SAPIENZA
UNIVERSITÀ DI ROMA

Study of Matching on Bethe Lattice

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea Magistrale in Fisica

Candidate

Gianmarco Perrupato
ID number 1615014

Thesis Advisor

Prof. Giorgio Parisi

Co-Advisor

Dr. Gabriele Sicuro

Academic Year 2017/2018

Study of Matching on Bethe Lattice
Master thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: perrupato.1615014@studenti.uniroma1.it

Abstract

In the present work we discuss some connections between Statistical physics and optimization. In particular, we analyze some matching problems on regular weighted graphs. We are interested in the random instances of the problems, where the weights are independently and identically distributed according to a given law, and the graphs are drawn from the ensemble of random regular graphs. The property of such ensemble of being locally-tree-like in the thermodynamic limit suggests to use the cavity method, whose predictions are compared with numerical simulations. We study the asymptotic costs and the finite size corrections of the standard matching. For valences $z = 3, 4$ the cavity estimates are in good-agreement with the numerical simulations. In order to study the finite size corrections with the cavity method we conjecture that such corrections receive a contribution due to the cost of the cycles of the graph. For $z = 3, 4$ the cavity estimates of such costs are compatible with the numerical simulation, performed with a Markov-Chain algorithm. Some observation corroborate the presence of cycle-dependent terms in the finite size corrections. To better investigate the role played by cycles we carry out some numerical simulations about the fractional and the loopy fractional matching. These seems to suggest that some properties of the fully connected topology still hold at finite connectivity.

Contents

Introduction

Suppose for a moment to be the governor of N villages in the desert, and to be responsible for building a water system: each village must be uniquely associated with one of the N oases of the region. *How can you perform this matching in such a way that the total transport cost is minimum?* The problem is simple in its formulation, but quite difficult to solve, indeed there are $N!$ ways to match villages and oases, and the problem is to select the cheapest one among them.

The *Matching problem* is a classical combinatorial optimization problem, of which there are several different versions. The one of the villages in the desert is the so-called *assignment* problem.

All the versions discussed in this thesis can be formulated in terms of *graph theory*, and belong to the so-called P computational complexity class, i.e., roughly speaking, can be solved in a time that scales as a polynomial in the number of nodes of the graph. We study such problems in the presence of two sources of quenched randomness: a link disorder, and a topological disorder. The first one means that the weights associated with the edges of the graphs, i.e. the lengths of the roads connecting villages to oases in the assignment case, are independent and identically distributed random variables. The second one means that the graphs are drawn according to the uniform distribution over the set of all regular graphs with a given degree.

In the 80s it has been recognized that several optimization problem, including the matching problem, can be turned into the statistical mechanics language by simply calling the cost to be minimized energy, and the optimum solution the ground state of the system. At the light of this analogy, we compare some predictions obtained with the replica symmetric (RS) cavity method at zero temperature with numerical simulations. The first quantity analyzed is the ground state energy in the thermodynamic limit. The accordance between the predictions and the simulations corroborates the correctness of the RS assumption. In a second moment the study focuses on the $1/N$-correction to the asymptotic cost at finite sizes of the system. In this context the fundamental idea underlying the cavity approach is to conjecture that this correction receives topological contributions due to the presence of cycles in the graph, an then to try to reduce the calculation of the finite size corrections to that of the average cost of the cycles of a given size. An algorithm, based on the generation of Markov chains in the space of random regular graphs, is presented: it allows to measure the cost of the cycles of a given length on a generic ensemble of random graphs.

The work is divided in five chapters. In chapter 1 some definitions and concepts

about graph theory, and optimization on graphs are presented. It is introduced the idea of locally tree likeness, that is an important property of the graph ensemble on which we want to focus. In chapter 2 some ideas about statistical mechanics, with particular regard to the mean field approximations, are presented. In chapter 3 the replica and the cavity method are introduced, and it is discussed the first example of application of the cavity method. Chapter 4 deals with the relation between statistical mechanics and the matching problems, focusing on some classical results that turns to be useful in chapter 5. The last chapter is about matching on random regular graphs, and addresses the studies summarized at the beginning of the introduction.

Chapter 1

Graphs and Optimization

1.1 Key definitions

1.1.1 What is a graph?

A *graph* G is an ordered pair $G = (V; E)$ comprising a set V of vertexes, or nodes, together with a set $E \subset V \times V$ of edges. A *subgraph* of a graph G is a graph $G' = (V'; E')$ such that $V' \subseteq V$ and $E' \subseteq E$. Each edge has either one or two vertexes associated with it, called its endpoints. An edge is said to connect its endpoints.

A graph is bipartite if V can be partitioned in two sets W and Z, called bipartite *parts* of the graph, that satisfy:

$$V = W \cup Z, \tag{1.1}$$

and such that there is no edge joining vertexes in the same set. To denote a bipartite graph one often writes $G = (W, Z; E)$.

For a generic graph $G = (V; E)$ we will denote by $i, j \dots$ the elements of V, and by (i, j) an edge whose endpoints are i and j. Edges of the form (i, i) are called *self-loops*, or simply loops, while edges that connect the same pair of vertexes are called *multi-edges*. A graph without self-loops and multi-edges is said *simple*. In what follows when we write graph we will always mean simple graph, unless otherwise specified, and we will refer to the general case as *multigraph*.

It is useful to define the *neighborhood* of a vertex i, denoted by ∂i, as the set of vertexes that have an edge in common with i, $\partial i \equiv \{j \in V : (i, j) \in E\}$. Every node i is characterized by the number of its neighbors, called *degree* or *valence*, that is denoted by $|\partial i|$ or k_i; a vertex with degree zero is called *isolated*, and a vertex with degree one is called *leaf*. When the degree of every vertex is finite, we say that the graph is locally finite.

We say that a graph is *weighted* if equipped with function $w : E \to \mathbb{R}$, that associates with each edge $e \in E$ a real number, that we call *weight*.

Given a graph $G = (V; E)$, the graph density of G is defined as:

$$D_G = \frac{2|E|}{|V|\,(|V| - 1)}. \tag{1.2}$$

If $D_G \approx 1$, G is considered *dense*; otherwise if $D_G \ll 1$, G is said *sparse*.

Cycles and matrix representation

Every graph G with N nodes can be represented by an $N \times N$ symmetric matrix $\boldsymbol{M}$, called *adjacency matrix*, whose elements M_{ij} are defined as follows:

$$M_{ij} = \begin{cases} 1 & \text{if there is an edge joining } i \text{ to } j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

Every feature of G, e.g. the number of cycles, is encoded in its adjacency matrix. A *walk* of length ℓ on a graph G is a sequence of vertexes $w = (v_0, v_1, ..., v_\ell)$ with $(v_m, v_{m+i}) \in E$. A walk is said to be *simple* if it is not self-intersecting. A *cycle* of length ℓ is a simple walk of length ℓ that is *closed*, i.e. $v_0 = v_\ell$. Usually a walk in which all vertexes and all edges are distinct is called *path*, then a cycle can be defined equivalently as a closed path. Since cycles play an important role in the finite size corrections to disorder systems defined on graphs, as we will be show in chapter 3, it is worth discussing in details how their number in a given graph is related with the adjacency matrix. The following proposition holds.

Proposition 1.1. *Let $\boldsymbol{M}$ be the adjacency matrix of a graph $G = (V; E)$, $|V| = N$, and let ℓ be a positive integer. Then the entry in the (i, j)-position of the matrix $\boldsymbol{M}^\ell$ is the number of walks of length ℓ from the vertex i to the vertex j in G. It follows that*

$$\text{Number of closed walks of length } \ell = \sum_{i=1}^{N} \left(\boldsymbol{M}^\ell\right)_{ii} \equiv \operatorname{Tr} \boldsymbol{M}^\ell. \tag{1.4}$$

A proof of Proposition 1.1 can be given by induction on the length ℓ [BBB93]. By imposing the walks to be non self-intersecting in Equation 1.4, it follows that the number of cycles of length ℓ of a graph with adjacency matrix $\boldsymbol{M}$ is:

$$\text{Number of cycles of length } \ell = \sum_{i=1}^{N} \sum_{j=1}^{N} \cdots \sum_{n=1}^{N} \mathcal{K}(i, j, ..., n) \underbrace{M_{ij} M_{jk} \cdots M_{ni}}_{\ell \text{ factors}}, \tag{1.5}$$

where:

$$\mathcal{K}(i, j, ..., n) = \prod_{\substack{p,q \in \\ \{i,...,n\}}} (1 - \delta_{pq}) \,. \tag{1.6}$$

In Equation 1.6 the product is over all the pairs p, q belonging to the set of indices $\{i, j, ..., n\}$, and δ_{pq} is the Kronecker delta. Two cycles of length ℓ are equivalent if the respective vertex sequences $(v_1, ..., v_\ell)$, and $(v'_1, ..., v'_\ell)$ can be turned one into the other by a transformation:

$$v_i \mapsto v'_{\ell+1-i}, \tag{1.7}$$

or by a cyclic permutation of the v_i's. Then if one is interested in the number of non equivalent cycles of length ℓ, has to divide Equation 1.5 by 2ℓ, to take into account that every cycle can be walked clockwise or counterclockwise, and that each of the ℓ nodes of the cycle can be equivalently chosen as the starting one.

Other important concepts whose definition is based on that of walk are *distance* and *connection*. The distance between two nodes can be easily defined as the length of the shortest walk joining them. A graph is said to be connected if for any two

nodes is defined a distance. A special case is that of *fully connected* or *complete* graph, where each node is equidistant from the others; usually a complete graph with N nodes is denoted by K_N.

A graph without cycles is said a *forest*; a connected forest is called *tree*.

1.1.2 Graph Ensembles

Graphs are useful tools for modeling a broad spectrum of problems arising, for example, in the study of complex physical systems, combinatorial optimization, social networks, the World Wide Web, and Internet [Mej10, Cal07]. One common feature to most of such complex networks is that they are large, and then they could be utterly impossible to handle [vdH16]. Nevertheless, the specific graph under investigation is often a realization of an ample class of possible graphs: the properties of the graph are therefore less informative of the typical ones of such a class of possible graphs, to which the specific graphs belongs. For this reasons, we introduce the concept of *random graph ensemble*.

A random graph ensemble $\mathbb{G} = (\mathcal{G}, \mathbb{P})$ is a set $\mathcal{G}$ of graphs together with a probability law $\mathbb{P}$ defined over it [BFK+01]. In what follows, when we write that a graph G belongs to an ensemble $\mathbb{G} = (\mathcal{G}, \mathbb{P})$ it is understood, with an abuse of language, that G is the outcome of an extraction over $\mathcal{G}$ with a probability law $\mathbb{P}$. Given a graph ensemble, one can compute the average value of a generic observable $A[G]$:

$$\mathbb{E}[A] \equiv \sum_{G \in \mathcal{G}} \mathbb{P}[G] A[G]. \tag{1.8}$$

We shall now introduce some random graph ensembles commonly used in statistical physics as well as in other disciplines [New18].

The *Erdős-Rényi ensemble*, $\mathbb{G}_{\mathrm{ER}}(N, p)$, which arises by taking N vertices, and placing an edge between any distinct pair of them with some fixed probability p, is the first random graph ensemble that has been studied. Denoting by M the number of edges of a graph G, the probability of extracting G from $\mathbb{G}_{\mathrm{ER}}(N, p)$, is given by:

$$\mathbb{P}[G] = p^M (1-p)^{\binom{N}{2} - M}. \tag{1.9}$$

Furthermore the degree of a node i is a binomial random variable that can be written as:

$$\mathbb{P}[|\partial i| = k] = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \tag{1.10}$$

because $N-1$ nodes can be connected to i. Since the average number of neighbors scales as Np, in order to maintain a finite connectivity in the large N limit, p have to scale as $\frac{1}{N}$. By setting $Np = c$, with c positive real number, we have that:

$$\lim_{N \to \infty} \mathbb{P}[|\partial i| = k] = \frac{c^k}{k!} e^{-ck}, \tag{1.11}$$

and then $|\partial i|$ becomes a Poisson random variable.

Another example is the *ensemble of random regular graphs* $\mathbb{G}_{\mathrm{RRG}}(N, z)$. The measure of this ensemble is uniform over all the graphs with N vertexes with the same number z of neighbors. It is important to note that in order $\mathbb{G}_{\mathrm{RRG}}(N, z)$ to be

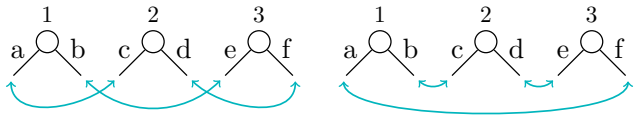
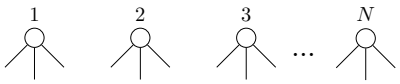


Figure 1.1. The stubs are lettered to identify them. The matching on the left is different from that on the right.

different from the empty set, Nz must be an even number, indeed for every graph it is immediate to see that:

$$\sum_{i=1}^{N} k_i = 2|E|. \tag{1.12}$$

An uniform sampling of $\mathbb{G}_{\text{RRG}}(N, z)$ can be obtained starting from the so-called *configurational model* [Wor99] as follows. Let us consider N vertexes, and suppose to draw z lines, usually called "stubs" or half-edges, emanating from each, as it is shown below for $z = 3$:



Let us call W the set of all the stubs. At this point, we pair the half-edges at random, i.e., we choose uniformly at random what is called a *matching* of the elements of W. The result of this procedure is a *configuration*. Each configuration corresponds to a multigraph belonging to a new ensemble $\mathbb{G}^*_{\text{RRG}}(N, z) = (\mathcal{G}^*_{\text{RRG}}, \mathbb{P}^*_{\text{RRG}})$, that is defined on the set of all regular multigraphs with N nodes with degree z: $\mathbb{G}^*_{\text{RRG}}$ contains elements that are not simple, i.e., that may have self-loops and multi-edges. It is important to note that while all matchings of stubs appear with equal probability in the model, in general it is not guaranteed that all multigraph topologies appear with equal probability, because more than one matching can correspond to the same topology: for example, in Figure 1.1 both the left and the right pairings, that are different matchings of W, produce a triangle.

If a configuration does not contain self-loops and multi-edges, i.e. if it constitutes a simple graph, one can generate all the matchings that correspond to it by taking anyone matching for that graph and permuting the half-edges at each vertex in every possible way. Then the number of matchings $|M(G)|$ corresponding to each simple graph G is $(z!)^N$. It can be easily proved that the configurational model provides a non uniform sampling of the set of all regular multigraphs [New18]. However if G is a graph, as seen above, $|M(G)|$ does not depend upon G, and then each graph can occur with the same probability. Therefore repeating the pairing procedure until the resulting graph has no self-loops and multi-edges gives a uniform sample of the random regular graph ensemble $\mathbb{G}_{\text{RRG}}$. It turns out that the probability to extract a regular graph G from $\mathbb{G}^*_{\text{RRG}}(N, z)$ has the form [Ell11]:

$$\mathbb{P}[G \in \mathbb{G}_{\text{RRG}}(N, z)] \to e^{-\frac{(z^2-1)}{4}} \text{ as } N \to \infty. \tag{1.13}$$

Then this sampling procedure of $\mathbb{G}_{\text{RRG}}$ quickly slows down with increasing z. In the next section a faster method is discussed for higher values of z, that allows us to construct a uniform sampling of the ensemble of regular graphs.

The configurational model can be easily generalized to the case of a generic degree sequence $\{k_i\}_{i=1}^N$, where k_i is the degree of the i-th node and, as before, $\sum_{i=1}^N k_i$ is even. In this more general context a new ensemble, $\mathbb{G}^*_{\text{Conf}}(N, \{k_i\}_{i=1}^N)$, is defined over the set $\mathcal{G}^*_{\text{Conf}}$ of all multigraphs with the given degree sequence through the pairing procedure discussed in the regular case. Moreover by a straightforward generalization of the method discussed to sample regular graphs, also in this case it is possible to construct a uniform sampling of the set of simple graphs with degrees $\{k_i\}_{i=1}^N$.

A useful property that characterizes graphs with a given degree sequence, and underlies some numerical simulations discussed in chapter 5, is given by the *Havel-Hakimi* theorem. Firstly it is useful to define *graphical sequences*: a sequence $\{k_i\}_{i=1}^N$ of non-negative integers is called graphical if it is a degree sequence of some graph.

Theorem 1.1 (Havel-Hakimi)**.** *Let K be the sequence $k_1, k_2, ..., k_N$, with $N \geq 2$ and $k_1 \geq k_2 \geq ... \geq k_N$. Let K' be the sequence obtained from K by discarding k_1, and subtracting 1 from each of the next k_1 entries of K, i.e. K' is:*

$$k_2 - 1, k_3 - 1, ..., k_{k_1+1} - 1, k_{k_1+2}, ..., k_N \tag{1.14}$$

then K is graphical if and only if K' is graphical.

Thanks to Theorem 1.1 one can easily construct a graph from a given sequence of integers, if it is graphic, following the steps of Algorithm 1.

Algorithm 1 Havel Hakimi (degree sequence $\{k_1, ..., k_N\}$)

```
S ← {k_i}_{i=1}^N;
while ∃ k_{s_i} ∈ S, k_{s_i} ≠ 0 do
    reorder S into non increasing order;
    if k_{s_1} > |S| − 1 then return the sequence is not graphic;
    end if
    for all i, 2 ≤ i ≤ k_{s_1} + 1 do
        k_{s_i} ← k_{s_i} − 1;
        connect the node s_1 to the node s_i;
    end for
    remove k_{s_1} from S;
end while
```

Fast generation of random graphs

From Equation 1.13 it is clear that, if one is interested in generating uniformly a large number of regular graphs with high degree, the procedure discussed in the previous section may be inefficient. In this section we briefly discuss an heuristic algorithm that, at least empirically, has turned out to be the most efficient one [MZ03,MKI+03].

More specifically the problem we want to address is the following: given a degree sequence that is graphical, we want to generate uniformly at random a simple connected graph having exactly this degree sequence; we are specifically interested

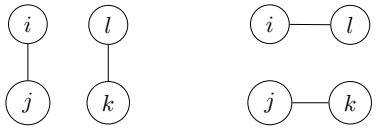


Figure 1.2. Example of edge swapping: the edges in the graph on the left are exchanged in such a way to create the graph on the right.

in the case of connected graphs. The generation process proposed by the *Markov chain Monte-Carlo* algorithm is composed of two main steps:

1. **Realize the given sequence**: generate a single simple connected graph that matches the degree sequence,
2. **Shuffle the edges** to make it random, while keeping it connected and simple.

The first step can be achieved in several different ways. The simplest one consists in generating G_0 by using the configurational model, discussed in the previous section. In this case one has to repeat the pairing procedure until a connected graph with the prescribed degree sequence is found. A different approach is to generate the first graph by using the Havel-Hakimi theorem (Algorithm 1): in linear time Havel-Hakimi generates a graph G' with the given degree sequence. If G' is connected then we can set $G_0 = G'$, otherwise one can, for example, make random edge swappings on G' until the resulting graph is not connected.

Let us now discuss the second step. Consider the following Markov Chain. Suppose to have a connected graph G_t that realizes the given degree sequence. The graph G_{t+1} is determined by a *swapping* procedure: we pick two edges at random, and then we swap them as shown in Figure 1.2, obtaining another graph G' with the same degrees. If G' is simple and connected, we consider the swap as valid: $G_{t+1} = G'$. Otherwise, we reject the swap: $G_{t+1} = G_t$. It is worth noting that this procedure preserves the degree of each node and therefore the degree sequence.

The set of states of this Markov chain is the space S of all simple connected graphs with the given degree sequence, and the initial state G_0 is the graph obtained by the first step. It is immediate to see that the transition $G_t \to G_{t+1}$ has probability $\frac{1}{|E|(|E|-1)}$, with $|E|$ number of edges, if there exists an edge swap that transforms G_t in G_{t+1}. If there are no such swap, the transition has probability 0.

Theorem 1.2. *This Markov chain is irreducible* [Tay81]*, symmetric and aperiodic* [MZ03].

A consequence of Theorem 1.2 is that the Markov chain converges to the uniform distribution on S, i.e., the set of all graphs having the desired properties. It is possible to define on S a notion of distance between two graph as the number of swappings required to turn one graph into the other. In [Wil99] it has been proved that the diameter of S, i.e. the distance between the two farthest graphs, is equal to $|E|$. An empirical result is that this Markov chain converges after $\mathcal{O}(|E|)$ swaps [VL05], however no precise formal results are known up to now. It is important to note that in order to do $\mathcal{O}(|E|)$ swaps one may have to process much more transitions. However it can be proved that the probability that a random edge swap is valid is

at least $\frac{\rho}{2z(z+1)}$, where z is the average degree, and ρ is the fraction of all possible pairs of vertices which have distance greater than or equal to 3. As ρ tends to grow with the size of the graph in most cases of interest, one may expect that the number of transitions requested to converge is also $\mathcal{O}(|E|)$.

Local weak convergence

One of the most important and useful properties of some sparse random graph ensembles for analytical investigations is the locally tree-likeness. In these ensembles, in a sense that is hinted in this section, almost any finite neighborhood of a node looks like a random tree in the large graph limit. Firstly it's necessary to introduce some definitions.

