, or more such components (right-hand side). Figure reproduced from Ref. [68].

Below the threshold given by Eq. (2.1),

$$\langle k^2 \rangle - 2 \langle k \rangle > 0, \tag{2.16}$$

there is no giant component and all connected components are finite trees, i.e., they form a forest. Then cutting any edge results in two finite parts, thus $u = 1$. In general, however, the probabilities $P_1(x)$ might not add up to one. If an edge belongs to the giant component then cutting it will either result in two infinite parts, or will not split the component at all, accounting for $u < 1$.

The fraction of links connected to a finite part of any size on at least one of their ends is given by u. Now, if an edge belongs to a small component of size s then both its edges lead to finite components. The probability of such situation is generated by $[H_1(x)]^2$, where we have used again the "powers" property of generating functions. The actual probability that a randomly chosen edge belongs to a finite part of arbitrary size, or equivalently the probability that it does not belong to the giant component, is given by u^2.

It follows from Eqs. (2.11) and (2.15) that u is a fixed point of the function $G_1(x)$, that is a point where the function is equal to its own argument,

$$u = G_1(u). \tag{2.17}$$

Looking at the definition of G_1 given by Eq. (2.7) it is easy to note that this equation is always satisfied by $u=1$. However, when condition (2.16) is fulfilled it has a non-trivial solution smaller than 1 as well [68]. As argued, this signals the appearance of the giant component.

Another quantity to which we will often refer to is the probability that a randomly chosen vertex belongs to a small component of any size,

$$h \equiv H_0(1). \tag{2.18}$$

Similarly as for u, it is not necessarily the case that $h = 1$ because there may be a giant component in the network. These two quantities are related by Eq. (2.13):

$$h = G_0(u). \tag{2.19}$$

2.3 Degree distribution in connected components

The average degree of vertices inside the giant component and in the finite remainder can be easily calculated using the results from the previous section. Recall that u^2 is the probability that a randomly chosen edge belongs to a finite component, and h is the probability that so does a random vertex. Consequently, the fraction of edges and vertices in the giant component is given by $1-u^2$ and $1-h$, respectively. Thus, the expressions for the average number of links $\langle L^{(g)} \rangle$ and vertices $\langle V^{(g)} \rangle$ inside the giant component read

$$\langle L^{(g)} \rangle = (1 - u^2)\, L, \tag{2.20}$$
$$\langle V^{(g)} \rangle = (1 - h)\, V, \tag{2.21}$$

and in the finite remainder,

$$\langle L^{(f)} \rangle = u^2 L, \tag{2.22}$$
$$\langle V^{(f)} \rangle = h\, V. \tag{2.23}$$

The average degree in the giant component $z^{(g)}$ and in the rest of the graph $z^{(f)}$ are given by:

$$z^{(g)} = \left\langle \frac{2L^{(g)}}{V^{(g)}} \right\rangle \sim 2\frac{\langle L^{(g)} \rangle}{\langle V^{(g)} \rangle} = z\, \frac{1-u^2}{1-h}, \tag{2.24}$$

$$z^{(f)} = \left\langle \frac{2L^{(f)}}{V^{(f)}} \right\rangle \sim 2\frac{\langle L^{(f)} \rangle}{\langle V^{(f)} \rangle} = z\, \frac{u^2}{h}, \tag{2.25}$$

where $z = 2L/V$ is the average degree of the whole graph.

As it was already pointed out, the giant connected component is not a tree. It contains on average $\langle n_l \rangle$ independent loops, given by

$$\langle n_l \rangle = \langle L^{(g)} \rangle - \langle V^{(g)} \rangle + 1 = \left[\frac{z}{2}(1-u^2) - 1 + h\right] V + 1. \tag{2.26}$$

Because all the remaining connected components are trees, this is also the total number of loops in the whole graph.

The average number of finite connected components $\langle n_c \rangle$ can be calculated using information that they form a forest. The number of links in a forest is $L^{(f)} = V^{(f)} - n_c$, yielding

$$\langle n_c \rangle = \left(h - u^2 \frac{z}{2}\right) V. \tag{2.27}$$

Once $\langle n_c \rangle$ in known, we can derive the formula for the average size $\langle s \rangle$ of the finite connected component,

$$\langle s \rangle = \left\langle \frac{V^{(f)}}{n_c} \right\rangle \sim \frac{\langle V^{(f)} \rangle}{\langle n_c \rangle} = \frac{2h}{2h - u^2 z}. \tag{2.28}$$

At the next step we derive formulas for the degree distribution in the giant component $p_k{}^{(g)}$ and in the small components $p_k{}^{(f)}$. Each vertex belongs either to the giant component or to the finite ones, thus once we have found the distribution in either of them we automatically get the other one from the relation

$$p_k = (1-h)p_k^{(g)} + hp_k^{(f)}. \tag{2.29}$$

The degree distribution in the giant component $p_k{}^{(g)}$ has been already calculated by the authors of Ref. [20] using the maximum entropy principle. However, we find it instructive to re-derive their results applying the generating function method from the previous chapter.

The key idea is to work only on the non-giant components of the graph. We will use a tilde to distinguish the quantities applying to them from the whole-graph counterparts introduced before. The vertex degree distribution and the excess degree distribution in the finite components are generated by:

$$\widetilde{G}_0(x) = \sum_{k=0}^{\infty} p_k^{(f)} x^k, \tag{2.30}$$

$$\widetilde{G}_1(x) = \frac{\widetilde{G}_0'(x)}{\widetilde{G}_0'(1)}. \tag{2.31}$$

Let $\widetilde{H}_1(x)$ be the generating function for the probability $\widetilde{P}_1(s)$ that a randomly chosen vertex from the non-giant part of the graph belongs to component of size s, and $\widetilde{H}_0(x)$ be the generating function for the probability $\widetilde{P}_0(s)$ that an edge leads to such a component. Using similar arguments as before we can show that these functions satisfy

$$\widetilde{H}_1(x) = x\,\widetilde{G}_1\big[\widetilde{H}_1(x)\big], \tag{2.32}$$

$$\widetilde{H}_0(x) = x\,\widetilde{G}_0\big[\widetilde{H}_1(x)\big]. \tag{2.33}$$

The probabilities $\widetilde{P}_0(s)$ and $\widetilde{P}_1(s)$ are actually conditional probabilities: $\widetilde{P}_0(s)$ is the probability that a vertex belongs to a finite component of size s provided that it belongs to some finite component, and $\widetilde{P}_1(s)$ is the probability that a link leads into a finite component of size s provided that it leads into a finite component at all. Thus we can relate them to the previously introduced probabilities $P_0(s)$ and $P_1(s)$ by:

$$\widetilde{P}_0(s) = \frac{P_0(s)}{h}, \tag{2.34}$$

$$\widetilde{P}_1(s) = \frac{P_1(s)}{u}. \tag{2.35}$$

It follows immediately that

$$H_1(x) = \sum_s P_1(s) = \sum_s u\widetilde{P}_1(s) = u\widetilde{H}_1(x), \tag{2.36}$$

$$H_0(x) = \sum_s P_0(s) = \sum_s h\widetilde{P}_0(s) = h\widetilde{H}_0(x). \tag{2.37}$$

We solve Eqs. (2.30) and (2.31) for $p_k{}^{(f)}$ assuming the following form of the vertex degree and the excess degree probabilities inside the non-giant components,

$$p_k^{(f)} = \frac{p_k a^k}{G_0(a)}, \quad q_k^{(f)} = \frac{q_k a^k}{G_1(a)}, \tag{2.38}$$

where $G_0(a)$ and $G_1(a)$ ensure the proper normalization of the distributions. The corresponding generating functions are then

$$\widetilde{G}_0(x) = \frac{G_0(ax)}{G_0(a)}, \tag{2.39}$$

$$\widetilde{G}_1(x) = \frac{G_1(ax)}{G_1(a)}. \tag{2.40}$$

Using the above relation for $G_1(x)$ we can rewrite Eq. (2.32),

$$a\widetilde{H}_1(x) = ax\frac{G_1\big[a\widetilde{H}_1(x)\big]}{G_1(a)}. \tag{2.41}$$

Comparing now Eq. (2.41) to Eq. (2.11) we see that both are consistent if

$$a\widetilde{H}_1(x) = H_1\left[\frac{ax}{G_1(a)}\right]. \tag{2.42}$$

Inserting this into (2.36) we get

$$aH_1(x) = u\,H_1\left[\frac{ax}{G_1(a)}\right], \tag{2.43}$$

because of Eq. (2.17), which can be solved by putting $a=u$.

It can be easily checked that this solution also satisfies Eq. (2.37). Substituting the expression for $\widetilde{G}_0(x) = G_0(ax)/G_0(a)$ into the relation (2.33) and using Eqs. (2.36), (2.19), and (2.13) we get

$$h\widetilde{H}_0(x) = hx\tilde{G}_0\big[\widetilde{H}_1(x)\big] = hx\frac{G_0\big[u\widetilde{H}_1(x)\big]}{G_0(u)} = hx\frac{G_0\big[H_1(x)\big]}{h} = H_0(x). \tag{2.44}$$

We have shown that our initial assumption given by Eqs. (2.38) was indeed correct and is satisfied by $a = u$. The degree distribution in the non-giant components is therefore

$$p_k^{(f)} = \frac{p_k u^k}{h}, \tag{2.45}$$

where we have substituted $G_0(u) = h$. The formula for the degree distribution in the giant component $p_k^{(g)}$ can be obtained from the relation (2.29),

$$p_k^{(g)} = p_k \frac{1-u^k}{1-h}. \tag{2.46}$$

In the limit $u \to 1$ it reduces to

$$\lim_{u\to 1} p_k^{(g)} = \lim_{u\to 1} p_k \frac{1-u^k}{1-G_0(u)} = \lim_{u\to 1} p_k \frac{ku^{k-1}}{\sum_k k p_k u^{k-1}} = \frac{k}{\langle k\rangle} p_k. \tag{2.47}$$

In this limit the connected giant cluster is a tree. Recall that Eq. (2.17) has always the solution $u=1$, which becomes the only one when $G_1'(1) = 1$. This is equivalent to

$$\sum_k k p_k^{(g)} = \sum_k \frac{k^2}{\langle k\rangle} p_k = 2. \tag{2.48}$$

2.4 Correlations inside the giant component

The joint probability distribution $p_{q,r}{}^{(g)}$ inside the giant connected component can be derived using the relation that the total number of links in a graph $n_{q,r}(G)$ is the sum of the number of such links inside the giant component $n_{q,r}{}^{(g)}(G)$ and their number in the non-giant part of the graph $n_{q,r}{}^{(f)}(G)$,

$$n_{q,r}(G) = n_{q,r}^{(g)}(G) + n_{q,r}^{(f)}(G). \tag{2.49}$$

We further assume that there are no correlations between vertex degrees in the graph as a whole and in its finite connected components. This allows us to factorize the edge probabilities $p_{q,r}$ and $p_{q,r}{}^{(f)}$. Recall that edges are represented as opposite directed pairs, i.e.,

$$\sum_{q,r} n_{q,r} = 2L. \tag{2.50}$$

From Eq. (1.38) we have then

$$p_{q,r} = \left\langle \frac{n_{q,r}}{2L} \right\rangle = qr \left\langle \frac{n_q}{2L} \right\rangle \left\langle \frac{n_r}{2L} \right\rangle, \tag{2.51}$$

$$p_{q,r}^{(f)} = \left\langle \frac{n_{q,r}{}^{(f)}}{2L^{(f)}} \right\rangle = qr \left\langle \frac{n_q{}^{(f)}}{2L^{(f)}} \right\rangle \left\langle \frac{n_r{}^{(f)}}{2L^{(f)}} \right\rangle. \tag{2.52}$$

Plugging these relations into Eq. (2.49) and using Eqs. (2.20) and (2.22) for $L^{(g)}$ and $L^{(f)}$, respectively, we get

$$\begin{aligned} p_{q,r}^{(g)} &= \left\langle \frac{n_{q,r}{}^{(g)}}{2L^{(g)}} \right\rangle = \frac{1}{1-u^2} \left(\left\langle \frac{n_{q,r}}{2L} \right\rangle - u^2 \left\langle \frac{n_{q,r}{}^{(f)}}{2L^{(f)}} \right\rangle \right) \\ &= \frac{qr}{1-u^2} \left(\left\langle \frac{n_q}{2L} \right\rangle \left\langle \frac{n_r}{2L} \right\rangle - u^2 \left\langle \frac{n_q{}^{(f)}}{2L^{(f)}} \right\rangle \left\langle \frac{n_r{}^{(f)}}{2L^{(f)}} \right\rangle \right) \\ &= \frac{qr}{z^2(1-u^2)} \left(\left\langle \frac{n_q}{V} \right\rangle \left\langle \frac{n_r}{V} \right\rangle - \frac{1}{u^2} \left\langle \frac{n_q{}^{(f)}}{V} \right\rangle \left\langle \frac{n_r{}^{(f)}}{V} \right\rangle \right). \end{aligned} \tag{2.53}$$

Equations (2.23) and (2.45) imply that

$$\frac{n_k{}^{(f)}}{V} = \frac{p_k^{(f)} V^{(f)}}{V} = p_k u^k, \tag{2.54}$$

so the final formula reads

$$p_{q,r}^{(g)} = \frac{q p_q r p_r}{z^2} \frac{1}{(1-u^2)} \left(1 - \frac{u^q u^r}{u^2}\right). \tag{2.55}$$

We may use this result to obtain $\bar{k}^{(g)}(q)$ from Eq. (1.53),

$$\begin{aligned} \bar{k}^{(g)}(q) &= \frac{\sum_r r\, p_{q,r}^{(g)}}{\sum_r p_{q,r}^{(g)}} = \frac{\langle k^2\rangle - \langle k^2\rangle^{(f)} h u^{q-2}}{\langle k\rangle - \langle k\rangle^{(f)} h u^{q-2}} \\ &= \frac{\langle k^2\rangle}{z} \frac{1}{1-u^q} \left(1 - \frac{\langle k^2\rangle^{(f)}}{\langle k^2\rangle} \frac{z}{z^{(f)}} u^q\right), \end{aligned} \tag{2.56}$$

where the last step follows from the relation between z and $z^{(f)}$ given by Eq. (2.25).

It is now interesting to look what happens to the correlations when u takes its extreme values. On one end when $u \to 0$ all the edges are contained inside the giant part, and the probability of finding an edge leading to a finite connected component vanishes. This corresponds to a graph consisting of one giant connected component and $p_0 V$ isolated vertices only. In the $u \to 0$ limit $\bar{k}^{(g)}(q)$ does not depend on q,

$$\lim_{u\to 0} \bar{k}^{(g)}(q) = \frac{\langle k^2\rangle}{z}, \tag{2.57}$$

so correlations disappear. Since this limit suppresses the formation of leaves, this indicates that leaves might play an important role in correlations inside the giant component. In Sec. 2.7 we investigate this issue further and provide examples backing up this statement.

Conversely, in the opposite limit, when $u \to 1$, the giant connected cluster is a tree, and the above formulas take the form

$$\lim_{u\to 1} p_{q,r}^{(g)} = \frac{q p_q r p_r}{z^2} \lim_{u\to 1} \frac{1-u^{q+r-2}}{1-u^2} = \frac{q p_q r p_r}{z^2} \frac{q+r-2}{2}, \tag{2.58}$$

and

$$\begin{aligned} \lim_{u\to 1} \bar{k}^{(g)}(q) &= \lim_{u\to 1} \frac{\langle k^2\rangle}{z} \frac{1}{1-u^q} \left(1 - \frac{\sum_k k^2 p_k u^k / h}{\langle k^2\rangle} \frac{z}{z u^2/h} u^q\right) \\ &= \frac{1}{z} \lim_{u\to 1} \frac{\langle k^2\rangle - u^{q-2} \sum_k k^2 p_k u^k}{1-u^q} \\ &= \frac{1}{z} \lim_{u\to 1} \frac{(q-2) u^{q-3} \sum_k k^2 p_k u^k + u^{q-2} \sum_k k^3 p_k u^{k-1}}{q u^{q-1}} \\ &= \frac{1}{zq} \big[(q-2)\langle k^2\rangle + \langle k^3\rangle\big]. \end{aligned} \tag{2.59}$$

This proves that the correlations between nearest neighbors do survive in the tree-limit.

2.5 Examples

While deriving the above formulas we have made several assumptions: (i) the vertex degrees are uncorrelated, (ii) the measured quantities are self-averaging, and of course (iii) all the derivations are only valid in the large-V limit. To verify to what extent those assumptions are met and, more importantly, to check the magnitude of the finite size effects, we have compared our analytical predictions to the results of MC simulations of moderate-sized graphs containing $V = 5000$ vertices. The three different random graph models considered included: classic ER graphs, graphs with exponential degree distribution, and scale-free networks governed by a power-law distribution.

The simulations were performed using the Metropolis algorithm introduced in Sec. 1.6.2, as proposed by the authors of Refs. [35] and [24]. It applies the Markov chain MC technique to recursively generate graphs with probabilities proportional to their statistical weights

$$P(\mu) = \prod_{i=1}^{V} w_{q_i}, \tag{2.60}$$

where w_{q_i} is the one-point weight depending on the i-th vertex degree. The acceptance ratio $A(\mu \to \nu)$ for the transformation from graph μ into ν given by Eq. (1.81) takes in this case the following form

$$A(\mu \to \nu) = \min\left\{1, \frac{P(\nu)}{P(\mu)}\right\}. \tag{2.61}$$

We have simulated graphs from the canonical ensemble in which the number of nodes V and links L is constant. The elementary move used to transform the graph consisted of rewiring an edge performed as explained below. First, an edge (i, j) and a vertex k were chosen at random. Then the edge was detached from j and attached back to k, unless $k = i$ or $k = j$. Clearly, such moves do conserve the total number of edges L, and the number of vertices does not change anyway. The only change is in the degrees of vertices j and k: $q_j \to q_j - 1$ and $q_k \to q_k + 1$. Therefore, the probability of accepting a move given by Eq. (2.61) reads

$$A(\mu \to \nu) = \min\left\{1, \frac{w_{q_j-1} w_{q_k+1}}{w_{q_j} w_{q_k}}\right\}. \tag{2.62}$$

One can show (see, e.g., Ref. [21]) that for the canonical ensemble of graphs the desired degree distribution p_k can be obtained in the asymptotic limit simply by choosing the one-points weights to be $w_q = q!\, p_q$. This allows us to generate random graphs with arbitrary degree distributions in a straightforward manner.

The results presented in this chapter were obtained using custom implementations based on Ref. [34] and the program from Ref. [73].

2.5.1 Erdős-Rényi graphs

Let the classic random ER graphs serve as the first example. From the discussion in Sec. 1.3.1 we know that their vertex degrees follow the Poisson distribution

$$p_k = \mathrm{e}^{-z}\frac{z^k}{k!}, \tag{2.63}$$

hence the generating function $G_0(x)$ defined by Eq. (2.2) is given by

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k = \mathrm{e}^{-z}\sum_{k=0}^{\infty}\frac{(zx)^k}{k!} = \mathrm{e}^{z(x-1)}. \tag{2.64}$$

The second generating function $G_1(x)$ can be then readily obtained from Eq. (2.7),

$$G_1(x) = \frac{G_0'(x)}{z} = \mathrm{e}^{z(x-1)}. \tag{2.65}$$

We see that both these functions are equal in this case, $G_0(x) = G_1(x)$. It follows from Eqs. (2.11) and (2.13) that $H_1(x) = H_0(x)$, so $h = u$ with h being the closest to one (from below) positive solution of Eq. (2.17),

$$h = \mathrm{e}^{z(h-1)}. \tag{2.66}$$

When $u = h$ the expressions for the average degree inside the giant component $z^{(g)}$ and in the rest of the graph $z^{(f)}$ given by Eqs. (2.24) and (2.25) simplify to

$$z^{(g)} = z(1+h), \qquad z^{(f)} = zh. \tag{2.67}$$

These are plotted against the average degree of the whole graph in Fig. 2.2 together with data obtained from MC simulations marked as points. The deviation of MC results for a finite size system from the asymptotic formulas does not exceed statistical errors of the simulations, so the agreement between them is perfect. In Fig. 2.3 we similarly compare the simulation results of the degree distribution inside the giant component $p_k{}^{(g)}$ to the analytical formula given by Eq. (2.46). Again, the agreement is very good without any noticeable finite-size effects.

At this point we find it instructive to derive the expression for the degree distribution in the non-giant part of the graph $p_k^{(f)}$ in a simpler and more elegant way than in Sec. 2.3. Namely, we can use the fact that when we omit the giant component from our considerations we are left with a graph with hN vertices and h^2L links on average. Since there are no further restrictions, we can assume that this graph is an ER graph as well. This means that its degree distribution is again a Poissonian, but now with the mean $z^{(f)} = hz$,

$$p_k^{(f)} = \mathrm{e}^{-z^{(f)}}\frac{\left(z^{(f)}\right)^k}{k!} = \mathrm{e}^{-hz}\frac{z^k h^k}{k!}. \tag{2.68}$$

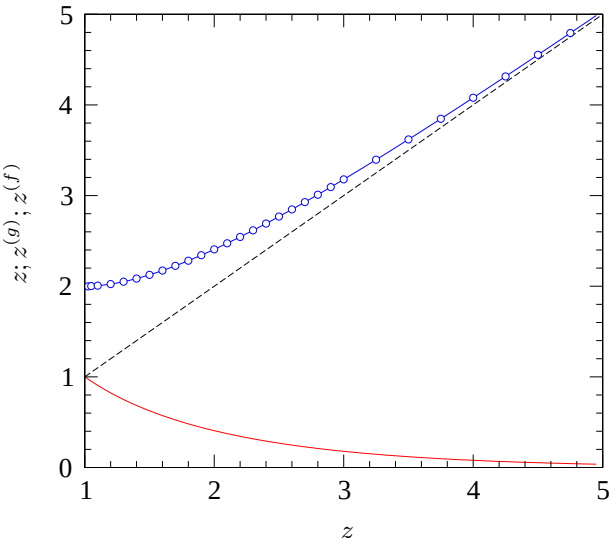

Figure 2.2: Average degree z of full ER graphs (dashed line), their giant connected component $z^{(g)}$ (upper blue solid line), and the rest $z^{(f)}$ (lower red solid line) as a function of z. Circles represent the results of MC simulations.

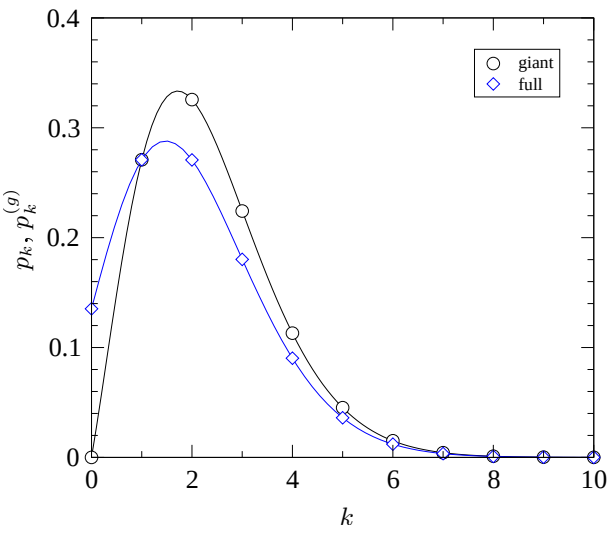

Figure 2.3: Degree distributions p_k and $p_k^{(g)}$ of ER graphs with average degree $z = 2$. Circles and diamonds mark the results of MC simulations for the giant component and for the full graph, respectively. Solid lines represent the corresponding analytical solutions.

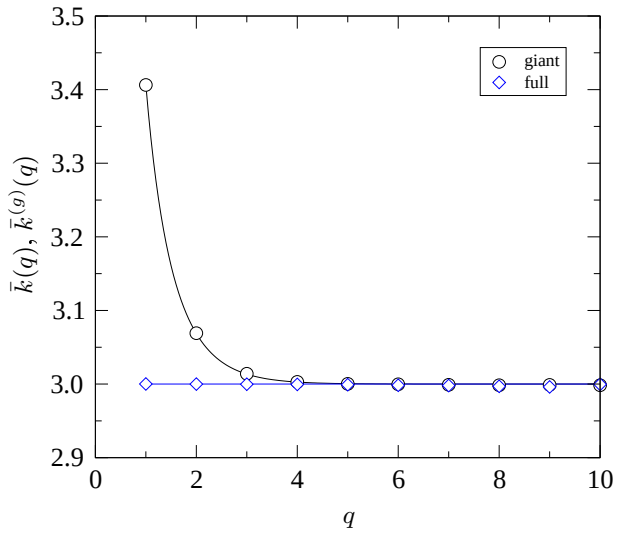

Figure 2.4: Average neighbor degree $\bar{k}(q)$ of ER graphs with average degree $z = 2$. Circles and diamonds mark the results of MC simulations for the giant component and for the full graph, respectively. Solid lines represent the corresponding analytical solutions.

Using the relation (2.66) we obtain the known formula (2.45) with $u = h$,

$$p_k^{(f)} = \mathrm{e}^{-z}\mathrm{e}^{-z(h-1)}\frac{z^k h^k}{k!} = p_k h^{k-1}. \tag{2.69}$$

Correlations in the giant component are characterized by $\bar{k}^{(g)}(q)$. To obtain the corresponding formula we first calculate the degree distribution's second moment

$$\begin{aligned}\left\langle k^2\right\rangle &= \sum_{k=0} k^2 \mathrm{e}^{-z}\frac{z^k}{k!} = \sum_{k=1} k\mathrm{e}^{-z}\frac{z^k}{(k-1)!} = \sum_{k=0}(k+1)\mathrm{e}^{-z}z\frac{z^k}{k!} \\ &= z\left(\sum_k k p_k + \sum_k p_k\right) = z(z+1).\end{aligned} \tag{2.70}$$

The non-giant part of the graph follows the Poisson distribution with the mean $z^{(f)}$, thus $\langle k^2\rangle^{(f)} = z^{(f)}(z^{(f)}+1)$. Then from Eq. (2.56) we get

$$\bar{k}^{(g)}(q) = \frac{z+1}{1-h^q}\left(1 - \frac{zh+1}{z+1}h^q\right). \tag{2.71}$$

A representative plot of the correlation functions $\bar{k}(q)$ and $\bar{k}^{(g)}(q)$ for graphs with $z = 2$ is presented in Fig. 2.4. One can clearly see the appearance of correlations in the giant connected component, as advocated in the introduction. The agreement of the MC results (marked in the graph by points) with the asymptotic formula is again very good.