A *rooted graph* is a graph together with the specification of a particular vertex o, called the root. An *isomorphism* from a rooted graph (G, o) to a rooted graph (G', o') is a bijection $\gamma : V \to V'$ that preserves:

- the root: $\gamma(o) = o'$,
- the edges: $(i, j) \in E \iff (\gamma(i), \gamma(j)) \in E'$.

When such a γ exists we say that (G, o) is isomorphic to (G', o'), and we write $(G, o) \equiv (G', o')$. We let $\mathcal{G}_\star$ denote the set of connected rooted graphs that are locally finite, i.e., each node has a finite number of neighbors. Given $(G, o) \in \mathcal{G}_\star$ and $d \in \mathbb{N}$, we let $[G, o]_d$ denote the rooted subgraph obtained from (G, o) by keeping only those vertexes whose distance from o is at most d, and all the edges between them. In $\mathcal{G}_\star$, a sequence $\{(G_n, o_n) : n \in \mathbb{N}\}$ is said to converge locally to (G, o) if for every radius $d \in \mathbb{N}$, there exists $n_d \in \mathbb{N}$ such that:

$$n \geq n_d \Longrightarrow [G_n, o_n]_d \equiv [G, o]_d. \tag{1.15}$$

The idea one wants to capture with this definition is that for large n, G_n looks very much like G in an arbitrarily large neighborhood of the root of G. What one observes is that the space $\mathcal{G}_\star$ can be endowed with a function $d_{\mathcal{G}_\star}$ that associates to each two rooted graphs belonging to it a distance [vdH16]:

$$d_{\mathcal{G}_\star}\left((G, o), (G', o')\right) = \frac{1}{R^\star + 1}, \quad \text{where } R^\star = \sup\left\{r : [G, o]_r \equiv [G', o']_r\right\}, \tag{1.16}$$

which metrizes the notion of convergence of Equation 1.15. This fact turns $\mathcal{G}_\star$ into a complete separable metric space [Sal11], and then the theory of the weak convergence of probability measures [Sag13] can be used.

Definition 1.3 (Local weak convergence of random graphs)**.** Let G_n be a random graph obtained from the ensemble $\mathbb{G}_\star = (\mathcal{G}_\star, \mathbb{P}_\star)$ and let (G, o) be a random rooted graph obtained from the same ensemble. Then we say that G_n converges in distribution in the local weak sense to (G, o) when:

$$\mathbb{E}_n\left[A(G_n, o_n)\right] \xrightarrow{n\to\infty} \mathbb{E}\left[A(G, o)\right], \tag{1.17}$$

where the expectation on the right-hand side of Equation 1.17 is w.r.t. (G, o), for every observable $A : \mathcal{G}_\star \to \mathbb{R}$, while the expectation $\mathbb{E}_n$ is w.r.t. the random vertex o_n, and the random graph G_n [AS04, vdH16].

Although analogous considerations can be made for different ensembles of sparse graphs, in the following we focus on the RRG one, that underlies all the models studied in this thesis. In the thermodynamic limit $N \to \infty$, graphs belonging to the random z-regular graph ensemble admit as local weak limit the infinite tree of valence z. This behavior here is a direct consequence of the fact that short cycles are asymptotically rare in random z-regular graphs, i.e., cycles have a vanishing density in the large graph limit, as it is discussed in the next section.

Cycles

The number of short cycles of given length in random regular graphs are random variables that become independent in the large graph limit. Given a set of random variables Z_i, $i \in I \subset \mathbb{N}$ with I finite, they are asymptotically independent Poisson random variables with means λ_i if their joint probability distribution tends to that of independent Poisson random variables whose means are fixed numbers λ_i. In formulas:

$$\mathbb{P}\left(\bigwedge_{\ell \in I} \{Z_\ell = r_\ell\}\right) \xrightarrow{N\to\infty} \prod_{\ell \in I} e^{-\lambda_\ell} \frac{\lambda_\ell^{r_\ell}}{r_\ell!}, \tag{1.18}$$

for every fixed set of non-negative integers r_ℓ.

Let us now suppose that $Z_\ell^{(N)}$, with $\ell \geq 3$, is the random variable that represents the number of cycles of length ℓ in a random z-regular graph with N nodes. For fixed $n \geq 3$, it is proved [Bol80, Wor80, Wor81] that $Z_\ell^{(N)}$, $3 \leq \ell \leq N$, are asymptotically independent Poisson random variables with means:

$$\lambda_\ell = \frac{(z-1)^\ell}{2\ell}, \tag{1.19}$$

then the average density of cycles of a given length asymptotically vanishes, and this is at the basis of the locally tree-like structure of the ensemble. Equation 1.19 can be argued by looking at the configurational model. The number a_ℓ of ℓ-cycles is in one to one correspondence with the number of sets of ℓ edges $\{e_1, ..., e_\ell\}$, such that there exists a sequence of ℓ distinct vertexes $(v_1, ..., v_\ell)$ with e_i connecting an end emanating from v_i to an end emanating from v_{i+1}, with addition modulo ℓ. Each set of such edges can be obtained from 2ℓ different sequences of edges. The total number of this kind of sequences can be found by counting all the ways they can be realized by the connection of stubs:

$$a_\ell = z^\ell (z-1)^\ell \frac{N!}{(N-\ell)!} \approx \frac{z^\ell (z-1)^\ell}{2\ell} N^\ell. \tag{1.20}$$

For any set S of ℓ pairwise disjoint edges, the probability that they all appear in a random matching is given by the ratio between the number of matchings that contain S and the total number of matchings:

$$\frac{(zN-2\ell)!!}{(zN-1)!!} = \frac{1}{(zN-1)(zN-3)\cdots(zN-2\ell+1)} \approx \frac{1}{(zN)^\ell}, \tag{1.21}$$

and then Equation 1.19 is recovered.

1.2 Optimization problems and graphs

There exist many problems that can be expressed in the graph language, and whose great practical and intrinsic interest motivates the development of efficient algorithms [KV18]. The attempt of most of these algorithms is to optimize, that is maximize or minimize, some quantity [Cla07]. In this section some definitions and examples about optimization theory are introduced. As the study of matching problems is the main subject of this thesis, a particular attention is payed to these.

An *instance*, or *input*, of an *optimization problem* is a pair $(\mathscr{F}, \mathscr{C})$, where $\mathscr{F}$ is the set of feasible solutions, and $\mathscr{C}$, called cost function, is a mapping:

$$\mathscr{C} : \mathscr{F} \longmapsto \mathbb{R}. \tag{1.22}$$

The problem is to find (if exists) an element $f \in \mathscr{F}$ for which:

$$\mathscr{C}(f) \leq \mathscr{C}(g) \quad \forall g \in \mathscr{F}. \tag{1.23}$$

Such element f is called a *globally optimal solution* to the given instance. An *optimization problem* is a set I of instances. Problems in which $\mathscr{F}$ is a countable set, possibly infinite, are called *combinatorial optimization problems*. To be concrete it is useful to discuss some classical examples.

The traveling salesman problem (TSP)

In an instance of the TSP we are given a complete graph on N vertexes K_N, and the distances between every pair of N nodes in the form of an $N \times N$ symmetric matrix $\boldsymbol{D}$, with $D_{ij} \in \mathbb{R}^+$. Let us define a *tour* as a closed path that visits every city exactly once. We can take:

$$\mathscr{F} = \{\text{all permutations } \pi \text{ of } N \text{ objects}\}. \tag{1.24}$$

Denoting by $N + 1 = 1$, the cost function is the map:

$$\mathscr{C} : \pi \longmapsto \sum_{j=1}^{N} D_{\pi(j)\pi(j+1)}. \tag{1.25}$$

A similar formulation of the problem can be given on a generic connected graph G. It is worth noting that in the complete case the number of possible solutions of the problem $|\mathscr{F}|$ is given by $\frac{N!}{2N} = \frac{(N-1)!}{2}$, therefore trying to solve an instance by a brute force approach, i.e. listing all possible solutions and then selecting the optimal one, would require a rapidly increasing time with the instance size.

Minimum spanning tree

A *spanning tree* $T = (V, E')$ of a graph $G = (V; E)$ is a subgraph that contains all the vertexes of G, and is a tree. Let us associate, as above, to each edge (i, j) of G a weight $w_{ij} \in \mathbb{R}^+$. The problem is to find a spanning tree of G, $T = (V, E')$, that has minimal total length of its edges. In this case we have:

$$\begin{aligned} &\mathscr{F} = \{\text{all spanning trees of } G\} \\ &\mathscr{C} : T = (V, E') \in \mathscr{F} \longmapsto \sum_{(i,j)\in E'} w_{ij}. \end{aligned} \tag{1.26}$$

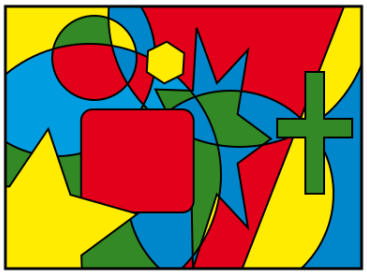


Figure 1.3. A map divided into regions can be seen as a graph in which each node is a region and each edge is a border. The four color theorem states that, given any separation of a plane into contiguous regions, no more than four colors are required to color the regions of the map so that no two adjacent regions have the same color [AH89].

The graph q-coloring problem (q-Col)

Suppose to have a palette of q colors identified by an integer $c \in 1, ..., q$. Given a graph $G = (V; E)$ the problem is to assign to each vertex $v \in V$ a color c_v in such a way that the cost function

$$\mathscr{C}[\{c_i\}_{i\in V}] = \sum_{(i,j)\in E} \delta_{c_i c_j} \tag{1.27}$$

is minimized. With this definition the best possible coloring, if it exists, is that of zero cost, in which all adjacent vertexes have different colors, as in Figure 1.3.

Satisfiability (SAT)

Let us consider a bipartite graph $G = (V; E)$. Therefore there exist two sets W, Z, with $|W| = M$ and $|Z| = N$, such that $V = W \cup Z$ and if $(a, i) \in E$ then $a \in W$, $i \in Z$. In the following, we index with $a, b, \ldots$ the elements of W and with $i, j, \ldots$ the elements of Z. Suppose now that to each node of G is associated with a binary variable, that is indicated with $y_a \in \{-1, 1\}$ for the nodes belonging to W and with $x_i \in \{-1, 1\}$ for the others, and that the relations

$$y_a = \prod_{i\in\partial a} \frac{1 - J_{(i,a)} x_i}{2} \quad \forall a \in W, \tag{1.28}$$

hold, where the $J_{(i,a)}$'s are constants belonging to $\{-1, 1\}$, fixed with the instance of the problem and associated to the edges of the graph. The *Satisfiability* problem consists in finding the configuration $\{x_i\}_{i=1}^{N} \in \{-1, 1\}^N$, such that:

$$\mathscr{C} = M - \sum_{a=1}^{M} y_a \tag{1.29}$$

is minimized. It is worth noting that the domain of feasible solutions, $\mathscr{F} = \{-1, 1\}^N$, is exponential in the size of Z, $|\mathscr{F}| = 2^N$.

1.2.1 Matching problems

A *matching* $\mathcal{M}$ of a graph $G = (V; E)$ is a subset of the edge set with the property that no two edges of $\mathcal{M}$ share the same node. A matching $\mathcal{M}$ of G is said maximum if, given another matching $\mathcal{M}'$, $|\mathcal{M}'| \leq |\mathcal{M}|$. By definition a matching can never have more than $|V|/2$ edges. If V is even and the cardinality of a matching is exactly $|V|/2$ the matching is said *perfect*. Suppose that each edge e of G is associated with a weight $w_e \in \mathbb{R}^+$. The *minimum matching* problem consists, given a graph $G = (V; E)$, in finding the perfect matching $\mathcal{M}$ such that:

$$\mathscr{C}[\mathcal{M}] \equiv \sum_{e \in \mathcal{M}} w_e = \sum_{e \in E} m_e w_e \tag{1.30}$$

is minimum. In Equation 1.30 $m_e : E \to \{0, 1\}$ is an application that associates 1 with edges belonging to $\mathcal{M}$ and 0 with other edges. Then for the minimum matching problem $\mathscr{F}$ is the set of all possible perfect matchings of G, i.e., the set of all possible sequences $\{m_e\}_{e \in E}$ that satisfy:

$$\sum_{k \in \partial i} m_{(k,i)} = 1 \quad \forall i \in V. \tag{1.31}$$

If we take $w_e = 1 \ \ \forall e$, enlarge $\mathscr{F}$ to the set of all matchings, and take:

$$\mathscr{C}[\mathcal{M}] = -\sum_{e \in \mathcal{M}} w_e, \tag{1.32}$$

the problem quite often is called simply *matching problem*, or maximum cardinality matching problem, that is the problem of finding a maximum cardinality matching $\mathcal{M}$ on a given graph G [PS98]. In what follows, unless otherwise specified, we will denote by matching problem, or standard matching problem, the minimum weight perfect matching problem on a generic graph.

If, instead of a generic graph, it a bipartite graph $G = (W, Z; E)$ is considered, with $|W| = |Z| = N$, the minimum matching problem is known as the *Assignment problem*. Furthermore if G is complete, it is worth noting that the matching can be represented by a permutation of N objects, and therefore there exist $N!$ different assignments.

Properties of matchings

In this section some important properties of matchings on graphs are discussed. A complete and wide discussion can be found in [LP09].

The *Tutte*'s theorem characterizes graphs in which a perfect matching is possible:

Theorem 1.4 (Tutte)**.** *A graph, $G = (V; E)$, has a perfect matching if and only if for every subset U of V, the subgraph induced by $V \setminus U$, i.e., obtained by removing the vertexes of U from G, has at most $|U|$ connected components with an odd number of vertexes* [LP09].

A necessary condition for the existence of a perfect matching is that the number of vertexes is even; however, because of Theorem 1.4, this condition is not sufficient, as it is shown in Figure 1.4. *Hall's marriage theorem* provides an intuitive formulation of Theorem 1.4 for bipartite graphs:

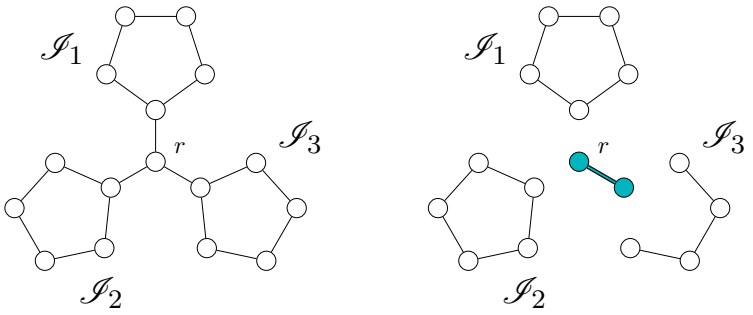


Figure 1.4. Suppose that a perfect matching exists for the graph on the left. Then the root node r has to be matched to one of its three neighbors, e.g. that on the right. However by this procedure $\mathscr{I}_1$ and $\mathscr{I}_2$ are left with an odd number of vertexes, then a perfect matching cannot exist.

Theorem 1.5 (Hall's marriage)**.** *Let G be a bipartite graph with parts W and Z. For a set X of vertexes in W, let ∂X denote the neighborhood of X in G, i.e. the set of all vertexes in Z adjacent to some element of X. Then there is a matching that entirely covers W if and only if for every subset X of W:*

$$|X| \leq |\partial X|. \tag{1.33}$$

If the graph is a tree, an interesting property holds on the number of possible matchings:

Proposition 1.2. *Every tree $T = (V; E)$ has at most one perfect matching.*

Proposition 1.2 can be easily proved by induction on the number of vertexes of the tree.

At this point let us discuss a strategy that usually is adopted to solve the maximum cardinality matching. A first important important definition is the following:

Definition 1.6 (Alternating and augmenting paths)**.** Let $G = (V; E)$ be a graph and $\mathcal{M}$ be a matching in G. Then a path $v_1, ..., v_k$ is said to be *alternating* respect to $\mathcal{M}$ if successive edges of the path alternately belong to $\mathcal{M}$ and $E \setminus \mathcal{M}$. An alternating path is said to be an *augmenting path* if the first and the last vertexes in it are *unmatched*.

An important characterization of maximum matching is given by the following result:

Theorem 1.7 (Berge's Theorem)**.** *If $G = (V; E)$ is a graph and $\mathcal{M}$ a matching in G, then $\mathcal{M}$ is a maximum matching if and only if G has no augmenting paths with respect to $\mathcal{M}$.*

Given an augmenting path P with respect to a matching $\mathcal{M}$, it is possible to improve $\mathcal{M}$ by inverting the edges along the path: matched edges are changed to unmatched ones and vice versa. By doing so, the cardinality of the matching is increased by 1, as shown in Figure 1.5. All known algorithms for maximum matchings are based on the idea to take a starting matching, e.g. the empty one, and repeatedly discover augmenting paths until a maximum matching is found.

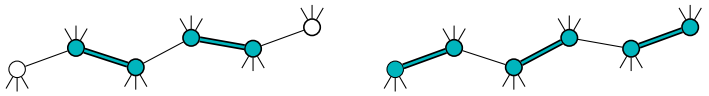


Figure 1.5. The path on the left is an example of augmenting path. The colored edges belong to the matching. If the path is reversed, as shown in the figure on the right, the cardinality of the matching is increased by 1.

The Hungarian algorithm

In this section we would like to sketch some ideas underling a classical combinatorial optimization algorithm, the Hungarian algorithm [Kuh05], that, in its original formulation, solves the assignment problem in time $\mathcal{O}(N^4)$. All discussed matching problems belong to the so-called P complexity class, i.e., to the class of optimization problems for which deterministic algorithms running in polynomial time respect to the input size are available [LP09, CLRS09].

The Hungarian algorithm is based on two fundamental considerations. The first one is that by choosing two vectors $\underline{\theta}, \underline{\lambda} \in \mathbb{R}^N$ the optimal solution is left unchanged under the following *gauge transformation*:

$$\begin{aligned} w_{i,j} &\longmapsto w_{ij} - \theta_i - \lambda_j, \\ \sum_{i,j} w_{ij} m_{ij} &\longmapsto \sum_{i,j} w_{ij} m_{ij} - \sum_i \theta_i - \sum_j \lambda. \end{aligned} \tag{1.34}$$

Indeed from Equation 1.34 it follows that the total cost of any feasible solution will change of the same quantity $-\sum_i \theta_i - \sum_i \lambda_i$, an therefore an optimal assignment will remain optimal. A gauge transformation $(\underline{\theta}, \underline{\lambda})$ is said *proper* if it preserves the positive sign of the weights.

The second important consideration is the following. Firstly let us observe that, thanks to the first consideration, all the weights can be taken non negative without loss of generality; let us consider the set S of all edges with zero weight, if any. Let us denote by $G' = (V; S)$ the graph whose nodes are that of G, and whose edge set is S. Thanks to the first consideration once again, if a perfect matching in $G' = (V; S)$ exists, it follows that such matching is the optimal solution for the assignment problem on G. The idea of the Hungarian algorithm is to manipulate the weights w_{ij} through suitable proper gauge transformations of $\mathbf{W} = (w_{ij})_{ij}$, in order to make more and more zeros appear, until $G' = (V; S)$ has a perfect matching. It is worth noting that, since in G' all the edges have the same weight, in that case the problem reduces to a maximum cardinality one. At the end the value of the optimal solution will be given by the sum of the original weights associated with the edges belonging to the perfect matching. The existence of such a gauge that allows to reduce the assignment problem to an unweighted matching problem on bipartite graph is guaranteed by the following result:

Theorem 1.8 (König-Egerváry)**.** *There exists a proper gauge $(\underline{\theta}^\star, \underline{\lambda}^\star)$ such that the cost of the optimal assignment $\mathcal{M}^\star$ is given by:*

$$\mathscr{C}[\mathcal{M}^\star] = \sum_i \theta_i^\star + \sum_i \lambda_i^\star. \tag{1.35}$$

Chapter 2

Equilibrium Statistical Mechanics

2.1 Some important ideas

The main aim of statistical mechanics grounds on the attempt to derive the thermodynamic properties of macroscopic bodies starting from the fundamental laws that describe their microscopic components [Hua09, LL68]. For a classical Hamiltonian system with $2N$ degrees of freedom, given an initial configuration $(q,p) = (q_1...q_N, p_1...p_N)$, the set of generalized positions $q_i(t)$ and generalized momenta $p_i(t)$ can be found in principle by solving the Cauchy problem for the Hamilton equations [LL60]:

$$\begin{cases} \dfrac{dq_i}{dt} = \dfrac{\partial \mathcal{H}}{\partial p_i}, \quad \dfrac{dp_i}{dt} = -\dfrac{\partial \mathcal{H}}{\partial q_i} & \forall\ i = 1, ..., N \\ q_i(0) = q_i, \ \ p_i(0) = p_i, \end{cases} \tag{2.1}$$

where $\mathcal{H}$ is the Hamiltonian of the system. As for a macroscopic object the order of magnitude of N is typically $\approx 10^{23}$, any attempt to derive the thermodynamic properties by solving Equation 2.1 would be vain. It is worth emphasizing that, even if a general solution of the equations of motion were known, it would be quite impossible to impose the initial conditions, because the time required by the best computer ever built would be larger than the recorded human history. The only way to face the problem is to follow a probabilistic approach. Given an arbitrary observable $\mathcal{O}(q,p)$, the fundamental hypothesis of equilibrium statistical mechanics is that, under suitable conditions [Par88], the time average $\overline{\mathcal{O}}$:

$$\overline{\mathcal{O}} = \lim_{t\to+\infty} \frac{1}{t}\int_0^t d\tau\ \mathcal{O}(q(\tau), p(\tau)) \ \text{ is equal to } \ \frac{1}{\mathcal{Z}}\int dqdp\ \mathcal{O}(q,p)\, e^{-\beta\mathcal{H}(q,p)}, \tag{2.2}$$

where:

$$\psi(q,p) = \frac{e^{-\beta\mathcal{H}(q,p)}}{\mathcal{Z}} \tag{2.3}$$

is the *canonical distribution*, $\mathcal{Z}$ is known as *partition function* and, by using scales in which temperature and energy have the same units, $\beta = T^{-1}$ is the inverse of the absolute temperature. It is well known that the partition function plays a central

role in the computation of many physical quantities of interest; the reader is referred to [Mus10] for a wide-ranging discussion of many examples.