2.5.2 Exponential degree distribution

The second relatively simple example are graphs whose vertex degrees k are exponentially distributed and proportional to $\mathrm{e}^{-\lambda k}$, where $\lambda > 0$ is some constant parameter. The properly normalized distribution is then

$$p_k = (1 - \mathrm{e}^{-\lambda})\mathrm{e}^{-\lambda k}, \tag{2.72}$$

with the average degree given by

$$z = (1 - \mathrm{e}^{-\lambda}) \sum_{k=0}^{\infty} k\mathrm{e}^{-\lambda k} = \frac{1}{\mathrm{e}^{\lambda} - 1}. \tag{2.73}$$

The generating functions of this distribution are

$$G_0(x) = (1 - \mathrm{e}^{-\lambda}) \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda k} x^k = \frac{1 - \mathrm{e}^{-\lambda}}{1 - x\mathrm{e}^{-\lambda}}, \tag{2.74}$$

$$G_1(x) = \frac{G_0'(x)}{z} = \left(\frac{1 - \mathrm{e}^{-\lambda}}{1 - x\mathrm{e}^{-\lambda}} \right)^2 = G_0^2(x), \tag{2.75}$$

which implies $u = h^2$, where h is the non-trivial solution of Eq. (2.19),

$$h = \frac{1 - \mathrm{e}^{-\lambda}}{1 - h^2\mathrm{e}^{-\lambda}}. \tag{2.76}$$

This solution exists only in the giant component regime. The left-hand side of the condition (2.16) evaluates to

$$\sum_{k=0}^{\infty} k(k-2)p_k = \frac{3 - \mathrm{e}^{\lambda}}{(\mathrm{e}^{\lambda} - 1)^2}, \tag{2.77}$$

from which we find that the giant cluster appears when $\lambda < \ln 3$. Then the solution of Eq. (2.76) reads

$$h = \frac{1}{2} \left(\sqrt{4\mathrm{e}^{\lambda} - 3} - 1 \right), \quad \text{for} \quad 1 < \mathrm{e}^{\lambda} < 3. \tag{2.78}$$

In Fig. 2.5 we plot the average degree of the whole graph z, and its giant and non-giant parts, $z^{(g)}$ and $z^{(f)}$, respectively, against the inverse of the exponential parameter λ, $\kappa = \lambda^{-1}$. As in the previous example of ER graphs, we do not find any visible deviations of the MC simulation results (rendered by points) from the theoretical predictions (solid lines).

The degree distribution inside the giant component $p_k^{(g)}$ and its correlation function $\bar{k}^{(g)}(q)$ together with their whole-graph counterparts are presented in Figs. 2.6 and 2.7, respectively, on the example of graphs with $\lambda = 2/3$. Also in this case we do not trace any finite-size effects. Although vertex degrees in the graphs as a whole are uncorrelated, there are strong disassortative correlations inside the giant component, similarly as in the case of ER graphs.

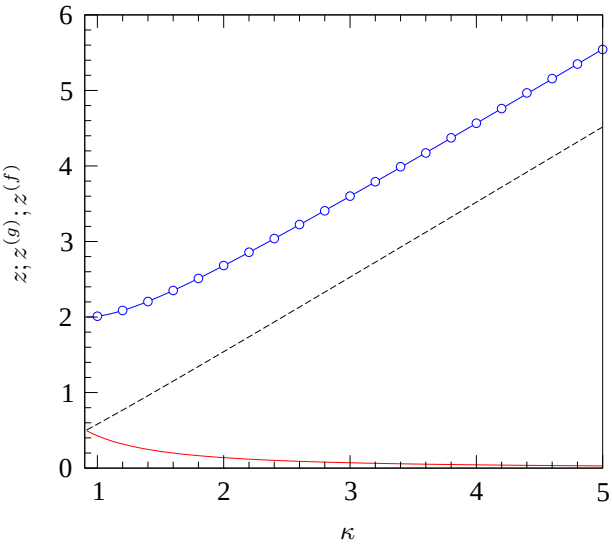

Figure 2.5: Average degree z of graphs with exponential degree distribution (dashed line), average degree z_g of their connected component (upper solid line), and average degree of the rest z_f (lower solid line) as a function of $\kappa = \lambda^{-1}$. Circles mark the results of MC simulations.

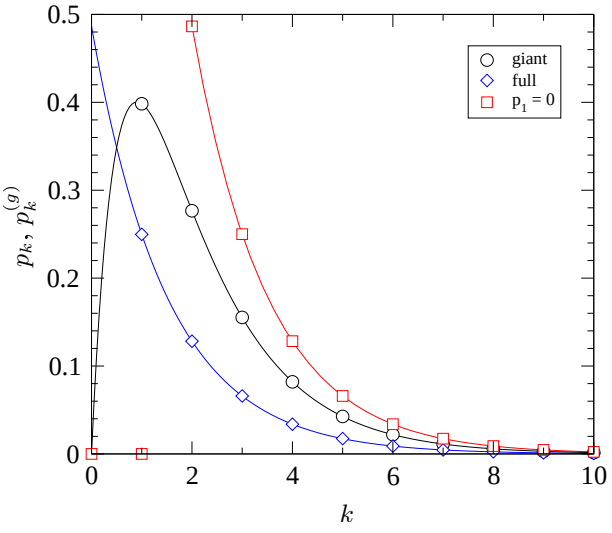

Figure 2.6: Degree distribution of graphs with exponential degree distribution with $\lambda = 2/3$. Circles and diamonds mark the results of MC simulations for the giant component and for the full graph, respectively. Squares stand for the special case of connected graphs without leaves described in Sec. 2.7. Solid lines represent the analytical solutions.

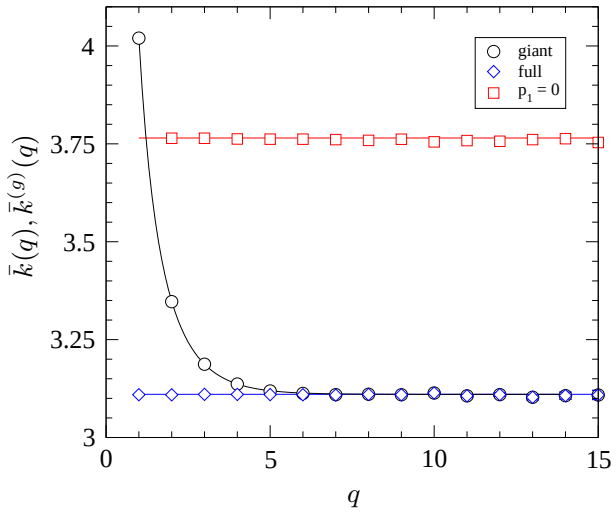

Figure 2.7: Average neighbor degree $\bar{k}(q)$ of graphs with exponential degree distribution with $\lambda = 2/3$. Circles and diamonds mark the results of MC simulations for the giant component and for the full graph, respectively. Squares stand for the special case of connected graphs without leaves described in Sec. 2.7. Solid lines represent the analytical solutions.

2.5.3 Scale-free graphs

The last, but probably the most interesting case are scale-free graphs whose degree distribution follows a power-law function of the form

$$p_k \propto k^{-\beta}, \tag{2.79}$$

where β is a characteristic degree exponent. While studying them we have to consider two scenarios: $2 < \beta \leq 3$, and $\beta > 3$. In the first one we expect correlations between node degrees, as pointed out in Refs. [35, 72, 74]. This invalidates the derivation of $p_{q,r}{}^{(g)}$ given by Eq. (2.55). Additionally, due to unbounded degree fluctuations in the limit of infinite network size, the second moment $\langle k^2 \rangle$ diverges and so $\bar{k}(q)$ and $\bar{k}^{(g)}(q)$ are not defined. Because our aim was to investigate the correlations appearing solely as an effect of the connectedness of graphs, we have decided to concentrate only on the $\beta > 3$ case. The former one is, however, equally interesting and merits further investigation. One line of pursuit is to use the modified configuration model algorithm proposed in Ref. [72] to generate uncorrelated graphs with heavy tails. Then one should obtain predictions at least for the joint probability $p_{q,r}$ which does not contain any divergences. Another possibility would be to use the V-dependent "cut-off" distribution instead of the "full" distribution $p_k \propto k^{-\beta}$, as proposed in Ref. [75]. This would yield V-dependent results, but may not be feasible analytically. For $\beta < 2$ already the first moment of

the distribution p_k is not defined so the generating function approach fails completely.

In finite-size scale-free networks a bound for the possible maximal degree exists. This natural cut-off is the value of degree k_c above which one expects to find at most one vertex, as defined by Dorogovtsev *et al.* [5, 75],

$$V \int_{k_c}^{\infty} p_k \, \mathrm{d}k \sim 1. \tag{2.80}$$

This leads to $k_c(V) \sim V^{1/(\beta-1)}$. We can use this value to estimate the asymptotic behavior of $\langle k^2 \rangle$:

$$\langle k^2 \rangle \approx \sum_k k^2 p_k - \int_{k_c(V)}^{\infty} k^2 p_k \, \mathrm{d}k \approx \langle k^2 \rangle_{\infty} - cV^{-\frac{\beta-3}{\beta-1}}, \tag{2.81}$$

where c is some constant. When $\beta > 3$ the second moment $\langle k^2 \rangle$ is finite, but for β close to 3 Eq. (2.81) converges very slowly. Thus, although there are no correlations, at least in the infinite-size limit [72, 74], we do expect strong finite-size effects near $\beta = 3$. Additionally, MC measurements of degree distribution at large k are always affected by the strong fluctuations due to the poor statistics in this region. The characteristic value above which the fluctuations become strong can be estimated to be around $k_f \sim V^{1/\beta}$ [5, 75]. In simulations one can improve this behavior by increasing the number of runs, nevertheless, one cannot overcome the natural cut-off barrier k_c.

To observe these effects we have simulated a system at $\beta = 13/4$ when $\langle k^2 \rangle$ approaches its asymptotic value as $V^{-1/9}$. The results of our simulations of graphs with 5000 vertices are presented in Figs. 2.8 and 2.9. As expected the data for p_k and $p_k^{(g)}$ distributions show strong cut-off effects around $k_c \approx 44$, but for smaller values of k the agreement with theoretical predictions is rather good.

Looking at the results for $\bar{k}(q)$ we notice two features: (i) Data for the full graph show a deviation from a straight line, indicating the presence of some correlations due to heavy tails. (ii) Data for the giant connected component show a very strong effect of correlations. The agreement with theoretical values is very poor, so we have not included them in the picture. This is due to the described cut-off effect on $\langle k^2 \rangle$. We can obtain a better agreement if we use the actual value of $\langle k^2 \rangle$ measured in simulations in Eq. (2.56) instead of its infinite-volume limit.

2.6 Connected graphs

Finally, we come to the calculation of the properties of the maximally random connected graphs. To this end we assume that the ensemble of the giant connected components of the maximal entropy graphs with distribution p_k

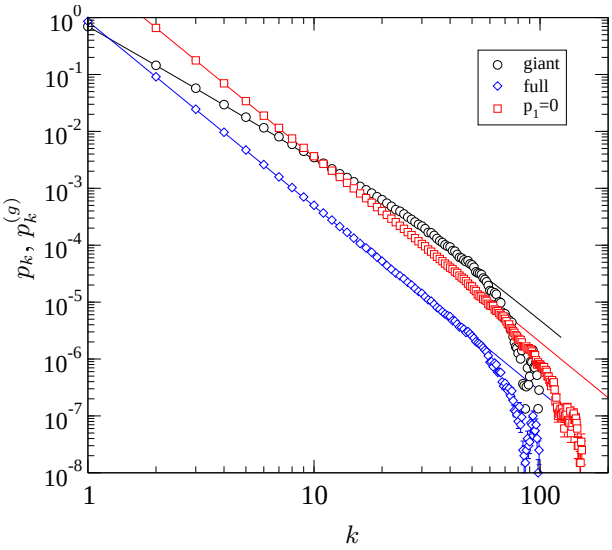

Figure 2.8: Degree distribution of scale-free graphs with $\beta = 3.25$. Circles and diamonds mark the results of MC simulations for the giant component and for the full graph, respectively. Squares stand for the special case of connected graphs without leaves described in Sec. 2.7. Solid lines represent the analytical solutions.

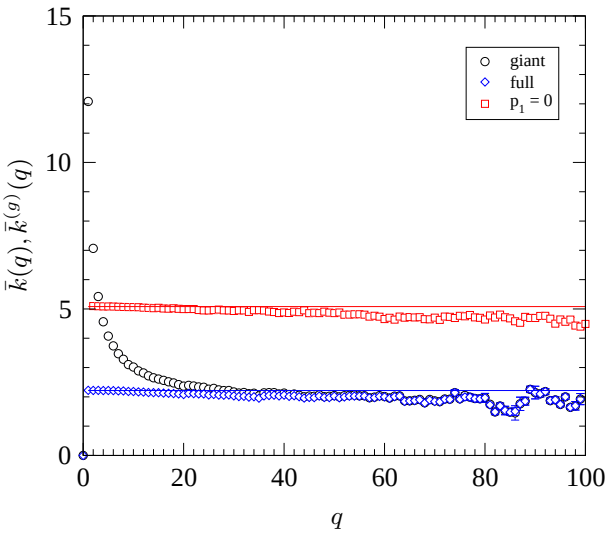

Figure 2.9: $\bar{k}(q)$ for scale-free graphs with $\beta = 3.25$. Circles and diamonds mark the results of MC simulations for the giant component and for the full graph, respectively. Squares stand for the special case of connected graphs without leaves described in Sec. 2.7. Solid lines represent the analytical solutions.

is a maximal entropy ensemble of connected graphs having distribution $p_k^{(g)}$. We neglect the fluctuations in the number of vertices and links in the giant component as they decay in the large-size limit. This seems to be a plausible assumption as there are no additional constraints except connectivity. A more detailed argumentation is provided in Appendix A.1.

On the above assumption we can infer the properties of the maximal entropy connected random graphs with distribution $p_k^{(g)}$ from the properties of the giant components of the maximal entropy random graphs having degree distribution p_k, related to $p_k^{(g)}$ by Eq. (2.46),

$$p_k^{(g)} = p_k \frac{1-u^k}{1-h}. \tag{2.82}$$

This is an opposite situation to what we had before, when we derived the properties of the giant connected component in graphs with some definite degree distribution p_k. Now p_k is unknown and we have to do the reverse—knowing $p_k^{(g)}$ we are seeking for the full graphs' distribution p_k. We can achieve this in a straightforward way by inverting Eq. (2.82) for $k > 0$,

$$p_k = p_k^{(g)} \frac{1-h}{1-u^k}, \tag{2.83}$$

and expressing u by the known distribution $p_k^{(g)}$. Once we find the relation for u we could use the normalization condition to calculate h and the probability of isolated vertices p_0. From Eq. (2.17) we have

$$u = \frac{G_0'(u)}{G_0'(1)} = \frac{\sum_{k=1}^{\infty} p_k^{(g)} \frac{k u^{k-1}}{1-u^k}}{\sum_{k=1}^{\infty} p_k^{(g)} \frac{k}{1-u^k}}. \tag{2.84}$$

Unfortunately, there is no closed-form solution to the above equation. Nevertheless, we can show by the following arguments that is does have a solution, which for a given degree distribution $p_k^{(g)}$ can be found numerically by an iterative procedure.

Let us first rewrite Eq. (2.84) in a more friendly form,

$$\sum_{k=1}^{\infty} p_k^{(g)} k u \frac{1-u^{k-2}}{1-u^k} \equiv g(u) = 0. \tag{2.85}$$

Now recall that the possible values of u are within the range $0 \leq u \leq 1$, where the boundary values correspond to a graph consisting only of a single connected part and possibly a bunch of isolated vertices, or one without the giant component at all, respectively. The values of $g(u)$ in these extreme points can be calculated yielding

$$g(0) = -p_1^{(g)}, \tag{2.86}$$

and

$$\lim_{u\to 1} g(u) = \sum_{k=1}^{\infty} p_k^{(g)}(k-2). \tag{2.87}$$

Because $p_1^{(g)}$ is non-negative and $\sum_{k=1}^{\infty} p_k^{(g)} = z^{(g)} \geq 2$, for connected graphs $g(0)$ is non-positive, and $g(1)$ is non-negative. It follows that equation $g(u) = 0$ given by (2.85) does have a solution within the allowed range of u.

Once we find u we can calculate h from the normalization of the distribution p_k,

$$1 = p_0 + (1-h)\sum_{k=1}^{\infty} \frac{p_k^{(g)}}{1-u^k}, \tag{2.88}$$

and the condition (2.19),

$$h = p_0 + (1-h)\sum_{k=1}^{\infty} \frac{u^k p_k^{(g)}}{1-u^k}. \tag{2.89}$$

As it turns out, however, these two equations are not independent. This can be seen by subtracting Eq. (2.89) from Eq. (2.88). Since

$$\sum_{k=1}^{\infty} \frac{p_k^{(g)}}{1-u^k} - \sum_{k=1}^{\infty} \frac{u^k p_k^{(g)}}{1-u^k} = \sum_{k=1}^{\infty} p_k^{(g)} \frac{1-u^k}{1-u^k} = 1, \tag{2.90}$$

we are eventually left with a tautology. Thus p_0 can be treated as a free parameter and set equal to zero, $p_0 = 0$. Then from Eq. (2.88) we get the final result

$$h = 1 - \left(\sum_{k=1}^{\infty} \frac{p_k^{(g)}}{1-u^k} \right)^{-1}. \tag{2.91}$$

2.6.1 Simulating connected graphs

The procedure described in the previous section may be actually used to generate connected random graphs in an efficient way. Typically, when performing MC simulations of connected graphs with degree distribution $p_k{}^{(g)}$ one would try to simulate them directly by checking on their connectivity after every MC step. This is an $O(V)$ operation as it entails searching through the entire structure. A more optimal approach is to generate graphs with distribution p_k given by Eq. (2.83) and use the giant connected component. Although this still requires calculating the connected parts, now this needs to be done only once before each measurement rather than every elementary move.

As an example, we have generated connected maximally random graphs with Poisson degree distribution

$$p_k^{(g)} = \begin{cases} 0 & \text{for } k = 0, \\ \dfrac{1}{\mathrm{e}^{z^{(g)}} - 1} \dfrac{{z^{(g)}}^k}{k!} & \text{for } k > 0, \end{cases} \tag{2.92}$$

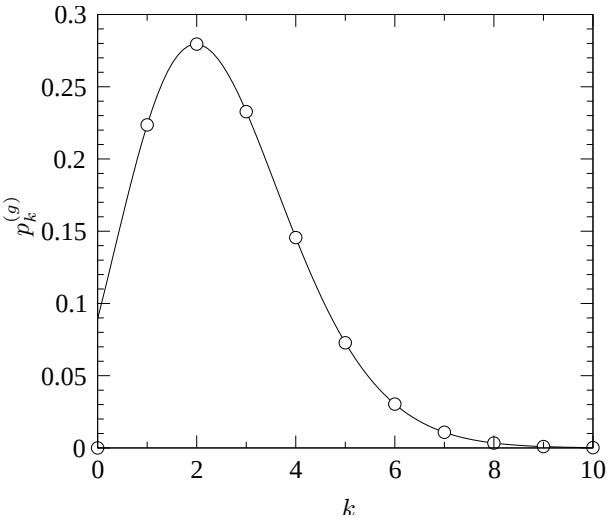

Figure 2.10: Degree distribution $p_k{}^{(g)}$ in the giant connected component. Circles mark the results of MC simulation, while the solid line denotes the desired distribution (2.92).

with $z^{(g)} \approx 2.50$. Note that because of $p_0 = 0$ the normalization is now $(\mathrm{e}^{z^{(g)}} - 1)^{-1}$ instead of e^{-z}. Solving Eq. (2.84) numerically we find that in this case $u \approx 0.1209$, $h \approx 0.0341$, and $z \approx 2.45$. Using the program from Ref. [73] we have simulated a maximally random graph with $5000/(1-h) \approx 5177$ vertices and 6342 links with degree distribution p_k as given by Eq. (2.83). In a draw of 10 000 independent graphs the average measured size of the giant component was 5000.24 ± 0.25 with standard deviation ≈ 20. The results vs. the distribution $p_k{}^{(g)}$ from Eq. (2.92) are presented in Fig. 2.10. As apparent from the plot, the measured degree distribution in the giant connected component agrees very well with the desired one.

2.7 Uncorrelated connected graphs

As stated at the beginning of this chapter and in Sec. 2.4, the nature of the correlations observed in connected random graphs is pure structural and is related to the presence of leaves. If this mechanism is the only one that induces correlations, we expect them to vanish in ensembles where leaves are forbidden.

By setting $p_1 = 0$ we exclude most of the finite connected components. What is left is one giant connected component and $p_0 V$ isolated vertices. This is a simple consequence of the fact that finite connected components are trees—but there are no trees without leaves, except the degenerated ones consisting of a single vertex. Then $u = 0$, since all edges are inside

the giant component and cutting none of them would split it. Moreover, $h = p_0$ because the only vertices from the non-giant part of the graph are these with degree $k = 0$. Therefore, if we are interested in connected graphs, it is sufficient to set $p_0 = 0$. Then all the small single-vertex components vanish and we obtain a graph containing only the giant component, i.e., a connected graph.

In order to verify our assumptions and compare them with the previous results we have carried out MC simulations of the same ensembles as in the examples from Secs. 2.5.2 and 2.5.3, but with p_1 explicitly set to zero. The exponential degree distribution without leaves is defined by

$$p_k = \begin{cases} 0 & \text{for } k \in \{0, 1\}, \\ \dfrac{1 - \mathrm{e}^{-\lambda}}{\mathrm{e}^{-2\lambda}}\, \mathrm{e}^{-\lambda k} & \text{for } k > 1. \end{cases} \tag{2.93}$$

The results for $\lambda = 2/3$ ($z \approx 3.055$) are presented in Figs. 2.6 and 2.7 (squares). As predicted, vertices are uncorrelated in stark contrast to the $p_1 > 0$ case plotted in the same figures. It is worth noting that now the giant component fills on average 99.9% of the whole graph.

We have also performed simulations of the scale-free distribution $p_k \propto 1/k^{13/4}$ without leaves. The results are presented in Figs. 2.8 and 2.9 (squares). We see that correlations are very much suppressed compared to the general case when we permit leaves (presented in the same figures). The slight remaining correlation is due to long tails, as explained in Sec. 2.5.3.

Chapter 3

Generic random trees

This chapter is based on the paper [2] *Long-range disassortative correlations in generic random trees* by P. Bialas and A. K. Oleś, Phys. Rev. E **81**, 041136 (2010).

In the previous chapter we have calculated correlations in connected random graphs concentrating on the joint nearest-neighbor degree probability distribution. We have showed that the degrees of adjoining nodes are indeed correlated and that these correlations are of disassortative type. It is, however, interesting how these correlations propagate to larger distances and whether they survive in the thermodynamic limit.

Unfortunately, the developed formalism of generating functions does not allow a simple derivation of the distance dependent correlation functions. Nevertheless, if we restrict ourselves to connected acyclic graphs, that is, random trees, it is possible to obtain the desired results applying field theory methods described in Sec. 1.5. We take advantage of the chief characteristic of trees, namely the existence of a unique path between any two vertices. It allows us to construct the partition function of trees with two vertices of specific degrees marked at some given distance apart, which would be impossible in the presence of cycles. We can use it to extract the joint probability distribution $\tilde{p}_{q,r}(l)$ introduced in Sec. 1.4.3, which serves us to define the connected correlation functions. As it turns out, correlations remain disassortative at all distances and vanish only as an inverse second power of the distance.

The following section formally introduces the model using the idea from Sec. 1.5.2 of defining the ensemble of maximal entropy random trees by means of a minifield theory. We show how the one-point vertex degree probability distribution p_q can be obtained in this formalism, and proceed on in Sec. 3.2 deriving the distance dependent joint probability distribution

$\tilde{p}_{q,r}(l)$. We use it to calculate the average neighbor degree $\bar{k}_l(q)$, and define the connected correlator $\tilde{p}^c_{\bar{q},\bar{r}}(l)$. Our approach is illustrated in Sec. 3.3 by the examples of ER trees and scale-free trees. In the last section we verify our results for scale-free trees numerically using MC simulations.

3.1 Vertex degree distribution

Consider a canonical ensemble of all labeled trees $\mathcal{T}(V)$ with V vertices. The normalized statistical weight of a tree is given by the product of the configuration space weight $1/V!$ and the functional weight defined as a product of one-point weights w_{q_i} depending on vertex degrees q_i,

$$P(T) = \Omega_V^{-1} \frac{1}{V!} \prod_{i \in T} w_{q_i}, \tag{3.1}$$

where the normalization is set by the partition function

$$\Omega_V \equiv \frac{1}{V!} \sum_{T \in \mathcal{T}(V)} \prod_{i \in T} w_{q_i}. \tag{3.2}$$

The configurational weight is only a matter of convention. As discussed in Sec. 1.3.2, it is introduced primarily to account for the same statistical weight of trees that differ from each other only by a permutation of vertex labels.

We prove in Appendix A.2 that the probability measure (3.1) generates a maximal-entropy ensemble with some definite degree distribution p_q. An important property of the weight $P(T)$ is that it factorizes into a product of individual vertices' weighs, so it does not introduce any explicit correlations between them. Thus any observed correlations arise due to the fact that we restrict ourselves to a specific set of graphs, and not from the measure itself.

We will use the Feynman diagram expansion technique presented in Sec. 1.5 to calculate Ω_V. To this end we introduce the grand-canonical ensemble generating function $\Omega(\mu)$ defined by the discrete Laplace transform of the canonical ensemble's partition function,

$$\Omega(\mu) = \sum_{V=1}^{\infty} \mathrm{e}^{-\mu V} \Omega_V, \tag{3.3}$$

where μ is the chemical potential associated with the nodes. $\Omega(\mu)$ can be calculated from the perturbative expansion of the zero-dimensional field theory generating function given by Eq. (1.60):

$$\mathrm{e}^{\frac{1}{\kappa}\Omega(\mu)} = \frac{1}{\sqrt{2\pi\kappa}} \int \mathrm{d}\phi \exp\left[\frac{1}{\kappa}\left(-\frac{1}{2}\phi^2 + \mathrm{e}^{-\mu} \sum_{q=1} \frac{w_q}{q!}\phi^q\right)\right]. \tag{3.4}$$

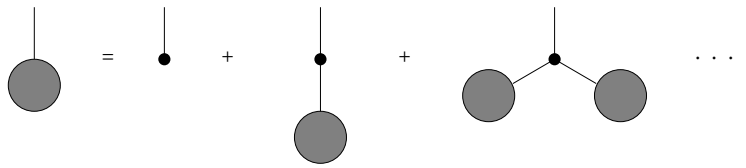

Figure 3.1: Graphical representation of Eq. (3.6). Each gray bubble corresponds to the sum over planted trees given by the partition function $Z(\mu)$. The branch without a vertex (small black circle) at one end is the stem.

We have incorporated the degree-one vertices into the sum by setting $J = \mathrm{e}^{-\mu} w_1$. The expansion in $\mathrm{e}^{-\mu}$ of the above integral generates Feynman diagrams with desired weights and symmetry factors, as described in Sec. 1.5.2. It contains, however, not only trees but all graphs, including ones which are not connected or contain loops. As showed in Sec. 1.5.4 we can restrict the expansion to connected graphs by taking its logarithm. To obtain just trees we use another expansion in κ. According to Feynman rules for the expression (3.4), each edge in the graph introduces a factor κ, and each vertex a factor $1/\kappa$. Together they contribute by a factor κ^{L-V}, where L is the number of edges in the graph. The number of independent loops in the graph is given by $L - V + 1$, thus the expansion in powers of κ is a loop expansion: the leading term groups graphs with no loops, the second one graphs with one loop, and so on.

In the limit $\kappa \to 0$ only tree graphs survive and their contribution is given by the Laplace's approximation (see Appendix B) of the integral in Eq. (3.4). The condition of the stationary value of the action is given by

$$\frac{\mathrm{d}}{\mathrm{d}\phi}\left(-\frac{1}{2}\phi^2 + \mathrm{e}^{-\mu} \sum_{q=1}^{\infty} \frac{w_q}{q!} \phi^q \right) = 0. \tag{3.5}$$

Denoting now by $Z(\mu)$ the solution of the above equation,

$$Z(\mu) = \mathrm{e}^{-\mu} \sum_{q=1}^{\infty} \frac{w_q}{(q-1)!} Z^{q-1}, \tag{3.6}$$

we get from Eq. (3.4) for the grand-canonical partition function

$$\Omega(\mu) = \mathrm{e}^{-\mu} \sum_{q=1}^{\infty} \frac{w_q}{q!} Z^q(\mu) - \frac{1}{2} Z^2(\mu). \tag{3.7}$$

The generating function of trees with one external node marked, called *planted* trees, is given by $\partial\Omega(\mu)/\partial J$. Since

$$Z(\mu) = \frac{\partial \Omega(\mu)}{\partial J}, \tag{3.8}$$

it is evident that $Z(\mu)$ is the partition function of the ensemble of planted trees. Equation (3.6) has a nice graphical interpretation presented in Fig. 3.1, which is often referred to in the literature.

The properties of planted trees and their critical behavior were calculated in Refs. [46, 76, 77]. The model has two geometrically distinct phases: the so-called *generic* or *tree* phase, and the *bush* phase. Let us define the positive definite function

$$F(Z) \equiv \sum_{q=1}^{\infty} \frac{w_q}{(q-1)!} Z^{q-2}, \tag{3.9}$$

and rewrite Eq. (3.6) as

$$\mathrm{e}^{\mu} = F(Z). \tag{3.10}$$

The weights w_q are all non-negative by assumption, which implies that $F'(Z)$ can have at most one zero for positive Z. It will take its minimal value there, since $F''(Z) > 0$ as well. It follows that Eq. (3.10) does not have any solution for $\mu < \mu_0$, where μ_0 corresponds to the minimum given by $F'(Z_0) = 0$. Consequently, the partition function $Z(\mu)$ must have a singularity at μ_0.

In our calculations we restrict ourselves to the generic phase. In this case the minimum of $F(Z)$ is inside its domain and in the vicinity of Z_0 the function is given by

$$F(Z) = F(Z_0) + \frac{1}{2} F''(Z_0)(Z - Z_0)^2 + \cdots, \tag{3.11}$$

from which, after inversion, we get the singular behavior of the partition function around μ_0,

$$Z(\mu) \approx Z_0 - Z_1 \sqrt{\Delta\mu}, \tag{3.12}$$

where

$$\Delta\mu \equiv \mu - \mu_0, \tag{3.13}$$

and

$$Z_1 = \sqrt{\frac{2F(Z_0)}{F''(Z_0)}}. \tag{3.14}$$

The vertex degree distribution of this model was already calculated in Ref. [78] using the correspondence with the balls in boxes model. We re-derive it here following a different method presented in Ref. [2]. This approach is especially useful because it can be easily extended to the case of two-point correlations studied in Ref. [79].

Let us denote by $\Omega(q;\mu)$ the partition function of the rooted grand-canonical ensemble of trees with one vertex of degree q marked,

$$\Omega(q;\mu) = w_q \frac{\partial \Omega(\mu)}{\partial w_q} = \mathrm{e}^{-\mu} \frac{w_q}{q!} Z^q(\mu). \tag{3.15}$$

The graphical interpretation of this equation is presented in Fig. 3.2. As before, gray bubbles correspond to the sum over planted trees given by the

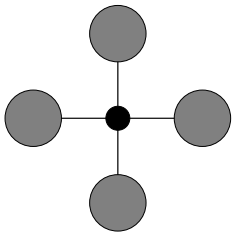

Figure 3.2: Graphical representation of the partition function $\Omega(q,\mu)$ given by Eq. (3.15) for $q = 4$. Each gray bubble corresponds to the sum over planted trees given by the partition function $Z(\mu)$, and the central black circle represents the root.

partition function $Z(\mu)$. The smaller black circle in the middle represents the root, which contributes a weight factor $\mathrm{e}^{-\mu}w_4$. The additional inverse factorial 1/4! comes from the fact that the relative position of branches in the compound tree is irrelevant. We will use $\Omega(q;\mu)$ to calculate the degree distribution p_q, which is proportional to the canonical partition function $\Omega_V(q)$. Using the expansion (3.12) Eq. (3.15) can be rewritten as

$$\Omega(q;\mu) \approx \mathrm{e}^{-\mu}\frac{w_q}{q!}Z_0^q\left(1-\frac{Z_1}{Z_0}\sqrt{\Delta\mu}\right)^q \approx \mathrm{e}^{-\mu}\frac{w_q}{q!}Z_0^q\exp\left(-q\frac{Z_1}{Z_0}\sqrt{\Delta\mu}\right), \tag{3.16}$$

where the second approximation follows from the assumption that $Z_1\sqrt{\Delta\mu}$ is only a small correction to Z_0, that is, $Z_1\sqrt{\Delta\mu} \ll Z_0$. The canonical partition function $\Omega_V(q)$ can be now retrieved from the above equation using the inverse Laplace transform,

$$\mathcal{L}^{-1}\left\{A\mathrm{e}^{-B\sqrt{\mu-\mu_0}}\right\} = \frac{1}{2\sqrt{\pi}}\,\frac{AB}{V^{3/2}}\,\exp\left(-\frac{B^2}{4V}+\mu_0 V\right), \tag{3.17}$$

which in the leading order of V yields

$$\Omega_V(q) = C_V\frac{w_q}{(q-1)!}Z_0^{q-1}. \tag{3.18}$$

Please note that the additional factor $\mathrm{e}^{-\mu}$ appearing in Eq. (3.16) introduces only a $V \to V-1$ shift, which is irrelevant in the large-V limit. Since all q-independent factors can be fixed by normalization, we did not bother to include them in Eq. (3.18) but rather denote them by C_V.

As already mentioned, $p_q = \frac{1}{N}\Omega_V(q)$, where N can be found from the normalization condition,

$$N = \sum_q \Omega_V(q) = C_V\sum_q \frac{w_q}{(q-1)!}Z_0^{q-1} = C_V Z_0 F(Z_0). \tag{3.19}$$

Thus, the final formula for the desired degree distribution reads

$$p_q = \frac{1}{F(Z_0)} \frac{w_q Z_0^{q-2}}{(q-1)!}. \tag{3.20}$$

3.2 Correlation functions

Our aim is to calculate the distance-dependent connected correlation function $\tilde{p}^c_{\bar{q},\bar{r}}(l)$ introduced in Sec. 1.4.3,

$$\tilde{p}^c_{\bar{q},\bar{r}}(l) = \sum_{q,r} q\, r\, \tilde{p}^c_{q,r}(l), \tag{3.21}$$

defined by means of the connected two-point probability

$$\tilde{p}^c_{q,r}(l) = \tilde{p}_{q,r}(l) - \tilde{p}_q(l)\, \tilde{p}_r(l). \tag{3.22}$$

To this end we have to derive the formula for the joint probability distribution $\tilde{p}_{q,r}(l)$ given by Eq. (1.34). We proceed by introducing the partition function $\Omega_l(q, r; \mu)$ of all trees with two points of degrees q and r marked such that the distance between these points is l. Note that since only trees are considered here, there is just one path joining any two vertices.

Applying the approach from the previous section we can use $Z(\mu)$ to express $\Omega_l(q, r; \mu)$ as follows (cf. Ref. [44]). Consider a chain of $(l + 1)$ vertices. Then the first and the last vertex, to which we will refer to as the head and the tail, respectively, are at distance l apart. If we now fix their degrees to be q and r then all the configurations from the ensemble generated by $\Omega_l(q, r; \mu)$ can be obtained by attaching trees to the vertices along the path joining the head and the tail in all possible ways.

The partition function $\Omega_l(q, r; \mu)$ is a product of the contributions from the end-point vertices, which have a definite degree, and the $(l - 1)$ vertices along the path connecting them, in which case we have to sum over all possible degrees greater or equal two. Thus one finds

$$\Omega_l(q, r; \mu) = \frac{\mathrm{e}^{-\mu} w_q}{(q-1)!}\, Z^{q-1} \left[\mathrm{e}^{-\mu} \sum_{k=2}^{\infty} \frac{w_k}{(k-2)!}\, Z^{k-2} \right]^{l-1} \frac{\mathrm{e}^{-\mu} w_r}{(r-1)!}\, Z^{r-1}, \tag{3.23}$$

where for the sake of clarity we have omitted the argument of the $Z(\mu)$ function. The factorials appearing in the above equation are due to the fact that the considered trees are non-planar and so the relative position of the attached branches does not matter. Equation (3.23) has a neat graphical form, which for $q = 4$ and $r = 3$ is presented in Fig. 3.3.

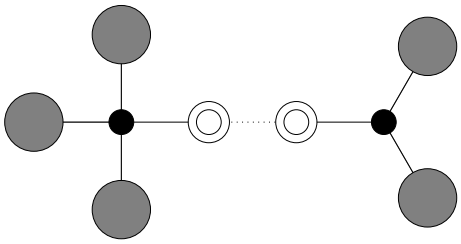

Figure 3.3: Graphical representation of the partition function $\Omega_l(q, r; \mu)$ given by Eq. (3.23) for $q = 4$ and $r = 3$. Gray bubbles correspond to the partition function $Z(\mu)$ and the smaller black circles mark the vertices with degrees q and r; double circles represent the $l - 1$ vertices along the path connecting them in which we sum over all possible insertions of the $Z(\mu)$ function (see Fig. 3.4).

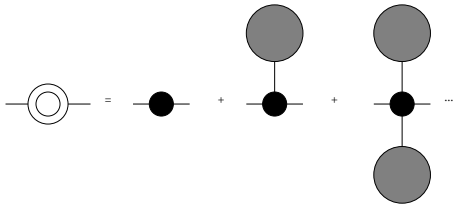

Figure 3.4: Double circles denote the summation over all possible insertions of the $Z(\mu)$ function (gray bubbles).

The term containing the sum over all the possible insertions of the $Z(\mu)$ function (see Fig. 3.4) can be rewritten in the following way:

$$\mathrm{e}^{-\mu} \sum_{k=2}^{\infty} \frac{w_k}{(k-2)!} \, Z^{k-2} = \mathrm{e}^{-\mu} \frac{\partial}{\partial Z} \left[Z F(Z) \right] = \mathrm{e}^{-\mu} \left[F(Z) + Z \frac{\partial F(Z)}{\partial Z} \right]. \tag{3.24}$$

We can now use relation (3.10) to eliminate $F(Z)$ from the above equation. Its differentiation with respect to μ gives

$$\mathrm{e}^{\mu} = \frac{\partial}{\partial \mu} F[Z(\mu)] = \frac{\partial F(Z)}{\partial Z} \frac{\partial Z(\mu)}{\partial \mu}, \tag{3.25}$$

leading to

$$\frac{\partial F(Z)}{\partial Z} = \mathrm{e}^{\mu} \frac{1}{Z'(\mu)}. \tag{3.26}$$

Inserting Eqs. (3.10) and (3.26) into Eq. (3.24) we finally obtain

$$\mathrm{e}^{-\mu} \sum_{k=2}^{\infty} \frac{w_k}{(k-2)!} \, Z^{k-2} = 1 + \frac{Z(\mu)}{Z'(\mu)}. \tag{3.27}$$

Using the above formula, Eq. (3.23) can be rewritten as

$$\Omega_l(q,r;\mu) = \mathrm{e}^{-2\mu}\frac{w_q}{(q-1)!}\frac{w_r}{(r-1)!}\, Z^{q+r-2}\left(1+\frac{Z}{Z'}\right)^{l-1}. \tag{3.28}$$

Its leading order approximation using the expansion given by Eq. (3.12) reads

$$\begin{aligned}\Omega_l(q,r;\mu) \approx \mathrm{e}^{-2\mu}\frac{w_q}{(q-1)!}\frac{w_r}{(r-1)!}Z_0^{q+r-2}\left(1-\frac{Z_1}{Z_0}\sqrt{\Delta\mu}\right)^{q+r-2}\\ \times\left(1-2\frac{Z_0}{Z_1}\sqrt{\Delta\mu}\right)^{l-1},\end{aligned} \tag{3.29}$$

which can be further approximated by

$$\begin{aligned}\Omega_l(q,r;\mu) \approx \mathrm{e}^{-2\mu}\frac{w_q}{(q-1)!}\frac{w_r}{(r-1)!}Z_0^{q+r-2}\\ \times\exp\left\{-\left[(q+r-2)\frac{Z_1}{Z_0}+2(l-1)\frac{Z_0}{Z_1}\right]\sqrt{\Delta\mu}\right\}.\end{aligned} \tag{3.30}$$

Similarly as we did in the previous section, we can now use the inverse Laplace transform given by Eq. (3.17) to derive the expression for the canonical partition function, which in the leading order in V reads

$$\Omega_l(q,r;V) \propto \frac{w_q}{(q-1)!}\frac{w_r}{(r-1)!}Z_0^{q+r-2}\left[(q+r-2)\frac{Z_1}{Z_0}+2(l-1)\frac{Z_0}{Z_1}\right]. \tag{3.31}$$

Recall now that $Z_1^2 = 2F(Z_0)/F''(Z_0)$. We have then

$$\begin{aligned}\frac{Z_0^2}{Z_1^2} = \frac{Z_0^2F''(Z_0)}{2F(Z_0)} = \frac{1}{2F(Z_0)}\sum_{q=2}(q-2)(q-3)\frac{w_qZ_0^{q-2}}{(q-1)!}\\ = \frac{1}{2}\sum_q(q-2)(q-3)\,p_q = \frac{1}{2}\big(\langle q^2\rangle - 5\,\langle q\rangle + 6\big).\end{aligned} \tag{3.32}$$

Using the fact that the average degree for trees equals asymptotically $\langle q\rangle = 2$ we get the final result

$$\frac{Z_0^2}{Z_1^2} = \frac{1}{2}\big(\langle q^2\rangle - 4\big). \tag{3.33}$$

Using this relation and fixing the normalization we obtain from Eq. (3.31) the expression for the joint probability distribution

$$\tilde{p}_{q,r}(l) = p_qp_r\,\frac{(q+r-2)+(\langle q^2\rangle-4)\,(l-1)}{2+(\langle q^2\rangle-4)\,(l-1)}. \tag{3.34}$$

Summing now over r we get

$$\tilde{p}_q(l) = p_q\,\frac{q+(\langle q^2\rangle-4)\,(l-1)}{2+(\langle q^2\rangle-4)\,(l-1)}. \tag{3.35}$$

Using the above results we get the expression for the connected probability (3.22),

$$\tilde{p}^{c}_{q,r}(l) = -p_q p_r \frac{(q-2)(r-2)}{\left[2 + (\langle q^2\rangle - 4)(l-1)\right]^2}. \tag{3.36}$$

Summing $\tilde{p}^{c}_{q,r}(l)$ over q and r we finally get the connected correlation function (1.48),

$$\tilde{p}^{c}_{\bar{q},\bar{r}}(l) = -\frac{\left(\langle q^2\rangle - 4\right)^2}{\left[2 + (\langle q^2\rangle - 4)(l-1)\right]^2}. \tag{3.37}$$

As evident from the above result, the correlations are non-zero and fall off with the distance as l^{-2}.

Similarly, the average degree of vertices at some distance l from a given vertex of degree q decreases with distance. If we assume self-averaging we will obtain asymptotically from the definition (1.41) that

$$\bar{k}_l(q) = \sum_r \left\langle \frac{r\, n_{q,r}(l)}{n_q(l)} \right\rangle \sim \sum_r \frac{r\, \langle n_{q,r}(l)\rangle}{\langle n_q(l)\rangle} = \sum_r \frac{r\tilde{p}_{q,r}(l)}{\tilde{p}_q(l)}. \tag{3.38}$$

Using the previous results given by Eqs. (3.34) and (3.35) we then get

$$\bar{k}_l(q) = 2 + \frac{\langle q^2\rangle - 4}{q + (\langle q^2\rangle - 4)(l-1)}. \tag{3.39}$$

3.3 Examples

In the derivation of the quantities from the previous section we did not assume any specific form of the weights $\{w_q\}$, so we expect that the obtained formulas hold for any choice of the one point-weights, provided that we stay in the generic phase. In the following sections we illustrate how to apply these general formulas to some examples with specific weights.

The first, and probably the simplest model are trees with uniformly weighted nodes for which $w_q = 1$ regardless of the vertex degree q. They correspond to the classical Erdős-Rényi graphs discussed in Sec. 2.5.1 of the previous chapter. As the second example we take scale-free trees, following a power law degree distribution. Albeit their second moment diverges, the integrated correlation functions do not vanish. We then compare our results against numerical simulations and find that they correspond quite well, taking into account finite size effects.

3.3.1 Erdős-Rényi trees

In the model of uniformly weighted trees we assume that the one point weights do not depend on vertex degrees and are all equal $w_q = 1$, implying that all trees from the ensemble share the same weight. In this case

$$F(Z) = \frac{1}{Z}\mathrm{e}^{Z}. \tag{3.40}$$

It is minimized by $Z_0 = 1$, for which $F(Z_0) = \mathrm{e}$. Inserting these values into Eq. (3.20) we immediately get the degree distribution

$$p_q = \mathrm{e}^{-1}\frac{1}{(q-1)!}. \tag{3.41}$$

To calculate the joint probability distribution and the connected correlation function we first need to find the second moment $\langle q^2\rangle$. In the previous section we have shown that

$$\frac{Z_0^2}{Z_1^2} = \frac{1}{2}\big(\langle q^2\rangle - 4\big). \tag{3.42}$$

Knowing the formula for $F(Z)$ we can calculate the left-hand side of the above equation directly,

$$\frac{Z_0^2}{Z_1^2} = \frac{Z_0^2 F''(Z_0)}{2F(Z_0)} = \frac{1}{2}. \tag{3.43}$$

Combining Eqs. (3.42) and (3.43) one finds $\langle q^2\rangle = 5$. Using this result we get for the connected two-point probability

$$\tilde{p}^c_{q,r}(l) = -\frac{1}{\mathrm{e}^2}\frac{(q-2)(r-2)}{(l+1)^2}\frac{1}{(q-1)!}\frac{1}{(r-1)!}, \tag{3.44}$$

and for the associated correlation function

$$\tilde{p}^c_{\bar{q},\bar{r}}(l) = -\frac{1}{(l+1)^2}. \tag{3.45}$$

The average degree of neighbors at distance l from a given vertex is given by

$$\bar{k}_l(q) = 2 + \frac{1}{q+l-1}. \tag{3.46}$$

3.3.2 Scale-free trees

Scale-free trees are a subset of a larger ensemble of scale-free graphs. These are networks whose degree distribution is governed by a power-law function $p_q \propto q^{-\beta}$. We have already investigated the nearest-neighbor correlations in scale-free graphs in Sec. 2.5.3.

The ensemble of scale-free trees corresponds to the planar graphs studied in Refs. [76] and [77], and is generated by

$$w_q = q^{-\beta}(q-1)!. \tag{3.47}$$

In such case $F(Z)$, given by Eq. (3.9), is defined in terms of a power series known as the polylogarithm function $\mathrm{Li}_\beta(Z)$,

$$F(Z) = \sum_{q=1}^{\infty} q^{-\beta} Z^{q-2} = \frac{1}{Z^2}\sum_{q=1}^{\infty}\frac{Z^q}{q^\beta} = \frac{1}{Z^2}\,\mathrm{Li}_\beta(Z), \tag{3.48}$$

which is minimized by Z_0 fulfilling the condition

$$2\,\mathrm{Li}_\beta(Z_0) = \mathrm{Li}_{\beta-1}(Z_0). \tag{3.49}$$

The radius of convergence of the polylogarithm function $\mathrm{Li}_\beta(Z)$ equals one. At the boundary, for $Z = 1$ and $\beta > 1$, it reduces to the Riemann zeta function,

$$\mathrm{Li}_\beta(1) = \zeta(\beta). \tag{3.50}$$

Another property of the polylogarithm is that if $\alpha < \beta$, then $\mathrm{Li}_\alpha(Z) > \mathrm{Li}_\beta(Z)$. Thus, Eq. (3.49) has a solution for $\beta < \beta_c$, where $\beta_c \approx 2.47875$ is given by

$$2\zeta(\beta_c) = \zeta(\beta_c - 1). \tag{3.51}$$

At the critical value of $\beta = \beta_c$ the partition function $Z(\mu)$ does no longer scale according to Eq. (3.12), so in principle we cannot use the Laplace transform given by Eq. (3.17) anymore. However, as shown in Ref. [80] the large-V behavior remains unchanged and we expect our formula to hold in the large-V limit. From Eq. (3.20) we read-off the degree distribution

$$p_q = \frac{Z_0^{q-2}}{F(Z_0)} q^{-\beta}. \tag{3.52}$$

If now $Z_0 = 1$ then the degree distribution depends on q only by the $q^{-\beta}$ term and is scale-free. This corresponds to the critical value β_c.

The second moment, which can be calculated from Eqs. (3.42) and (3.43),

$$\left\langle q^2 \right\rangle = Z_0^2 \frac{F''(Z_0)}{F(Z_0)} + 4 = 10 + \frac{1}{\mathrm{Li}_\beta(Z_0)} \left[\mathrm{Li}_{\beta-2}(Z_0) - 5\,\mathrm{Li}_{\beta-1}(Z_0)\right], \tag{3.53}$$

diverges as $\beta \to \beta_C$. Formula (3.36) leads for $l > 1$ to

$$\lim_{V\to\infty} \tilde{p}^c_{q,r}(l) = 0, \tag{3.54}$$

suggesting that the correlations vanish in the large-V limit. However, this limit (3.54) is not uniform and the integrated correlation functions do not disappear:

$$\lim_{V\to\infty} \tilde{p}^c_{\bar{q},\bar{r}}(l) = -\frac{1}{(l-1)^2}, \tag{3.55}$$

and

$$\lim_{V\to\infty} \bar{k}_l(q) = 2 + \frac{1}{l-1}. \tag{3.56}$$

Please note that the above results are universal and valid for any kind of scale-free trees with $\beta < 3$.

3.3.3 Monte Carlo simulations

Analytical results obtained in the previous sections are valid only in the strict $V \to \infty$ limit and it is clear that for finite trees these formulas will not hold for any l. We expect that they are valid only at distances much smaller than the average distance $\langle l \rangle$ defined as

$$\langle l \rangle = \frac{1}{V^2} \sum_l l \, \langle n(l) \rangle \, . \tag{3.57}$$

Here $\langle n(l) \rangle$ is the ensemble average counting the number of all pairs of vertices at distance l apart,

$$\langle n(l) \rangle = \sum_{T \in \mathcal{T}_V} P(T) \; n(l), \tag{3.58}$$

where $P(T)$ is the statistical weight of a tree defined by Eq. (3.1). The normalization factor $1/V^2$ in Eq. (3.57) compensates the number of all possible vertex pairs.

According to Ref. [80] the scaling of $\langle l \rangle$ with the size of the graph is described by the Hausdorff or fractal dimension d_H,

$$\langle l \rangle \sim V^{1/d_H}. \tag{3.59}$$

For the generic trees discussed $d_H = 2$. In the case of scale-free trees considered in example 2 we expect (cf. Ref. [19])

$$d_H = \frac{1}{\gamma}, \qquad \gamma = \frac{\beta_c - 2}{\beta_c - 1}, \tag{3.60}$$

which results in $d_H \approx 3$. As already argued in Sec. 2.5.3, the volume dependence in scale-free trees manifests itself by a cut-off in the degree distribution p_q.

Due to a higher value of the Hausdorff dimension d_H we expect scale-free trees to be more prone to finite size effects than the generic ones. They are a good candidate to verify whether our approach and assumptions made are legitimate. If the analytical formulas are well reproduced by MC simulations for scale-free trees, then one should expect similar or even better agreement for the generic ones.

To check the size dependence we have performed MC simulations of the ensemble described in Sec. 3.3.2. We have used an algorithm similar to "baby-universe surgery" (see Ref. [81]). The basic move consisted of picking an edge at random and cutting it. Then the smaller of the two resulting trees was grafted on some random vertex of the bigger one. The most time consuming part of the algorithm was to find which tree was smaller. To save time, the two trees were traversed simultaneously until one of them was filled

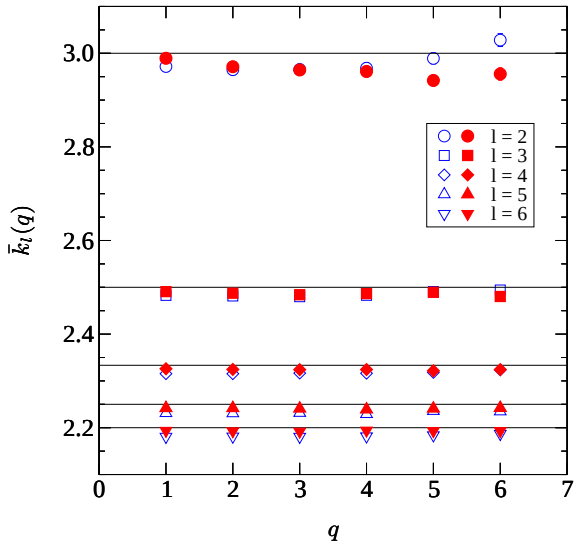

Figure 3.5: Average degree of neighbors $\bar{k}_l(q)$ at distances from $l = 2$ to 6 in trees with 64 000 (empty symbols) and 128 000 (filled symbols) vertices. Each symbol denotes different l; straight lines are the predictions given by Eq. (3.56).

completely. Additionally, to pick the attachment point from the bigger tree efficiently, the vertices of the trees were marked during the traversal. This move was supplemented with moves consisting of cutting off leaf nodes and attaching them to some other parts of the tree. This was much faster as it did not require traversing the tree. However, the autocorrelation time for such moves alone was much higher, especially for the scale-free trees. Because those trees are at the phase transition between the generic and the crumpled phase [77], the autocorrelation time is high even for the described tree-grafting algorithm.