An interesting and useful property of the distribution of Equation 2.3 is that it can be derived from a *variational principle*; in particular it minimizes the so-called *free energy functional* $\mathcal{F}[P]$:

$$\mathcal{F}[P] = E[P] - \frac{S[P]}{\beta}, \tag{2.4}$$

$$E[P] = \int dqdp\ P(q,p)\mathcal{H}(q,p), \quad S[P] = -\int dqdp\ P(q,p)\log P(q,p),$$

subject to the positivity and the normalization constraints, i.e., $P(q,p) \geq 0$, and $\int dqdp\ P(q,p) = 1$. In Equation 2.4 $S[P]$ is the *entropy* of P [Gne17], and $E[P]$, when P is equal to the canonical distribution, is the average energy of the system described by the Hamiltonian $\mathcal{H}(q,p)$.

The variational principle can be easily checked when the set of all possible configurations is countable, that could happen in a system obtained as an approximation of a Hamiltonian one, by dividing its phase space into cells. In this case all the configurations are identified by an index k, and Equation 2.3, $\mathcal{Z}$ and Equation 2.4 become respectively:

$$\psi_k = \frac{e^{-\beta\mathcal{H}_k}}{\mathcal{Z}}, \quad \mathcal{Z} = \sum_k e^{-\beta\mathcal{H}_k}, \quad \mathcal{F}[P] = \sum_k P_k\mathcal{H}_k + \frac{1}{\beta}\sum_k P_k \log P_k. \tag{2.5}$$

The constrained minimum of Equation 2.4 can be reduced to the problem of finding the minimum of the Lagrangian:

$$\mathcal{L} = \mathcal{F}[P] + \lambda\left(\sum_k P_k - 1\right), \tag{2.6}$$

where the Lagrangian multiplier λ has to be chosen in such a way that the minimum of $\mathcal{L}$ satisfies $\sum_k P_k = 1$. The extremality condition implies:

$$\frac{\partial\mathcal{L}}{\partial P_k} = 0 \Rightarrow \mathcal{H}_k + \frac{1}{\beta} + \lambda + \frac{1}{\beta}\log P_k = 0 \quad \forall k, \tag{2.7}$$

and, by using λ for the normalization, it follows:

$$P_k = \psi_k = \frac{e^{-\beta H_k}}{\mathcal{Z}}, \quad \mathcal{Z} = \sum_k \psi_k e^{-\beta H_k}, \tag{2.8}$$

that is the minimum, being the second derivative $\frac{\partial^2\mathcal{L}}{\partial P_k^2}$ greater than zero:

$$\frac{\partial^2\mathcal{L}}{\partial P_k^2} = \frac{1}{\beta P_k} > 0\ \forall\, k. \tag{2.9}$$

The formulation in terms of a variational principle, as will be introduced in section 2.4, is the starting point for the discussion of many approximate methods.

2.2 Ferromagnetic systems: the Ising model

An important success of statistical mechanics is represented by its application to the study of phase transitions on magnetic systems [Wol00]. For it seems that the essential features of cooperative phenomena do not depend on the details of the interaction of electrons with nuclei, or between molecules, often modelings of the true system are used [Nis05, Bru67]. The standard procedure consists in trying to construct the simplest model that still preserves the interesting physical properties of the system, by picking from all the degrees of freedom just a small set of variables. Such a model usually depends on a few parameters that can be fitted experimentally to reproduce the right physics.

In many magnetic materials the electrons responsible for their magnetic behavior are localized near the atoms of a lattice, and the force, which tends to orient the spins, is the exchange interaction, whose range is comparable with the lattice spacing [Par88]. One of the most popular and prototypical models that describe this situation is the Ising model. We consider a periodic lattice with N sites, in dimension d. To each site i one attaches a "spin" x_i that can take only two values, $x_i \in \{-1, 1\}$, and represents the magnetization produced by the site; therefore there are 2^N configurations, each one defined by a vector of the form $\mathcal{C} = (x_1, ..., x_N)$. Let us assume that only pairs of neighbouring spins interact, and that the energy of a configuration $\mathcal{C}$ is:

$$\mathcal{H}(\mathcal{C}) = -J \sum_{(i,k)} x_i x_j - \sum_i h_i x_i, \tag{2.10}$$

where the sum over i and k runs over all possible nearest neighbours of the lattice, and the second term takes into account the presence of a site-dependent external magnetic field h_i. The coupling constant J rules the strenght of the interaction between spins. Moreover if $J > 0$ the energy is lower for parallel spins, and the system is said to be ferromagnetic; if $J < 0$ the system is antiferromagnetic, and nearby spins tends to be antiparallel.

An exact solution of the model [Bré10, Mus10], the exact calculation of the free energy (Equation 2.4) and of the correlation function between variables $\langle x_i x_j \rangle$,

$$\langle x_i x_j \rangle = \frac{1}{\mathcal{Z}} \sum_{\mathcal{C}} x_i x_j e^{-\beta \mathcal{H}(\mathcal{C})}, \tag{2.11}$$

is available only in one dimension, or in two dimensions at zero magnetic field: the other cases require an approximate approach. Of all different methods, the class of mean fields approximations (MFA) [Kad09], that will be introduced in section 2.4, play an important role due to the wideness of its applicability range, expanded to all the systems that can be studied with the formalism of equilibrium statistical mechanics.

In the following, we focus on systems with a countable set of possible states. Often in the study of such kind of models, statistical mechanics meets graph theory, whereas the lattices over which the systems are defined can be considered as graphs and graph theory language can be useful. For this reason, before discussing the

MFA, in section 2.3 the factor graph formalism is introduced, a useful tool for the study of models defined on generic graphs that allows to represent graphically the probability distribution associated to the model.

Spontaneous magnetization

For dimension $d \geq 2$, ferromagnetic systems like the ferromagnetic Ising model have spontaneous magnetization for $h_i = 0$ below a critical temperature, called Curie temperature. Such a behavior is associated to a probability measure in the space of configurations that is not invariant under the symmetry group of the Hamiltonian of the system at zero external field. At first sight spontaneous symmetry breaking seems to be in conflict with Equation 2.3: if $\mathcal{H}$ is invariant, ψ must also be invariant. This argument, although correct at finite volume, fails for an infinite system in which Equation 2.3 is only formal and the spontaneous symmetry breaking occurs because of the appearance of real zeros in the partition function, and therefore of non-analyticities in the free energy density, a fact that can only take place in the termodynamical limit. The presence of such non-analytic behavior is related to the presence of more than one equilibrium state (in the case of ferromagnets, two equilibrium states corresponding to the two different magnetizations). According to the Ehrenfest classification, if the free energy is differentiable $(k-1)$ times, but not k times, the singularity point is called a *transition* of order k: it is common in the community to use the terminology "second order transition" for any transition of order $k \geq 2$. The description of second order transition through the theory of criticality and critical exponents it is one of the major successes of statistical mechanics. The reader is referred to [Par88, Bré10] for a complete discussion about this topic.

2.3 Factor graphs

Abstracting from the Ising example, assume the configuration of a magnetic-like system to be determined by a set of N variables $x = (x_1, ..., x_N)$, taking values in a finite alphabet $\mathcal{X}$, $x_i \in \mathcal{X}$ $\forall i$. Suppose the canonical distribution $\psi(\underline{x})$ of the system can be factorized as:

$$\psi(\underline{x}) = \frac{1}{\mathcal{Z}} \prod_{a=1}^{M} \psi_a(\underline{x}_{\partial a}), \tag{2.12}$$

where $\underline{x}_{\partial a} \equiv \{x_i | i \in \partial a\}$, and ∂a is a set of indices belonging to the set $[N]$ of all possible subsets of integer i such that $1 \leq i \leq N$; the $\psi_a(\underline{x}_{\partial a})$'s moreover are real non negative function such that $\forall S \subseteq \partial a$ it is impossible to find two real non negative functions f, g that satisfy:

$$\psi_a(\underline{x}_a) = f(\underline{x}_S) g(\underline{x}_{\bar{S}}), \tag{2.13}$$

where $\overline{S}$ is the complementary of S, i.e. $S \cup \overline{S} = \partial a$. In the Ising case for example, the fact that only neighboring variables interact allows to factorize the distribution as:

$$\psi(\underline{x}) = \frac{1}{\mathcal{Z}} \prod_{i=1}^{N} e^{\beta h_i x_i} \prod_{\langle k,l \rangle} e^{\beta J x_k x_l}, \tag{2.14}$$

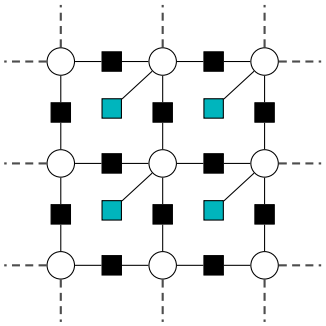
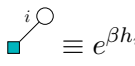


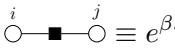
Figure 2.1. Representation of the factor graph associated with an Ising model defined on a square bidimensional lattice.

where we stressed the fact that i and j are neighbours using the notation $\langle i, j \rangle$, and every factor $e^{\beta h_i x_i}$ and $e^{\beta J x_i x_j}$ corresponds to a single interaction: the former between the external magnetic field h_i and x_i, and the latter between x_i and x_j. It is possible to provide a graphical representation of a joint probability distribution like (2.14), by associating with it a so-called *factor graph*. The factor graph of a generic distribution of the form (2.12) is a graph F_ψ that contains two types of nodes: N *variable nodes*, each one associated with a variable x_i, and M *function nodes*, each one associated with a function ψ_a. To distinguish variable from function nodes, the former will be represented by circles, and the latter by squares. Every factor graph is bipartite: an edge can join only a variable node to a function node; in particular the variable node i is joined to the function node a if and only if $i \in \partial a$. Then the parts of F_ψ are a variable nodes set V, and a function nodes set F. A probability distribution like (2.12) together with its factor graph is usually called *graphical model*.

A factor graph representation of an Ising model defined on a bidimensional square lattice is given in Figure 2.1, where the following associations between the factors of Equation 2.14 and the graph's components are made:



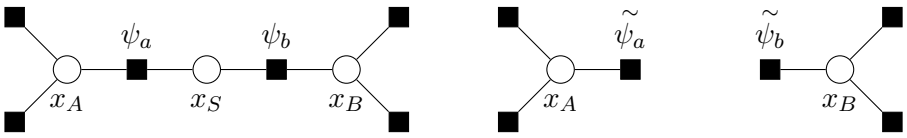
$$\equiv e^{\beta h_i x_i},$$



$$\equiv e^{\beta J x_i x_j}.$$

The generic factor graph is made up of several connected components. It is important to note that variables belonging to disconnected components are statistically independent: indeed in this case, by definition of function node, the joint probability distribution of all the variables is factorized in the product of the probability distributions of the connected components. This fact underlies the way the *conditional independence* structure of the variables is encoded in the graph.

Definition 2.1 (Conditional independence)**.** Let A, B, $S \subseteq [N]$ be three disjoint subsets of the variable nodes, and denote by $\underline{x}_A$, $\underline{x}_B$ and $\underline{x}_S$ the corresponding sets of variables. If the conditional probability distribution $\psi(\underline{x}_A, \underline{x}_B | \underline{x}_S)$ can be written



(a) The unique path connecting A to B passes through S.

(b) Conditioning upon x_S creates two disconnected components.

Figure 2.2. Example of conditional independence between variables.

as:

$$\psi(\underline{x}_A, \underline{x}_B|\underline{x}_S) = \psi(\underline{x}_A|\underline{x}_S)\psi(\underline{x}_B|\underline{x}_S), \tag{2.15}$$

then the variables $\underline{x}_A$, $\underline{x}_B$ are said to be conditionally independent with respect to S.

Graphically the conditional independence of $\underline{x}_A$ and $\underline{x}_B$ is reflected in the absence of paths on the factor graph joining a node of A to a node of B without passing through S [MM09]. Let us consider, for example, a tree factor graph. If i is a variable node, and $A = \{j \in V \setminus \{i\} \colon \exists a \in F \text{ such that } i, j \in \partial a\}$, then $\forall j, k \in A$, x_j and x_k are conditionally independent with respect to i. This property underlies the exactness of the Bethe method (subsection 2.4.3) for graphical models defined on trees. In Figure 2.2a the three sets A, B and S contain each one only a single variable node; conditioning upon a variable x_i is equivalent to eliminating the corresponding variable node from the graph, and modifying the adjacent function nodes accordingly: then by conditioning on $\underline{x}_S$ the factor graph becomes the same as that of Figure 2.2b, where the factors associated with the function nodes a and b, that are the adjacent function nodes to the removed variable, are different.

2.4 Mean Field Approximations

Quite often the computation of partition functions poses a formidable problem from a mathematical point of view. For this reason, it is important to develop approximate methods that allow us to analyze the most relevant physical aspects of the original problem.

2.4.1 The naïve mean field approximation

Suppose that a system S is defined by a set of N variables $\underline{x} = (x_1, ..., x_N)$, with x_i taking values in a finite alphabet $\mathcal{X}$ $\forall i$, and by a probability distribution $\psi(\underline{x})$ of the form (2.3) that we write as follows:

$$\psi(\underline{x}) = \frac{1}{\mathcal{Z}} \prod_{a=1}^{M} \psi_a(\underline{x}_{\partial a}). \tag{2.16}$$

The heart of the mean field approach relies on the attempt to ensure that some variables, that in the starting model are correlated, become independent in the approximate one. In the naïve MFA, that is the simplest MFA, all the correlations

between variables are removed. This is achieved by replacing $\psi(\underline{x})$ with a distribution $p(\underline{x})$ of the form:

$$p(\underline{x}) = \prod_{i=1}^{N} b_i(x_i), \tag{2.17}$$

i.e., $p(\underline{x})$ belongs to the set $\mathfrak{F}_S$ of all possible completely factorized distributions defined on the phase space of S. In order to choose exactly $p(\underline{x})$, the fundamental idea is to apply in $\mathfrak{F}_S$ the variational principle (2.4), i.e. to look in this space for the $p(\underline{x})$ that minimizes the free energy functional $\mathcal{F}$. If the x_i's are Ising-like variables, e.g. $x_i \in \{-1, 1\}$, the b_i's, called believes, can be expressed without loss of generality in the form:

$$b_i(x_i) = \frac{1 + m_i x_i}{2},$$

where $m_i \in \mathbb{R}$ is a so-called *mean field parameter*. As $\underline{m} = (m_1, ..., m_N)$ defines uniquely a probability distribution in $\mathfrak{F}_S$, the minimization of the free energy functional in $\mathfrak{F}_S$ for a system with a binary alphabet becomes the search for the global minimum of a function $f(\underline{m})$ of N variables m_i:

$$f(m_1, ..., m_N) = \mathcal{F}\left[p\left(\underline{x}; \underline{m}\right)\right] = \langle \mathcal{H}(\underline{x}) \rangle_p - \frac{1}{\beta} S\left[p\left(\underline{x}; \underline{m}\right)\right], \tag{2.18}$$

where the notation $p(\,\cdot\,; \underline{m})$ emphasizes that p is fixed by the realization of the mean field parameters, and $\langle \mathcal{H}(\underline{x}) \rangle_p$ indicates the expected value of $\mathcal{H}$ with respect to p. In order to be concrete let us consider the example of an Ising model defined on a square lattice in d dimensions; in this case the energy $\mathcal{H}$ takes the form:

$$\mathcal{H} = -\frac{1}{2} \sum_{i,j} J_{ij} x_i x_j - \sum_{i=1}^{N} h_i x_i, \tag{2.19}$$

where the first summation is extended to all the pairs (i, j), $1 < i, j < N$, and:

$$J_{ij} = \begin{cases} 1 & \text{if } |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.20}$$

The extremality conditions for the minimum, obtained by substituting (2.17) in (2.18), and by imposing $\frac{\partial f}{\partial m_i} = 0$, take the following form:

$$m_i = \tanh\left[\beta \left(\sum_j J_{ij} m_j + h_i\right)\right] \quad \forall i, \tag{2.21}$$

that is a system of N coupled self-consistent equations for $\underline{m}$. The system (2.21) provides the best guess for the mean field parameters $\underline{m}^\star$ in this approximation. Remarkably enough, for $h_i = 0$ it predicts a transition temperature β_c^{-1} below which a spontaneous magnetization appears. Indeed, for $\beta > \beta_c = 1/2d$ the free energy develops two minima, corresponding to two opposite magnetization values.

Solving mean-field expressions like Equation 2.21 in general is considerably simpler than computing the partition function and, once the fixed point is known, it is possible to estimate several physical observables of the system. For more details the reader is referred to [Par88].

2.4.2 Region-based approximations

In order to go beyond the naïve MFA it is necessary to reintroduce some correlations in the approximated model. A possible way to do that is by the region-based approximation of the free energy [YFW05]. The idea is to divide the starting system into regions suitably chosen, to compute the single-region free energies, and then to approximate the total free energy with a combination of single-region contributions.

Definition 2.2 (Region)**.** A region R of a factor graph is a pair (V_R, F_R), where V_R denotes a set of variable nodes and F_R a set of function nodes, such that if a function node a belongs to F_R, all the variable nodes neighboring a are in V_R.

The choice of the set $\mathcal{R}$ of regions into which we have to break up the factor graph is in principle arbitrary, but we require that no function or variable nodes are excluded from all the regions, in order not to reduce the size of the phase space of S. Suppose to choose a set of regions $\mathcal{R}$; let us denote by $\{P_R\}$ a set of probability distributions, one for each $R \in \mathcal{R}$, such that $\forall R$ P_R is a function of the variables belonging to R only. It is possible to define a so-called region-based free energy functional $\mathcal{F}_\mathcal{R}$, that associates with a given $\{P_R\}$ a real number as follows:

$$\begin{aligned}\mathcal{F}_\mathcal{R}[\{P_R\}] &\equiv \sum_{R\in\mathcal{R}} c_R \sum_{\underline{x}_R} P_R(\underline{x}_R) E_R(\underline{x}_R) + \sum_{R\in\mathcal{R}} c_R \sum_{\underline{x}_R} P_R(\underline{x}_R) \log P_R(\underline{x}_R) \\ &\equiv U_\mathcal{R}[\{P_R\}] - S_\mathcal{R}[\{P_R\}],\end{aligned} \tag{2.22}$$

where:

$$E_R = -\sum_{a\in F_R} \log \psi_a(\underline{x}_{\partial a}), \tag{2.23}$$

and it has been set $\beta \equiv 1$ for simplicity of notation. It is important to note that each node can belong to more than one region: the c_R constants are introduced to assure that every factor and variable node contributes just one time to $\mathcal{F}_\mathcal{R}$; this is guaranteed by the conditions:

$$\sum_{R\in\mathcal{R}} c_R \mathbb{I}(i \in V_R) = \sum_{R\in\mathcal{R}} c_R \mathbb{I}(a \in F_R) = 1 \quad \forall\, a \in V_R, \quad \forall\, i \in F_R, \tag{2.24}$$

where $\mathbb{I}(\cdot)$ is the indicator function:

$$\mathbb{I}(\text{condition}) = \begin{cases} 1 & \text{if the condition is satisfied} \\ 0 & \text{otherwise.} \end{cases} \tag{2.25}$$

It is worth noting that even if we chose as $\{P_R\}$ the set of marginals $\mu_R(\underline{x}_R)$ of the true probability distribution $\psi(\underline{x})$, in general $\mathcal{F}_\mathcal{R}$ (2.22) is different from the free energy functional (2.4) computed on $\psi(\underline{x})$. Indeed in this case the average region-based energy $U_\mathcal{R}$ is equal to the average true energy, but in general the region-based entropy $S_\mathcal{R}$ is different from the true one. In Proposition 2.1 are discussed some examples in which $\mathcal{F}_\mathcal{R}$ is equal to $\mathcal{F}$.

Proposition 2.1. *If the distributions $\{P_R\}$ are equal to the corresponding exact marginal probabilities $\{\mu_R\}$, in formulas:*

$$P_R(\underline{x}_R) = \sum_{\underline{x}_{V\setminus R}} \psi(\underline{x}) \equiv \mu_R(\underline{x}_R) \quad \forall R \in \mathcal{R}, \tag{2.26}$$

then we can state the following:

(i) the average region-based energy:

$$U_{\mathcal{R}}(\{P_R\}) = \sum_{R \in \mathcal{R}} c_R E_R(P_R) \tag{2.27}$$

is exact, i.e., it is equal to the energy term of the free energy of the real distribution.