In order to observe finite size effects we have simulated trees of various sizes up to 128 000 vertices. We have concentrated on the average neighbor degree distribution $\bar{k}_l(q)$ and the connected correlation function $p^c_{\bar{q},\bar{r}}(l)$. To verify the assumption of self-averaging we have measured these quantities for individual graphs and only then took their average.

Figure 3.5 shows the measured distance dependent average vertex degree $\bar{k}_l(q)$, as defined in Eq. (1.41),

$$\bar{k}_l(q) = \sum_r \left\langle \frac{r\, n_{q,r}(l)}{n_q(l)} \right\rangle , \tag{3.61}$$

as a function of q for various values of l. The asymptotic solution (3.56) is plotted with solid lines. The most severe finite size effects occur at short

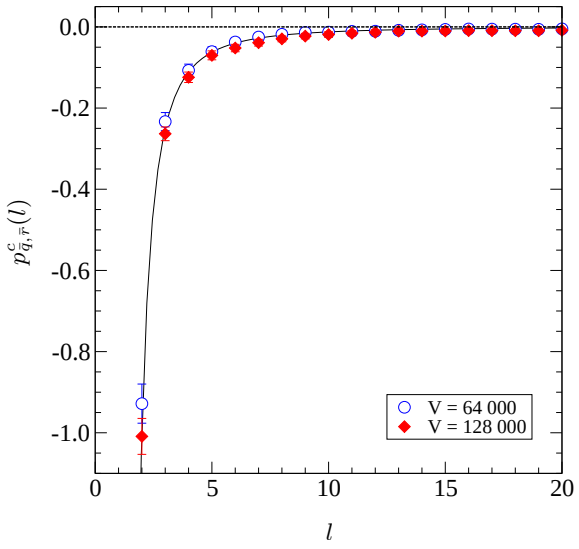

Figure 3.6: Connected correlation function $p^c_{\bar{q},\bar{r}}(l)$ for scale-free trees with 64 000 (empty circles) and 128 000 (filled diamonds) vertices. The solid line represents the prediction given by Eq. (3.55).

distances for $l = 2$. This is to be expected: for finite V the average $\langle q^2 \rangle$ is also finite and so introduces an small correction to the denominator of Eq. (3.56). The relative influence of this deviation diminishes with distance, but scales linearly with degree q. Thus, the agreement of MC data with asymptotic formulas improves with distance and for larger l is rather decent.

The results for the connected correlation function

$$p^c_{\bar{q},\bar{r}}(l) = \sum_{q,r} \left\langle \frac{q\, r\, n_{q,r}(l)}{n(l)} \right\rangle - \left[\sum_{q} \left\langle \frac{q\, n_q(l)}{n(l)} \right\rangle \right]^2 \tag{3.62}$$

are presented in Fig. 3.6. It is well reproduced by our result (3.55) despite the finite size and the fact that the probabilities in Eq. (3.62) are calculated on individual graphs rather than on the whole ensemble.

Chapter 4

Ising spins on regular random graphs

This chapter is based on the paper [3] *Correlation functions of Ising spins on thin graphs* by P. Bialas and A. K. Oleś, e-print arXiv:1104.4030 (2011).

Although Ising spins on random graphs have already been studied before (see, e.g., Refs. [49, 50]), and different aspects of the systems have been brought up, there has been little contribution to the topic of correlations in them. In this chapter we introduce and solve the Ising spin model on an ensemble of regular random graphs. By regular random graphs we do actually mean pseudographs or multigraphs whose all vertices have the same degree, say q, but which are otherwise completely random. In particular, they may contain multiedges and loops. They correspond to the Feynman diagrams of a ϕ^q field theory and we will exploit this analogy in our calculations. This model is intended to serve as a preparation to the study of a more interesting and complicated strictly geometrical model of random graphs displaying a phase transition described in the following chapter.

Some of the presented results have already been obtained by other authors, but we re-derive them here for completeness. We shall show that in the large-size limit, the leading-order behavior of the Ising model on q-regular random graphs agrees with the Bethe solution valid for infinite Cayley trees with coordination number q, called Bethe lattices. This name was coined by Kurata et al. [82] and Katsura et al. [83] who showed that the properties of the Ising model on an infinite Cayley tree, for which the surface effect is ignored, can be exactly calculated by the Bethe approximation (see Ref. [84]). This solution has a divergent isothermal susceptibility under zero

external field at the critical temperature T_B given by Bethe, below which spontaneous magnetization of the system occurs.

The Bethe solution differs substantially from the solution on finite Cayley trees. It has been shown (see, e.g., Refs. [85–87]) that the spontaneous magnetization per spin in a ferromagnetic Ising model on such trees studied in the large-size limit vanishes at any non-zero temperature, and yet the susceptibility per spin diverges below a critical temperature, which is less than T_B. The reason for this discrepancy is that in a finite tree the number of sites on its surface is of the same order as the total number of sites in the whole system. Because of this, the quantities obtained by taking the thermodynamic limit of the finite tree are different from those of an infinite one, for which there is no contribution from the surface. Although correlation functions on finite Cayley trees have already been studied before (see Refs. [88, 89]), such systems differ fundamentally from the infinite ones, as mentioned above. In particular, they do not exhibit a phase transition observed for the Ising spins on regular random graphs, as we shall show.

The first section introduces and formally defines the model, which is then solved in Sec. 4.2 by calculating its partition function. Specific properties of the system, like the magnetization and susceptibility, are calculated in Secs. 4.2.2 and 4.2.3, respectively. The formulas for the spin-spin correlation functions are derived in Sec. 4.3 using the similarity to Bethe lattices in the infinite volume limit. The last Section 4.4 addresses the influence of spins on the geometry.

4.1 Definition of the model

The Ising model is a classic example of a complex system capable of a phase transition (see, e.g., Ref. [36] or [27]). It was originally proposed by W. Lenz in 1920 as a simple model of a ferromagnet. The one-dimensional problem was first solved by E. Ising [90], while the exact solution for the $d = 2$ case, in which an order-disorder transition is observed, was derived by Onsager [91].

The model consists of a set of discrete variables s_i, called spins, which are either $+1$ or -1, also referred to as the *up* ($\uparrow$) and *down* ($\downarrow$) states. In the original formulation the spins are arranged on a regular lattice, and each spin interacts with its nearest neighbors. However, generalized models may include long-range interactions and more choices for spin values.

In our approach the spins are placed in the vertices of q-regular pseudographs with coordination number q equal or greater than three. Since there are no other restrictions concerning the geometry and the graphs are totally random in other aspects, we will refer to them shortly as random q-regular graphs, bearing in mind that from now on under the term graphs we actually mean pseudographs. In the presence of an externally imposed

field B the system's Hamiltonian is

$$H = -\frac{1}{2}\sum_{i,j} J_{ij} s_i s_j - B\sum_i s_i. \tag{4.1}$$

The subscripts label vertices and J_{ij} is the interaction energy between adjacent spins: $J_{ij} \neq 0$ if two nodes are connected, or $J_{ij} = 0$ otherwise. The sign of J_{ij} determines the type of interaction—in ferromagnets $J_{ij} > 0$ and the neighboring spins try to align parallel to one another, whereas in the antiferromagnetic case $J_{ij} < 0$ and anti-parallel orientation is favored. Each pair of vertices is counted twice in Eq. (4.1), hence the 1/2 factor.

From now on we focus only on ferromagnetic interactions in the absence of any external field B. The Hamiltonian (4.1) becomes then

$$H = -\frac{1}{2}\sum_{i,j} A_{ij} s_i s_j, \tag{4.2}$$

where $\mathbf{A}$ is the adjacency or connectivity matrix of the underlying graph, and we have set $J_{ij} = 1$ for ferromagnetic interaction.

The partition function of the considered Ising model defined on graphs with a fixed number of vertices is given by the sum over all q-regular graphs with n vertices and all the possible arrangements of spins on each graph G,

$$Z_n = \sum_{G\in\mathcal{G}_{fl}} \sum_{s_1,\ldots,s_n} \mathrm{e}^{-\beta H(G;s_1,\ldots,s_n)}. \tag{4.3}$$

In fact, the above defined ensemble $\mathcal{G}_{fl}$ is a collection of fully labeled pseudographs, i.e., graphs which have all vertices and links labeled and may include multiple- and self-connections. All such graphs are equiprobable and thus have the same configuration space weight. Some of the label permutations define the same topologies, so when unlabeled graphs are considered the weights become distinct. It can be shown (see, e.g., Ref.[24]) that these weights are identical to the inverse of the symmetry factors $s(G)$ of the corresponding Feynman diagrams appearing in the perturbative expansions of a minifield theory. Thus, the partition function of Ising spins organized on an ensemble of Feynman diagrams $\mathcal{G}$ of a ϕ^q field theory is

$$Z_n = \sum_{G\in\mathcal{G}} \frac{1}{s(G)} \sum_{s_1,\ldots,s_n} \mathrm{e}^{-\beta H(G;s_1,\ldots,s_n)}, \tag{4.4}$$

and we may use the perturbation theory approach described in Sec. 1.5 to enumerate them.

4.2 Solution

The ensemble of q-regular random graphs is identical to ϕ^q Feynman diagrams of the zero-dimensional field theory described in Sec. 1.5. Thus, if ϕ_+ and ϕ_- fields are associated with the "up" and "down" or $+1$ and -1 spins, respectively, the requisite ensemble can be generated from the Feynman diagram expansion of the grand-canonical partition function

$$Z(\mu) = \frac{1}{2\pi} \int \mathrm{d}\phi_+ \mathrm{d}\phi_- \exp\left[-\frac{1}{2} \boldsymbol{\phi}^T \Delta^{-1} \boldsymbol{\phi} \; + \mathrm{e}^{-\mu} \frac{1}{q!} \left(\phi_+^q + \phi_-^q \right) \right], \tag{4.5}$$

where

$$\boldsymbol{\phi} \equiv \begin{pmatrix} \phi_+ \\ \phi_- \end{pmatrix}, \tag{4.6}$$

and Δ is the transfer matrix,

$$\Delta \equiv \begin{pmatrix} \sqrt{g} & \frac{1}{\sqrt{g}} \\ \frac{1}{\sqrt{g}} & \sqrt{g} \end{pmatrix}. \tag{4.7}$$

We use the positive coupling constant g throughout the calculations in accordance to the convention from Refs. [48, 50]. It is related to $\beta = 1/k_B T$ by

$$\sqrt{g} = \mathrm{e}^{\beta}. \tag{4.8}$$

In the following we will set $k_B = 1$ and regard

$$T = 2 \, \frac{1}{\ln g} \tag{4.9}$$

as the generalized temperature of the system.

The grand-canonical partition function $Z(\mu)$ generates q-regular graphs of all possible sizes. It is related to the generating function of graphs with a specific number of vertices n by the discrete Laplace transform,

$$Z(\mu) = \sum_n \mathrm{e}^{-\mu n} Z_n. \tag{4.10}$$

In the following we will be interested in the ensemble of graphs in the asymptotic limit $n \to \infty$. The desired partition function Z_n can be calculated in the large-size limit using the Laplace's integral approximation technique described in Appendix B. We start by performing binomial expansion of the ϕ^q terms in Eq. (4.5),

$$\begin{aligned} \exp\left[\mathrm{e}^{-\mu} \frac{1}{q!} \left(\phi_+^q + \phi_-^q \right) \right] &= \sum_{n=0}^{\infty} \frac{1}{n!} \left[\mathrm{e}^{-\mu} \frac{1}{q!} \left(\phi_+^q + \phi_-^q \right) \right]^n \\ &= \sum_{n=0}^{\infty} \mathrm{e}^{-\mu n} \frac{1}{(q!)^n} \sum_{k=0}^{n} \frac{\phi_+^{qk} \, \phi_-^{q(n-k)}}{k!(n-k)!}. \end{aligned} \tag{4.11}$$

The sum over n reflects the various sizes of graphs in the grand-canonical ensemble and k is the number of "up" spins. Using relation (4.10) we can extract from $Z(\mu)$ the canonical partition function

$$Z_n = \frac{1}{2\pi}\frac{1}{(q!)^n}\sum_{k=0}^{n}\frac{1}{k!(n-k)!}\int \mathrm{d}\phi_+\mathrm{d}\phi_-\ \mathrm{e}^{-\frac{1}{2}\boldsymbol{\phi}^T\Delta^{-1}\boldsymbol{\phi}}\ \phi_+^{qk}\ \phi_-^{q(n-k)}. \tag{4.12}$$

For convenience, from now on we will use the ratio of positive and negative spins given by $z = k/n$ rather than k. Furthermore, we rescale the fields by extracting the size dependent factor,

$$\phi_\pm = \sqrt{n}\ \psi_\pm. \tag{4.13}$$

Since we are interested in the asymptotic limit of $n \to \infty$, we may replace the sum over k in Eq. (4.12) with an integral over z and rewrite it in the form

$$Z_n = \frac{n}{2\pi}A_n\int_0^1 \mathrm{d}z\ B_n(z)\int \mathrm{d}\psi_+\mathrm{d}\psi_-\ \mathrm{e}^{nf(\psi_+,\psi_-)}, \tag{4.14}$$

where

$$f(\psi_+,\psi_-) = -\frac{1}{2}\boldsymbol{\psi}^T\Delta^{-1}\boldsymbol{\psi} + q\left[z\ln\psi_+ + (1-z)\ln\psi_-\right], \tag{4.15}$$

and the field independent factors are

$$A_n = \exp\left[n\left(\frac{q}{2}\ln n - \ln q!\right)\right], \tag{4.16}$$

$$B_n(z) = \frac{1}{(zn)![(1-z)n]!}. \tag{4.17}$$

Let $(\overline{\psi}_+, \overline{\psi}_-)$ be the maximum of $f(\psi_+, \psi_-)$ given by

$$\frac{\partial f(\overline{\psi}_+, \overline{\psi}_-)}{\partial \overline{\psi}_\pm} = 0, \tag{4.18}$$

satisfying $H(\overline{\psi}_+, \overline{\psi}_-) > 0$, where $H(\overline{\psi}_+, \overline{\psi}_-)$ is the value of the determinant of the Hessian matrix of f given by Eq. (B.14). The field integral in Eq. (4.14) can be now asymptotically approximated using Eq. (B.16),

$$\int \mathrm{d}\psi_+\mathrm{d}\psi_-\ \mathrm{e}^{nf(\psi_+,\psi_-)} \sim \mathrm{e}^{nf(\overline{\psi}_+,\overline{\psi}_-)}\ \frac{2\pi}{n\sqrt{H(\overline{\psi}_+, \overline{\psi}_-)}}, \tag{4.19}$$

yielding for the partition function

$$Z_n \sim A_n\int_0^1 \mathrm{d}z \exp\left[nf(\overline{\psi}_+, \overline{\psi}_-) + \ln B_n(z) - \frac{1}{2}\ln H(\overline{\psi}_+, \overline{\psi}_-)\right]. \tag{4.20}$$

We are now to find the expression for $f(\overline{\psi}_+, \overline{\psi}_-)$. The condition for the maximum (4.18) is a system of two quadratic equations. It has in general four solutions, from which only the one positive in both ψ_+ and ψ_- has physical meaning,

$$\begin{cases} \overline{\psi}_+(z) = \sqrt{q/2}\, g^{-\frac{3}{4}} \left[1 + 2z\left(g^2 - 1\right) + \sqrt{1 + 4\left(g^2 - 1\right)(1-z)z}\right]^{\frac{1}{2}}, \\ \overline{\psi}_-(z) = \overline{\psi}_+(1-z). \end{cases}$$

A straightforward way to proceed would be to plug the above result directly into Eq. (4.15). However, there is a better way: we may rewrite Eq. (4.18) in matrix form

$$\Delta^{-1} \begin{pmatrix} \overline{\psi}_+ \\ \overline{\psi}_- \end{pmatrix} = q \begin{pmatrix} z/\overline{\psi}_+ \\ (1-z)/\overline{\psi}_- \end{pmatrix}, \tag{4.21}$$

and notice that

$$\overline{\boldsymbol{\psi}}^T \Delta^{-1} \overline{\boldsymbol{\psi}} = q, \tag{4.22}$$

which simplifies $f(\overline{\psi}_+, \overline{\psi}_-)$ to

$$f(\overline{\psi}_+, \overline{\psi}_-) = q \left[z \ln \overline{\psi}_+ + (1-z) \ln \overline{\psi}_- - \frac{1}{2} \right]. \tag{4.23}$$

The logarithm of the $B_n(z)$ term appearing in Eq. (4.20) can be approximated using the Stirling formula

$$\ln n! = n \ln n - n + \frac{1}{2} \ln 2\pi n + O(n^{-1}), \tag{4.24}$$

yielding

$$\begin{aligned} \ln B_n(z) &= -\ln(zn)! - \ln[(1-z)n]! \\ &\approx -n\left[z \ln z + (1-z)\ln(1-z)\right] - \frac{1}{2}\ln z(1-z) \\ &\quad -n(\ln n - 1) - \ln 2\pi n. \end{aligned} \tag{4.25}$$

Gathering the results given by Eqs. (4.23) and (4.25) we eventually obtain the partition function Z_n in the leading order in n,

$$Z_n \sim \mathrm{e}^{-nC_n} \int_0^1 \mathrm{d}z\, \mathrm{e}^{-nF(z)}, \tag{4.26}$$

where the leading-order approximation of $F(z)$ is given by

$$F(z) \sim z \ln z + (1-z)\ln(1-z) - q\left[z \ln \overline{\psi}_+ + (1-z)\ln \overline{\psi}_-\right], \tag{4.27}$$

and C_n encapsulates the z independent terms,

$$C_n = \ln q! - (1 - \ln n)(1 - q/2). \tag{4.28}$$

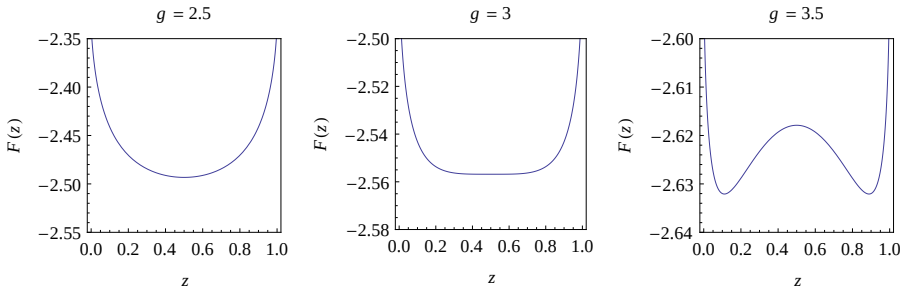

Figure 4.1: Plots of the $F(z)$ function for $q = 3$ and three different values of g: (a) for $g = 2.5$, (b) at the transition $g = 3$, (c) and for $g = 3.5$.

The next-to-leading corrections to $F(z)$ are calculated in Appendix C.

We know from statistical physics that the partition function of a system is proportional to the exponent of its free energy. We also know that the number of regular random graphs scales with the number of vertices n faster than e^n. Therefore, the free energy per spin, which is proportional to $-\ln(Z_n)/n$, diverges, so our system has no thermodynamic limit. Nevertheless, because the volume factor does not depend on z it does not contribute to the expectation values given by

$$\langle \mathcal{O} \rangle = \frac{1}{Z_n}\, \mathrm{e}^{-nC_n} \int_0^1 \mathrm{d}z\, \mathcal{O}(z)\, \mathrm{e}^{-nF(z)} = \frac{\int_0^1 \mathrm{d}z\, \mathcal{O}(z)\, \mathrm{e}^{-nF(z)}}{\int_0^1 \mathrm{d}z\, \mathrm{e}^{-nF(z)}}. \tag{4.29}$$

This average can be calculated in the leading order using the familiar Laplace approximation.

The function $F(z)$ is symmetric around $z = 1/2$, that is,

$$F(z) = F(1 - z), \tag{4.30}$$

which directly corresponds to the up-down symmetry: the counter-configuration can be simply obtained by simultaneously flipping all spins. For small values of the coupling parameter $F(z)$ has a minimum at this point (see Fig. 4.1(a)). When g increases the one minimum splits into two minima, and the former minimum in the middle becomes now a maximum (Fig. 4.1(c)). Exactly at the transition $F(z)$ has an inflection point given by $F''(1/2) = 0$. This solved for g yields the value of the coupling parameter at which the transition occurs

$$g = \frac{q}{q-2}. \tag{4.31}$$

In the next section we shall show that the minima of the $F(z)$ function correspond to the magnetization of the system and the above value of g defines the critical coupling at which the order-disorder phase transition occurs.

4.2.1 Monte Carlo simulations

In the following sections we study in detail some specific properties of the model, like the magnetization, susceptibility, and correlation functions. The analytic formulas describing them are derived in the large-size limit, so formally they apply only to infinite systems. These solutions are supplemented with data obtained from numerical simulations of finite size-systems. The performed MC simulations allow us to verify to which extent the assumptions made are correct and can be used to systems of finite size.

We have simulated the Ising spin model on q-regular random graphs using a hybrid Metropolis-type algorithm accounting for both the evolution of the spins and the changes in the underlying geometry. The elementary step of the algorithm consists of two parts. In the first switching part (see Ref. [92]) a pair of edges, say A-B and C-D, is selected randomly with uniform probability from all the edges in the graph. Next, their endpoints are cross-exchanged to give A-D and B-C. Such moves preserve the degrees of the vertices involved assuring that the graph remains regular. They are always possible, since we permit multiedges and loops. The resulting local energy change between the initial configuration μ and the final configuration ν is

$$E_\nu - E_\mu = -(s_A s_D + s_B s_C) + (s_A s_B + s_C s_D), \tag{4.32}$$

where $s_A, \ldots, s_D$ are the spins at the corresponding sites. Following Eq. (1.81), we do always accept the move when $E_\nu - E_\mu < 0$, and transitions to higher energy configurations occur with the probability

$$A(\mu \to \nu) = \exp\left\{\beta\ [s_A(s_D - s_B) + s_C(s_B - s_D)]\right\}. \tag{4.33}$$

The second part is the classic single-spin-flip dynamics move (cf. Ref. [54]). It consists of reversing one randomly chosen spin s_i with the probability

$$A(\mu \to \nu) = \min\left\{1, \exp\left(-2\beta s_i \sum_{j \neq i} A_{ij} s_j\right)\right\}, \tag{4.34}$$

where the sum runs over all nearest neighbors of site i (recall that $\mathbf{A}$ is the adjacency matrix of the graph). Because the weight of an edge does not change when the spins at both its endpoints are flipped simultaneously we omit in the above calculation self-connections by summing over j distinct from i.

4.2.2 Magnetization

The net magnetization M of the system is defined by the difference in the number of up and down spins,

$$M \equiv k - (n - k) = 2k - n = n\,(2z - 1) = n\,m, \tag{4.35}$$

where m is the per spin magnetization associated with the order parameter,

$$m = (2z - 1). \tag{4.36}$$

Its average value can be calculated from

$$\langle m \rangle = \frac{\int_0^1 \mathrm{d}z (2z-1)\, \mathrm{e}^{-nF(z)}}{\int_0^1 \mathrm{d}z\, \mathrm{e}^{-nF(z)}} \tag{4.37}$$

using either Laplace's approximation around the minimum of $F(z)$ in the $n \to \infty$ limit, or by direct numerical integration for specific n.

Let z_0 be the minimum of the $F(z)$ function. Then Taylor's expansion of $F(z)$ around z_0 up to quadratic order equals

$$F(z) \approx F(z_0) + \frac{1}{2}(z - z_0)^2 F''(z_0), \tag{4.38}$$

where we have used the fact that $F'(z_0)$ vanishes at the extremum. The denominator of Eq. (4.37) evaluates to

$$\begin{aligned} \int_0^1 \mathrm{d}z\, \mathrm{e}^{-nF(z)} &\sim \mathrm{e}^{-nF(z_0)} \int_{-\infty}^{+\infty} \mathrm{d}z\, \mathrm{e}^{-\frac{1}{2}(z-z_0)^2 nF''(z_0)} \\ &= \mathrm{e}^{-nF(z_0)} \sqrt{\frac{2\pi}{nF''(z_0)}}, \end{aligned} \tag{4.39}$$

while the numerator to

$$\begin{aligned} \int_0^1 \mathrm{d}z\, (2z-1)\, \mathrm{e}^{-nF(z)} &\sim \mathrm{e}^{-nF(z_0)} \int_{-\infty}^{+\infty} \mathrm{d}z\, (2z-1)\, \mathrm{e}^{-\frac{1}{2}(z-z_0)^2 nF''(z_0)} \\ &= (2z_0 - 1)\, \mathrm{e}^{-nF(z_0)} \sqrt{\frac{2\pi}{nF''(z_0)}}, \end{aligned} \tag{4.40}$$

yielding

$$\langle m \rangle = 2z_0 - 1. \tag{4.41}$$

Therefore, in the large-size limit magnetization is defined by the value of z minimizing the free energy function of the system. From now on, we will use the letter m alone to denote the average magnetization per spin, and omit the brackets $\langle \cdot \rangle$ around it, which is a common practice in literature.