(ii) if all the regions R are disjoint, then also $S_{\mathcal{R}}$ is exact, i.e., it is equal to the entropy term of the free energy of the real distribution.

(iii) if the exact probability distribution $\psi(\underline{x})$ is uniform, then $S_{\mathcal{R}}$ is exact.

In summary the idea underling this approximation is to break up the factor graph into a set of regions that include every function and variable nodes, and to write the overall free energy as the sum of the free energies of all the regions, with a focus on subtracting out the excess contribution produced by the overlaps.

After choosing the appropriate set of regions, depending on which correlations we want to take into account, following the same path of the naïve MFA, the idea is to find the best guess for the probability distributions $\{P_R\}$ by minimizing the region-based free energy functional. However, if we want to approximate with $\{P_R\}$ the set of the corresponding true marginals of the model, there is a set of necessary constraints to impose:

- each P_R has the form of a probability function, then it must be normalized to one, and satisfy $0 \leq P_R(\underline{x}_R) \leq 1$ for any state $\underline{x}_R$;
- for every pair of regions R_1 and R_2 that share a set O of variable nodes, all the marginals in O must be consistent:

$$\sum_{\underline{x}_{R_1 \setminus O}} P_{R_1}(\underline{x}_{R_1}) = \sum_{\underline{x}_{R_2 \setminus O}} P_{R_2}(\underline{x}_{R_2}). \tag{2.28}$$

In general, increasing the size of the regions improves the approximation one obtains by minimizing the region-based free energy [YFW03]. As shown in Proposition 2.1 the energetic term is exact by construction, then there is an improvement arising in the entropic term, that becomes increasingly accurate as the regions become larger. In the limit where a single region covers all the nodes in the system clearly the region-based approximation becomes exact.

2.4.3 The Bethe approximation

The Bethe approximation (BA) belongs to the class of the region-based approximations. It is more refined than the naïve MFA, since it reintroduces correlations between variables entering the same interaction. In the BA we take the regions included in $\mathcal{R}$ to be of two different types: large regions and small regions. The set $\mathcal{R}_L$ of large regions is in one-to-one correspondence with the set of the function nodes, as each element belonging to $\mathcal{R}_L$ contain exactly one function node and all the variable nodes neighboring it. Otherwise $\mathcal{R}_S$, the set of small regions, coincides

with the set of the variable nodes of the graphs. The c_R's are given, for each region $R \in \mathcal{R}$, by:

$$c_R = 1 - \sum_{S \in \mathcal{S}(R)} c_S, \tag{2.29}$$

where $\mathcal{S}(R)$ is the set of all the regions S such that the set of variable and factor nodes in R are a subset of those in S. Moreover $\{P_R\}$ is composed by a set of distributions $P_a(\underline{x}_{\partial a})$, one for each function node of the graph, and by a set of distributions $P_i(x_i)$, one for each variable node. Then the region-based energy (2.22) in the Bethe case is:

$$U_{\text{Bethe}} = -\sum_{a \in F} \sum_{\underline{x}_{\partial a}} P_a(\underline{x}_{\partial a}) \log \psi_a(\underline{x}_{\partial a}), \tag{2.30}$$

and the Bethe entropy:

$$S_{\text{Bethe}} = -\sum_{a \in F} \sum_{\underline{x}_{\partial a}} P_a(\underline{x}_{\partial a}) \log P_a(\underline{x}_{\partial a}) - \sum_{i \in V} (1 - |\partial i|) \sum_{x_i} P_i(x_i) \log P_i(x_i). \tag{2.31}$$

The consistency constraints (subsection 2.4.2) become:

$$\sum_{x_i} P_i(x_i) = \sum_{\underline{x}_{\partial a}} P_a(\underline{x}_{\partial a}) = 1, \ \sum_{\underline{x}_{\partial a \setminus i}} P_a(\underline{x}_{\partial a}) = P_i(x_i), \ \forall i \in \partial a \text{ and } a \in F, \tag{2.32}$$

where the normalization condition on $P_a(\underline{x}_{\partial a})$ is automatically fixed by the others, and then can be omitted. In order to find the minimum of $\mathcal{F}_{\text{Bethe}}$, subject to the constraints (2.32), we define the Lagrangian $\mathcal{L}$:

$$\mathcal{L} = \mathcal{F}_{\text{Bethe}} + \sum_i \gamma_i \left[\sum_{x_i} P_i(x_i) - 1 \right] + \sum_{x_i} \sum_{(i,a) \in E} \lambda_i^a(x_i) \left[\sum_{\underline{x}_{\partial a \setminus i}} P_a(\underline{x}_{\partial a}) - P_i(x_i) \right], \tag{2.33}$$

where the superscript on the generic Lagrange multiplier λ is a function node index, and the subscript is a variable node index. Without taking into account the normalization conditions, every $P_a(\underline{x}_{\partial a})$ is fixed by $|\mathcal{X}|^{|\partial a|}$ parameters, and $P_i(x_i)$ by $|\mathcal{X}|$ ones. Then the extremality condition for (2.33) consists in setting to zero the derivatives of $\mathcal{L}$ with respect to the $M|\mathcal{X}|^{|\partial a|}$ possible values of the $P_a(\underline{x}_{\partial a})$'s, and to the $N|\mathcal{X}|$ possible values of the $\{P_i(x_i)\}$'s:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial P_a(\underline{x}_{\partial a})} = -\left[\log \dfrac{\psi_a(\underline{x}_{\partial a})}{P_a(\underline{x}_{\partial a})} + 1 \right] + \displaystyle\sum_{i \in \partial a} \lambda_i^a(x_i) \equiv 0, \\ \dfrac{\partial \mathcal{L}}{\partial P_i(x_i)} = (1 - |\partial i|) \left[\log P_i(x_i) + 1 \right] + \gamma_i - \displaystyle\sum_{a \in \partial i} \lambda_i^a(x_i) \equiv 0, \end{cases} \tag{2.34}$$

from which it follows:

$$P_a(\underline{x}_{\partial a}) = \psi_a(\underline{x}_{\partial a}) \exp\left(1 - \sum_{i \in \partial a} \lambda_i^a(x_i) \right), \tag{2.35}$$

$$P_i(x_i) = \exp\left(\frac{\sum_{a \in \partial i} \lambda_i^a(x_i) - \gamma_i}{1 - |\partial i|} - 1 \right) = \frac{1}{Z_i} \exp\left(- \frac{\sum_{a \in \partial i} \lambda_i^a(x_i)}{|\partial i| - 1} \right), \tag{2.36}$$

where Z_i, or equivalently γ_i, is fixed by the normalization condition on P_i, and the other Lagrange multipliers $\lambda_i^a(x_i)$ by the condition of local consistency:

$$\sum_{\underline{x}_{\partial a \setminus i}} \psi_a(\underline{x}_{\partial a}) \exp\left(1 - \sum_{i \in \partial a} \lambda_i^a(x_i)\right) = \frac{1}{Z_i} \exp\left(-\frac{\sum_{a \in \partial i} \lambda_i^a(x_i)}{|\partial i| - 1}\right) \tag{2.37}$$

Equation 2.37 defines a set of $|E|(|\mathcal{X}|-1)$ coupled equations that fix the $|E|(|\mathcal{X}|-1)$ $\lambda_i^a(x_i)$ multipliers.

There exists a particular class of graphical model whose topology alone guarantees the exactness of the Bethe region-based decomposition of the free energy: the class of tree graphical models, i.e., graphical models defined on tree graphs. The fact that Equation 2.30 and Equation 2.31 define the true free energy on a tree is a direct consequence of the following property:

Proposition 2.2. *In a tree graphical model the joint probability distribution $\psi(\underline{x})$ of all the variables can be written in terms of the marginals $\mu_a(\underline{x}_{\partial a})$ and $\mu_i(x_i)$ as:*

$$\psi(\underline{x}) = \prod_{a \in F} \mu_a(\underline{x}_{\partial a}) \prod_{i \in V} \mu_i(x_i)^{1-|\partial i|} \tag{2.38}$$

Proof. Equation 2.38 can be proved by induction on the cardinality $|F| \equiv M$ of the set of function nodes. For $M = 1$, just one interaction, $|\partial i| = 1\ \forall i$, and then $\psi(\underline{x})$ is equal to its marginal on the set of the variable nodes linked to the only function node: Equation 2.38 holds. Now assume that the property is valid for any tree graphical model with $|F| \leq M$, and consider a specific factor graph T' with $|F| = M + 1$. T' can be constructed attaching a single function node a to a factor graph with M function nodes. The fundamental observation is that the only way to attach a function node to a tree, with the constraint of preserving in the final graph the tree topology, is by connecting it to a single variable node i, conversely there would appear cycles. Denoting by $\underline{x}$ the set of variables of the T tree, and by $\underline{x}'$ those of T', it follows that:

$$\psi_{T'}(\underline{x}') = \psi_T(\underline{x})\psi(\underline{x}_{\partial a \setminus i}|x_i) = \psi_T(\underline{x})\frac{\mu_a(\underline{x}_{\partial a})}{\mu_i(x_i)}, \tag{2.39}$$

that is a consequence of the fact that $\underline{x}_{\partial a \setminus i}$ and $\underline{x}$ are conditionally independent variables. Using this result together with the induction hypothesis we obtain:

$$\psi_{T'}(\underline{x}') = \prod_{b \in F(T)} \mu_b(\underline{x}_{\partial b}) \prod_{\substack{j \in V(T), \\ j \neq i}} \mu_j(x_j)^{1-|\partial j|} \mu_i(x_i)^{1-\left(|\partial i|-1\right)} \frac{\mu_a(\underline{x}_{\partial a})}{\mu_i(x_i)}, \tag{2.40}$$

that is equal to Equation 2.38. □

Thanks to Proposition 2.2, the energetic and the entropic terms of the exact free energy,

$$\mathcal{F}[\psi(\underline{x})] \equiv -\sum_{\underline{x}} \psi(\underline{x}) \log \prod_{a \in F} \psi_a(\underline{x}_{\partial a}) + \sum_{\underline{x}} \psi(\underline{x}) \log \psi(\underline{x}), \tag{2.41}$$

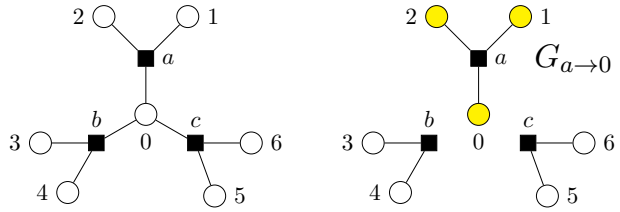


Figure 2.3. On the left an example of small tree. On the right the subtree $G_{a\to 0}$ is represented with yellow leaves.

can be rewritten, respectively, as:

$$\begin{aligned}
&\sum_{\underline{x}} \psi(\underline{x}) \log \prod_{a\in F} \psi_a(\underline{x}_{\partial a}) = \sum_{a\in F} \sum_{\underline{x}_{\partial a}} \mu_a(\underline{x}_{\partial a}) \log \psi_a(\underline{x}_{\partial a}), \\
&\sum_{\underline{x}} \psi(\underline{x}) \log \psi(\underline{x}) = \sum_{\underline{x}} \psi(\underline{x}) \left(\sum_{a\in F} \log \mu_a(\underline{x}_{\partial a}) + \sum_{i\in V} (1-|\partial i|) \log \mu_i(x_i) \right).
\end{aligned} \tag{2.42}$$

Substituting (2.41) in (2.42) one obtains:

$$\mathcal{F}[\psi(\underline{x})] = -\sum_{a\in F} \sum_{\underline{x}_{\partial a}} \mu_a(\underline{x}_{\partial a}) \log \frac{\psi_a(\underline{x}_{\partial a})}{\mu_a(\underline{x}_{\partial a})} + \sum_{i\in V} \sum_{x_i} (1-|\partial i|) \mu_i(x_i) \log \mu_i(x_i), \tag{2.43}$$

that proves the exactness of the BA for models defined on trees, since (2.43) is equal to the free energy derived with the BA (2.30, 2.31).

It is possible to derive on trees an iterative algorithm (section 2.4.4) known as Belief propagation (BP) that, in time linear in the number of nodes, computes the exact marginals of the graphical model distribution, provided the factor nodes have bounded degree, and the alphabet size is bounded as well [MM09]. In other words when there are no cycles, i.e., on trees, BP finds the global minima of the Bethe free energy. Moreover it turns out that a set $\{P_R\}$ is a fixed point of the iterative procedure prescribed by BP algorithm in any graph if and only if it is a local stationary point of the Bethe free energy [YFW03].

In section 2.4.4 belief propagation is introduced with the aim of putting the Bethe approximation in a different perspective. At the end a less algorithmic approach is discussed, that allows one to study systems defined on random graphical models.

2.4.4 Cavity Method

Belief Propagation

A typical problem to solve on a graphical model is the computation of the marginal distributions of a set of variables: this allows, for example, to compute the free energy in the Bethe approximation. When the factor graph has no cycles, such kind of computation can be performed efficiently by a crucial property: the distributivity of the sum with respect to the product [AM00]. Let us see how to exploit this basic property on the simple example reproduced in Figure 2.3. We begin by assuming

that we want to compute the marginal at node 0 on the tree G:

$$\mu(x_0) \propto \sum_{\underline{x}_{V\setminus 0}} \prod_{\ell \in F} \psi_\ell(\underline{x}_{\partial \ell}). \tag{2.44}$$

Node 0 is the root of three subtrees of G, that we can distinguish by the name of the factor node neighbor of 0 that they contain, namely $G_{a\to 0} = (V_{a\to 0}, F_{a\to 0}; E_{a\to 0})$, $G_{b\to 0} = (V_{b\to 0}, F_{b\to 0}; E_{b\to 0})$, $G_{c\to 0} = (V_{c\to 0}, F_{c\to 0}; E_{c\to 0})$, as shown in Figure 2.3. We can rewrite the right side of Equation 2.44 as:

$$\begin{aligned}
&\sum_{\underline{x}_{V\setminus 0}} \prod_{\ell \in F_{a\to 0}} \psi_\ell(\underline{x}_{\partial \ell}) \prod_{m \in F_{b\to 0}} \psi_m(\underline{x}_{\partial m}) \prod_{n \in F_{c\to 0}} \psi_n(\underline{x}_{\partial n}) \\
&\propto \left\{ \sum_{\underline{x}_{V_{a\to 0}\setminus 0}} \prod_{\ell \in F_{a\to 0}} \psi_\ell(\underline{x}_{\partial \ell}) \right\} \left\{ \sum_{\underline{x}_{V_{b\to 0}\setminus 0}} \prod_{m \in F_{b\to 0}} \psi_m(\underline{x}_{\partial m}) \right\} \left\{ \sum_{\underline{x}_{V_{c\to 0}\setminus 0}} \prod_{n \in F_{c\to 0}} \psi_n(\underline{x}_{\partial n}) \right\} \\
&\propto \mu_{a\to 0}(x_0)\mu_{b\to 0}(x_0)\mu_{c\to 0}(x_0).
\end{aligned} \tag{2.45}$$

The second line is obtained from the first by using the distributive property. In the last step we denoted by $\mu_{a\to 0}(x_0)$, $\mu_{b\to 0}(x_0)$ and $\mu_{c\to 0}(x_0)$ the marginals with respect to the factor graphs, respectively, $G_{a\to 0}$, $G_{b\to 0}$ and $G_{c\to 0}$. The problem of computing a marginal with respect to G is then reducible to the one of computing marginals with respect to subgraphs of G. This procedure can be repeated recursively. Consider the subtree $G_{a\to 0}$. This can be decomposed into the factor node a, plus the subtrees $G_{1\to a}$ and $G_{2\to a}$. Thanks to the distributive property again, it follows that:

$$\begin{aligned}
\mu_{a\to 0}(x_0) &\propto \sum_{\underline{x}_{V_{a\to 0}\setminus 0}} \prod_{\ell \in F_{a\to 0}} \psi_\ell(\underline{x}_{\partial \ell}) \\
&\propto \sum_{x_1, x_2} \psi_a(\underline{x}_{\partial a}) \left\{ \sum_{\underline{x}_{V_{1\to a}\setminus 1}} \prod_{m \in F_{1\to a}} \psi_m(\underline{x}_{\partial m}) \right\} \left\{ \sum_{\underline{x}_{V_{2\to a}\setminus 2}} \prod_{n \in F_{2\to a}} \psi_n(\underline{x}_{\partial n}) \right\} \\
&\propto \sum_{x_1, x_2} \psi_a(\underline{x}_{\partial a}) \nu_{1\to a}(x_1)\nu_{2\to a}(x_2),
\end{aligned} \tag{2.46}$$

where the function $\nu_{1\to a}(x_1)$ and $\nu_{2\to a}(x_2)$ are defined by Equation 2.46 with the prescription of normalization to one. These arguments did not use the specific structure of the factor graph in Figure 2.3, but they instead hold for any tree. Namely given a tree G, and a directed edge $a \to i$ (factor-to-variable) or $i \to a$ (variable-to-factor) we can define the subgraphs $G_{a\to i}$ or $G_{i\to a}$ as above and the corresponding so-called "*cavity marginals*" of the variable i: $\mu_{a\to i}(x_i)$ and $\nu_{i\to a}(x_i)$.

Definition 2.3 (cavity marginals)**.** The functions $\mu_{a\to i}(x_i)$ and $\nu_{i\to a}(x_i)$ are cavity marginals, i.e., marginal probability distributions on *cavity graphs*. This means that $\mu_{a\to i}(x_i)$ would be the true marginal of x_i if a hole were dug in the starting graph, by removing from the definition of the joint probability distribution that defines the model all the $\{\psi_b\}_{b\in \partial i\setminus a}$'s. Analogously $\nu_{i\to a}(x_i)$ is a cavity marginal because it would be the true marginal of x_i in a graph obtained from the starting one by removing the function node labeled with a.

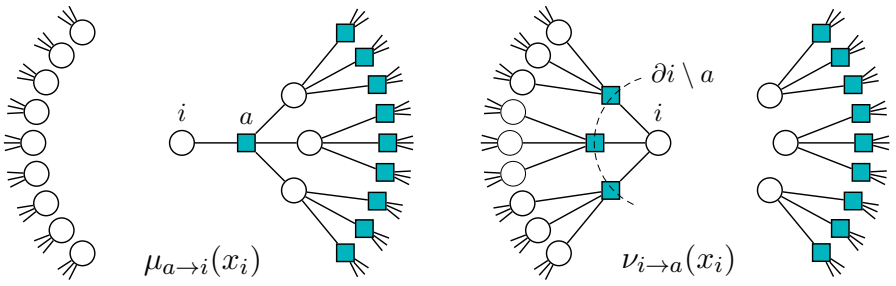


Figure 2.4. The marginal distribution of the variable i is $\mu_{a\to i}(x_i)$ on the cavity graph on the left, and $\nu_{i\to a}(x_i)$ on that on the right. The dashed line on the graph on the right is used to underline the neighborhood of i.

From Equation 2.45 and Equation 2.46, rewritten for a generic tree G and generic nodes, it follows that the cavity marginals must satisfy the following equations, that are called *Belief propagation* equations:

$$\begin{aligned} \nu_{i\to a}(x_i) &\propto \prod_{b\in\partial i\setminus a} \mu_{b\to i}(x_i) \\ \mu_{a\to i}(\underline{x}_i) &\propto \sum_{\underline{x}_{\partial a\setminus i}} \psi_a(\underline{x}_{\partial a}) \prod_{k\in\partial a\setminus i} \nu_{k\to a}(x_k). \end{aligned} \qquad \forall\,(i,a)\in E \tag{2.47}$$

Once the cavity marginals are known, the marginal of a variable x_i is simply:

$$\mu(x_i) \propto \prod_{b\in\partial i} \mu_{a\to i}(x_i). \tag{2.48}$$

It is worth noting that the cavity marginals are the unique solution of the Equation 2.47, from the recursive argument discussed above.

We have derived Equation 2.47 for a generic tree, but one may ask what happens if we use the cavity marginals, and in particular Equation 2.48, to estimate the true marginals of a model defined on a generic factor graph with cycles. Firstly in the general case it has to be discussed how to solve Equation 2.47. The main way to solve BP equations is by an iterative procedure, defined by the *Belief propagation algorithm*. Belief propagation is an iterative "message-passing" algorithm: the messages are probability distributions that take value in the space of all possible probability distributions over the single-variable alphabet $\mathcal{X}$, and are sent from one node to another, connected to them by a link. For each edge (i,a), where i is a variable node and a a function node, at the *t-th* iteration are defined two messages: one for the variable-to-function direction $\nu^{(t)}_{i\to a}(x_i)$, and another message $\mu^{(t)}_{a\to i}(x_i)$ for the other direction. The messages are updated over time according to the following rule:

$$\begin{aligned} \nu^{(t+1)}_{i\to a}(x_i) &\propto \prod_{b\in\partial i} \mu^{(t)}_{b\to i}(x_i) \\ \mu^{(t+1)}_{a\to i}(\underline{x}_i) &\propto \sum_{\underline{x}_{\partial a\setminus i}} \psi_a(\underline{x}_{\partial a}) \prod_{k\in\partial a\setminus i} \nu^{(t)}_{k\to a}(x_k), \end{aligned} \qquad \forall\,(i,a)\in E. \tag{2.49}$$

When $\partial i \setminus a$ is an empty set, $\nu_{j\to a}(x_j)$ is chosen as the uniform distribution. Similarly if $\partial a \setminus j$ is empty then $\mu_{a\to j}(x_j) = \psi_a(x_j)$. In tree-graphical model the following theorem holds [MM09].

Theorem 2.4. *Consider a tree-graphical model such that the distance between any two nodes is at most t_*, then:*

1. *irrespective of the initial condition, the BP update equations (2.49) converge after at most t_* iterations: for any edge (i,a), and any $t > t_*$, $\nu^{(t)}_{i\to a} = \nu^*_{i\to a}$, $\mu^{(t)}_{i\to a} = \mu^*_{i\to a}$;*
2. *the fixed point messages provide the exact marginals.*

Then, by exploiting the distributive property on trees, it is possible to compute marginals in a time $T = \mathcal{O}(t^*)$, instead of spending $T = \mathcal{O}\left(|\mathcal{X}|^N\right)$, where N is the number of variable nodes, for a brute force summation of Equation 2.44. Moreover, even on factor graphs with cycles, often BP estimates get to be fairly accurate. For example in the case of the assignment problem on generic bipartite graph, BP has been proven to yield exact results [BSS05]. However for graphical models with cycles BP is not guaranteed to converge in general. Then, because the BP fixed points correspond to Bethe free energy minima as stated in subsection 2.4.3, one could simply choose to minimize the Bethe free energy directly. Such free energy minimization are slower than BP algorithm, but they are at least guaranteed to converge [YFW03]; on the other hand, empirical analysis indicates that the failure of BP convergence is a clue that the results from minimizing the Bethe approximation will also be quite inaccurate [WT01, YFW03].

Random Graphical Models: the random-instance BP

The Belief propagation algorithm can be run on the factor graph associated with a given graphical model, in order to try to compute the fixed point messages of Equation 2.47. However, in chapter 5, we will be interested in the study of a graphical model not for a single realization of the problem, but for an ensemble of possible instances. Then, for example, there will not be a given factor graph, but an ensemble of factor graphs, over which one is interested in compute average properties of the solution. We will introduce the concept of disorder in the next chapter. For the moment, let us simply suppose that we want to average our results on a graph ensamble and, possibly, on some additional randomness in parameters appearing in the messages. Since in this frame the messages themselves become random variables, their probability distribution become the quantity of interest [MM09]. This distribution can be characterized in the large system limit starting from the BP equations Equation 2.47, as we shall now see.

Firstly a *random graphical model* is an ensemble of probability distributions on $x = (x_1, ..., x_N)$ which have the form:

$$\psi(\underline{x}) = \prod_{a\in F} \psi_a(\underline{x}_{\partial a}), \tag{2.50}$$

where the factor graph is denoted as usual as $G = (V, F; E)$ and the factors ψ_a possibly depend on some random parameter J. For any possible factor-node degree $|\partial a| = k$, we are given a list of feasible factors $\psi^k(x_1, ..., x_k; J)$, marked by a parameter J, and a distribution P_J^k over the set of all possible realization of J. Once the degree is extracted according with the degree distribution of the ensemble, the corresponding parameter is drawn with distribution P_J^k, and the function $\psi_a(\cdot)$ is taken to be equal to $\psi^k(\,\cdot\,; J_a)$. Then the random graphical model is fully characterized by the graph ensemble, that fixes the distribution of the variable and function nodes degrees, the set of distributions P_J^k, and the list of possible factors, denoted by $\{\psi^k(\,\cdot\,; J_a)\}$.

Now let us indicate the BP messages updating rules (2.47) for a given realization of the random graphical model in a compact way as

$$\begin{aligned} \Psi_{i\to a}(\{\mu_{b\to i}^{(t)} : b \in \partial i \setminus a\}) &= \Psi_{i\to a}(\mu_1^{(t)}, ..., \mu_{|\partial a|-1}^{(t)}) \\ \Phi_{a\to i}(\{\nu_{j\to b}^{(t)} : j \in \partial a \setminus i\}) &= \Phi_{a\to i}(\nu_1^{(t)}, ..., \nu_{|\partial i|-1}^{(t)}). \end{aligned} \tag{2.51}$$

Even if not explicitly indicated, both updating rules in Equation 2.51 depend on the realization of the parameters $\{J\}$. For locally tree like graph ensembles it can be proved the following result:

Theorem 2.5 (Density evolution)**.** *Let $t \geq 0$ and let (i, a) be a uniformly random edge in the factor graph. Then, as $N \to \infty$, the messages $\mu_{a\to i}^{(t)}$ and $\nu_{i\to a}^{(t)}$ converge in distribution to the random variables $\mu^{(t)}$ and $\nu^{(t)}$, respectively, defined through the following so-called density evolution equations* [MM09, DMS13]:

$$\begin{aligned} \nu^{(t+1)} &\stackrel{d}{=} \Psi_\ell(\mu_1^{(t)}, ..., \mu_{\zeta-1}^{(t)}), \\ \mu^{(t+1)} &\stackrel{d}{=} \Phi_k(\nu_1^{(t)}, ..., \nu_{\kappa-1}^{(t)}). \end{aligned} \tag{2.52}$$

Here the symbol $y \stackrel{d}{=} z$ indicates that y and z are equal in distribution. Moreover the messages $\mu_b^{(t)}$, with $b \in \{1, ..., \ell\}$, are independent random copies of $\mu^{(t)}$, and $\nu_j^{(t)}$, with $j \in \{1, ..., \kappa\}$, are independent random copies of $\nu^{(t)}$. If the degree distribution of the variable nodes is p_z and the degree distribution of the factor nodes is q_k, then the degree distribution of ℓ and κ are

$$\hat{p}_\zeta = \frac{\zeta p_\zeta}{\sum_{z=1}^\infty z p_z}, \quad \hat{q}_\kappa = \frac{\kappa q_\kappa}{\sum_{k=1}^\infty k q_k} \tag{2.53}$$

The numerical solution of the density evolution equations (2.52) takes a crucial role, as will be shown in chapter 5 in our study of the matching problem, because they allow to compute observables of the random graphical model.

Chapter 3

Disordered Systems

In this chapter some basic concepts about disordered systems are presented, in order to lay the ground for the discussion about the connection between statistical mechanics and combinatorial optimization in the next chapters.

3.1 Disorder, frustration and self-averageness

In Section 2.2 we introduced the Ising model, in which the Hamiltonian of the system is fixed by the coupling constants J, and by a set of external fields h_i acting on the spins. When the external field is site independent the problem is considerably simplified by the translational invariance of the system. Unfortunately this symmetry is not realized in many real cases due to the presence of *disorder*: vacancies in the arrangement of atoms, competition between ferromagnetic and antiferromagnetic interactions and lattice irregularities are just a few examples. Some types of disorder can be modelized by considering an Hamiltonian in which a set of random parameters $\{J_1, J_2, ..., J_M\} \equiv \{J\}$ is introduced:

$$\mathcal{H} = \mathcal{H}_J\left(\underline{\sigma}\right). \tag{3.1}$$

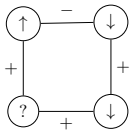
As in Equation 3.1, each quantity that depends on the parameters $\{J\}$ will be denoted with a subscript J. In Equation 3.1 $\underline{\sigma}$ specifies the configuration of the system, and the generic $J_k \in \{J\}$ is randomly extracted from a certain probability distribution that specifies the disorder. A remarkable case is that of *quenched disorder*, in which disorder can be considered fixed on the time scale over which the degrees of freedom of the system fluctuate: the thermodynamic properties of the system have to be computed at a fixed instance of the disorder. In the following discussions we will be interested in this kind of quenched-disorder calculation.

A key feature of disorder is that it is a possible source of *frustration*: it becomes impossible to satisfy all the couplings at the same time, as it would be in a ferromagnetic system; in other words there exist a set of local constraints in conflict each other. A paradigmatic example can be found in the *Edward-Anderson* model (EAM). The EAM belongs to the class of spin glass models, i.e. the class of spin systems

with random interactions. Its Hamiltonian can be written in the form:

$$\mathcal{H} = \sum_{(i,j)} J_{ij}\sigma_i\sigma_j, \tag{3.2}$$

where $\sigma_i \in \{-1, 1\}$, the sum is performed over the nearest neighbors of an hypercubic lattice in D dimension, and J_{ij} are independent and identical distributed random variables. In this case the following situation can arise:



(3.3)

where the "+" and "−" signs next to the edges represent the signs of the couplings between spins. In Figure 3.3 it is clear that a configuration of the spin on the left that minimizes at the same time the energy contribution of all edges separately does not exist. Then in a frustrated system the ground state cannot be found by a search that minimizes locally all the constraints as, for example, in the translational invariant Ising model. This fact in general makes considerably more difficult the search for the ground state configuration, and the ground state energy.

Self-averageness

By constructing an Hamiltonian with a set of random parameters $\{J\}$, as in Equation 3.2, for a generic size of the system each observable depends on the realization of $\{J\}$, including, for example, the free energy density:

$$\frac{\mathcal{F}_J^{(N)}}{N} \equiv f_J^{(N)} = -\frac{1}{\beta N}\log\sum_{\{J\}} e^{-\beta\mathcal{H}_J(\underline{\sigma})}, \tag{3.4}$$

where N is the size of the system. As it is utterly impossible to measure the precise realization of the quenched parameters on a physical sample, the idea is to average over all possible realizations, with the aim of trying to extract informations representative of all the samples. Then for example, one tries to compute the expected value over disorder of $f_J^{(N)}$,

$$f^{(N)} \equiv \overline{f_J^{(N)}} = -\frac{1}{\beta N}\overline{\log \mathcal{Z}_{\mathcal{J}}}, \tag{3.5}$$

where the bar denotes the average over disorder. It turns out that, if the form of the interaction is not too pathological, in the thermodynamic limit the free energy density assumes the same value for each set of quenched parameters which has non vanishing probability [MPV87]. This property is called *self-averageness* and in formulas reads:

$$\lim_{N\to\infty} \Pr\left(\left|f_J^{(N)} - \overline{f_J^{(N)}}\right| > \epsilon\right) = 0, \quad \text{for } \epsilon > 0, \tag{3.6}$$

therefore sample-to-sample fluctuations go to zero in the limit of large system, and then one can expect that a theoretical calculation of the mean value of f_J over the

whole ensemble gives the same answer as the experiments. For models with short range interactions a general argument allows to argue that the sample-to-sample fluctuations are proportional to $1/\sqrt{N}$:

$$\overline{\left(f_J^{(N)} - f^{(N)}\right)^2} = \mathcal{O}\left(\frac{1}{N}\right). \tag{3.7}$$

Indeed let us divide our d-dimensional system in a number of macroscopic sub-systems. The total free energy will be the sum of the free energies of the sub-systems, plus a contribution coming from the interactions at the interfaces between the sub-systems, that, in the large system size limit, scale, respectively, as N and $N^{\frac{d-1}{d}}$. Once we compute the free energy density, the surface contribution can be neglected, and the sub-systems free energies are independent random variables: therefore we can apply the central limit theorem to the sum to get the scaling in Equation 3.7 [FH93, CC05].

3.2 The replica trick

The computation of observables for disordered systems requires therefore to average over the quenched disorder the logarithm of the sample-dependent partition function (3.5). A famous and powerful technique, called *replica trick*, is usually adopted for this task. It consists in computing this average by an analytic continuation of the average of the partition function of n uncoupled *replicas* of the initial system [MPV87].

Let us firstly denote by $\mathcal{P}(\{J\})$ the probability distribution of the quenched parameters, and introduce the following quantities:

$$\mathcal{Z}_n = \sum_{\{J\}} \mathcal{P}(\{J\}) (\mathcal{Z}_J)^n, \qquad f_n = -\frac{1}{\beta n N} \log \mathcal{Z}_n, \tag{3.8}$$

where $(\mathcal{Z}_J)^n$ is the average over disorder of the *replicated* partition function. At this point the fundamental observation is that:

$$\lim_{n\to 0} f_n \equiv f_0 = \overline{f_J}, \tag{3.9}$$

where the normalization condition of the probability distribution have been used, and the relation:

$$x^n \approx 1 + n \log x \quad \text{for} \quad n \approx 0 \quad \text{and} \quad x > 0. \tag{3.10}$$

The replica trick consists in computing f_n for integer n, using the fact that in this case we can write:

$$(\mathcal{Z}_J)^n = \sum_{\underline{\sigma}^1} \sum_{\underline{\sigma}^2} \cdots \sum_{\underline{\sigma}^n} \exp\left\{ -\sum_{a=1}^{n} \beta \mathcal{H}_J(\underline{\sigma}^a) \right\}, \tag{3.11}$$

this allows to obtain f_n for integer n. The next step is to extent n to an analytic function of n, and finally to compute $f_0 = f$.

This method finds a paradigmatic application in the solution of the infinite range spin glass, the *Sherrington-Kirkpatrick* model (SKM), defined by the Hamiltonian:

$$\mathcal{H}_J = \sum_{i<j} J_{ij} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i, \tag{3.12}$$

where $\sigma_i \in \{-1, 1\}$, h is a uniform magnetic field, and the couplings are random variables that, just to be concrete, we can extract from a Gaussian distribution of the form:

$$\mathcal{P}(\{J\}) = \sqrt{\frac{N}{2\pi}} e^{-N\frac{J^2}{2}}. \tag{3.13}$$

In Equation 3.13 the variance has been taken equal to $\frac{1}{N}$ in order to guarantee the extensivity of the energy; in fact in this case all the couples of variables interact, and then the term of interaction between spins in $\mathcal{H}_J$ has $\mathcal{O}(N^2)$ addends. It can be found [Nis01] that the average replicated partition function can be written as

$$\mathcal{Z}_n = \exp\left(\frac{\beta^2 n N}{4}\right) \int \prod_{a<b} dq_{ab} \exp\left(-\frac{\beta^2 N}{2} \sum_{a<b} q_{ab}^2 + N \log \Xi[\mathbf{q}]\right), \tag{3.14}$$

where:

$$\Xi[\mathbf{q}] = \sum_{\sigma^1,\dots,\sigma^n} \exp\left(\beta^2 \sum_{a<b} q_{ab}\sigma^a\sigma^b + \beta h \sum_a \sigma^a\right), \tag{3.15}$$

and $\mathbf{q}$ is an $n \times n$ symmetric matrix with elements $(\mathbf{q})_{ab} = q_{ab}$. The exponent of the integrand in Equation 3.14 is proportional to N, so that it is possible to evaluate the integral by the Laplace method [Mil06].
In SKM the saddle point condition $\forall a, b$ with $a \neq b$ reads:

$$q_{ab} = \frac{1}{\beta^2}\frac{\partial \log \Xi[\mathbf{q}]}{\partial q_{ab}} = \frac{1}{\Xi} \sum_{\sigma^1,\dots,\sigma^n} \sigma^a\sigma^b \exp\left(\beta^2 \sum_{a<b} q_{ab}\sigma^a\sigma^b + \beta h \sum_a \sigma^a\right). \tag{3.16}$$

The simplest way to try to find a solution of Equation 3.32 is through the *replica-symmetric* (RS) assumption, which consists in considering all the replicas equivalent:

$$q_{ab} \equiv q \;\; \forall a, b \;\text{ with }\; a \neq b. \tag{3.17}$$

This assumption is motivated by the fact that the function at the exponent of the integrand in Equation 3.14 is left invariant when we exchange some of the lines or the rows of the matrix $\mathbf{q}$, and therefore the group of permutations of n elements is a symmetry of the problem. However this approach gives results at variance with the computer simulations below a certain temperature [KS78]; at $T = 0$ for example one obtains a negative entropy, when the entropy of the model must be non negative by definition. The origin of this pathological behavior was identified by Parisi [Par83], who showed in particular that an infinite-step sophisticated breaking of the replica symmetry is needed. For the sake of brevity, and because the problem we are interested in show no replica symmetry breaking, we will not discuss here Parisi's solution [MPV87].

3.3 A cavity calculation for the RS solution of the SKM

It is possible to derive the replica-symmetric solution of the SKM through a cavity calculation. Following the presentation of [Méz15], in this section this approach is

presented in the case of zero external magnetic field, $h = 0$, in order to give a first example of application of the cavity method.

The basic idea is to go from a SK system Σ_N composed of N spins to a Σ_{N+1} system that has $N + 1$ spins, assuming that in the thermodynamic limit there is no difference between observables computed in both systems. Let us denote by σ_0 the spin added to the system of N spins to create Σ_{N+1}, and by J_{0j} the coupling connecting it to the generic spin σ_j in Σ_{N+1}. If we write the set of J_{0i} as $\{J_0\}$, and that of J_{ij}, with $i, j \neq 0$, as $\{J\}$, the probability distributions of disorder in Σ_N and Σ_{N+1} read, respectively:

$$\begin{aligned} \mathcal{P}_N(\{J\}) &= \prod_{i<j} P_N(J_{ij}) \propto \exp\left(-\frac{N}{2}\sum_{i<j} J_{ij}^2\right) \\ \mathcal{P}_{N+1}(\{J\},\{J_0\}) &= \prod_{j,i<j} P_{N+1}(J_{ij},J_{0j}) \propto \exp\left(-\frac{N+1}{2}\left[\sum_{i<j} J_{ij}^2 + \sum_j J_{0j}^2\right]\right), \end{aligned} \tag{3.18}$$

and the probability distribution of a certain configuration of spins in Σ_{N+1} is given by the canonical distribution:

$$\psi_{N+1}(\underline{\sigma},\sigma_0) \propto \exp\left(-\beta\mathcal{H}_N(\underline{\sigma}) + \beta\sigma_0\sum_{j=1}^{N} J_{0j}\sigma_j\right), \tag{3.19}$$

where $\underline{\sigma} = \{\sigma_1, ..., \sigma_N\}$, and $\sum_j J_{0j}\sigma_j \equiv \phi^c$ is the so-called cavity field acting on σ_0.

It is useful at this point to find the probability distribution of ϕ^c, a computation that can be done by evaluating all its the moments. The working assumption of the cavity method at the RS level can be stated by saying that the susceptibility:

$$\chi = \frac{1}{N}\sum_{i<j}(\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle) \tag{3.20}$$

has to be finite. This is equivalent to say that two randomly chosen spins are uncorrelated with probability that tends to one in the thermodynamic limit, as could happen if the system were defined on a tree (subsection 2.4.4). By this hypothesis we start with the computation of the probability distribution of the cavity field in Σ_N, $\mathcal{P}_N(\phi^c)$: it is worth noting that in this case ϕ^c can be defined as the field measured at site 0 once σ_0 has been removed from Σ_{N+1}. Let us consider:

$$\langle\phi^c\rangle_N = \sum_i J_{0i}\langle\sigma_i\rangle_N \xrightarrow{N\to\infty} \phi, \tag{3.21}$$

$$\langle(\phi^c)^2\rangle_N - \langle\phi^c\rangle_N^2 = \sum_{i,j} J_{0i}J_{0j}(\langle\sigma_i\sigma_j\rangle_N - \langle\sigma_i\rangle_N\langle\sigma_j\rangle_N), \tag{3.22}$$

where the thermal average has been denoted by angle brackets, and ϕ has been used to indicate the thermodynamic limit of $\langle\phi^c\rangle_N$. Because the sum in Equation 3.20 involves $\mathcal{O}(N^2)$ terms, χ will be finite as long as the connected correlation of σ_i and

σ_j, namely, $\langle\sigma_i\sigma_j\rangle_N - \langle\sigma_i\rangle_N\langle\sigma_j\rangle_N$, is of order $1/N$. Then the sum in Equation 3.22 will be dominated by the terms with $i = j$, and for N large holds:

$$\begin{aligned} \text{for } i = j: \quad \left\langle (\phi^c)^2 \right\rangle_N - \langle \phi^c \rangle_N^2 &= \sum_i J_{0i}^2 \left(1 - \langle\sigma_i\rangle_N^2\right) \\ &= 1 - \frac{1}{N}\sum_i \langle\sigma_i\rangle_N^2 = 1 - \overline{\langle\sigma_i\rangle_N^2} = 1 - q, \end{aligned} \tag{3.23}$$

where in the second equality we have used $J_{ij} \sim 1/\sqrt{N}$, while in the third we have substituted the sum over all sites with the average over the disorder at a single site, due to the self-averageness of the magnetization. Finally, we have used the definition of the so-called Edwards-Anderson order parameter $\overline{\langle\sigma_i\rangle_N^2} = q$. Using similar reasoning one finds that all odd moments bigger than the first are zero, while all even moments are given by the expression:

$$\left\langle (\phi^c)^{2p} \right\rangle = (2p-1)!!(1-q)^p, \tag{3.24}$$

that are the moments of a gaussian distribution with variance $1 - q$. Therefore the probability distribution of the cavity field is:

$$\mathcal{P}_N(\phi^c) \propto \exp\left(-\frac{(\phi^c - \phi)^2}{2(1-q)}\right). \tag{3.25}$$

Now let us consider the joint probability distribution of ϕ^c and σ_0 in the Σ_{N+1} system:

$$\mathcal{P}_{N+1}(\phi^c, \sigma_0) \propto \exp\left(-\frac{(\phi^c - \phi)^2}{2(1-q)} + \beta\phi^c\sigma_0\right). \tag{3.26}$$

The expectation value of the spin σ_0 in the Σ_{N+1} system is:

$$\langle\sigma_0\rangle_{N+1} = \tanh(\beta\phi) = \tanh\left(\beta\sum_{i=1}^{N} J_{0i}\langle\sigma_i\rangle_N\right), \tag{3.27}$$

where the average is taken with respect to the probability density 3.26, integrating over the cavity field ϕ^c. It is interesting to note that the cavity field is not a self-averaging quantity, in fact ϕ has non-vanishing fluctuations. By computing its moments, as before, it is possible to derive the probability distribution $P(\phi)$ of ϕ. The averaged field is:

$$\overline{\phi} = \sum_i \overline{J_{0i}\langle\sigma_i\rangle_N} = 0, \tag{3.28}$$

which is equal to zero because the average of the couplings J is zero. The average squared field reads:

$$\overline{\phi^2} = \sum_{ij} \overline{J_{0i}J_{0j}\langle\sigma_i\rangle_N\langle\sigma_j\rangle_N} = \frac{1}{N}\sum_i \langle\sigma_i\rangle_N^2 = q. \tag{3.29}$$

By computing all the higher-order moments, it is possible to show that all the odd moments are zero, while all the even ones obey a similar relation to that seen in Equation 3.24. We can thus conclude that ϕ is Gaussian-distributed:

$$\mathcal{P}(\phi) = \frac{1}{\sqrt{2\pi q}}\exp\left(-\frac{\phi^2}{2q}\right). \tag{3.30}$$

Moreover, from the definition of q:

$$q = \overline{\langle \sigma_0 \rangle^2}_{N+1}, \tag{3.31}$$

by using Equation 3.30 and Equation 3.26 it follows the self-consistent equation for the order parameter q:

$$q = \int \frac{d\phi}{\sqrt{2\pi q}} \exp\left(-\frac{\phi^2}{2q}\right) \tanh(\beta\phi)^2. \tag{3.32}$$

It turns out that the replica trick at the replica-symmetric level is equivalent to the cavity method introduced in subsection 2.4.4. For this reason one often refers to the cavity method of subsection 2.4.4 as replica-symmetric cavity method.