In high temperatures, which correspond to the values of the coupling constant below the value given by Eq. (4.31), the function $F(z)$ has only one maximum at $z_0 = 1/2$. On average, half of the spins are pointing upwards and the other half downwards resulting in the average magnetization being zero (see Fig. 4.2). When the temperature is lowered, the minimum splits and moves away from $z = 1/2$ causing the magnetization to become non-zero. The system spontaneously chooses either of the two orientations,

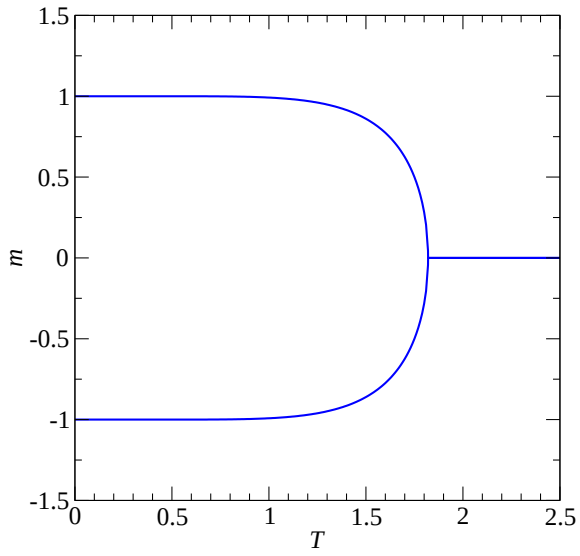

Figure 4.2: Average magnetization per spin m of the Ising model on 3-regular random graphs as function of the generalized temperature $T = 2/\ln g$. At the critical temperature T_c the system changes continuously from the ordered to the disordered phase.

which is called spontaneous symmetry breaking. In fact, the observed behavior is typical of the order-disorder phase transition occurring at the critical temperature T_c, which corresponds to the critical value of the coupling parameter g_c given by Eq. (4.31),

$$g_c = \frac{q}{q-2}. \tag{4.42}$$

When $g < g_c$ the system is in the disordered phase. Intuitively, since all spins are completely randomly oriented, they cancel each other. After passing the critical point, ordering in the system appears and the magnetization becomes finite.

The critical coupling value (4.42) agrees with the results obtained in Refs. [82, 83] on Bethe lattices. This correspondence was already noticed in Refs. [48, 50] which suggests that asymptotically the Ising model on random graphs is equivalent to the Bethe lattice. This might appear surprising, since our model allows loops and multiple edges, which are absent in the Bethe lattice. The similarity is even more interesting considering that the two classes of models are solved by seemingly completely different methods. However, as argued in Ref. [93] the observed identity can be explained by the fact that the ratio of the equations for ϕ_+ and ϕ_- fields maximizing the

action in Eq. (4.5) (called by the authors saddle-point equations),

$$\begin{cases} \phi_+ &= \dfrac{1}{(q-1)!}\left(\sqrt{g}\,\phi_+^{q-1} + \dfrac{1}{\sqrt{g}}\,\phi_-^{q-1}\right), \\ \phi_- &= \dfrac{1}{(q-1)!}\left(\dfrac{1}{\sqrt{g}}\,\phi_+^{q-1} + \sqrt{g}\,\phi_-^{q-1}\right), \end{cases} \tag{4.43}$$

reproduces the fixed points of the Bethe recursion relation. Intuitively, because the number of loops scales slower than n, it does not contribute to the leading order of our asymptotic solution.

Expressions for the magnetization in the ordered phase can be obtained for specific q from the zeros of $F'(z)$. For $q = 3$ we find

$$m = \frac{g}{g-2}\sqrt{\frac{g-3}{g+1}}, \qquad g > 3, \tag{4.44}$$

and in the $q = 4$ case

$$m = \frac{g}{g^2-2}\sqrt{g^2-4}, \qquad g > 2. \tag{4.45}$$

4.2.3 Susceptibility

The susceptibility per site,

$$\chi = \frac{1}{n}\left(\langle M^2\rangle - \langle M\rangle^2\right), \tag{4.46}$$

can be calculated in a similar fashion to the magnetization using the leading order approximation of the corresponding integrals,

$$\langle M^2\rangle = n^2\frac{\int_0^1 \mathrm{d}z(2z-1)^2\,\mathrm{e}^{-nF(z)}}{\int_0^1 \mathrm{d}z\,\mathrm{e}^{-nF(z)}} \approx \frac{4n}{F''(z_0)} + \left[n(2z_0-1)\right]^2. \tag{4.47}$$

Combining the above formula with the previous result for magnetization given by Eq. (4.41) we get

$$\chi = \frac{4}{F''(z_0)} = \frac{4}{F''\left[\frac{1}{2}(1+m)\right]}. \tag{4.48}$$

In the symmetric phase where $m = 0$ one obtains

$$\chi = \left(1 - \frac{q}{2}\,\frac{g-1}{g}\right)^{-1}, \tag{4.49}$$

while in the broken symmetry phase the calculations for $q = 3$ lead to

$$\chi = \frac{4g}{(g-3)(g-2)^2(g+1)}. \tag{4.50}$$

The derivation of a corresponding formula for $q = 4$ is straightforward.

When comparing analytical results to MC simulations we do actually use a modified definition of susceptibility

$$\widetilde{\chi} = \frac{1}{n}\left(\langle M^2\rangle - \langle |M|\rangle^2\right) \tag{4.51}$$

instead of χ given by Eq. (4.46). This is motivated by the fact that the average magnetization $\langle M\rangle$ is not well defined in numerical simulations. On a finite lattice it is in principle zero across the whole range of g. Nevertheless, in the broken symmetry phase its measured value will in general depend on the algorithm used and the duration of the simulation. The average absolute value of magnetization is given by the integral

$$\langle |M|\rangle = n\frac{\int_0^1 \mathrm{d}z|(2z-1)|\,\mathrm{e}^{-nF(z)}}{\int_0^1 \mathrm{d}z\,\mathrm{e}^{-nF(z)}}, \tag{4.52}$$

and its approximation in the symmetric phase is given by

$$\frac{1}{n}\langle |M|\rangle^2 = \frac{2}{\pi}\chi. \tag{4.53}$$

In the broken symmetry phase $\langle |M|\rangle = n\,m$. We plot the resulting expression for $\widetilde{\chi}$ (dashed line) together with data obtained from MC simulations in Fig. 4.3.

Instead of performing saddle-point approximation of the integrals (4.47) and (4.52), we can integrate them numerically. The results are plotted in Fig. 4.3 with solid lines. As one can see the agreement is very good already for small systems, even if we keep only the leading order terms given by Eq. (4.27). Including higher-order terms does not improve the result significantly.

4.3 Correlations

The solution of the model coincides in the leading order of n with the Bethe solution for Ising spins defined on infinite Cayley trees, or Bethe lattices, with coordination number equal q [48]. Qualitatively this corresponds to graphs with predominately long cycles, which can be locally considered as tree-like. Such approximation allows us to derive the correlation functions using the transfer matrix approach originally proposed for branched polymers [79] assuming that there is only one path between the two vertices of interest some distance r apart.

The partition function of the Ising model on Cayley trees can be obtained from that of the model defined on an ensemble of planted trees. These are trees with an additional external phantom node attached to the root, which can be thought of as a handle by which the trees are glued together—we

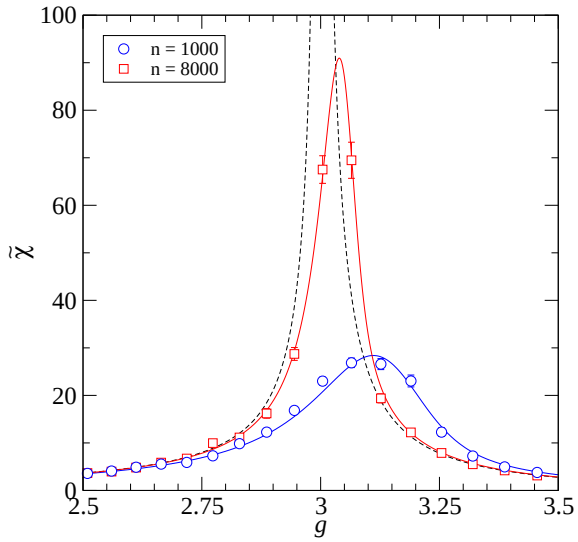

Figure 4.3: $\widetilde{\chi}$ as a function of the coupling g for $q=3$. Circles (blue) render MC data for graphs of size $n = 1000$ and squares (red) for those ones of size $n = 8000$ vertices. The error bars show statistical uncertainty which increases near the transition. The solid lines plot the corresponding results of numerical integration whereas the saddle-point approximation is given by the dashed line.

shall refer to it as the stem. The partition function of planted trees may be distinguished into Φ_+ and Φ_- depending on the value of the spin carried by the external node. The original partition function Z is then given by the partition functions of q such trees connected by their stems,

$$Z = \frac{1}{q!}\left(\Phi_+^q + \Phi_-^q\right). \tag{4.54}$$

The relative ordering of the individual components is irrelevant, hence the $1/q!$ factor.

The partition functions $\Phi_\pm$ of planted trees decorated by Ising spins can be recursively defined by:

$$\begin{cases} \Phi_+ &= \dfrac{1}{(q-1)!}\left(\sqrt{g}\,\Phi_+^{q-1} + \dfrac{1}{\sqrt{g}}\,\Phi_-^{q-1}\right), \\ \Phi_- &= \dfrac{1}{(q-1)!}\left(\dfrac{1}{\sqrt{g}}\,\Phi_+^{q-1} + \sqrt{g}\,\Phi_-^{q-1}\right). \end{cases} \tag{4.55}$$

A graphical representation of these equations is depicted in Fig. 4.4, in which the generating functions for Cayley trees with coordination number equal three are shown. The bright bubbles correspond to Φ_+, the dark ones to

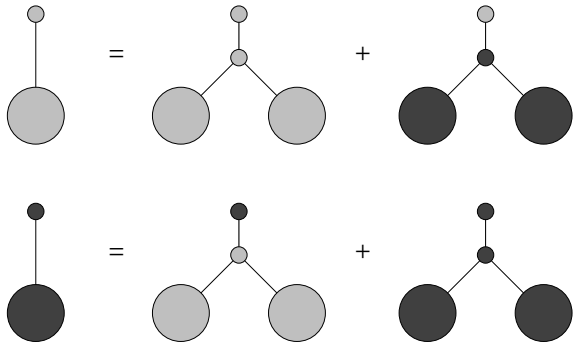

Figure 4.4: Graphical representation of the generating equations (4.55) for coordination number $q=3$. The bright bubbles correspond to Φ_+ and the dark ones to Φ_-. The bright and dark points are the vertices carrying the $+$ and $-$ spins, respectively.

Φ_-, and the bright and dark points stand for vertices carrying the "up" and "down" spins, respectively. In each of $\Phi_\pm$ the root can be connected to a vertex carrying either an "up" or a "down" spin, to which $q-1$ components $\Phi_\pm$ are attached.

Eqs. (4.55) are analogous to those of Ising spins arranged on branched polymers [79]. The difference is that here the probability of ending the branch is zero, which formally means that the considered trees are infinite. Hence, there is no chemical potential associated with each vertex.

The solution of Eqs. (4.55) depends on the side of the criticality. For $g \leq g_c$ the system is in the symmetric phase and $\Phi_+ = \Phi_-$. Then both equations reduce to one,

$$\Phi_\pm = \frac{1}{(q-1)!}\left(\sqrt{g} + \frac{1}{\sqrt{g}}\right)\Phi_\pm^{q-1}, \tag{4.56}$$

which can be readily solved yielding

$$\Phi_\pm^{q-2} = \frac{\sqrt{g}}{1+g}\,(q-1)! = \frac{1}{2\cosh\beta}(q-1)!. \tag{4.57}$$

When $g > g_c$ the symmetry is broken and the functions Φ_+ and Φ_- become distinct, resulting in a non-zero net magnetization. Then, however, the system cannot be solved analytically for general q. A few sample solutions for specific values of q ranging from 3 up to 6 are summarized in Tab 4.1. Magnetization in the symmetry-broken phase is given by [47],

$$m = \frac{\Phi_+^q - \Phi_-^q}{\Phi_+^q + \Phi_-^q}. \tag{4.58}$$

q	$\Phi_\pm$
3	$\dfrac{\sqrt{g}}{g-1}\left(1 \pm \sqrt{\dfrac{g-3}{g+1}}\right)$
4	$\left[\dfrac{3\sqrt{g}\left(g \pm \sqrt{g^2-4}\right)}{g^2-1}\right]^{\frac{1}{2}}$
5	$\left\{\dfrac{6\sqrt{g}}{g^2-1}\left[3g+1-A \pm \sqrt{-10-2A+2g\,(g+A)}\right]\right\}^{\frac{1}{3}}$ where $A = \sqrt{5+g\,(g+2)}$
6	$\left\{\dfrac{30\sqrt{g}}{g^2-1}\left[3g-\sqrt{4+g^2} \mp \sqrt{-12+2g\big(g+\sqrt{4+g^2}\big)}\right]\right\}^{\frac{1}{4}}$

Table 4.1: Analytical solutions of Eqs. (4.55) in the symmetry-broken phase for a few sample values of the branching factor q.

If we now substitute into the above equation values from Tab. 4.1, we will obtain expressions identical to those calculated in Sec. 4.2 for $q = 3$ and $q = 4$ given by Eqs. (4.44) and (4.45).

To proceed with correlation functions let us consider a chain of $r+1$ vertices along the path joining points i and j, depicted in Fig. 4.5. To each of the $r-1$ intermediate vertices $q-2$ rooted labeled trees are attached, which can be freely permuted in $(q-2)!$ ways. The endpoints are connected to $q-1$ such trees. The spin of each vertex may be either "up" or "down", and every link is weighted by the usual two-point weight $\left(\sqrt{g}\right)^{s_i s_j}$, where s_i and s_j are the spins carried by its ends. The sum over all possible arrangements of spins on the chain can be expressed in matrix notation by

$$Z(r)_{ij} = \frac{1}{[(q-1)!]^2}\Phi_x^{q-1}\left(\overbrace{\Delta\mathbf{M}\cdots\Delta\mathbf{M}}^{r-1}\Delta\right)_{ij}\Phi_y^{q-1}, \tag{4.59}$$

where

$$\mathbf{M} \equiv \frac{1}{(q-2)!}\begin{pmatrix}\Phi_+^{q-2} & 0\\ 0 & \Phi_-^{q-2}\end{pmatrix}. \tag{4.60}$$

It will be convenient to transform Eq. (4.59) into a more compact form using matrices

$$\mathbf{M}^{\frac{1}{2}} = \frac{1}{\sqrt{(q-2)!}}\begin{pmatrix}\Phi_+^{\frac{q-2}{2}} & 0\\ 0 & \Phi_-^{\frac{q-2}{2}}\end{pmatrix}, \tag{4.61}$$

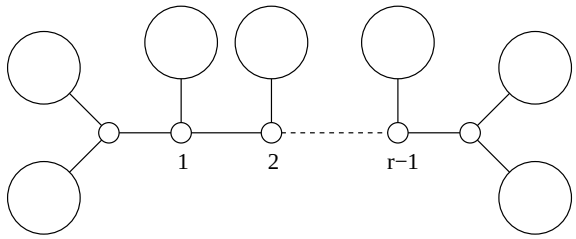

Figure 4.5: Graphical representation of the correlation function $Z(r)_{ij}$ given by Eq. (4.59) for $q=3$. Bubbles depict $\Phi_\pm$ and points stand for individual vertices connected by links associated with Δ. Only the first and the last vertex have specific spins and each of them contributes by a $\frac{1}{2}\Phi_\pm^2$ factor. The spin of the $r-1$ vertices along the path connecting them can be either "up" or "down" which is covered by the matrix $\mathbf{M}$.

and

$$\widetilde{\mathbf{Q}} \equiv \mathbf{M}^{\frac{1}{2}} \Delta \mathbf{M}^{\frac{1}{2}}, \tag{4.62}$$

which allow us to rewrite $Z(r)_{ij}$ in the form

$$\begin{aligned} Z(r)_{ij} &= (q-2)!\, \frac{\Phi_x^{q-1} M_{xx}^{-\frac{1}{2}}}{(q-1)!} \left(\mathbf{M}^{\frac{1}{2}} \Delta \mathbf{M}^{\frac{1}{2}}\right)^r_{ij} \frac{M_{yy}^{-\frac{1}{2}} \Phi_y^{q-1}}{(q-1)!} \\ &= (q-2)!\, \frac{\Phi_x^{\frac{q}{2}}}{(q-1)!}\, Q^r_{ij}\, \frac{\Phi_y^{\frac{q}{2}}}{(q-1)!}. \end{aligned} \tag{4.63}$$

The matrix power in Eq. (4.63) can be now calculated by diagonalizing $\mathbf{Q}$. If λ_1 and λ_2 are the eigenvalues of $\mathbf{Q}$, and the corresponding normalized eigenvectors are (a, b) and $(-b, a)$, respectively, then

$$\mathbf{Q} = \mathbf{P} \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \mathbf{P}^{-1}, \tag{4.64}$$

where

$$\mathbf{P} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{4.65}$$

is the matrix composed of the eigenvectors of $\mathbf{Q}$. For normalized eigenvectors $\det \mathbf{P} = 1$ and we have

$$\mathbf{P}^{-1} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \tag{4.66}$$

It follows that the matrix power $\mathbf{Q}^r$ can be expressed by the above

$$\mathbf{Q}^r = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} \lambda_1^r & \\ & \lambda_2^r \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \tag{4.67}$$

Technical details concerning the calculation of eigenvalues and eigenvectors of $\mathbf{Q}$ are given in Appendix D. The final results for λ_1 and λ_2 are:

$$\lambda_1 = q - 1 \,, \tag{4.68}$$

$$\lambda_2 = (q-1)\left[\frac{\sqrt{g}}{(q-1)!}(\Phi_+^{q-2} + \Phi_-^{q-2}) - 1\right] \equiv (q-1)\,\widetilde{\lambda}_2 \,, \tag{4.69}$$

and the eigenvector is given by

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \begin{pmatrix} \Phi_+^{\frac{q}{2}} \\ \Phi_-^{\frac{q}{2}} \end{pmatrix} . \tag{4.70}$$

In the following we will concentrate on calculating the ordinary two-point spin-spin correlators, from which the expression for the connected correlation function will be derived. To proceed, let us recall the basic correlators introduced in Secs. 1.4.1 and 1.4.2,

$$G^{ss}(r) = \frac{1}{n}\left\langle \sum_{i,j} s_i s_j \delta_{d(i,j),r} \right\rangle , \tag{4.71}$$

$$G^{s1}(r) = \frac{1}{n}\left\langle \sum_{i,j} s_i \delta_{d(i,j),r} \right\rangle , \tag{4.72}$$

$$G^{11}(r) = \frac{1}{n}\left\langle \sum_{ij} \delta_{d(i,j),r} \right\rangle . \tag{4.73}$$

These are useful in defining the distance-dependent averages (1.26) and (1.28):

$$\langle ss(r)\rangle = \frac{G^{ss}(r)}{G^{11}(r)}, \tag{4.74}$$

$$\langle s(r)\rangle = \frac{G^{1s}(r)}{G^{11}(r)}. \tag{4.75}$$

For the connected correlation function we will rely on the definition (1.29),

$$\begin{aligned} G_c^{ss}(r) &\equiv \frac{1}{G^{11}(r)}\left\langle \sum_{i,j} (s_i - \langle s\rangle)\,(s_j - \langle s\rangle)\,\delta_{d(i-j),r} \right\rangle \\ &= \langle ss(r)\rangle - 2m\,\langle s(r)\rangle + m^2, \end{aligned} \tag{4.76}$$

because it has the property of integrating to susceptibility,

$$\sum_r G_c^{ss}(r) G^{11}(r) = \chi. \tag{4.77}$$

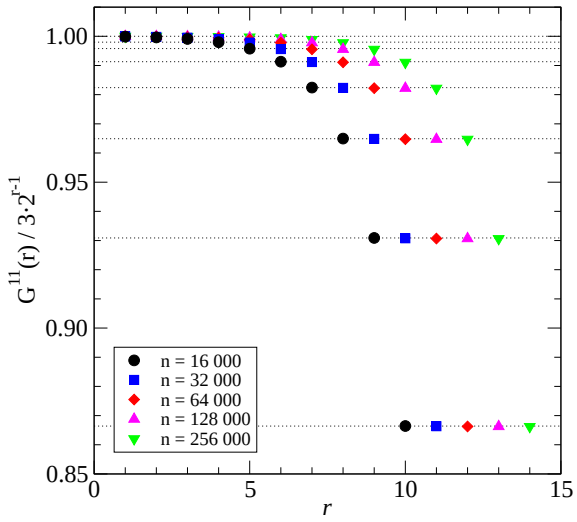

Figure 4.6: Ratio of the correlation function $G^{11}(r)$ obtained from MC simulations of ensembles of various sizes with $q=3$ at $g_c = 3$ to its infinite volume limit $q\,(q-1)^{r-1}$.

The $G^{11}(r)$ correlation function is a volume-factor: it is the average number of vertices at the distance r from some given vertex. On an infinite Cayley tree with fixed geometry it equals

$$G^{11}(r) = q\,(q-1)^{r-1}. \tag{4.78}$$

Figure 4.6 presents the ratio of $G^{11}(r)$ measured in MC simulations to its infinite volume limit (4.78). Please note the scaling relation evident from the graph:

$$2G^{11}(r;n) = G^{11}(r+1;2n). \tag{4.79}$$

This can be also written as

$$G^{11}(r) = n\,\mathcal{F}\left(\frac{2^r}{n}\right) = n\,\widetilde{\mathcal{F}}(r\ln 2 - \ln n). \tag{4.80}$$

The volume factor scales with distance like $G^{11}(r) \propto 2^r$, which indicates that the Hausdorff dimension d_H of the system, as defined in Eq. (3.59), is infinite, $d_H = \infty$.

The two-point correlation functions (4.71), (4.72), and (4.73) can be

expressed using $Z(r)_{ij}$ as:

$$G^{11}(r) = \frac{1}{Z}\sum_{ij} Z(r)_{ij}, \tag{4.81}$$

$$G^{1s}(r) = \frac{1}{Z}\sum_{ij} s_i Z(r)_{ij}, \tag{4.82}$$

$$G^{ss}(r) = \frac{1}{Z}\sum_{ij} s_i s_j Z(r)_{ij}. \tag{4.83}$$

Substituting the previously derived expression (4.63) for $Z(r)_{ij}$ and performing the summation over spins we readily obtain

$$G^{11}(r) = \frac{1}{Z}\,\frac{\lambda_1^r(a\Phi_+^{\frac{q}{2}} + b\Phi_-^{\frac{q}{2}})^2 + \lambda_2^r(b\Phi_+^{\frac{q}{2}} - a\Phi_-^{\frac{q}{2}})^2}{(q-1)!(q-1)}, \tag{4.84}$$

$$G^{1s}(r) = \frac{1}{Z}\,\frac{\lambda_1^r(a^2\Phi_+^q - b^2\Phi_-^q) + \lambda_2^r(b^2\Phi_+^q - a^2\Phi_-^q)}{(q-1)!(q-1)}, \tag{4.85}$$

$$G^{ss}(r) = \frac{1}{Z}\,\frac{\lambda_1^r(a\Phi_+^{\frac{q}{2}} - b\Phi_-^{\frac{q}{2}})^2 + \lambda_2^r(b\Phi_+^{\frac{q}{2}} + a\Phi_-^{\frac{q}{2}})^2}{(q-1)!(q-1)}. \tag{4.86}$$

The above correlation functions can be calculated using the values of λ_1, λ_2, a, and b, given by Eqs. (4.68), (4.69), and (4.70) yielding

$$G^{11}(r) = \frac{1}{Z}\,\frac{(q-1)^{r-1}}{(q-1)!}\,\left(\Phi_+^q + \Phi_-^q\right), \tag{4.87}$$

$$G^{1s}(r) = \frac{1}{Z}\,\frac{(q-1)^{r-1}}{(q-1)!}\,\left(\Phi_+^q - \Phi_-^q\right), \tag{4.88}$$

$$G^{ss}(r) = \frac{1}{Z}\,\frac{(q-1)^{r-1}\left(\Phi_+^q - \Phi_-^q\right)^2 + 4\Phi_+^q\Phi_-^q\widetilde{\lambda}_2^r}{(q-1)!\left(\Phi_+^q + \Phi_-^q\right)}. \tag{4.89}$$

If we now recall the definition of Z (4.54) we will finally obtain for $G^{11}(r)$ the same result as previously given by Eq. (4.78), which proves the consistency of the method used. Similarly, the rest of spin-spin correlation functions reduce to:

$$G^{1s}(r) = q(q-1)^{r-1}\,\frac{\Phi_+^q - \Phi_-^q}{\Phi_+^q + \Phi_-^q} = m\,G^{11}(r), \tag{4.90}$$

$$\begin{aligned} G^{ss}(r) &= q(q-1)^{r-1}\left[\frac{4\Phi_+^q\Phi_-^q}{(\Phi_+^q + \Phi_-^q)^2}\widetilde{\lambda}_2^r + \left(\frac{\Phi_+^q - \Phi_-^q}{\Phi_+^q + \Phi_-^q}\right)^2\right] \\ &= G^{11}(r)\left[\frac{4\Phi_+^q\Phi_-^q}{(\Phi_+^q + \Phi_-^q)^2}\widetilde{\lambda}_2^r + m^2\right]. \end{aligned} \tag{4.91}$$

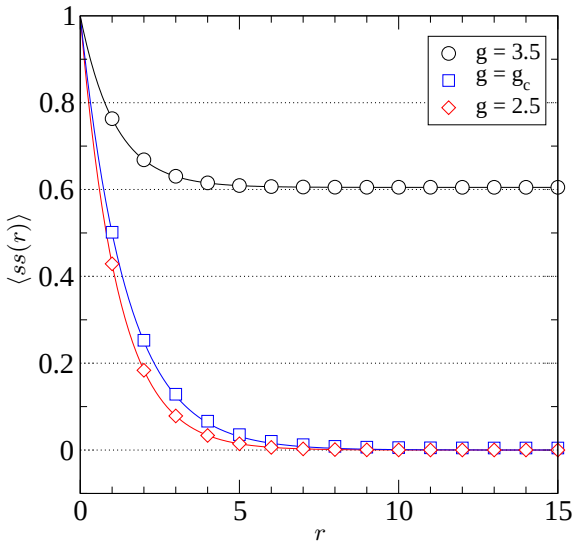

Figure 4.7: Correlation function $\langle ss(r)\rangle$ for $q=3$. Symbols render MC data for graphs of size $n=256000$ in the symmetric phase for $g=2.5$ (diamonds), at the transition $g_c=3$ (squares), and in the symmetry-broken phase for $g=3.5$ (circles). Solid lines plot the analytical predictions given by Eqs. (4.92) and (4.94).

In the symmetric phase, when $m = 0$ and $\Phi_+ = \Phi_-$ are given by Eq. (4.57), we immediately get

$$\langle ss(r)\rangle = \frac{G^{ss}(r)}{G^{11}(r)} = \left(\frac{g-1}{g+1}\right)^r = \tanh^r \beta, \tag{4.92}$$

which agrees with the result obtained in Ref. [88] and coincides with the correlation function of an Ising spin chain (cf. Ref. [94]). It is easy to check that the relation (4.77) is satisfied by the above function.