Chapter 4

Random Matching Problems and Statistical Mechanics

In this chapter, a class of matching problems in the presence of disorder is introduced. In particular, we review some results already known in the literature, obtained by using techniques of statistical mechanics of disordered systems.

Optimization and disorder

In chapter 1 some optimization problems defined on graphs have been introduced. Given a definite instance of a certain problem, one can search for the optimal solution by running specific algorithms. For example, given a bipartite graph with parts of the same cardinality, and a set of positive weights to associate with its edges, one can use the Hungarian algorithm to find the optimal assignment (section 5.4.1).

Instead of solving single inputs, it may also be interesting to equip the set of the instances of a given optimization problem with a probability law, and then to study the average properties of the problem over such set. Besides its intrinsic interest, this *random-instance* approach can also give some insight into the qualitative understanding of the problem and suggest new algorithms and heuristics [MPV87]. Statistical mechanics has played a central role in this kind of investigations, thanks to many precious tools developed for the study of disordered systems and phase transitions, like the replica theory (section 3.2) and the cavity method (subsection 2.4.4). The fundamental observation [KGV87, MP85, FA86] is to read the cost function as an energy, the optimal solution as the ground state of a physical system, and to interpret the average over the instances as a sample average. This correspondence is so virtuous because it turns out that the typical complications that characterize disordered systems are the same that one usually encounters in the study of random-instance optimization problems: the presence of *quenched disorder*, whose realizations are represented by the instances of the problem, and *frustration*, that typically arises from the presence of non-local constraints. Such constraints prevent the ground state from being found through local research algorithms, i.e. through a local minimization of the energy. Let us consider, for example, the matching case: a local search for the minimum cost configuration would match an unmatched node a with the unmatched neighbor b that satisfies $w((a,b)) = \min_{n\in\partial a} w((a,n))$, where $w((a,b))$ is the weight associated with the edge (a,b). However, as in the case of spin glasses, this strategy

Table 4.1. Comparison between some keywords of combinatorial optimization and statistical mechanics.

Optimization	Statistical mechanics
instance	sample
cost function	energy
optimal configuration	ground state
optimal cost	ground state energy

in general is not guaranteed to find the ground state. The Table 4.1 compares some key words of statistical mechanics and combinatorial optimization.

The reader is referred to [MPV87] for a review of the main results obtained by using statistical mechanics in order to study random instances of classical combinatorial optimization problems, like TSP and SAT. In what follows we focus on the matching problem. Only for this chapter we will denote by $2N$, instead that by N, the number of nodes of the graphs.

4.1 The random link matching problem

One of the first variants of the matching problem that has been formulated in terms of statistical physics is the so-called *random link matching problem* (RM). It is a minimum weight matching (chapter 1) on the complete graph $K_{2N} = (V; E)$, in which the weights[1] w_e of the links $e \in E$ are independent and identically distributed random numbers, drawn from a given probability law $\rho(w)$. Recalling the definition of Section 5.4.1, the Hamiltonian of the system, $\mathcal{H}$, can be defined as follows:

$$\mathcal{H} = \sum_{e \in E} m_e w_e, \tag{4.1}$$

where:

$$m_e \in \{0, 1\} \ \ \forall e \in E, \ \ \sum_{k \in \partial i} m_{(i,k)} = 1 \ \ \forall i \in V. \tag{4.2}$$

In this context the weights associated with the edges of the graph represent the unique source of quenched disorder: averaging over all possible inputs of the problems is equivalent to average over all possible realizations of the weights.

4.1.1 Thermodynamic limit

For low temperatures the length scale of a link belonging to the matching is fixed by the typical distance of the nearest neighbor of a given node [VM84]. Being the topology fully connected, this distance scales with N, and moreover in the large N

[1] In what follows sometimes it will be useful to talk about the weights as "lengths", and then, for example, an expression like "the distance of the nearest neighbor of a given node" has to be intended as "the weight of the link with the smallest weight between those incident to the given node".

limit it is expected to tend to zero. Indeed let us consider the probability distribution of the distance of the nearest neighbor $n(w)$:

$$n(w) = \frac{\rho(w)\left[1 - \int_0^w d\ell\rho(\ell)\right]^{N-2}}{\int_0^\infty dw\rho(w)\left[1 - \int_0^w d\ell\rho(\ell)\right]^{N-2}}. \tag{4.3}$$

For large N only values near to 1 of the term in brackets, that correspond to $w \approx 0$, contribute to $n(w)$. For this reason the only relevant property of the weights distribution $\rho(w)$ in the thermodynamic limit is its behaviour around $w = 0$. Let us suppose that the distribution of the weights has the form:

$$\rho(w) \overset{w\to 0}{\approx} w^r, \quad r > -1. \tag{4.4}$$

At this point we want to deduce how the typical weight belonging to the matching scales with N. As each vertex has $N-1$ neighbors, one may expect that $w^\star$ satisfies:

$$\int_0^{w^\star} dw\rho(w) = \mathcal{O}\left(\frac{1}{N}\right), \tag{4.5}$$

and then, since for large N we know that $w^\star \to 0$, we can use Equation 4.4 and Equation 4.5 to deduce that $w^\star = \mathcal{O}(N^{-\frac{1}{r+1}})$. This fact implies that the energy scales as $N^{1-\frac{1}{r+1}}$. Hence, to obtain an extensive energy in the large N limit, a possibility is to define a new inverse temperature $\hat{\beta}$ in such a way that:

$$\beta = \hat{\beta} N^{\frac{1}{r+1}}. \tag{4.6}$$

Thus, in principle, one has to study the thermodynamic limit at fixed $\hat{\beta}$, and then take the limit $\hat{\beta} \to \infty$ to study the properties of the ground state.

In [MP85, MP87] it has been used the replica method to study the bipartite and non bipartite case. In particular, starting from the partition function

$$Z(\hat{\beta}) = \sum_{\{m_e\}} \mathbb{I}\left(\sum_{k\in\partial i} m_{(i,k)} = 1\right) e^{-\hat{\beta} N^{\frac{1}{r+1}} \mathcal{H}[\{m_e\}]}, \tag{4.7}$$

Mézard and Parisi evaluated the average optimal cost for a distribution of the form (4.4) in the thermodynamical limit using the replica trick:

$$\mathcal{E}_r = -\lim_{\substack{\hat{\beta}\to+\infty \\ N\to+\infty}} \lim_{n\to 0} \frac{1}{\hat{\beta} n N} \log \overline{Z^n(\hat{\beta})}, \tag{4.8}$$

under the replica symmetric assumption (section 3.2). In particular, for $r = 0$ they found

$$\mathcal{E}_0 = \frac{\pi^2}{12}. \tag{4.9}$$

The validity of the replica symmetric assumption, corroborated by numerical simulations, has been proved in [MP87, PR02] by a study of the stability of the replica symmetric saddle point. In particular they showed that the spectrum of the Hessian matrix contains only nonnegative eigenvalues at any temperature in the fully connected bipartite and non bipartite case for distributions with $r = 0$. The asymptotic

result of Equation 4.9 can be obtained with the cavity method [MM09], with a calculation that is a special case of the one that we will present in chapter 5 for a finite connectivity case.

In 2001 Aldous [Ald01] provided a rigorous treatment of the random link matching problem on complete bipartite graph, in the limit $N \to \infty$, and for independently and exponentially distributed random weights, confirming in this case the prediction of Equation 4.9.

4.1.2 Finite size corrections

In [MP87, PR02, CDMS17] some RM problems have been studied with the replica theory, analyzing the terms beyond the leading order of the averaged energy density expansion in powers of $1/N$. The sub-leading corrections, unlike the thermodynamic average energy density, depend on the Maclaurin expansion of the probability distribution of the weights up to, at least, the second term [PR02].

In [PR02] for the complete non bipartite graph, in the case of weights distributed according to the uniform law on the interval $[0, 1]$, Parisi and Ratiéville found:

$$\mathcal{E}^N_{\text{flat}} = \frac{\pi^2}{12} + \frac{1}{N}\left(-\frac{\zeta(3)}{2} + \sum_{p=0}^{\infty} \frac{I_{2p+1}}{2p+1}\right) + o\left(\frac{1}{N}\right), \tag{4.10}$$

where:

$$I_p = \int_1^{\infty} \frac{dC}{2C} \int \prod_{i=1}^{p} \frac{dx_i}{x_i + C}, \tag{4.11}$$

and the integration with respect to $x_1, ..., x_p$ has to be performed over the domain defined by $\forall i,\ \ x_i \geq 0$ and $x_i x_{i+1} \leq 1, x_p x_1 \leq 1$. The first four terms are:

$$\begin{aligned} I_1 &= \frac{\zeta(2)}{4} \approx 0.411234 \\ I_2 &= \frac{\zeta(3)}{2} \approx 0.601028 \\ I_3 &= \frac{3\zeta(4)}{16} \approx 0.202936 \\ I_4 &= 4\zeta(5) - \frac{\pi^2\zeta(3)}{3} \approx 0.193102. \end{aligned} \tag{4.12}$$

In the expressions above $\zeta(x)$ is the Riemann zeta function. In the same work, by using the flat distribution in the bipartite case, they also found:

$$\mathcal{E}^N_{\text{flat}} = \frac{\pi^2}{12} - \frac{1}{2N}(1 + 2\zeta(3)) + o\left(\frac{1}{N}\right). \tag{4.13}$$

In 1998 Parisi conjectured [Par98] an exact formula for the averaged optimal cost of the bipartite RM as a function of the size N of the system at zero temperature, $\mathcal{E}^N(\hat{\beta} \to \infty)$, in the case of weights independent and identically distributed according to:

$$\rho(w) = e^{-w} \quad w > 0 \quad \text{(case r=0)}. \tag{4.14}$$

The conjecture is the following:

$$\mathcal{E}^N = \frac{1}{2}\sum_{j=1}^{N}\frac{1}{j^2}. \tag{4.15}$$

By taking the limit $N \to \infty$ in Equation 4.15, the result of Equation 4.9 is recovered:

$$\lim_{N\to\infty}\mathcal{E}^N = \frac{\zeta(2)}{2}. \tag{4.16}$$

Then the bipartite case with exponential distribution (4.14) up to the first subleading correction is:

$$\mathcal{E}^N_{\text{exp}} = \frac{\pi^2}{12} - \frac{1}{2N} + o\left(\frac{1}{N}\right). \tag{4.17}$$

The Parisi's conjecture given in Equation 4.15 has been independently proved by Linusson and Wästlund [LW04] and Nair, Prabhakar, and Sharma [NPS05] in 2004.

4.1.3 Anomalous Scaling

If we interpret the term with the sum in Equation (4.11) as a contribution due to the cycles of all possible lengths, one could expect that the next to sub-leading term of the expansion (4.10) is a consequence of the fact that at finite N cycles of all possible lengths cannot appear in the graph. In [LPS17] an heuristic argument to extract the scaling with N of this correction is discussed. Firstly it is useful to estimate the integral in (4.11) for large p, i.e., for large cycle lengths: indeed one expects the terms corresponding to large cycle lengths to receive a correction due to the finite size of the system. For large p:

$$0 \leq \int \frac{dC}{2C}\int\prod_{i=1}^{p}\frac{dx_i}{x_i + C} \leq \int \frac{dC}{2C}\left(\log\frac{1+C}{C}\right)^p = \frac{1}{p} + o\left(\frac{1}{p}\right), \tag{4.18}$$

where the last passage follows considering that the only contribution of the term in brackets comes from the region in which $\log(1 + 1/C) \approx 1/C$. Now let us observe that a random path of length p on the complete graph K_N has a probability of order p/N of intersecting itself in the next step. Therefore, for a random path of length p, the total probability of intersection is of order p^2/N. This suggests that the maximum length of a cycle p scales as $\sqrt{N}$, and then, in order to correct the sum in (4.11), one could try to regularize the summand with a function of the form $f\left(p/\sqrt{N}\right)$, where $f(0)$ is a constant different from zero, and $\lim_{x\to\infty} f(x) = 0$, as we cannot have cycles of length $p > N$. Then the regularized version of the sum (4.11) takes the form:

$$\sum_p \frac{1}{p^2} f\left(\frac{p}{\sqrt{N}}\right) \approx a + \frac{b}{\sqrt{N}}, \tag{4.19}$$

where the last passage follows by approximating the sum with an integral. Therefore, from this argument, one may expect an anomalous scaling of the sub-sub-leading correction to the cost of the form $N^{-3/2}$, that in fact is seen numerically in [LPS17].

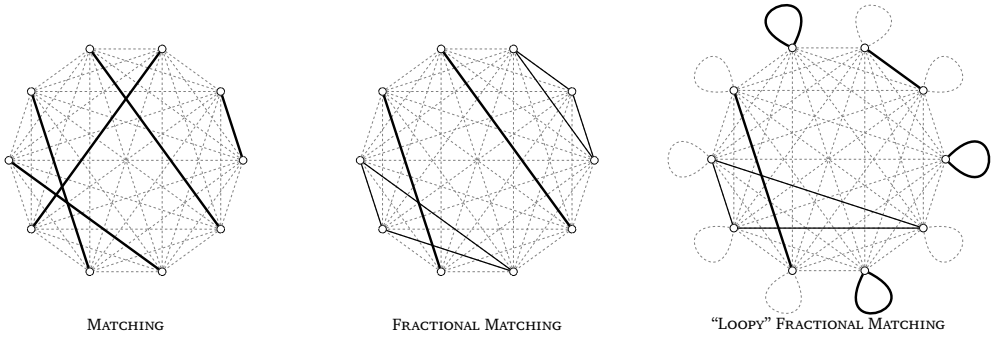


Figure 4.1. On the left, complete graph with a matching on it. In the center, the same graph with a fractional matching on it: in this case, cycles are allowed. On the right, the graph obtained allowing loops on each vertex, with a "loopy" fractional matching on it. Thin lines correspond to edges with an occupation number equal to $1/2$, while thick lines correspond to edges with an occupation number equal to 1. Picture from [LMPS18].

4.2 The fractional and loopy fractional matching

In this section we introduce two other versions of the matching problem, the *fractional matching* and the *loopy fractional matching*, and then we discuss some results of [LMPS18] that will be useful in chapter 5.

Let us introduce firstly the random link fractional matching. Consider a complete graph $K_{2N} = (V; E)$, and suppose to independently associate with each edge a non-negative random variables w_e, $e \in E$, drawn from a probability density $\rho(w)$, as in the previously discussed RM case. In the random fractional matching problem (RFM) we search for the set of occupation numbers m_e, $e \in E$, that minimizes the cost:

$$\mathcal{H}[\{m_e\}] = \sum_{e \in E} m_e w_e, \tag{4.20}$$

with the constraints:

$$m_e \in [0, 1] \;\; \forall\, e \in E, \;\; \sum_{k \in \partial i} m_{(i,k)} = 1 \;\; \forall i \in V. \tag{4.21}$$

The difference with the matching problems discussed up to now is the domain of definition of the occupation numbers, that in this case take values in a real interval. It is possible to show that in optimal matchings $\mathcal{M}$ the occupation numbers are such that $m_e \in \{0, 1/2, 1\}$, and therefore each optimal $\mathcal{M}$ contains only two kinds of edges: those belonging to odd cycles of the graph, and those that do not share their endpoints [Wï0].

In a variation of the random fractional matching, called the random "loopy fractional" matching (RLFM), an additional non-negative weight w_v is associated with each vertex $v \in V$ of the graph. Each weight w_v is a random variable extracted independently from all the other weights with the same distribution. The cost now is:

$$\mathcal{H}[\{m_e\}, \{m_v\}] = \sum_{e \in E} m_e w_e + 2 \sum_{v \in V} m_v w_v, \tag{4.22}$$

with the constraints:

$$m_e \in [0,1] \ \forall\, e \in E, \ \ m_v \in [0,1] \ \forall\, v \in V, \ \sum_{k \in \partial v} m_{(v,k)} + 2m_v = 1 \ \ \forall v \in V. \quad (4.23)$$

The RLFM then is equivalent to a RFM in a complete graph in which each node has a self-loop that can be occupied. Wästlund proved that, in the loopy fractional matching, in the optimal configuration $m_e \in \{0, 1/2, 1\}\ \forall e$, and $m_v \in \{0,1\}\ \forall v$ [WÏ0]. Note that any feasible configuration for the usual matching problem is feasible for the fractional matching problem; moreover, any feasible configuration for the fractional matching problem is feasible for the loopy fractional matching problem. Then $\forall N$ and for all distributions of the weights, certainly:

$$\mathcal{E}^{\mathrm{RLFM}} \leq \mathcal{E}^{\mathrm{RFM}} \leq \mathcal{E}^{\mathrm{RM}}. \quad (4.24)$$

By considering densities $\rho(w)$ with non-negative support and such that $\rho(w) = 1 - \mu w + o(w)$ for $w \to 0^+$, in [LMPS18] it is found with a replica approach:

$$\mathcal{E}^N_\eta(\mu) = \frac{\pi^2}{12} + \frac{1}{2N}\left[(\mu - 1)\zeta(3) + \frac{1-\eta}{4}\zeta(2)\right] + o\left(\frac{1}{N}\right), \quad (4.25)$$

where $\eta \in \{-1, 1\}$: $\eta = 1$ is the random loopy fractional matching case, and $\eta = -1$ is the fractional matching.

At this point let us denote by $\Delta\mathcal{E}^{\mathrm{RM}}_{\mathrm{flat}}/N$, $\Delta\mathcal{E}^{\mathrm{RFM}}_{\mu}/N$, and $\Delta\mathcal{E}^{\mathrm{RLFM}}_{\mu}/N$, respectively, the $\mathcal{O}(1/N)$ corrections of the RM, of the RFM, and of the RLFM. In order to compare these three corrections in the case of $\mu = 0$, we rewrite them together below:

$$\begin{aligned}
\Delta\mathcal{E}^{\mathrm{RM}}_{\mathrm{flat}} &= -\frac{\zeta(3)}{2} + \sum_{p=0}^{\infty} \frac{I_{2p+1}}{2p+1} = -\frac{\zeta(3)}{2} + \frac{\zeta(2)}{4} + \sum_{\substack{k \geq 3, \\ k \text{ odd}}}^{\infty} \frac{I_k}{k}, \\
\Delta\mathcal{E}^{\mathrm{RFM}}_{0} &= -\frac{\zeta(3)}{2} + \frac{\zeta(2)}{4} \\
\Delta\mathcal{E}^{\mathrm{RLFM}}_{0} &= -\frac{\zeta(3)}{2}.
\end{aligned} \quad (4.26)$$

Note that both $\Delta\mathcal{E}^{\mathrm{RFM}}_0$ and $\Delta\mathcal{E}^{\mathrm{RLFM}}_0$ can be obtained starting from $\Delta\mathcal{E}^{\mathrm{RM}}_{\mathrm{flat}}$: the first one by removing the series over k, and the second one by removing all the terms but $\zeta(2)/4$. This consideration could make one think that the third term of (4.26) is a sum of contributions of the odd cycles, that disappears when one allows odd cycles to belong to the matching, and that $I_1 = \zeta(2)/4$ is a vertex contribution, that disappears when self-loops appear in the feasible solutions. Note that (4.24), which is verified with the sign of equality in the thermodynamic limit, at the first order becomes a chain of strict inequalities, being the disappearing terms all positive.

Chapter 5

Matching on Random Regular Graphs

In chapter 4 some results about matching problems on fully connected topologies have been presented. Conversely, in this chapter we focus on a class of sparse graphs: those belonging to the ensemble $\mathbb{G}_{\mathrm{RRG}}(N, z)$ of regular graphs. An important property of $\mathbb{G}_{\mathrm{RRG}}(N, z)$ is the locally-tree-likeness in the limit $N \to \infty$. This property motivates the idea to use iterative approaches, like the cavity method (subsection 2.4.4), which are typical of statistical mechanics on tree-like structures [MP01]. For this reason, random regular graphs are usually called *Bethe lattices*.

We study some aspects of four different kinds of matching problems, already defined in chapter 1 and chapter 4: the standard matching problem, the fractional matching, the loopy fractional matching, and the assignment problem. The main aim is to study the optimal solutions in the presence of two sources of quenched randomness:

- a **topological disorder**, as the graphs are drawn from the ensemble of random regular graphs $\mathbb{G}_{\mathrm{RRG}}(N, z)$;
- a **link disorder**, as with each edge is associated a random weight. The weights are independent and identically distributed, analogously to the cases discussed in chapter 4.

Hence a realization of the problem is defined by a realization of the graph, and by a realization of the weights associated with its edges.

It is crucial to notice that the presence of cycles should play a determining role: indeed, as it happens for spin glasses, they contribute to frustrate the system (see, for example, Figure 5.1).

On RRGs we expect that the cycles do not produce effects that contribute to the thermodynamic limit, due to the fact that the presence of cycles is asymptotically suppressed. We will test this hypothesis through a comparison between numerical simulations and some predictions obtained with the cavity method at the replica symmetric level (Section 2.4.4). Conversely, as the density of the number of cycles of fixed length scales as $1/N$, one may expect that the first finite size correction to the thermodynamic cost density receives a contribution due the cycles. This

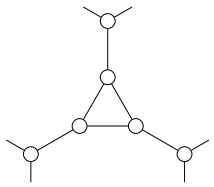


Figure 5.1. The triangle, as all the odd-cycles, produces frustration introducing non-local constraints in the system: the edges connecting the nodes of the triangle to the external nodes cannot all be excluded at the same time from a perfect matching.

idea certainly is also motivated by the observations made in chapter 4 for the fully connected case.

As we will show in this chapter, with the cavity method it is possible to estimate the costs of the cycles of a given length. After comparing these estimates with numerical simulations, performed with a method that takes inspiration from the edge-swapping algorithm of section 1.1.2, the idea is to try to gain a quantitative understanding of the role of the cycles in finite size effects on a sparse topology.

The chapter is organized as follows. In Sections 5.1 the definition of standard matching problem on RRGs is presented. In Sections 5.2 and 5.3 the cavity arguments for the computation of the asymptotic energy density and the cost of the cycles are discussed. Section 5.4 deals with some numerical simulations.

5.1 Problem definition

Let us start considering an instance G of $\mathbb{G}_{\mathrm{RRG}}(N, z)$. The first step for the formulation of the matching problem on G in terms of a statistical mechanics problem is to define the energy of the system:

$$\mathcal{H} = \sum_{g \in E} m_g w_g. \tag{5.1}$$

Equation 5.1 is equivalent to the standard cost of the matching (Equation 1.30): m_g is an occupation number associated with each edge $g \in E$, and it is equal to 1 or 0 depending if, respectively, g belongs or not to the matching; the w_g's are quenched positive weights associated with each edge of the graph, drawn independently from the same probability distribution $\rho(w)$, that we will assume to have the form:

$$\rho(w) = \frac{1}{z} e^{-\frac{w}{z}}. \tag{5.2}$$

A configurations of the system is defined by a set of occupation numbers $\{m_g\}_{g \in E}$. We are interested in the configurations that represent perfect matchings. However, since for finite N a perfect matching is not guaranteed to exist for all regular graphs (chapter 1), before calculating the thermodynamic limit, formally one has to relax the condition:

$$\sum_{k \in \partial i} m_{(k,i)} = 1, \quad \forall\, i \in V, \tag{5.3}$$

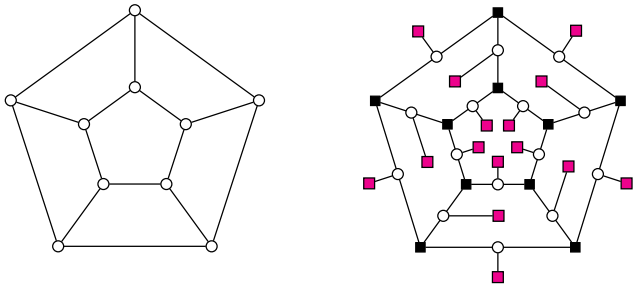


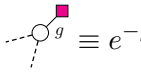
Figure 5.2. On the left it is given an example of regular graph with $N = 10$ nodes and valence $k = 3$. On the right it is reported the factor graph representing a model with probability distribution of the form (5.4), and defined on the graph on the left.

i.e., to relax the request that all the nodes of the graph have to be matched. This relaxation can be controlled introducing a Lagrange parameter γ as follows. Let us define a probability associated with each configuration of the system as:

$$\psi(\underline{m}) = \frac{1}{\mathcal{Z}} \prod_{k \in V} \mathbb{I}\left[\sum_{g \in \partial k} m_g \leq 1\right] \exp\left[-\beta\gamma\left(1 - \sum_{g \in \partial k} m_g\right)\right] \prod_{g \in E} \exp\left(-\beta m_g w_g\right). \tag{5.4}$$

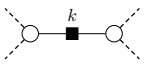
For a finite value of the coupling constant γ, matchings that are not perfect are energetically discouraged, but have non-zero measure. In the limit $\gamma \to \infty$ the hard constraint of perfect matching is recovered. In Equation 5.4 it has been introduced an artificial inverse temperature β, that at the end of the computations will be sent to infinity to study the ground state properties of the system. Indeed in the double limit $\beta, \gamma \to \infty$ it is easy to see that Equation 5.4 is concentrated on the minimum cost perfect matching.

It is possible to give a graphical representation of the probability distribution (5.4) in terms of factor graphs. Given the graph on which one wants to solve the matching problem, in order to construct the associated factor graph it is sufficient to substitute each node of the starting graph with a function node, to associate with each edge a variable node, and then to associate uniquely with each variable node an additional function node. The meaning of each node can be easily understood from the example in Figure 5.2: the variable nodes associated with each edge represent the occupation numbers m_g, the function nodes uniquely associated with each variable node, that we will call *soft nodes*, verify the following correspondence:



$$\equiv e^{-\beta m_g w_g},$$

and the function nodes corresponding to the nodes of the starting graph, that we will call *hard nodes*, verify the correspondence:



$$\equiv \mathbb{I}\left[\sum_{g \in \partial k} m_g \leq 1\right] \exp\left[-\beta\gamma\left(1 - \sum_{g \in \partial k} m_g\right)\right].$$

5.2 The cavity equation and the ground state energy

In this section we derive the density evolution equation for the standard matching, and an expression for the average optimal cost per node $\mathcal{E}$ for $N \to \infty$. As anticipated, we exploit a fundamental property of the random regular graph ensemble in the large graph limit, i.e. the fact that it is locally tree-like, using a cavity approach to the problem. Let us start from the fixed instance case. The first step is to write the equations for the cavity marginals (2.47) of a generic variable node m_e.

As each variable node e is associated with exactly two hard nodes, we use the following notation: if k is an hard node adjacent to a variable node e, the other hard node linked to e will be denoted by $\overline{k}_e$. The messages from variable nodes to hard function nodes have the form:

$$\nu_{e\to k} \propto \exp(-\beta m_e w_e)\mu^{-}_{\overline{k}_e\to e}, \tag{5.5}$$

instead the messages from hard[1] function nodes to variable nodes are:

$$\mu_{k\to e} \propto \sum_{m_{\partial k\setminus e}} \mathbb{I}\left[\sum_{g\in\partial k\setminus e} m_g + m_e \leq 1\right] \exp\left[-\beta\gamma\left(1 - \sum_{g\in\partial k} m_g\right)\right] \prod_{g\in\partial k\setminus e} \nu_{g\to k}, \tag{5.6}$$

where, for brevity, it has been omitted the argument of the messages:

$$\begin{aligned} \mu_{k\to e} &\equiv \mu_{k\to e}(m_e) \\ \nu_{e\to k} &\equiv \nu_{e\to k}(m_e). \end{aligned} \tag{5.7}$$

By substituting Equation 5.5 into Equation 5.6 one obtains the dependence of one function-to-variable message of a variable e from the function-to-variable messages of the variable nodes neighboring e:

$$\mu_{k\to e} \propto \sum_{m_{\partial k\setminus e}} \theta_e \exp\left[-\beta\gamma\left(1 - \sum_{g\in\partial k} m_g\right)\right] \prod_{g\in\partial k\setminus e} \exp(-\beta m_g w_g)\mu^{-}_{\overline{k}_g\to g}, \tag{5.8}$$

where, in order to lighten the notation, we called $\theta_e \equiv \mathbb{I}\left[\sum_{g\in\partial k\setminus e} m_g + m_e \leq 1\right]$.

It is important to note that in this discussion the thermodynamic limit is implicit in the use of the cavity equations, since only in this limit the tree-like approximation is exact.

As in this problem the alphabet of the variables is binary, $m_e \in \{0, 1\}$, each cavity marginal can be parameterized by a single number. Let us represent the function-to-variable cavity marginal as follows:

$$\mu_{k\to e} \propto \exp\left(\beta h_{k\to e} m_e\right), \tag{5.9}$$

where $h_{k\to e}$ is the parameter that uniquely defines $\mu_{k\to e}$; it can be understood as a "cavity field" acting on e. From Equation 5.9 it follows that the Bethe estimate

[1] Observe that there would be another equation for the messages sent from soft function nodes to variable nodes. However it can be omitted, since each soft node is connected only to one variable node. This fact implies that if s is a soft function node, the function-to-variable message is known, $\mu_{s\to e} \propto \exp\left(-\beta m_e w_e\right)$, then, in order to solve the fixed instance problem, i.e., to compute the marginals (2.48), it is sufficient to focus on Equation 5.5 and Equation 5.6.

of the marginal probability distribution $\psi(m_e)$ of a variable node m_e, expressed in terms of the cavity fields acting on it, takes the form:

$$\psi(m_e) \propto \exp\left[\beta\left(h_{k\to e} + h_{\overline{k}_e\to e} - w_e\right) m_e\right]. \tag{5.10}$$

Hence, by Equation 5.10, when $h_{k\to e} + h_{\overline{k}_e\to e} > w_e$, m_e is more likely to belong to the matching, vice versa when $h_{k\to e} + h_{\overline{k}_e\to e} < w_e$, m_e is more likely to be excluded from the matching.

Now let us rewrite Equation 5.8 in terms of the cavity fields. The starting point is the relation:

$$h_{k\to e} = -\frac{1}{\beta}\log\frac{\mu_{k\to e}(0)}{\mu_{k\to e}(1)}. \tag{5.11}$$

One can use Equation 5.8 to compute $\mu_{k\to e}(1)$ and $\mu_{k\to e}(0)$, and then use Equation 5.11 to obtain a recursive relations for the fields. It is found that:

$$\mu_{k\to e}(1) \propto \prod_{g\in\partial k\setminus e} \mu_{\overline{k}_g\to g}(0), \tag{5.12}$$

and that

$$\mu_{k\to e}(0) \propto \exp(-\beta\gamma)\prod_{g\in\partial k\setminus e}\mu_{\overline{k}_g\to g}(0) + \sum_{f\in\partial k\setminus e}\exp(-\beta w_f)\mu_{\overline{k}_f\to f}(1)\prod_{\substack{g\in\partial k\setminus e\\ g\neq f}}\mu_{\overline{k}_g\to g}(0), \tag{5.13}$$

from which Equation 5.11 takes the form:

$$h_{k\to e} = -\frac{1}{\beta}\log\left\{\exp(-\beta\gamma) + \sum_{f\in\partial k\setminus e}\exp\left[-\beta w_f + \beta h_{\overline{k}_f\to f}\right]\right\}. \tag{5.14}$$

By taking the double limit $\gamma\to\infty$ and $\beta\to\infty$:

$$\begin{aligned} h_{k\to e} &= -\lim_{\beta\to 0}\lim_{\gamma\to\infty}\frac{1}{\beta}\log\left\{\exp(-\beta\gamma) + \sum_{f\in\partial k\setminus e}\exp\left[-\beta w_f + \beta h_{\overline{k}_f\to f}\right]\right\} \\ &= -\lim_{\beta\to 0}\frac{1}{\beta}\log\left\{\sum_{f\in\partial k\setminus e}\exp\left[-\beta w_f + \beta h_{\overline{k}_f\to f}\right]\right\} = \min_{f\in\partial k\setminus e}\left(w_f - h_{\overline{k}_f\to f}\right). \end{aligned} \tag{5.15}$$

Equation 5.15 relates the cavity field at zero temperature acting on a generic variable node e to those acting on the variable nodes neighboring e, and, despite being more compact, it is equivalent to Equation 5.8. Up to now, the discussion regarded a given instance of the matching, i.e., an infinite regular tree with a given realization of the weights on the edges. Now we want to focus on the random instance case. As in the thermodynamic limit we are assuming to work on a z-regular tree, it can be used Theorem 2.5, which implies:

$$h^{(t+1)} \overset{d}{=} \min_i\left(w_i - h_i^{(t)}\right), \quad \text{with } 1\le i\le z-1. \tag{5.16}$$

Equation 5.16 is the density evolution for the matching case on a z-regular tree. In Equation 5.16 the cavity field, h, becomes a random variable. The fields $h_i^{(t)}$, with $1 \leq i \leq z-1$, are independent random copies of $h^{(t)}$, and w_i, with $1 \leq i \leq z-1$, are independent random weights distributed according to $\rho(w)$. Equation 5.16 is equivalent to the recursive equation:

$$\mathcal{P}_{t+1}\left(h^{(t+1)}\right) = \int \prod_{\ell=1}^{z-1} dw_\ell \rho(w_\ell) \int \prod_{\ell=1}^{z-1} dh_\ell \mathcal{P}_t\left(h_\ell^{(t)}\right) \delta\left(h^{(t+1)} - \min_i\left(w_i - h_i^{(t)}\right)\right) \tag{5.17}$$

where $1 \leq i \leq z-1$, $\delta(\cdot)$ is the Dirac delta function, and $\mathcal{P}_t$ is the probability distribution of $h^{(t)}$. Now suppose that for $t \to \infty$, $\mathcal{P}^{(t)}(h)$ converges to a fixed point $\mathcal{P}(h)$. Then the Bethe estimate of the average energy density $\mathcal{E}$ is simply found by keeping in mind the parameterization of Equation 5.10:

$$\mathcal{E} = \frac{z}{2}\int_0^\infty dw \rho(w) w \int dh_1 dh_2 \mathcal{P}(h_1)\mathcal{P}(h_2)\mathbb{I}\left[h_1 + h_2 - w \geq 0\right]. \tag{5.18}$$

Note that, without the factor $z/2$, the integral in Equation 5.18 gives the total energy per number of edges of the graph.

Population dynamics

We have not been able to find an analytical solution of Equation 5.17 in the limit $t \to \infty$. We solved therefore the equation sampling $\mathcal{P}(h)$ with a numerical method, in order to compute the energy density $\mathcal{E}$. The idea is to approximate the probability distribution $\mathcal{P}(h)$ through a sample of M independent identical distributed copies of h. As M becomes large, the empirical distribution of such sample should converge to the actual distribution of h [MM09]. We shall call the sample $\{h_i\} \equiv \{h_1, h_2, ..., h_M\}$ a *population*; from this the name *population dynamics*. The algorithm that samples $\mathcal{P}(h)$ and estimates $\mathcal{E}$ is described by the pseudocode (2). As inputs, it requires the population size M, and the maximum number of iterations T. After updating the cavity fields, the cost density is found by the following procedure. We draw z weights $w_1, ..., w_z$ according to the exponential distribution, z fields $h_1, ..., h_z$ from the population, and we consider as if they were the weights and fields of z incident edges, as represented for the case of $z = 3$ in Figure 5.3. Now if we denote by m the index:

$$m = \arg\min_{1 \leq i \leq z}(w_i - h_i), \tag{5.19}$$

then it is easy to see that w_m samples the distribution of the cost of z edges. As the number N of nodes is related to the number of edges through $|E| = \frac{zN}{2}$, the average over different extractions of w_m divided by two gives an estimate of the energy density $\mathcal{E}$. For $M = 10^7$ and $T = 500$ we found the energy densities written in Table 5.1. In Figure 5.4 the cavity estimates for the average energy densities $\mathcal{E}$ are reported. The data trend turns to be good-fitted by a function of the form:

$$f(z) = a + \frac{b}{z^2} + \frac{c}{z^4}, \tag{5.20}$$

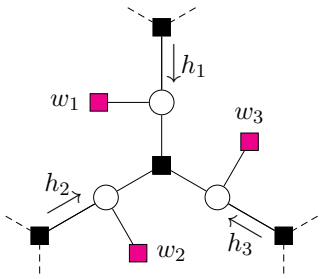


Figure 5.3. The weights w_i are written near by the corresponding soft node. The fields h_i are those that define the function-to-variable cavity marginals outgoing from the external function nodes.

z	$\mathcal{E}$	$\Delta\mathcal{E}$
3	0.83789	0.00002
4	0.82894	0.00002
5	0.82607	0.00001
6	0.82474	0.00001
7	0.82407	0.00002
8	0.82365	0.00002
9	0.82336	0.00001
10	0.82314	0.00002
20	0.82259	0.00001

Table 5.1. Cavity estimates of the average energy densities as a function of the degree z of the RRG ensemble, in the case of exponential distribution $\exp(-w/z)/z$.

with:

$$\begin{aligned} a &= 0.82241 \pm 0.00002 \\ b &= 0.069 \pm 0.004 \\ c &= 0.6291 \pm 0.1736. \end{aligned} \tag{5.21}$$

As expected from the results of chapter 4, a is compatible with $\pi^2/12$, the thermodynamic cost density on a fully connected graph.

5.3 Cost of the cycles

We assume that on $\mathbb{G}_{\mathrm{RRG}}(N, z)$ the average cost Γ_ℓ of a cycle of length ℓ for $N \to \infty$ is equal to:

$$\Gamma_\ell = \lim_{\beta\to\infty} \left\langle -\frac{1}{\beta} \log \frac{\mathcal{Z}_\ell}{\mathcal{Z}_0} \right\rangle_\rho , \tag{5.22}$$

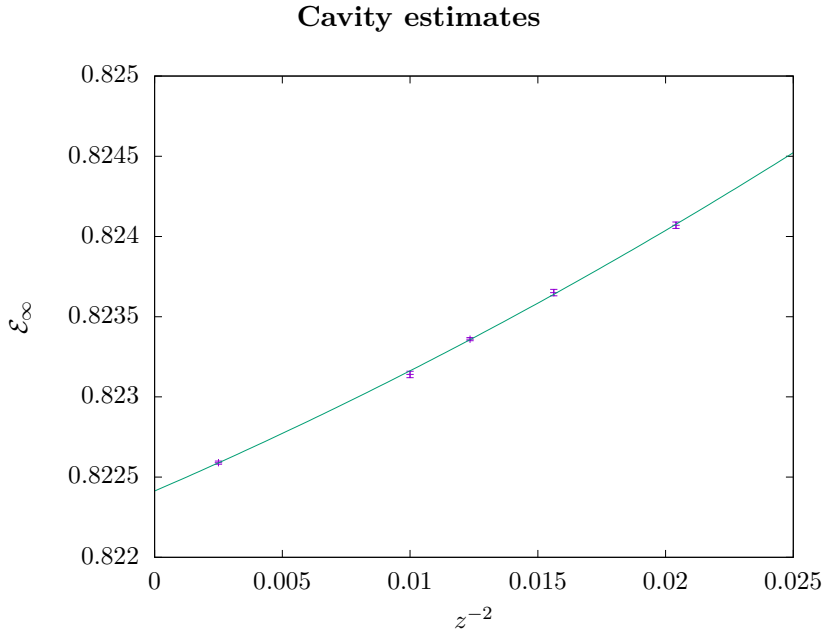


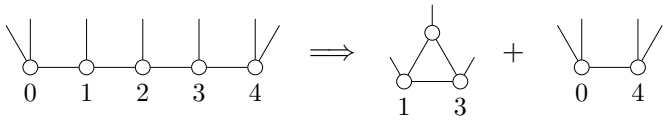
Figure 5.4. Cavity estimates of the average energy densities $\mathcal{E}$ in the thermodynamic limit for $z = 7, 8, 9, 10, 20$. The fitting function has the form reported in Equation 5.20.