Combining Eqs. (4.90) and (4.91) with the definition of the connected correlation function (4.76) we get

$$G_c^{ss}(r) = \frac{4\Phi_+^q \Phi_-^q}{(\Phi_+^q + \Phi_-^q)^2}\widetilde{\lambda}_2^r. \tag{4.93}$$

Using the values from Tab. 4.1 we obtain in the symmetry-broken phase for $q = 3$,

$$\langle ss(r)\rangle = \frac{4}{(g-2)^2(g+1)}\frac{1}{(g-1)^r} + m^2, \tag{4.94}$$

and for $q = 4$,

$$\langle ss(r)\rangle = \frac{4}{(g^2-2)^2}\frac{1}{(g^2-1)^r} + m^2. \tag{4.95}$$

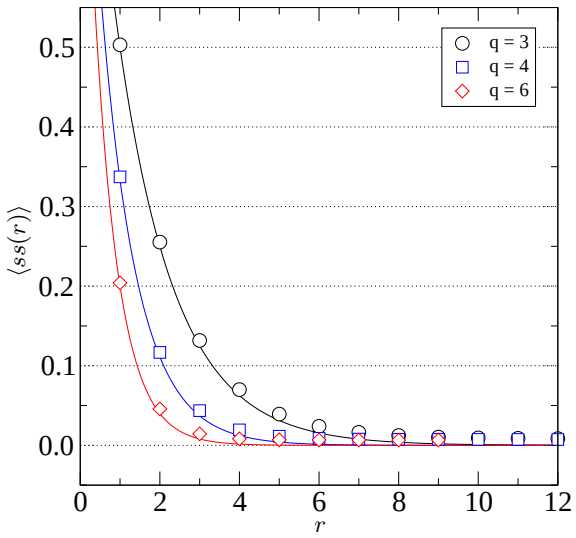

Figure 4.8: Correlation function $\langle ss(r)\rangle$ at the transition $g = g_c$ for different values of q. Points represent MC results for graphs of size $n = 64000$ while the solid lines plot the analytical predictions. Please note that for $q = 6$ the maximum encountered distance (the diameter of the graph) is $r = 9$.

Although similar formulas can be derived for higher values of q, they are much more complicated.

The range of fluctuations of the order parameter of the system are characterized by correlation length ξ defined by (see, e.g., Refs. [27, 36])

$$G_c^{ss}(r) \propto \mathrm{e}^{-r/\xi}. \tag{4.96}$$

The correlation length of Ising spins on Bethe lattices is finite for all values of g and does not diverge at the phase transition unlike in typical models studied in statistical physics. This might appear surprising, but it is an intrinsic property of the model related to the fact that the diameter of the system scales with the number of sites as $\ln(n)$. Accounting for it, we shall rather use the $G^{ss}(r)$ correlation function to extract the correlation length, for which ξ diverges at the critical point g_c as expected.

Figure 4.7 plots the correlation function $\langle ss(r)\rangle$ for $q = 3$ and different values of g. As one can see the agreement with MC results is very good. In Fig. 4.8 correlation functions for different q's at the respective transition points $g = g_c$ are compared. Again, MC results match the asymptotic predictions. The small discrepancies diminish with the increase of the graphs' size.

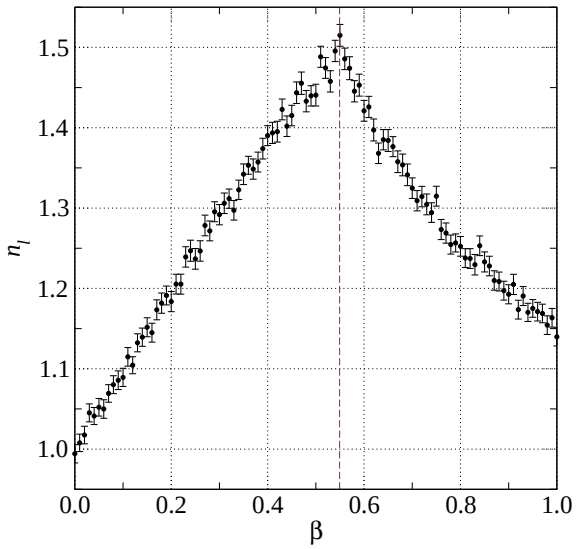

Figure 4.9: Average number of loops n_l in graphs of size $n=32000$ vertices and $q=3$. The black points represent MC data and the dashed red line marks the transition point $\beta=\beta_c$.

4.4 Influence of spins on geometry

On a finite tree spins can be integrated out exactly and the resulting factor does not depend on the shape of the tree. This means that the spins do not have any influence on the geometry. The situation changes when cycles are allowed, which is especially obvious for loops, i.e., links attached at both ends to the same vertex. Without any spins or $\beta = 0$ one can show combinatorially that the expected number of such loops on a graph equals $(q-1)/2$. For positive β, however, we should observe an enhancement as each loop contributes the e^{β} factor to the partition function contrary to links joining different vertices which can have different spins. For $\beta \to \infty$ all the spins have the same sign and again the geometry decouples.

In Fig. 4.9 we plot the average number of loops n_l as a function of β. The results agree qualitatively with the above scenario. Nevertheless, one should note that while looking pronounced this is still only a $1/n$ effect. The number of loops is independent of n and negligible in the large volume limit.

Chapter 5

Random graph model with a phase transition

In the previous chapter we have discussed an Ising spin model arranged on regular random graphs. The order parameter of that model is the spin value at each vertex and we have shown that the system undergoes a phase transition between the ordered and disordered phase. The topology of the underlying graphs is locally tree-like: in the large-volume limit the cycles become predominately long, so they have virtually no influence on the spin structure. Similarly, as the number of loops and multiple edges scales slower than n these local degeneracies are negligible as well. All this results in the behavior of the model identical to Bethe lattices with corresponding coordination number.

In this chapter we study a similar, but purely geometrical model. Geometrical in this context means that the geometry configuration itself determines the state of the system and no additional spin structure is introduced. Unlike in the Ising model previously discussed, we do not restrict ourselves to regular graphs anymore and allow vertices with varying number of neighbors. As a consequence, we may now associate the order parameter of the system with vertex degrees.

One possible choice is to distinguish between two types of vertices: ones with odd degrees and ones with even degrees. The simplest model implementing this concept are random graphs with vertex degrees equal 2, 3, and 4. We shall solve the model and show that in the leading order the solution coincides with the solution derived for the Ising model defined ϕ^3 Feynman diagrams from the previous chapter. Thus, asymptotically both models are thermodynamically equivalent and exhibit the same critical behavior.

Nevertheless, although the solutions of these two models coincide, qualitatively they are completely different, and so the nature of the phase transition observed: it consists in a rearrangement of the graph's geometrical

structure rather than the alignment of the spins. The proposed model is useful in the sense that there are only few models exhibiting an exclusively geometrical phase transition.

5.1 Definition of the model

Let us consider unlabeled graphs $\mathcal{G}(n)$ of size n and average vertex degree equal $\langle q \rangle = 3$. For a fixed number of vertices the number of edges is definite and equal $l = \frac{1}{2} \langle q \rangle n = \frac{3}{2} n$, which implies that n has to be even. Furthermore, we narrow the set of possible vertex degrees by allowing vertices with degrees 2, 3, and 4 only. There are no further restrictions considering the topology. In particular, multigraphs containing multiple edges and loops are covered by this definition. Similarly as in the previous chapter, we will refer to these graphs shortly as random graphs.

One could ask, why not just take graphs build of vertices having only two distinct degree values, say 3 and 4, for example? Then, however, it would be impossible to keep the number of vertices and links simultaneously constant leading to problems with numerical simulations and with the thermodynamic limit.

The smallest possible ensemble of the described type is that for $n = 2$ presented in Fig. 5.1. It consists of four graphs in total, three of which are connected, and one has two separate components. The next in size ensemble is that with $n = 4$, as for $n = 3$ it would be impossible to keep the desired average vertex degree equal 3. All of the fifty-five graphs forming the ensemble are depicted in Fig. 5.2, and the corresponding symmetry factors listed in Tab. 5.1.

The number of non-isomorphic shapes grows rapidly with the increase of the graphs' size n (by definition, *isomorphic* graphs differ from each other only by vertex and edge labels). We have checked that there are 1072 topologies made up of six vertices, but found it hard to enumerate all the possibilities in acceptable time for graphs as small as $n = 8$ already.

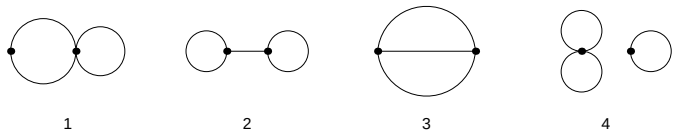

Figure 5.1: Ensemble of all graphs with $n = 2$ vertices. Topologies 1, 2, and 3 are connected, whereas topology 4 has two separate components. The corresponding symmetry factors are as follows: $s(1) = 4$, $s(2) = 8$, $s(3) = 12$, $s(4) = 16$.

So far we have considered unlabeled graphs. These are identical to the corresponding Feynman diagrams containing vertices of degree 2, 3, and 4, which are generated by the perturbative expansion of integrals appearing in the zero-dimensional field theory as discussed in Sec. 1.5. This analogy provides a useful way to enumerate them systematically and we will make use of it in the following section.

In the following we will be actually interested in the properties of the canonical ensemble of fully labeled random graphs $\mathcal{G}_{fl}(n)$ defined by the canonical partition function

$$Z(n) = \frac{1}{n!(2l)!} \sum_{G \in \mathcal{G}_{fl}(n)} w(G), \tag{5.1}$$

where $1/n!(2l)!$ accounts for the possible permutations of indices. The weights $w(G)$ of individual graphs depend on their properties and are discussed in detail later on. When it comes to pure topology, all the graphs from this ensemble share the same structural weight and are equiprobable. Each shape from the former unlabeled ensemble corresponds to a number of such labeled graphs. These two ensembles can be related by the symmetry factors $s(G)$ associated with Feynman diagrams (see Sec. 1.5.3). The partition function of the labeled ensemble (5.1) is equivalent to the partition function of the unlabeled topologies $\mathcal{G}(n)$, i.e., Feynman diagrams, weighted by symmetry factors,

$$Z(n) = \sum_{G \in \mathcal{G}(n)} \frac{1}{s(G)}\, w(G). \tag{5.2}$$

The main reason why it is convenient to work with labeled diagrams rather than bare topologies is that the former ones are the native representation of graphs in numerical simulations.

Now we come back to the weights $w(G)$. They depend solely on the graph's topology and are a product of one-point and two-point weights,

$$w(G) = \prod_{i \in G} w_{q_i} \prod_{\{j,k\} \in G} w_{q_j, q_k}, \tag{5.3}$$

where q_i is the i-th vertex degree, and by $\{j, k\}$ we mean the edge joining vertices i and j. The model is parametrized by two variables σ and g corresponding to one- and two-point weights, respectively.

The one-point weights are assigned to vertices in a way that each vertex of odd degree contributes by a σ factor, while every vertex having an even degree carries σ^{-1}. Formally,

$$w_q = \begin{cases} \sigma & \text{if} \quad q = 3, \\ \sigma^{-1} & \text{if} \quad q \in \{2, 4\}, \\ 0 & \text{otherwise.} \end{cases}$$

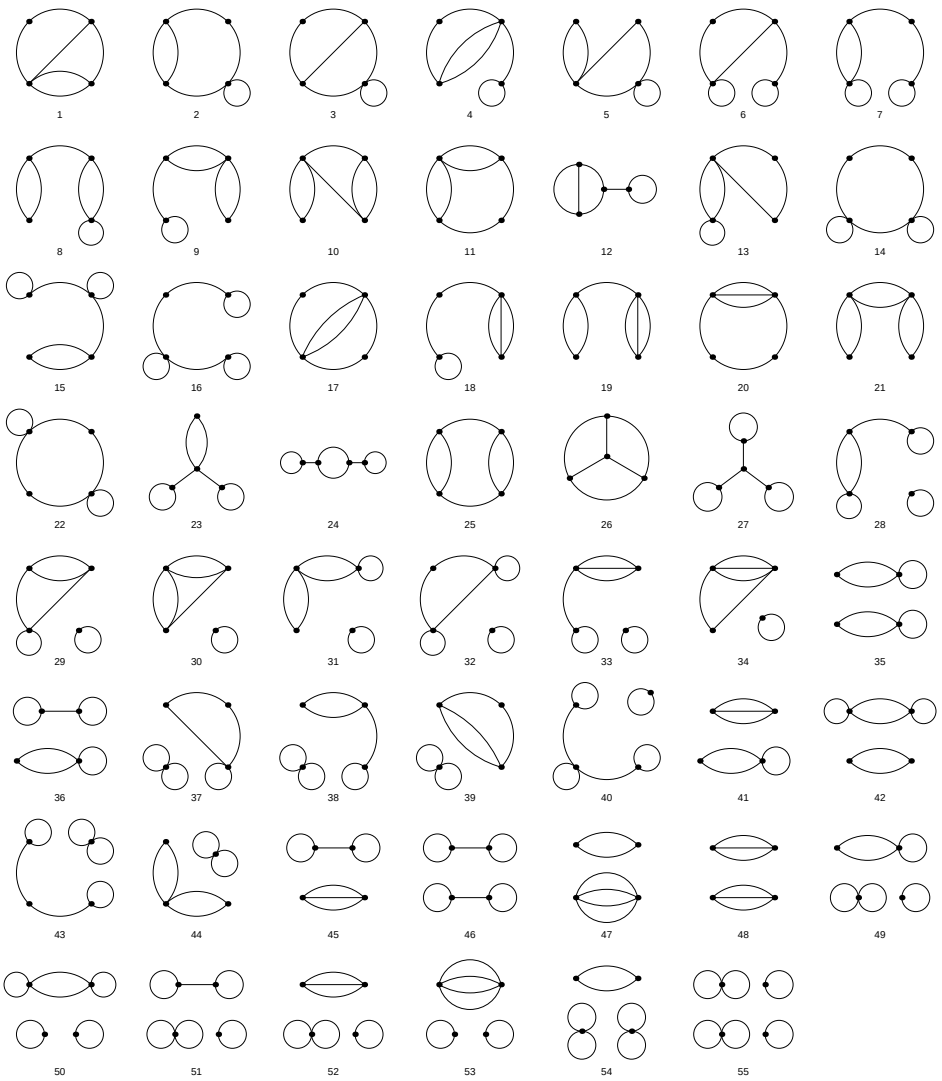

Figure 5.2: Ensemble of all graphs with $n = 4$ vertices. Topologies 1–27 are connected, 28–48 and 49–54 have two and three separate components, respectively, and the last one consists of 4 disjoint vertices. The corresponding symmetry factors are listed in Tab. 5.1. The diagrams were plotted and their corresponding symmetry factors calculated using the MATHEMATICA package FEYNARTS [95].

Symmetry factor s	**Topology No.**
2	1
4	2, 3, 4, 5, 6
8	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
12	18, 19, 20
16	21, 22, 23, 24, 25, 28, 29, 30, 31, 32
24	26, 33, 34
32	35, 36, 37, 38, 39, 40
48	27, 41
64	42, 43, 44, 49
96	45
128	46, 50, 51
192	47, 52
288	48
384	53
512	54
1024	55

Table 5.1: Symmetry factors of the graphs from the $n = 4$ ensemble in Fig. 5.2.

Such a choice narrows the possible degrees of graph's vertices down to 2, 3, and 4 only. From now on we will shortly refer to vertices having degree 3 as *odd vertices*, and those with even degrees as *even vertices*.

The two-point weights are symmetric and parametrized by the coupling strength g. Each edge connecting vertices of the same type, i.e., both having odd/even degrees, carries a $\sqrt{g}$ factor, whereas edges joining vertices of different parity are weighted by $\frac{1}{\sqrt{g}}$. This can be summarized by

$$w_{q_i,q_j} = g^{\frac{1}{2}-[(q_i+q_j) \bmod 2]}. \tag{5.4}$$

5.2 Solution

As it was put forward in the previous section, the graphs from the introduced ensemble are identical with Feynman diagrams appearing in zero-dimensional field theory. Therefore, the partition function of the model can be generated by similar means as the partition function of the Ising spins on

regular random graphs from the previous chapter given by Eq. (4.5). The difference is that now the fields do not distinguish between spins put on vertices but between vertex degrees. Thus, after relating the odd and even vertices to the ϕ_o and ϕ_e fields, respectively, the grand-canonical partition function $Z(\mu)$ is given by the perturbative expansion of the integral,

$$Z(\mu) = \frac{1}{2\pi} \int \mathrm{d}\phi_o \mathrm{d}\phi_e \, \exp\left\{ -\frac{1}{2} \boldsymbol{\phi}^T \Delta^{-1} \boldsymbol{\phi} + \mathrm{e}^{-\mu} \left[\frac{\sigma}{3!} \phi_o^3 + \frac{1}{\sigma} \left(\frac{1}{2!} \phi_e^2 + \frac{1}{4!} \phi_e^4 \right) \right] \right\}. \tag{5.5}$$

Here

$$\boldsymbol{\phi} \equiv \begin{pmatrix} \phi_o \\ \phi_e \end{pmatrix}, \tag{5.6}$$

and Δ defines the transfer matrix identical to the previous one (4.7),

$$\Delta \equiv \begin{pmatrix} \sqrt{g} & \frac{1}{\sqrt{g}} \\ \frac{1}{\sqrt{g}} & \sqrt{g} \end{pmatrix}. \tag{5.7}$$

Let n be the total number of vertices in a graph and k be the number of odd vertices. Following the scheme from Chapter 4 we rewrite the σ dependend part of the exponent in Eq. (5.5) as infinite series and perform binomial expansion of the resulting expression:

$$\begin{aligned}
&\exp\left[\frac{\sigma}{3!} \phi_o^3 + \frac{1}{\sigma} \left(\frac{1}{2!} \phi_e^2 + \frac{1}{4!} \phi_e^4 \right) \right] \\
&\quad = \sum_{n=0}^{\infty} \frac{1}{n!} \left[\frac{\sigma}{3!} \phi_o^3 + \frac{1}{\sigma} \left(\frac{1}{2!} \phi_e^2 + \frac{1}{4!} \phi_e^4 \right) \right]^n \\
&\quad = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\sigma}{3!} \phi_o^3 \right)^k \left[\frac{1}{\sigma} \left(\frac{1}{2!} \phi_e^2 + \frac{1}{4!} \phi_e^4 \right) \right]^{n-k} \\
&\quad = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \sigma^k \frac{1}{\sigma^{n-k}} \frac{1}{n!} \frac{n!}{k!(n-k)!} \frac{1}{6^k} \phi_o^{3k} \binom{n-k}{\frac{n-k}{2}} \left(\frac{1}{2!} \phi_e^2 \right)^{\frac{n-k}{2}} \left(\frac{1}{4!} \phi_e^4 \right)^{\frac{n-k}{2}} \\
&\quad = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \sigma^{2k-n} \frac{1}{k!} \left[\left(\frac{n-k}{2} \right)! \right]^{-2} \frac{1}{6^k \, 48^{\frac{n-k}{2}}} \, \phi_o^{3k} \, \phi_e^{3(n-k)}.
\end{aligned} \tag{5.8}$$

In the above we have used the fact that since vertices with even degrees are evenly distributed between the ones having degree two and those with degree four, the number of vertices in each of the classes equals $(n-k)/2$.

We may now extract the canonical partition function Z_n of graphs with a specific number of vertices n using the relation (4.10),

$$\begin{aligned}
Z_n = \frac{1}{2\pi} \sum_{k=0}^{n} \sigma^{2k-n} \frac{1}{k!} \left[\left(\frac{n-k}{2} \right)! \right]^{-2} \frac{1}{6^k \, 48^{\frac{n-k}{2}}} \\
\times \int_{-\infty}^{+\infty} \mathrm{d}\phi_o \mathrm{d}\phi_e \, \mathrm{e}^{-\frac{1}{2} \boldsymbol{\phi}^T \Delta^{-1} \boldsymbol{\phi}} \, \phi_o^{3k} \, \phi_e^{3(n-k)}.
\end{aligned} \tag{5.9}$$

Likewise for the Ising spin model from the previous chapter, it is convenient to switch to the rescaled fields,

$$\psi_{o,e} \equiv \frac{1}{\sqrt{n}}\,\phi_{o,e}, \tag{5.10}$$

and use the fraction of odd vertices,

$$z \equiv \frac{k}{n},$$

rather than their number k. In these new variables the partition function Z_n takes in the asymptotic limit $n \to \infty$ the form

$$Z_n = \frac{n}{2\pi} A_n \int_0^1 \mathrm{d}z\, B_n(z) \int \mathrm{d}\psi_o \mathrm{d}\psi_e\, \mathrm{e}^{n f(\psi_o,\psi_e)}, \tag{5.11}$$

where

$$f(\psi_o,\psi_e) = -\frac{1}{2}\,\boldsymbol{\psi}^T \Delta^{-1} \boldsymbol{\psi} + 3\left[z \ln \psi_o + (1-z)\ln \psi_e\right], \tag{5.12}$$

and

$$A_n = \exp\left(\frac{3}{2} n \ln n\right), \tag{5.13}$$

$$B_n(z) = \sigma^{2k-n}\,\frac{1}{k!}\left[\left(\frac{n-k}{2}\right)!\right]^{-2} \frac{1}{6^k\, 48^{\frac{n-k}{2}}}. \tag{5.14}$$

The function $f(\psi_o,\psi_e)$ given by Eq. (5.12) is identical to the corresponding function of the Ising model given by Eq. (4.15) for $q = 3$. Therefore, we may almost immediately write the final result for the partition function

$$Z_n \sim \mathrm{e}^{-nC_n} \int_0^1 \mathrm{d}z\, \mathrm{e}^{-nF(z)}, \tag{5.15}$$

where $F(z)$ is now given by

$$F(z) \sim -3\left[z \ln \overline{\psi}_o + (1-z)\ln \overline{\psi}_e\right] - \frac{1}{n} \ln B_n(z), \tag{5.16}$$

and $\overline{\psi}_o$, $\overline{\psi}_e$ denote the extremum of the f function. We still have to calculate the $\ln B_n(z)$ term in $F(z)$. Its leading order in n approximation using the Stirling formula (4.24) yields

$$\begin{aligned}\frac{1}{n} \ln B_n(z) \sim {} & -z \ln z - (1-z)\ln(1-z) + z\left(2\ln\sigma - \frac{1}{2}\ln 3\right) \\ & + 1 - \ln\left(2\sqrt{3}\,\sigma\, n\right).\end{aligned} \tag{5.17}$$

By plugging the above expansion into Eq. (5.16) and drawing outside the integral the terms which do not depend on z we get for $F(z)$ in the leading order,

$$F(z) \sim z \ln z + (1-z) \ln (1-z) - 3 \left[z \ln \overline{\psi}_o + (1-z) \ln \overline{\psi}_e \right] \\ -z \left(2 \ln \sigma - \frac{1}{2} \ln 3 \right), \tag{5.18}$$

and for the volume factor

$$C_n = - \ln \left[\pi \left(3\, n\, \mathrm{e}\, \sigma^2 \right)^{-\frac{n}{2}} \right]. \tag{5.19}$$

We shall now compare Eq. (5.18) to the corresponding expression for the Ising model $F_{\mathrm{Ising}}(z)$ with coordination number $q = 3$ from Chapter 4. Both models differ by

$$F_{\mathrm{Ising}}(z) - F(z) = z \left(2 \ln \sigma - \frac{1}{2} \ln 3 \right), \tag{5.20}$$

which is the $F(z) = F(1-z)$ symmetry breaking term. Using the analogy to spin systems we may write $\sigma = \mathrm{e}^h$ and interpret h as the externally applied magnetic field. When

$$\sigma_c = \sqrt[4]{3} \tag{5.21}$$

the difference (5.20) cancels and in the leading order in n approximation the considered model is equivalent to the previously discussed Ising model on 3-regular random graphs. However, unlike in the spin model, the symmetry is retained only in the leading order and broken by higher-order terms.

Figure 5.3 contains plots of the $F(z)$ function for various values of the g and σ parameters. We already know the critical coupling value from the Ising model: it is given by Eq. (4.42) and for $q = 3$ reads $g_c = 3$. The external field σ controls the strength of the asymmetric term in $F(z)$, which depends linearly on z. For $\sigma < \sigma_c$ it is an increasing function in z and a decreasing one for $\sigma > \sigma_c$. It contributes to the free energy function either by raising its values at the $z \to 1$ end (top row of Fig. 5.3) or by lowering it (bottom row of Fig. 5.3).

The order parameter of the model m is now the per spin difference in odd and even vertices in the system. It can be approximated in a similar fashion as in the case of the Ising model by the value of z_0 minimizing the $F(z)$ function,

$$m = 2z_0 - 1. \tag{5.22}$$

We plot the above expression for m in the (g, σ) parameter space in Fig. 5.4, where one immediately recognizes the phase transition with increasing g. A section of this surface in the plane of $\sigma = \sigma_c$ is shown in Fig. 5.5, where points obtained from MC simulations are additionally rendered. We see that

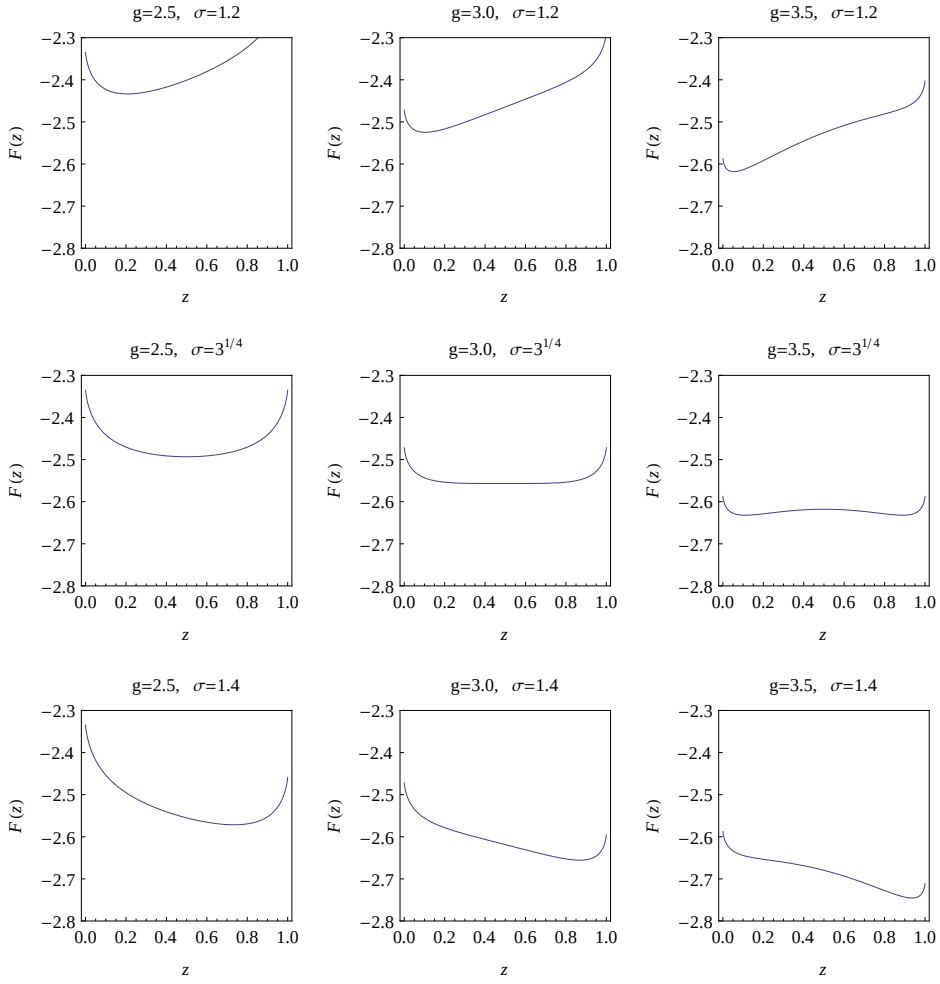

Figure 5.3: Plots of the $F(z)$ function in a (σ, g) parameter grid. σ changes between the rows: in the first one $\sigma = 1.2$, in the middle one $\sigma_c = \sqrt[4]{3}$, and in the last one $\sigma = 1.4$. Each of the rows contains plots for three different values of g: below the critical point for $g = 2.5$, at the transition for $g_c = 3$, and in the broken symmetry phase for $g = 3.5$.

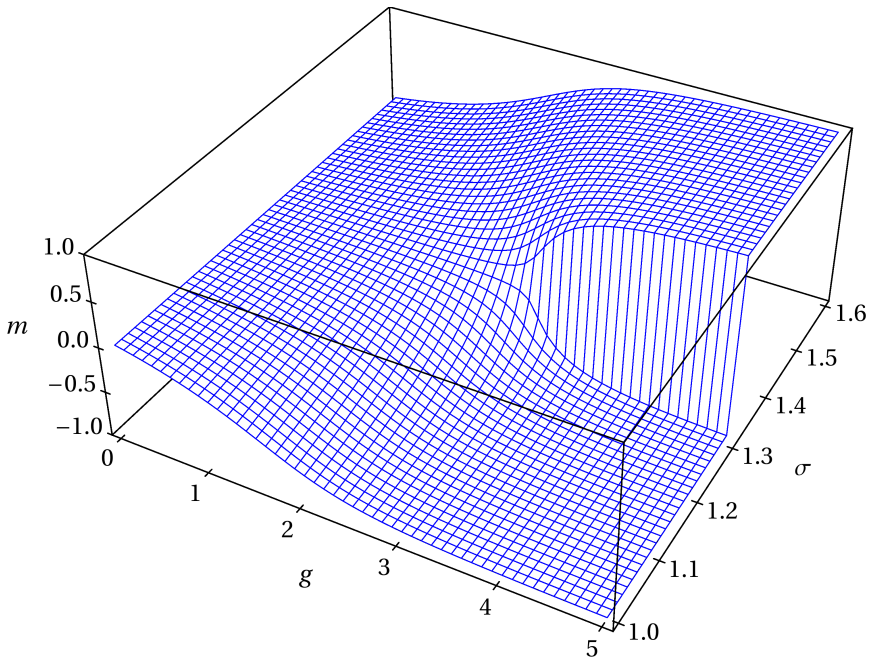

Figure 5.4: The relationship between the average magnetization m and the parameters g and σ. Figure 5.5 shows a section of this surface in the plane of $\sigma = \sigma_c$, whereas Fig. 5.6 shows two sections at $g = g_c$ and $g > g_c$.

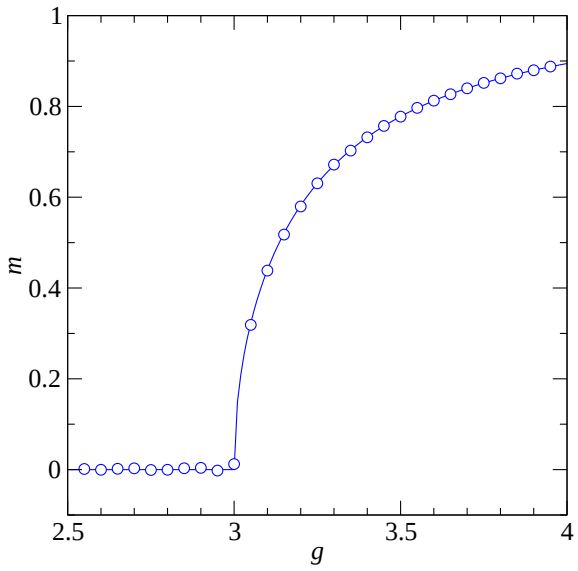

Figure 5.5: The magnetization per spin m as a function of g. This picture corresponds to the previous one for the Ising spin system (cf. Fig. 4.2) with the difference that now we plot the magnetization as a function of the coupling parameter g rather than the generalized temperature $T = 2/\ln g$. Points represent MC simulation results of a system of size $n = 64000$.

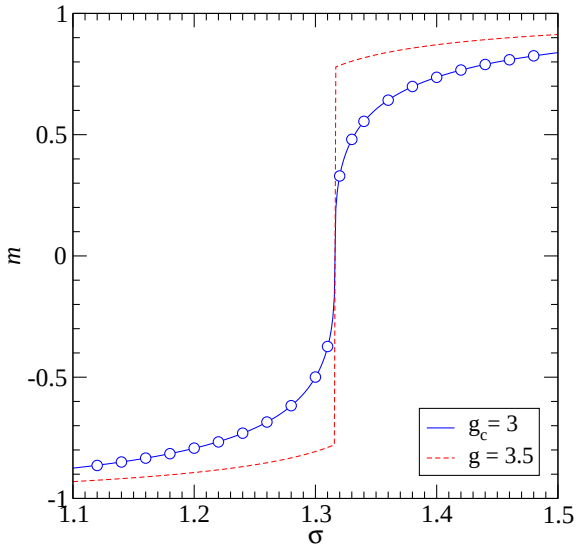

Figure 5.6: The magnetization per spin m as a function of σ. The solid line corresponds to the critical value of the coupling parameter $g_c = 3$, and the dashed line to $g = 3.5$. Points represent MC simulation results of a system of size $n = 64000$.

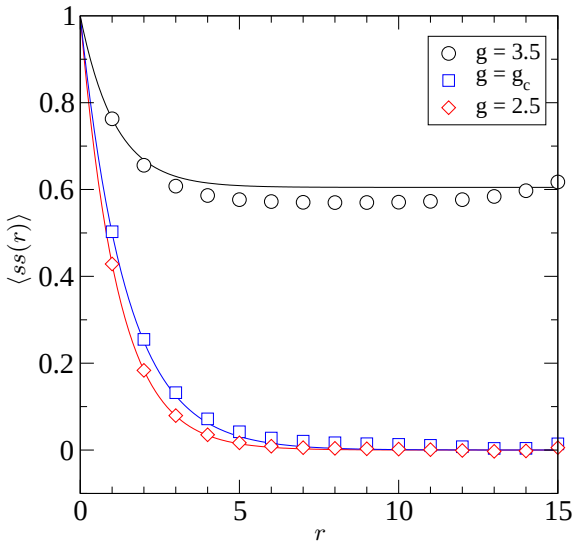

Figure 5.7: Correlation function $\langle ss(r)\rangle$. Symbols render MC simulation results for a system of size $n = 128000$. Solid lines plot the analytical predictions for the Ising model given by Eqs. (4.92) and (4.94).

at the critical value g_c the system changes continuously from the disordered to the ordered phase.

Above the critical coupling value g_c the magnetization m is a discontinuous function of σ. Figure 5.6 contains plots of the average magnetization m as a function of σ at the criticality $g = g_c$ and in the ordered phase for $g = 3.5$. When we vary σ the system changes suddenly at $\sigma_c = \sqrt[4]{3}$ from the even to the odd configuration. It is evident that σ plays the role of a switching parameter determining the type of ordering in the system: its values above σ_c select the odd phase, while the even phase is favored by $\sigma < \sigma_c$.

In the following section we shall compare the asymptotic approximation developed for the Ising spin model to numerical results obtained from MC simulations of the random graph model from this chapter.

5.3 Monte Carlo simulations

We have found that in the asymptotic limit of infinite system size and for $\sigma = \sigma_c$ the solution of the random graph model with a phase transition from the current chapter is identical to the solution of the Ising model arranged on regular random graphs. Therefore, we expect both models to exhibit the same behavior and we may infer the properties of the former one using the formulas developed for the Ising spins. This is especially useful in the case

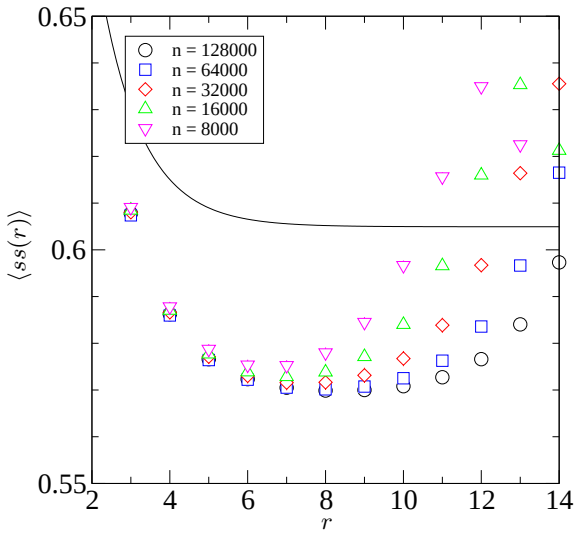

Figure 5.8: Correlation function $\langle ss(r)\rangle$ in the broken symmetry phase. Points render MC results for various system sizes, while the solid line denotes the analytic formula (4.94). Statistical errors are not larger than the size of the points.

of the quantities for which we do not know a simple derivation of analytic formulas, in particular the correlation functions. The method from the previous chapter applied to calculate them does not work anymore, as there is no way to impose the condition of the numbers of vertices with degrees 2 and 4 being equal. It is interesting to confront the asymptotic results derived in Chapter 4 with MC simulations of the current model and see to which extent these two match.

At this point we would like to stress that although the mathematical description of these two models is the same, their nature of the underlying systems is completely different. While for the Ising model the orientation of spins determines the behavior of the system, in the random graph model the geometrical rearrangement of the network itself is crucial.

In order to gain insight in the behavior of the correlation functions we have performed numerical simulations of the model using a Metropolis type algorithm, similar to these ones from the previous chapters. The elementary updates of the network were carried out using edge rewiring performed as follows. First, a random edge, say A-B, and a new attachment point C were drawn. In order to keep the constrains of the model, the edge was detached from its endpoint B and attached back to C only if the degree of B was greater than two, and the degree of the destination point C was smaller than four. If not the case, the move was skipped but still counted to ensure the detailed balance condition, similarly as in the situation when B = C, which

does not change the system at all. Actual updates were performed according to the Metropolis probability given by Eq. (1.81). The local change of the network's weight was calculated by iterating through the connections of the detachment point B and the attachment point C accounting for some special cases arising due to degeneracies in the form of loops and multiedges and situations when B and C are nearest neighbors.

We compare the numerical results obtained for the correlation function $\langle ss(r)\rangle$ of the current random graph model to the analytic formulas (4.92) and (4.94) derived for the Ising spin model from Chapter 4. In the disordered phase and at the transition, the asymptotic predictions of the previous spin model are rather well reproduced by the numerical simulations of the current geometrical one—see Fig. 5.7. However, in the ordered phase we notice some discrepancies; these are plotted for better visibility in a magnified scale in Fig. 5.8. It is evident from the plot that this deviation does not diminish with the increase of the size n of the system and tends asymptotically to a curve laying significantly below the theoretical one for the Ising model. Note that the convergence is rather slow since the scaling of the diameter of the system is logarithmic in n.

Summarizing, the observed differences between the spin and the geometrical model can be explained by the fundamentally different nature of these two models. In particular, in the geometric model of random graphs the symmetry is, unlike in the Ising model, not exact and broken by higher-order terms. In a spin system each configuration has its reversed counterpart obtained simply by flipping all the spins. In the geometrical model, on the other hand, such an analogy does no longer exist since there is no simple one-to-one correspondence between a given topology and one with the opposite ratio of odd to even vertices. Obviously, one cannot just exchange the odd vertices for the even ones and vice versa, since the connectivity of the resulting graph would be undetermined.

Chapter 6

Summary and Conclusions

In the first part of this dissertation we were primarily concerned with the study of structural correlations in connected random graphs with arbitrary degree distribution. We have calculated them analytically and showed that even though there are no explicit interactions between nodes in the system, their degree is correlated. These correlations are of disassortative type and are induced by the additional structural constraint of being connected.

In **Chapter 2**, in particular, we have derived analytic formulas for the joint probability distribution quantifying nearest-neighbor degree correlations in maximally random connected graphs by relating them to the giant connected component. This extends the results obtained previously by the authors of Refs. [20, 68, 69]. Our research strongly indicates that these correlations are related to the presence of leaves, i.e., nodes with degree one. We have found a very firm support for this conjecture: correlations disappear in MC simulations of connected graphs with explicitly forbidden leaves. This suggests that the only reason for the observed correlations is the absence of small two-vertex components and hedgehogs—components composed of a central vertex and a number of leaves attached to it. This has been already stated by Białas in Ref. [44], where the author has shown that in the grand-canonical ensemble of arbitrary-sized trees, in which hedgehogs are present, correlations vanish. This is a very interesting issue which clearly deserves further attention.

We have concluded **Chapter 2** by suggesting a novel and effective method of generating maximally random connected graphs with desired degree distribution ${p_k}^{(g)}$ based on the Metropolis algorithm. Namely, using the formulas from Sec. 2.6 we first calculate the degree distribution p_k of a disconnected ensemble, whose giant connected component will have the required degree distribution ${p_k}^{(g)}$. Then, instead of generating connected graphs directly, we simulate the larger disconnected ones and use their giant component. This is by far more efficient as we do not have to check each time whether the rewiring does not disconnect the graph, which is an

$O(N)$ operation. Therefore, the described procedure reduces the execution time drastically and allows one to simulate even large connected graphs in reasonable time.

In **Chapter 3** we have extended our study of nearest-neighbor correlations from the previous chapter to greater distances by specializing to random trees from the generic phase. We have calculated them exactly in the large-size limit and showed that they remain disassortative at all distances and vanish only as an inverse second power of the distance. The appearance of the long-range correlations in generic trees is puzzling, since normally power-like or scale-free behavior of correlations manifests itself near a phase transition. However, the studied trees are, except for the scale-free ones, not critical (cf. Refs. [77, 78, 96]). Although the mechanism of the correlations between adjoining vertices is rather easy to explain, it remains to be understood how this effect propagates to larger distances.

The correlations observed in connected random graphs are an example of the so-called *structural* or *kinematic correlations* which are a consequence of some global constraints. They should be contrasted with *dynamic correlations* whose origin are local two-point interactions between vertices and which can be generated by ascribing two-point weights to links [79]. This distinction might be vital, for example, in simplicial quantum gravity where degree-degree correlations correspond to curvature-curvature correlations (see, e.g., Ref. [38]). Since the simplicial manifolds are connected by definition, curvature correlations could be attributed to the above described mechanism rather than to some kind of gravitational interaction (cf. Refs. [44, 97]). We deeply believe that our results will help in clarifying such issues and in interpreting data obtained from MC simulations.

In the remainder of the thesis we have focused on critical phenomena occurring in random geometries. Our ultimate goal was to construct and investigate a random geometry model with a geometrical phase transition within its pattern of connections. To this end we have adopted connected random graphs with additional restrictions imposed on vertex degrees.

In order to gain experience and to train our intuition we have first considered in **Chapter 4** classic Ising spins in the environment of regular random graphs, i.e., maximally random (pseudo)graphs with fixed connectivity q. In the large-volume limit this model is formally equivalent to the Ising model on the Bethe lattice. Nevertheless, it has the advantage that it is well defined, has a genuine mean-field phase transition, and can be easily simulated using MC methods. We have calculated the partition function of the model illustrating the approach in which regular random graphs were related to corresponding Feynman diagrams appearing in the graphical perturbation expansion of integrals from quantum field theory. Additionally, we have derived expressions for correlation functions in both the symmetric and the symmetry-broken phase with less formal methods than the ones traditionally used (cf. Ref. [88]). From these calculations we got a picture of

the phase transition. The correlation function $\langle ss(r)\rangle$ does not exhibit any critical behavior: the correlation length is finite for any finite coupling g. Nevertheless, the volume factor $G^{11}(r)$ grows exponentially and this growth compensates the decay of the correlation function $\langle ss(r)\rangle$. As we increase r by one, the influence of the spins at this distance drops by a factor of $(g-1)/(g+1)$. At the same time, however, their number increases by a factor of $q-1$. When these rates are equal, that is, when

$$\frac{g+1}{g-1} = q-1, \tag{6.1}$$

all shells of spins contribute equally and we observe the phase transition. One can check that the above condition is equivalent to the expression (4.31) for the critical value

$$g_c = \frac{q}{q-2}. \tag{6.2}$$

Finally, we have moved on in **Chapter 5** to the random graph model with a geometrical phase transition. By dividing vertices according to their degree into odd and even classes and by identifying these two with the two possible spin orientations in the Ising model we have showed that thermodynamically the geometrical model is identical with the Ising model on top of regular random graphs. Unfortunately, we did not manage to calculate the correlations analytically as there is no easy way to adopt the method used to calculate them in the previous chapter to the model considered here. Despite this, we have compared the outcome of MC simulations to the previous analytic results obtained for the Ising model and found that the agreement of these two near the phase transition and in the symmetric phase is remarkably good. However, in the ordered phase the two systems exhibit some discrepancies. These can be accounted to the fundamental differences of the models. To be specific, the symmetry in the geometrical model is, unlike in the Ising model, not exact. It is broken by the higher-order terms of the $F(z)$ function, so formally $F(z) = F(1-z)$ only in the thermodynamic limit. Moreover, there is no one-to-one correspondence between shapes from the odd and even phases, as it is the case in the Ising spin system, where the counter-configuration is simply obtained by flipping all spins. It is clearly desirable to fully understand the observed behavior of the correlations within the geometrical model. Such understanding could be profitable in future research of some more sophisticated models.

In conclusion, the obtained results are clearly a valuable contribution to the evolving field of complex networks. Nevertheless, as we have pointed out, there are still open questions left which merit further studies. The work in this area is strongly motivated by the desire to better understand the structure of various real-world systems and processes occurring in them.

Appendix A

Maximal entropy ensembles

A.1 Giant connected components

We show by the following argumentation that the giant connected components of a maximal entropy ensemble of random graphs form themselves a maximal entropy ensemble. Let $P(G)$ define a maximal entropy ensemble $\mathcal{G}$ of graphs with V vertices, L links, and vertex degree distribution p_k. We assume that the probability $P(G)$ factorizes into a product of all the G's connected components' weights,

$$P(G) = \prod_{C \in G} P_c(C). \tag{A.1}$$

Furthermore, let $\mathcal{G}_c$ be the ensemble of all giant connected components. In the thermodynamic limit we can neglect the fluctuations and assume that all graphs in this ensemble have $V^{(g)}$ vertices and $L^{(g)}$ links, and that their degree distribution is given by ${p_k}^{(g)}$. Because of the property (A.1), the entropy of the whole ensemble $(\mathcal{G}, P)$,

$$S = -\sum_{G \in \mathcal{G}} P(G) \ln P(G), \tag{A.2}$$

consists of the entropy of the giant connected component ensemble $(\mathcal{G}_c, P_c)$ and the rest,

$$S = S^{(g)} + S^{(f)}. \tag{A.3}$$

Suppose now that there exists another measure P_c' defined on the ensemble $\mathcal{G}_c$ such that the entropy

$$S^{(g)'} = -\sum_{G \in \mathcal{G}_c} P_c'(G) \ln P_c'(G) \tag{A.4}$$

is greater than the original one $S^{(g)}$, but the vertex degree probability distribution remains unchanged. Then we could define a new probability on

the ensemble $\mathcal{G}$ such that

$$P'(G) = P'_c(C^{(g)}) \prod_{C \neq C^{(g)}} P_c(C), \tag{A.5}$$

where $C^{(g)}$ stands for the giant connected component of graph G. The degree distribution of the $(\mathcal{G}, P')$ ensemble would be the same as that of the $(\mathcal{G}, P)$ ensemble, but according to (A.3), its entropy would be greater. This contradicts the assumption that $(\mathcal{G}, P)$ is the maximal entropy ensemble. It follows that the ensemble of giant connected components must be a maximal entropy ensemble.

A.2 Generic random trees

For a given a set of weights w_q and a fixed number of vertices V the ensemble defined by the probability distribution $P(T)$ has a well defined degree distribution

$$p_q = \frac{1}{V} \sum_{T \in \mathcal{T}} n_q(T) P(T), \tag{A.6}$$

where $n_q(T)$ is the number of vertices with degree q in the tree T. In the asymptotic limit of large V this distribution is given by Eq. (3.20). Reference [20] contains a proof that the probability distribution $P(T)$ given by Eq. (3.1),

$$P(T) = \Omega_V^{-1} \frac{1}{V!} \prod_{i \in T} w_{q_i}, \tag{A.7}$$

maximizes entropy among all the measures producing the distribution p_q. Their arguments are repeated here for completeness.

We want to maximize the entropy

$$\mathcal{S} = -\sum_{T \in \mathcal{T}} P(T) \ln P(T) \tag{A.8}$$

under the constraints given by Eq. (A.6) and $\sum_T P(T) = 1$. This is a typical optimization problem which can be solved using the Lagrange multipliers method (see, e.g., Refs. [98] or [99]). Thus we are seeking a stationary point, i.e., a point where all the partial derivatives are zero, of the expression

$$-\sum_{T \in \mathcal{T}} P(T) \ln P(T) + \lambda \left[\sum_{T \in \mathcal{T}} P(T) - 1 \right] + \sum_{q=1}^{V} \lambda_q \left[\sum_{T \in \mathcal{T}} n_q(T) P(T) - p_q V \right], \tag{A.9}$$

where λ and λ_q are the Lagrange multipliers. Differentiating the above with respect to $P(T)$ we get for each $T \in \mathcal{T}$,

$$\ln P(T) = \lambda - 1 + \sum_q \lambda_q n_q(T), \tag{A.10}$$

leading to

$$P(T) = \mathrm{e}^{\lambda-1} \prod_q \mathrm{e}^{\lambda_q n_q(T)}. \tag{A.11}$$

We see that for the choice

$$\mathrm{e}^{\lambda-1} = \Omega_V^{-1} \frac{1}{V!}, \quad \text{and} \quad \mathrm{e}^{\lambda_q} = w_q, \tag{A.12}$$

we reproduce Eq. (A.7). It is, however, not obvious if this solution is unique. We prove that it is indeed as follows. Let us assume that we have another set of weights $\tilde{w}_q$ that produces the same probability distribution p_q, as given by Eq. (3.20),

$$p_q = \frac{1}{F(Z_0)} \frac{w_q Z_0^q}{(q-1)!} = \frac{1}{F(\tilde{Z}_0)} \frac{\tilde{w}_q \tilde{Z}_0{}^q}{(q-1)!}. \tag{A.13}$$

It follows that

$$\tilde{w}_q = \frac{F(\tilde{Z}_0)}{F(Z_0)} \left(\frac{Z_0}{\tilde{Z}_0} \right)^q w_q, \tag{A.14}$$

hence

$$\begin{aligned} \prod_{i \in T} \tilde{w}_{q_i} &= \prod_{i \in T} \frac{F(\tilde{Z}_0)}{F(Z_0)} \left(\frac{Z_0}{\tilde{Z}_0} \right)^{q_i} w_{q_i} \\ &= \left[\frac{F(\tilde{Z}_0)}{F(Z_0)} \right]^V \left(\frac{Z_0}{\tilde{Z}_0} \right)^{\sum_{i \in T} q_i} \prod_{i \in T} w_{q_i}. \end{aligned} \tag{A.15}$$

But this gives identical probability measure to the original one (A.7) because for every tree $\sum_{i \in T} q_i = 2(V-1)$, so the constant factor before the product is canceled by the normalization.

Appendix B

Laplace's integral approximation method

Laplace's approximation method is a technique used to estimate integrals of the form

$$I(n) = \int_a^b \mathrm{d}x \mathrm{e}^{nf(x)} \tag{B.1}$$

for large n, where the variable of integration x is real and $f(x)$ is some twice-differentiable real function independent of n.

The main idea behind this method is that if $f(x)$ has a unique global maximum at x_0 inside the interval $[a, b]$ then at this point the integrand will be exponentially larger than at any other points within the region of integration. Consequently, the dominant contribution to the integral will come from the values of x near x_0, and the contribution from the rest will be asymptotically negligible as $n \to \infty$. Thus, to evaluate the integral (B.1) approximately, it is crucial to approximate $f(x)$ accurately near x_0.

Assuming that $f(x)$ exhibits a unique global maximum at x_0 we start by replacing the original integral over $[a, b]$ with an integral over a small neighborhood of x_0 as $n \to \infty$,

$$I(n) \approx \int_{x_0-\epsilon}^{x_0+\epsilon} \mathrm{d}x \, \mathrm{e}^{nf(x)}. \tag{B.2}$$

For ϵ sufficiently small we can perform Taylor's expansion of $f(x)$ around x_0,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + O(x^3). \tag{B.3}$$

Since $f(x)$ has a maximum at x_0, the first derivative of f vanishes at x_0 and the above equation can be approximated up to quadratic order by

$$f(x) \approx f(x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2. \tag{B.4}$$

Plugging this into Eq. (B.2) we obtain

$$I(n) \approx \mathrm{e}^{nf(x_0)} \int_{x_0-\epsilon}^{x_0+\epsilon} \mathrm{d}x\, \mathrm{e}^{\frac{1}{2}f''(x_0)(x-x_0)^2}. \tag{B.5}$$

We may extend the limits of integration of the integral in Eq. (B.5) to go from $-\infty$ to $+\infty$ as the exponent in the integrand decays very fast away from the maximum. Recall that $f''(x_0) < 0$, so the resulting integral is a Gaussian integral which can be readily calculated,

$$\int_{-\infty}^{+\infty} \mathrm{d}x\, \mathrm{e}^{\frac{1}{2}f''(x_0)(x-x_0)^2} = \sqrt{\frac{2\pi}{n\,|f''(x_0)|}}. \tag{B.6}$$

Thus finally, to the leading order,

$$I(n) \sim \mathrm{e}^{nf(x_0)} \sqrt{\frac{2\pi}{n\,|f''(x_0)|}} \quad \text{as } n \to \infty. \tag{B.7}$$

The described procedure can be applied to determine the asymptotics of the integral

$$I(n) = \iint_{-\infty}^{+\infty} \mathrm{d}x\mathrm{d}y\, \mathrm{e}^{nf(x,y)}, \tag{B.8}$$

where $f(x, y)$ is now a real function of two arguments. Let (x_0, y_0) be the point at which $f(x, y)$ has a unique global maximum. Then the integral over the whole XY plane can be approximated by an integral over a small region around this point,

$$I(n) \approx \int_{x_0-\epsilon}^{x_0+\epsilon} \int_{y_0-\sigma}^{y_0+\sigma} \mathrm{d}x\mathrm{d}y\, \mathrm{e}^{nf(x,y)}, \tag{B.9}$$

where ϵ and σ are some small positive constants. We proceed by expanding $f(x, y)$ around (x_0, y_0) using Taylor series,

$$f(x,y) = \sum_{k=0}^{\infty} \left\{ \frac{1}{k!} \left[(x-x_0)\frac{\partial}{\partial x'} + (y-y_0)\frac{\partial}{\partial y'} \right]^k f(x',y') \right\}_{\substack{x'=x_0 \\ y'=y_0}}, \tag{B.10}$$

from which the approximation up to quadratic order follows,

$$\begin{aligned} f(x,y) \approx\ & f(x_0,y_0) + \Delta x \Delta y f_{xy}(x_0,y_0) \\ & + \frac{1}{2}\left[(\Delta x)^2 f_{xx}(x_0,y_0) + (\Delta y)^2 f_{yy}(x_0,y_0) \right], \end{aligned} \tag{B.11}$$

where $\Delta x \equiv x - x_0$, $\Delta y \equiv y - y_0$, and we have used the fact that the first derivative vanishes at the maximum. Substituting $f(x, y)$ back into Eq. (B.9) we get

$$I(n) \approx \mathrm{e}^{nf(x_0,y_0)} \iint \mathrm{d}x\mathrm{d}y \mathrm{e}^{n\left\{\Delta x\Delta y f_{xy}(x_0,y_0) + \frac{1}{2}\left[(\Delta x)^2 f_{xx}(x_0,y_0) + (\Delta y)^2 f_{yy}(x_0,y_0)\right]\right\}}. \tag{B.12}$$

Since the significant contribution to the latter integral comes only from a small neighborhood of (x_0, y_0), its limits of integration can be extended and it can be treated as a Gaussian-type integral. Let us introduce the shorthand notation using the Hessian matrix (see Ref. [98]),

$$\mathbf{H}(f) \equiv \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}, \tag{B.13}$$

and denote its determinant at the critical point by

$$H(x_0, y_0) \equiv \left| \begin{matrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{matrix} \right|_{\substack{x=x_0 \\ y=y_0}} = f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0). \tag{B.14}$$

Then, if $H(x_0, y_0) > 0$, the integral in Eq. (B.12) is convergent and equals

$$\frac{2\pi}{n\sqrt{H(x_0, y_0)}}. \tag{B.15}$$

This yields the final result

$$I(n) \sim \mathrm{e}^{nf(x_0, y_0)} \frac{2\pi}{n\sqrt{H(x_0, y_0)}}. \tag{B.16}$$

Appendix C

Next-to-leading corrections to $F(z)$

The next-to-leading corrections to the $F(z)$ function defined in Eq. (4.26) are of order n^{-1}. They are produced by the Hessian term and the non-leading terms in the Stirling approximation of $\ln B_n(z)$ given by Eq. (4.25).

The Hessian matrix of $f(\psi_+, \psi_-)$ given by Eq. (4.15) has the form

$$\mathbf{H} = -\Delta^{-1} - q \begin{pmatrix} z/\psi_+^2 & 0 \\ 0 & (1-z)/\psi_-^2 \end{pmatrix}, \tag{C.1}$$

and its determinant at $(\overline{\psi}_+, \overline{\psi}_-)$ equals

$$\begin{aligned} H(\overline{\psi}_+, \overline{\psi}_-) &= \frac{g}{g^2-1}\left[1 + q\sqrt{g}\left(\frac{z}{\overline{\psi}_+^2} + \frac{1-z}{\overline{\psi}_-^2}\right)\right] + q^2 \frac{z\,(1-z)}{\left(\overline{\psi}_-\overline{\psi}_+\right)^2} \\ &= \frac{g}{g^2-1}\frac{1 + 4(g^2-1)(1-z)z - \sqrt{1+4(g^2-1)(1-z)z}}{(g^2-1)(1-z)z}. \end{aligned} \tag{C.2}$$

Allowing for this correction and the contribution from the $O(1)$ term in the Stirling formula (4.25), the approximation of the $F(z)$ function reads

$$\begin{aligned} F(z) \approx\; & z\ln z + (1-z)\ln(1-z) - q\left[z\ln\overline{\psi}_+ + (1-z)\ln\overline{\psi}_-\right] \\ & + \tfrac{1}{2n}\left\{\ln H(\overline{\psi}_+, \overline{\psi}_-) + \ln\left[z\,(1-z)\right]\right\}. \end{aligned} \tag{C.3}$$

Appendix D

Eigensystem of matrix Q

To begin with, let us rewrite Eq. (4.55) in matrix form

$$\begin{pmatrix}\Phi_+\\ \Phi_-\end{pmatrix} = \frac{1}{(q-1)!}\begin{pmatrix}\sqrt{g}\Phi_+^{q-2} & \frac{1}{\sqrt{g}}\Phi_-^{q-2}\\ \frac{1}{\sqrt{g}}\Phi_+^{q-2} & \sqrt{g}\Phi_-^{q-2}\end{pmatrix}\cdot\begin{pmatrix}\Phi_+\\ \Phi_-\end{pmatrix} \equiv \mathbf{A}\cdot\begin{pmatrix}\Phi_+\\ \Phi_-\end{pmatrix}. \tag{D.1}$$

If now $\mathbf{v}$ is an eigenvector of matrix $\mathbf{A}$ with eigenvalue λ then from the defining relation

$$\lambda\mathbf{v} = \mathbf{A}\mathbf{v} \tag{D.2}$$

it follows that $\mathbf{A}$ has an eigenvalue equal one with corresponding eigenvector proportional to (Φ_+, Φ_-). We are interested in calculating eigenvectors of matrix $\mathbf{Q}$ defined in Eq. (4.62)

$$\mathbf{Q} = \mathbf{M}^{\frac{1}{2}}\Delta\mathbf{M}^{\frac{1}{2}} = \frac{1}{(q-2)!}\begin{pmatrix}\sqrt{g}\Phi_+^{q-2} & \frac{1}{\sqrt{g}}\left(\Phi_+\Phi_-\right)^{\frac{q-2}{2}}\\ \frac{1}{\sqrt{g}}\left(\Phi_+\Phi_-\right)^{\frac{q-2}{2}} & \sqrt{g}\Phi_-^{q-2}\end{pmatrix}. \tag{D.3}$$

This can be done by relating it to matrix A

$$Q_{ij} = (q-1)\,A_{ij}\left(\frac{\Phi_i}{\Phi_j}\right)^{\frac{q-2}{2}} \tag{D.4}$$

and transforming the eigenequation (D.2)

$$\lambda v_i = \sum_j A_{ij}v_j = \sum_j \frac{q-1}{q-1}\left(\frac{\Phi_i}{\Phi_j}\frac{\Phi_j}{\Phi_i}\right)^{\frac{q-2}{2}} A_{ij}v_j = \frac{1}{q-1}\sum_j Q_{ij}\left(\frac{\Phi_j}{\Phi_i}\right)^{\frac{q-2}{2}} v_j \tag{D.5}$$

yielding

$$(q-1)\lambda\Phi_i^{\frac{q-2}{2}} v_i = \sum_j Q_{ij}\Phi_j^{\frac{q-2}{2}} v_j. \tag{D.6}$$

This can be written in form of an eigenequation

$$\widetilde{\lambda}\widetilde{\mathbf{v}} = \mathbf{Q}\widetilde{\mathbf{v}} \tag{D.7}$$

where

$$\widetilde{\lambda} \equiv (q-1)\lambda, \quad \widetilde{v_i} \equiv \Phi_i^{\frac{q-2}{2}} v_i, \tag{D.8}$$

relate the eigenvalues and eigenvectors of the two matrices. The above implies that $\mathbf{Q}$ has an eigenvalue $\lambda_1 = q-1$ with the corresponding normalized eigenvector

$$\mathbf{v}_1 = \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \begin{pmatrix} \Phi_+^{\frac{q}{2}} \\ \Phi_-^{\frac{q}{2}} \end{pmatrix}. \tag{D.9}$$

The second eigenvector $\mathbf{v}_2$ perpendicular to $\mathbf{v}_1$ is

$$\mathbf{v}_2 = \begin{pmatrix} -b \\ a \end{pmatrix} = \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \begin{pmatrix} -\Phi_-^{\frac{q}{2}} \\ \Phi_+^{\frac{q}{2}} \end{pmatrix}. \tag{D.10}$$

By multiplying it by $\mathbf{Q}$ and applying Eq. (4.55) to the third line below we can easily obtain the second eigenvalue

$$\mathbf{Q} \cdot \mathbf{v}_2 = \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \frac{1}{(q-2)!} \begin{pmatrix} \sqrt{g}\Phi_+^{q-2} & \frac{1}{\sqrt{g}} \left(\Phi_+ \Phi_-\right)^{\frac{q-2}{2}} \\ \frac{1}{\sqrt{g}} \left(\Phi_+ \Phi_-\right)^{\frac{q-2}{2}} & \sqrt{g}\Phi_-^{q-2} \end{pmatrix} \begin{pmatrix} -\Phi_-^{\frac{q}{2}} \\ \Phi_+^{\frac{q}{2}} \end{pmatrix}$$

$$= \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \frac{1}{(q-2)!} \begin{pmatrix} -\Phi_-^{\frac{q}{2}} \left(\sqrt{g}\Phi_+^{q-2} - \frac{1}{\sqrt{g}} \Phi_+^{q-1}/\Phi_- \right) \\ \Phi_+^{\frac{q}{2}} \left(\sqrt{g}\Phi_-^{q-2} - \frac{1}{\sqrt{g}} \Phi_-^{q-1}/\Phi_+ \right) \end{pmatrix}$$

$$= \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \frac{q-1}{(q-1)!} \begin{pmatrix} -\Phi_-^{\frac{q}{2}} \left(\sqrt{g}\Phi_+^{q-2} + \sqrt{g}\Phi_-^{q-2} - \sqrt{g}\Phi_-^{q-2} - \frac{1}{\sqrt{g}} \frac{\Phi_+^{q-1}}{\Phi_-} \right) \\ \Phi_+^{\frac{q}{2}} \left(\sqrt{g}\Phi_-^{q-2} + \sqrt{g}\Phi_+^{q-2} - \sqrt{g}\Phi_+^{q-2} - \frac{1}{\sqrt{g}} \frac{\Phi_-^{q-1}}{\Phi_+} \right) \end{pmatrix}$$

$$= \frac{q-1}{\sqrt{\Phi_+^q + \Phi_-^q}} \begin{pmatrix} -\Phi_-^{\frac{q}{2}} \left[\frac{\sqrt{g}}{(q-1)!} \left(\Phi_+^{q-2} + \Phi_-^{q-2} \right) - 1 \right] \\ \Phi_+^{\frac{q}{2}} \left[\frac{\sqrt{g}}{(q-1)!} \left(\Phi_-^{q-2} + \Phi_+^{q-2} \right) - 1 \right] \end{pmatrix}$$

$$= (q-1) \left[\frac{\sqrt{g}}{(q-1)!} \left(\Phi_+^{q-2} + \Phi_-^{q-2} \right) - 1 \right] \frac{1}{\sqrt{\Phi_+^q + \Phi_-^q}} \begin{pmatrix} -\Phi_-^{\frac{q}{2}} \\ \Phi_+^{\frac{q}{2}} \end{pmatrix}, \tag{D.11}$$

so finally

$$\lambda_2 = (q-1) \left[\frac{\sqrt{g}}{(q-1)!} (\Phi_+^{q-2} + \Phi_-^{q-2}) - 1 \right]. \tag{D.12}$$

List of Author's Publications

[1] Bialas, P. and Oleś, A. K. (2008) Correlations in connected random graphs. *Phys. Rev. E*, **77**, 036124.

[2] Bialas, P. and Oleś, A. K. (2010) Long-range disassortative correlations in generic random trees. *Phys. Rev. E*, **81**, 041136.

[3] Bialas, P. and Oleś, A. K. (2011) Correlation functions of Ising spins on thin graphs. *eprint arXiv:1104.4030*.

Bibliography

[4] Albert, R. and Barabási, A.-L. (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.*, **74**, 47–97.

[5] Dorogovtsev, S. N. and Mendes, J. F. F. (2002) Evolution of networks. *Adv. Phys.*, **51**, 1079–1187.

[6] Newman, M. E. J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.

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

[8] Newman, M. E. J. (2010) *Networks: An Introduction*. Oxford University Press.

[9] Bornholdt, S. and Schuster, H. (2006) *Handbook of Graphs and Networks: From the Genome to the Internet*. John Wiley & Sons.

[10] Caldarelli, G. (2007) *Scale-Free Networks: Complex Webs in Nature and Technology*. Oxford Finance, Oxford University Press.

[11] Newman, M. E. J. (2003) Mixing patterns in networks. *Phys. Rev. E*, **67**, 026126.

[12] Maslov, S., Sneppen, K., and Zaliznyak, A. (2004) Detection of topological patterns in complex networks: Correlation profile of the Internet. *Physica A*, **333**, 529–540.

[13] Newman, M. E. J. and Park, J. (2003) Why social networks are different from other types of networks. *Phys. Rev. E*, **68**, 036122.

[14] Erdős, P. and Rényi, A. (1959) On random graphs, I. *Publ. Math.-Debrecen*, **6**, 290–297.

[15] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, **5**, 17–61.

[16] Erdős, P. and Rényi, A. (1961) On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, **38**, 343–347.

[17] Milgram, S. (1967) The small world problem. *Psychol. Today*, **2**, 60–67.

[18] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F. (2008) Critical phenomena in complex networks. *Rev. Mod. Phys.*, **80**, 1275–1335.

[19] Burda, Z., Correia, J. D., and Krzywicki, A. (2001) Statistical ensemble of scale-free random graphs. *Phys. Rev. E*, **64**, 046118.

[20] Bauer, M. and Bernard, D. (2002) Maximal entropy random networks with given degree distribution. *eprint arXiv:cond-mat/0206150*.

[21] Dorogovtsev, S., Mendes, J., and Samukhin, A. (2003) Principles of statistical mechanics of uncorrelated random networks. *Nucl. Phys. B*, **666**, 396–416.

[22] Burda, Z., Jurkiewicz, J., and Krzywicki, A. (2004) Statistical mechanics of random graphs. *Physica A*, **344**, 56–61.

[23] Park, J. and Newman, M. E. J. (2004) Statistical mechanics of networks. *Phys. Rev. E*, **70**, 066117.

[24] Bogacz, L., Burda, Z., and Wacław, B. (2006) Homogeneous complex networks. *Physica A*, **366**, 587–607.

[25] Bondy, J. and Murty, U. (1976) *Graph Theory with Applications*. Elsevier Science Publishing Co.

[26] Wilson, R. (1996) *Introduction to Graph Theory*. Pearson Education.

[27] Plischke, M. and Bergersen, B. (2006) *Equilibrium Statistical Physics*. World Scientific, 3rd edn.

[28] Bollobás, B. (2001) *Random Graphs*. Cambridge studies in advanced mathematics, Cambridge University Press, 2nd edn.

[29] Janson, S., Łuczak, T., and Ruciński, A. (2000) *Random Graphs*. Wiley-Interscience series in discrete mathematics and optimization, John Wiley.

[30] Drmota, M. (2009) *Random Trees: An Interplay between Combinatorics and Probability*. Springer.

[31] Baxter, R. (1982) *Exactly Solved Models in Statistical Mechanics*. Academic Press.

[32] Solomonoff, R. and Rapoport, A. (1951) Connectivity of random nets. *B. Math. Biol.*, **13**, 107–117.

[33] Gilbert, E. N. (1959) Random graphs. *Ann. Math. Stat.*, **30**, 1141–1144.

[34] Oleś, A. K. (2006) *Correlations in random networks*. Master's thesis, Jagiellonian University, Cracow, Poland (in Polish).

[35] Burda, Z. and Krzywicki, A. (2003) Uncorrelated random networks. *Phys. Rev. E*, **67**, 046118.

[36] Binney, J. J., Dowrick, N. J., Fisher, A. J., and Newman, M. E. J. (1992) *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford science publications, Clarendon Press.

[37] Bialas, P. (1997) Correlations in fluctuating geometries. *Nucl. Phys. B*, **53**, 739–742.

[38] de Bakker, B. V. and Smit, J. (1995) Two-point functions in 4D dynamical triangulation. *Nucl. Phys. B*, **454**, 343–354.

[39] Bialas, P., Burda, Z., Krzywicki, A., and Petersson, B. (1996) Focusing on the fixed point of 4D simplicial gravity. *Nucl. Phys. B*, **472**, 293–308.

[40] Pastor-Satorras, R., Vázquez, A., and Vespignani, A. (2001) Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.*, **87**, 258701.

[41] Vázquez, A., Pastor-Satorras, R., and Vespignani, A. (2002) Large-scale topological and dynamical properties of the Internet. *Phys. Rev. E*, **65**, 066130.

[42] Boguñá, M. and Pastor-Satorras, R. (2003) Class of correlated random networks with hidden variables. *Phys. Rev. E*, **68**, 036112.

[43] Serrano, M. A., Maguitman, A., Boguñá, M., Fortunato, S., and Vespignani, A. (2007) Decoding the structure of the WWW: A comparative analysis of Web crawls. *ACM Trans. Web*, **1**.

[44] Bialas, P. (1996) Long range correlations in branched polymers. *Phys. Lett. B*, **373**, 289–295.

[45] Newman, M. E. J. (2002) Assortative mixing in networks. *Phys. Rev. Lett.*, **89**, 208701.

[46] Jurkiewicz, J. and Krzywicki, A. (1997) Branched polymers with loops. *Phys. Lett. B*, **392**, 291–297.

[47] Bachas, C., de Calan, C., and Petropoulos, P. M. S. (1994) Quenched random graphs. *J. Phys. A*, **27**, 6121.

[48] Baillie, C. F., Johnston, D. A., and Kownacki, J. P. (1994) Ising spins on thin graphs. *Nucl. Phys. B*, **432**, 551–570.

[49] Baillie, C. F., Janke, W., Johnston, D. A., and Plecháč, P. (1995) Spin glasses on thin graphs. *Nucl. Phys. B*, **450**, 730–752.

[50] Baillie, C. and Johnston, D. (1996) Spin models on thin graphs. *Nucl. Phys. B.*, **47**, 649–652.

[51] Bessis, D., Itzykson, C., and Zuber, J. (1980) Quantum field theory techniques in graphical enumeration. *Adv. Appl. Math.*, **1**, 109–157.

[52] Srednicki, M. (2007) *Quantum Field Theory*. Cambridge University Press.

[53] Ryder, L. (1996) *Quantum Field Theory*. Cambridge University Press.

[54] Newman, M. E. J. and Barkema, G. T. (1999) *Monte Carlo Methods in Statistical Physics*. Clarendon Press.

[55] Metropolis, N. and Ulam, S. (1949) The Monte Carlo method. *J. Am. Stat. Assoc.*, **44**, 335–341.

[56] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.

[57] Priestley, M. B. (1981) *Spectral Analysis and Time Series*, vol. 1 of *Probability and mathematical statistics*. Academic Press.

[58] Madras, N. and Sokal, A. D. (1988) The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *J. Stat. Phys.*, **50**, 109–186.

[59] Sokal, A. (1997) Monte Carlo methods in statistical mechanics: Foundations and New Algorithms. DeWitt-Morette, C., Cartier, P., and Folacci (eds.), *Functional integration (Cargèse, 1996)*, vol. 361 of *NATO Adv. Sci. Inst. Ser. B Phys.*, pp. 131–192, Plenum.

[60] Salas, J. and Sokal, A. (1997) Dynamic critical behavior of the Swendsen-Wang algorithm: The two-dimensional three-state Potts model revisited. *J. Stat. Phys.*, **87**, 1–36.

[61] Berg, B. A. (2004) *Markov Chain Monte Carlo Simulations and Their Statistical Analysis: With Web-Based Fortran Code*. World Scientific.

[62] Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Ann. Stat.*, **7**, 1–26.

[63] Efron, B. (1981) Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589–599.

[64] Efron, B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans*, vol. 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM.

[65] Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, vol. 57 of *Monographs on statistics and applied probability*. Chapman & Hall.

[66] Gentle, J., Härdle, W., and Mori, Y. (2004) *Handbook of Computational Statistics: Concepts and Methods*. Springer.

[67] Molloy, M. and Reed, B. (1995) A critical point for random graphs with a given degree sequence. *Random Struct. Algor.*, **6**, 161–180.

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

[69] Fronczak, A., Fronczak, P., and Hołyst, J. A. (2005) How to calculate the main characteristics of random uncorrelated networks. *AIP Conf. Proc.*, **776**, 52–68.

[70] Albert, R., Jeong, H., and Barabási, A.-L. (1999) Internet: Diameter of the World-Wide Web. *Nature*, **401**, 130–131.

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

[72] Catanzaro, M., Boguñá, M., and Pastor-Satorras, R. (2005) Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, **71**, 027103.

[73] Bogacz, L., Burda, Z., Janke, W., and Waclaw, B. (2005) A program generating homogeneous random graphs with given weights. *Comput. Phys. Commun.*, **173**, 162–174.

[74] Boguñá, M., Pastor-Satorras, R., and Vespignani, A. (2004) Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B*, **38**, 205–209.

[75] Dorogovtsev, S. N., Mendes, J. F. F., and Samukhin, A. N. (2001) Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E*, **63**, 062101.

[76] Ambjørn, J., Durhuus, B., Fröhlich, J., and Orland, P. (1986) The appearance of critical dimensions in regulated string theories. *Nucl. Phys. B*, **270**, 457–482.

[77] Bialas, P. and Burda, Z. (1996) Phase transition in fluctuating branched geometry. *Phys. Lett. B*, **384**, 75–80.

[78] Bialas, P., Burda, Z., and Johnston, D. (1997) Condensation in the Backgammon model. *Nucl. Phys. B*, **493**, 505–516.

[79] Bialas, P. (2000) Matter correlations in branched polymers. *Nucl. Phys. B*, **575**, 645–660.

[80] Burda, Z., Erdmann, J., Petersson, B., and Wattenberg, M. (2003) Exotic trees. *Phys. Rev. E*, **67**, 026105.

[81] Ambjørn, J., Białas, P., Jurkiewicz, J., Burda, Z., and Petersson, B. (1994) Effective sampling of random surfaces by baby universe surgery. *Phys. Lett. B*, **325**, 337–346.

[82] Kurata, M., Kikuchi, R., and Watari, T. (1953) A theory of cooperative phenomena. III. Detailed discussions of the cluster variation method. *J. Chem. Phys.*, **21**, 434–448.

[83] Katsura, S. and Takizawa, M. (1974) Bethe lattice and the Bethe approximation. *Prog. Theor. Phys.*, **51**, 82–98.

[84] Bethe, H. A. (1935) Statistical theory of superlattices. *P. Roy. Soc. Lond. A Mat.*, **150**, 552–575.

[85] Matsuda, H. (1974) Infinite susceptibility without spontaneous magnetization. *Prog. Theor. Phys.*, **51**, 1053–1063.

[86] von Heimburg, J. and Thomas, H. (1974) Phase transition of the Cayley tree with Ising interaction. *J. Phys. C Solid State*, **7**, 3433.

[87] Stošić, T., Stošić, B. D., and Fittipaldi, I. P. (1998) Exact zero-field susceptibility of the Ising model on a Cayley tree. *J. Magn. Magn. Mater.*, **177-181, Part 1**, 185–187.

[88] Morita, T. and Horiguchi, T. (1975) Susceptibility and correlation function of the Ising model on the Cayley tree. *Prog. Theor. Phys.*, **54**, 982–998.

[89] Falk, H. (1975) Ising spin system on a Cayley tree: Correlation decomposition and phase transition. *Phys. Rev. B*, **12**, 5184–5189.

[90] Ising, E. (1925) Beitrag zur Theorie des Ferromagnetismus. *Z. Phys. A-Hadron Nucl.*, **31**, 253–258.

[91] Onsager, L. (1944) Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, **65**, 117–149.

[92] Milo, R., Kashtan, N., Itzkovitz, S., Newman, M. E. J., and Alon, U. (2004) On the uniform generation of random graphs with prescribed degree sequences. *eprint arXiv:cond-mat/0312028v2*.

[93] Johnston, D. A. and Plechac, P. (1997) Why loops don't matter. *eprint arXiv:cond-mat/9705160*.

[94] Bedeaux, D., Shuler, K. E., and Oppenheim, I. (1970) Decay of correlations. III. Relaxation of spin correlations and distribution functions in the one-dimensional Ising lattice. *J. Stat. Phys.*, **2**, 1–19.

[95] Hahn, T. (2001) Generating Feynman diagrams and amplitudes with FeynArts 3. *Comput. Phys. Commun.*, **140**, 418–431.

[96] Bialas, P., Burda, Z., and Jurkiewicz, J. (1998) Correlation functions and critical behaviour on fluctuating geometries. *Phys. Lett. B*, **421**, 86–92.

[97] Bialas, P., Burda, Z., Petersson, B., and Tabaczek, J. (1997) Appearance of mother universe and singular vertices in random geometries. *Nucl. Phys. B*, **495**, 463–476.

[98] Weisstein, E. (1999) *CRC Concise Encyclopedia of Mathematics*. CRC Press.

[99] Arfken, G. and Weber, H. (2005) *Mathematical Methods for Physicists*. Elsevier, 6th edn.