Algorithm 2 Population Dynamics (size N, iterations T)

Initialize $\{h_i^{(0)}\}$;
for $t = 1, ..., T$ **do**
  $\mathcal{E} \leftarrow 0$;
  for i=1,...,M **do**
    Draw $j(1), ..., j(z-1)$ uniformly in $\{1, ..., M\}$;
    Draw $w(1), ..., w(z-1)$ with a distribution $\rho(w)$;
    $h_i^{(t)} \leftarrow \min_k(w(k) - h_{j(k)}^{(t-1)})$, $1 \leq k \leq z-1$;

    Draw $j(1), ..., j(z)$ uniformly in $\{1, ..., M\}$;
    Draw $w(1), ..., w(z)$ with a distribution $\rho(w)$;
    $min \leftarrow \min_k(w(k) - h_{j(k)}^{(t-1)})$, $1 \leq k \leq z$;
    $index \leftarrow k^\star$, where $w(k^\star) - h(k^\star) == min$;
    $\mathcal{E} \leftarrow \mathcal{E} + w(index)$;
  end for
  $\mathcal{E} \leftarrow 0.5\,\mathcal{E}/M$;
  print($\mathcal{E}$);
end for

where the angle brackets denote the average over the weights, $\mathcal{Z}_0$ is the partition function of the matching on an infinite z-regular tree, and $\mathcal{Z}_\ell$ is the partition function of the matching on an infinite z-regular graph, that differs from the tree just for the presence of a single cycle of length ℓ. Let us observe that it is possible to turn the regular tree into the other graph by an edge swapping, as represented below in the case of $z = 3$ and $\ell = 3$:



(5.23)

For general ℓ and z the procedure is the same: it is possible to realize the transformation by choosing a path of length $\ell + 1$ on the tree, and by performing the edge-swapping:

$$(0,1),(\ell,\ell+1) \Rightarrow (0,\ell+1),(1,\ell), \tag{5.24}$$

where (i, j) is the edge connecting i to j.

Now suppose to choose a path of length $\ell + 1$ on the tree. Moreover, in order to lighten the factor graph description, let us denote by $\mathcal{L}$ the set of variable nodes, and by $\widehat{\mathcal{L}}$ the set of function nodes belonging to such path. First of all we focus on the case of a fixed instance of the weights. In order to do that, we want to write $\mathcal{Z}_0$ in terms of the cavity marginals of the variables belonging to $\partial\widehat{\mathcal{L}} \setminus \mathcal{L}$, and $\mathcal{Z}_\ell$ in terms of the same cavity marginals, but after the edge swapping represented in 5.23. In the following we write, for generic variable and function nodes, $\theta_a(m_{\partial a}) \equiv \mathbb{I}\left[\sum_{j\in\partial a} m_j - 1\right]$. The partition function of the tree, $\mathcal{Z}_0$, is:

$$\mathcal{Z}_0 = \sum_{m_{\partial\widehat{\mathcal{L}}}} \left[\prod_{i\in\mathcal{L}} e^{-\beta w_i m_i}\right] \left[\prod_{a\in\widehat{\mathcal{L}}} \theta_a(m_{\partial a}) \prod_{k\in\partial a\setminus\mathcal{L}} \nu_{k\to a}(m_k)\right]. \tag{5.25}$$

Let us denote by $\mathcal{L}'$ the set of variable nodes belonging to the cycle obtained after the edge-swapping, and by $\widehat{\mathcal{L}'}$ the set of function nodes belonging the same cycle. We call m_0 the variable node associated with the isolated edge. The partition function $\mathcal{Z}_\ell$ has the form:

$$\mathcal{Z}_\ell = \sum_{m_{\partial\widehat{\mathcal{L}}}} \mathcal{C}\left(m_{\partial\widehat{\mathcal{L}'}}\right) \mathcal{D}\left(m_{\partial 0\cup\partial\ell+1}\right), \tag{5.26}$$

where $\mathcal{C}$ is the cycle contribution:

$$\mathcal{C} = \left[\prod_{i\in\mathcal{L}'} e^{-\beta w_i m_i}\right] \left[\prod_{a\in\widehat{\mathcal{L}'}} \theta_a(m_{\partial a}) \prod_{k\in\partial a\setminus\mathcal{L}'} \nu_{k\to a}(m_k)\right], \tag{5.27}$$

and $\mathcal{D}$ is the isolated edge contribution,

$$\mathcal{D} = e^{-\beta w_0 m_0} \prod_{\substack{j\in\\ \partial 0\setminus\ell+1}} \nu_{j\to 0}(m_j)\theta_0(m_{\partial 0}) \prod_{\substack{j\in\\ \partial\ell+1\setminus 0}} \nu_{j\to\ell+1}(m_j)\theta_{\ell+1}(m_{\partial\ell+1}). \tag{5.28}$$

By using Equation 5.5 and Equation 5.11 of section 5.2, and dividing $\mathcal{Z}_\ell$ and $\mathcal{Z}_0$ by $\prod_{a\in\widehat{\mathcal{L}}}\prod_{i\in\partial a\setminus\mathcal{L}}\mu_{i\to a}(0)$, one obtains:

$$\frac{\mathcal{Z}_\ell}{\mathcal{Z}_0}=\frac{\sum_{m_{\partial\widehat{\mathcal{L}}}}\widetilde{\mathcal{C}}\left(m_{\partial\widehat{\mathcal{L}}'}\right)\widetilde{\mathcal{D}}\left(m_{\partial 0\cup\partial\ell+1}\right)}{\sum_{m_{\partial\widehat{\mathcal{L}}}}\left[\prod_{i\in\mathcal{L}}e^{-\beta w_i m_i}\right]\left[\prod_{a\in\widehat{\mathcal{L}}}\theta_a\prod_{k\in\partial a\setminus\mathcal{L}}\exp\left(-\beta\sum_{j\in\partial a\setminus\mathcal{L}}m_j\phi_{j\to a}\right)\right]}, \quad (5.29)$$

where:

$$\widetilde{\mathcal{C}}=\left[\prod_{i\in\mathcal{L}'}e^{-\beta w_i m_i}\right]\left[\prod_{a\in\widehat{\mathcal{L}}'}\theta_a\exp\left(-\beta\sum_{j\in\partial a\setminus\mathcal{L}'}m_j\phi_{j\to 0}\right)\right], \quad (5.30)$$

$$\widetilde{\mathcal{D}}=\theta_0\theta_{\ell+1}\exp\left(-\beta\sum_{j\in\partial 0}m_j\phi_{j\to 0}-\beta\sum_{j\in\partial\ell+1}m_j\phi_{j\to\ell+1}-\beta m_0 w_0\right), \quad (5.31)$$

and $\phi_{i\to j}=w_{ij}-h_{i\to(i,j)}$. Note that, given a variable m_e belonging to $\partial\widehat{\mathcal{L}}\setminus L$, we considered equal the cavity marginals of m_e in the case of the tree and of the tree with a cycle. This follows from the fact that the cavity graph obtained removing the function node a, $a\in\widehat{\mathcal{L}}\cap\partial e$, is the same in both cases, if we consider the cycle far from the isolated edge.

By substituting Equation 5.29 into Equation 5.22, and taking the limit $\beta\to\infty$, it follows that, at a given instance of the weights:

$$\lim_{\beta\to\infty}-\frac{1}{\beta}\log\frac{\mathcal{Z}_\ell}{\mathcal{Z}_0}=E_C-E_T, \quad (5.32)$$

The first addend is the contribution of the graph with the cycle:

$$E_C=\min_{c_0,c_{\mathcal{L}'}}\left[\sum_{a=1}^{\ell}c_a w_a+\sum_{a=1}^{\ell}\mathbb{I}(c_a+c_{n(a)}=0)\gamma_a+c_0 w_0+(1-c_0)(\phi_0+\phi_{\ell+1})\right], \quad (5.33)$$

where $c_0\in\{0,1\}$, and $c_{\mathcal{L}'}=(c_1,c_2,...,c_\ell)\in\{0,1\}^\ell$ is a set of occupation numbers that represents all possible matchings, perfect and not, of a graph consisting of only one cycle of length ℓ. The function $n(a)$ is such that $n(a)=a+1$, if $1\le a\le\ell-1$, and $n(a)=1$, if $a=\ell$. The other term, E_T, is the contribution of the tree:

$$E_T=\min_{c_{\mathcal{L}}}\left[\sum_{a=0}^{\ell}c_a w_a+\sum_{a=1}^{\ell}\mathbb{I}(c_{a-1}+c_a=0)\gamma_a+(1-c_0)\phi_0+(1-c_\ell)\phi_{\ell+1}\right], \quad (5.34)$$

where $c_{\mathcal{L}'}=(c_0,c_1,...,c_\ell)\in\{0,1\}^{\ell+1}$ is a set of occupation numbers that represents all possible matchings, perfect and not, of a graph consisting of only one path of length $\ell+1$. In Equation 5.33 and Equation 5.34, we denoted by γ_a the smallest of the $z-2$ differences $w_a-h_{j\to a}$, with $j\in\partial a\setminus\mathcal{L}'$; and we denoted by ϕ_i, where $i=0,\ell+1$, the smallest of the $z-1$ fields $w_j-h_{i\to j}$, with $j\in\partial i\setminus\mathcal{L}$.

It is possible to switch to the random instance case by considering $w_a,\gamma_a,\phi_0,\phi_{\ell+1}$ as random variables. The weights are independently distributed according to $\rho(w)$, and the two cavity fields $\phi_0,\phi_{\ell+1}$ are distributed according to the probability

distribution that satisfy Equation 5.16 in the limit $t \to \infty$; the law of the fields γ_a satisfies the following relation:

$$\gamma^{(t+1)} \stackrel{d}{=} \min_i \left(w_i - h_i^{(t)} \right), \quad \text{with } 1 \leq i \leq z-2, \tag{5.35}$$

taken in the limit $t \to \infty$. In Equation 5.35 the γ_i's are $z-2$ independent drawings from the same probability distribution.

In order to estimate $\Gamma_\ell = \langle E_C - E_L \rangle_\rho$, we used *population dynamics*, essentially by calling inside the loop in i of the pseudocode (2) the two functions EnergyChain(ℓ), and EnergyTree(ℓ). These two function, respectively, allow to estimate $\langle E_C \rangle_\rho$ and $\langle E_L \rangle_\rho$, and are schematically described in the pseudocode (3).

For a population of $M = 10^7$, a number $T = 10^3$ iterations, and $\ell = 3, 4, 5$, we obtained the data given in Table 5.2, reported in Section 5.4.2.

Algorithm 3 Cycles costs (ℓ)

function EnergyChain(ℓ)
  $\phi_0 \leftarrow Get\_h()$;
  $\phi_\ell \leftarrow Get\_h()$;
  Draw $w(0), ..., w(\ell)$ with a distribution $\rho(w)$;
  for i=1,...,ℓ **do**
    $\gamma(i) \leftarrow Get\_g()$;
  end for
  return the minimum E_C over all the configurations of $c_0, c_{\mathcal{L}'}$ of Equation 5.33;
end function

function EnergyTree(ℓ)
  $\phi_0 \leftarrow Get\_h()$;
  $\phi_\ell \leftarrow Get\_h()$;
  Draw $w(0), ..., w(\ell)$ with a distribution $\rho(w)$;
  for i=1,...,ℓ **do**
    $\gamma(i) \leftarrow Get\_g()$;
  end for
  return the minimum E_T over all the configurations of $c_{\mathcal{L}}$ of Equation 5.34;
end function

function *Get_h*()
  Draw $j(1), ..., j(z-1)$ uniformly in $\{1, ..., M\}$;
  Draw $w(0), ..., w(z-1)$ with a distribution $\rho(w)$;
  return $\min_k (w(k) - h_{j(k)}^{(t-1)})$, $1 \leq k \leq z-1$;
end function

function *Get_g*()
  Draw $j(1), ..., j(z-2)$; uniformly in $\{1, ..., M\}$;
  Draw $w(0), ..., w(z-2)$ with a distribution $\rho(w)$;
  return $\min_k (w(k) - h_{j(k)}^{(t-1)})$, $1 \leq k \leq z-2$;
end function

5.4 Numerical Simulations

The numerical simulations can be conceptually divided in two parts. Those of the first part, that we will compare with the cavity predictions, concern the standard matching problem. Those of the second part, for which we have no cavity predictions, concern the fractional matching, the loopy fractional matching, and the assignment problem.

We make use of the solver implemented in the Lemon Graph Library [DJK11]. It is based on Edmond's blossom algorithm [Edm09], that has $\mathcal{O}(|V||E|\log|V|)$ computational time complexity on a generic graph $G = (V; E)$.

5.4.1 Standard Matching

In this section we discuss some numerical results about the standard matching problem on random regular graphs with valences $z = 3, 4$.

To improve the uncertainties on the estimates of the energy, both for $z = 3$ and $z = 4$, it has been used the following expedient. Suppose we want to estimate the average, denoted with angle brackets, of a given quantity A. By linearity of the expected value, given another quantity B, it follows that:

$$\langle A\rangle = \langle A - B\rangle + \langle B\rangle. \tag{5.36}$$

Let us compute the variance of $A - B$:

$$\sigma^2_{A-B} = \left\langle (A - B - \langle A - B\rangle)^2 \right\rangle = \sigma^2_A + \sigma^2_B - 2\left\langle (A - \langle A\rangle)(B - \langle B\rangle)\right\rangle, \tag{5.37}$$

where $\sigma^2_\cdot$ denotes the variance of the distribution of the random variable specified in the subscript. From Equation 5.37 it follows that if A and B are positively correlated it is possible that $\sigma^2_{A-B} < \sigma^2_A$. Now let us suppose that we know exactly $\langle B\rangle$. In this case one can measure the quantity $A - B$, and add to it at the end the value of $\langle B\rangle$, obtaining a less fluctuating estimate of $\langle A\rangle$, if A and B are properly correlated. In the matching case we have chosen as B a heuristic cost: for a given instance, for each node we select the edge with the lowest weight incident to it and we define our heuristic cost as the sum of such weights. The expected value of the heuristic cost in case of exponential distribution and regular graphs of valence z, is simply 1. In this case this technique allowed to reduce the uncertainties up to a factor ≈ 1.2.

Valence $z = 3$

Let us start from the case $z = 3$. In Figure 5.5 we present the average optimal cost densities for $16 \le N \le 120$ in the case of exponential distribution of the weights (5.2). Each point corresponds to the average of the optimal cost of 10^6 independent realizations of the system. Each realization has been constructed by generating a regular graph with the configurational model (Section 1.1.2), and then by assigning to each edge of the graph a random weight.

In the fully connected case, as discussed in chapter 4, it is observed an anomalous finite size correction proportional to $N^{-3/2}$. To evaluate the presence of such an

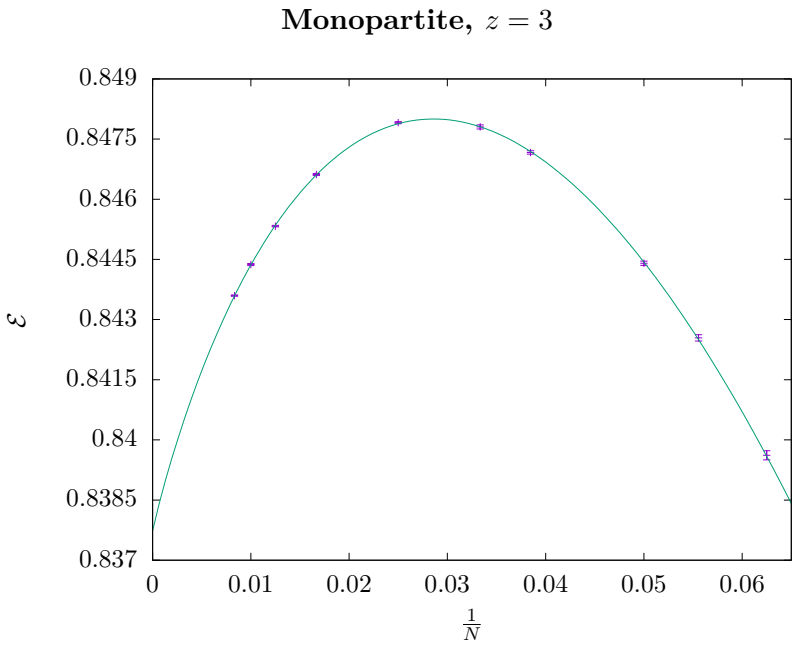


Figure 5.5. Average cost density $\mathcal{E}$ as a function of $1/N$ in the case of RRGs with valence $z = 3$ and link weights exponentially distributed (5.2). The solid line corresponds to the fit obtained using (5.38).

anomalous correction, we try to fit the data in Figure 5.5 with a function of the form:

$$\mathcal{E} = \mathcal{E}_3 + \frac{c_1}{N} + \frac{\widetilde{c}}{N^{\frac{3}{2}}} + \frac{c_2}{N^2}, \tag{5.38}$$

where it is found that:

$$\begin{aligned} \mathcal{E}_3 &= 0.8378 \pm 0.0001 \\ c_1 &= 1.11 \pm 0.03 \\ \widetilde{c} &= -4.8 \pm 0.3 \\ c_2 &= 1.5 \pm 0.9 \end{aligned} \tag{5.39}$$

The asymptotic cost density, $\mathcal{E}_3$, is in good agreement with the cavity prediction $\widehat{\mathcal{E}}_3$ (Table 5.1):

$$\widehat{\mathcal{E}}_3 = 0.83789 \pm 0.00002. \tag{5.40}$$

To improve the estimates of $c_1, \widetilde{c}, c_2$ we can try to fit $N(\mathcal{E}_3 - \widehat{\mathcal{E}_3})$ as a function of $1/\sqrt{N}$, as it is shown in Figure 5.6.

For a function of the form:

$$f(N) = c_1 + \frac{\widetilde{c}}{\sqrt{N}} + \frac{c_2}{N} \tag{5.41}$$

the best fit is determined by the parameters:

$$\begin{aligned} c_1 &= 1.096 \pm 0.006 \\ \widetilde{c} &= -4.63 \pm 0.07 \\ c_2 &= 1.4 \pm 0.2 \end{aligned} \tag{5.42}$$

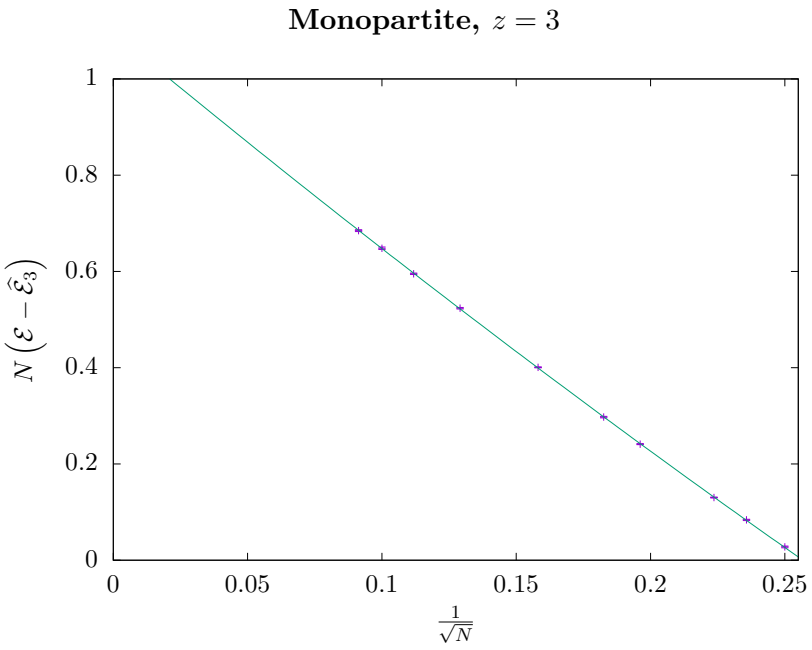


Figure 5.6. Behavior of $N\left(\mathcal{E} - \widehat{\mathcal{E}}_3\right)$ as a function of $1/\sqrt{N}$. The solid line is a quadratic fit in $1/\sqrt{N}$.

The fact that for large N the choice of the axis scaling linearizes the data, leads to conclude that even in the finite connectivity case there is a corrective term proportional to $N^{-3/2}$.

Denoting by $\mathcal{E}(G)$ the optimal cost per node corresponding to the instance G, for $z = 3$ we measured a quantity that will be useful in subsection 5.4.2: the connected correlation function $N\overline{\mathcal{E}(G)\mathcal{N}_\ell(G)}_c \equiv N\overline{\mathcal{E}(G)\mathcal{N}_\ell} - N\mathcal{E}\overline{\mathcal{N}_\ell}$, where $N\mathcal{E} = N\overline{\mathcal{E}(G)}$ is the average total energy and $\langle\mathcal{N}_\ell(G)\rangle$ is the average number of polygons of length ℓ, for some values of the number of nodes N. In Figure 5.7 are represented the measures for $\ell = 3, 4$ as functions of $1/N$, together with fitting functions of the form:

$$f_\ell(N) = a_\ell + \frac{b_\ell}{\sqrt{N}}, \tag{5.43}$$

where:

$$\begin{aligned} a_3 &= 0.136 \pm 0.001 \qquad & b_3 &= -0.8 \pm 0.1 \\ a_4 &= 0.0001 \pm 0.0007 \qquad & b_4 &= -0.86 \pm 0.07. \end{aligned} \tag{5.44}$$

Both for $\ell = 3$ and $\ell = 4$ the estimates are noisy, and require more statistics.

Valence $z = 4$

In Figure 5.8 the average optimal cost densities for $30 \leq N \leq 120$ are reported in the case of exponential distribution of the weights. Each point corresponds to the average of the optimal cost of 10^6 independent realizations of the system. Each realization has been constructed as in the case of $z = 3$, previously discussed.

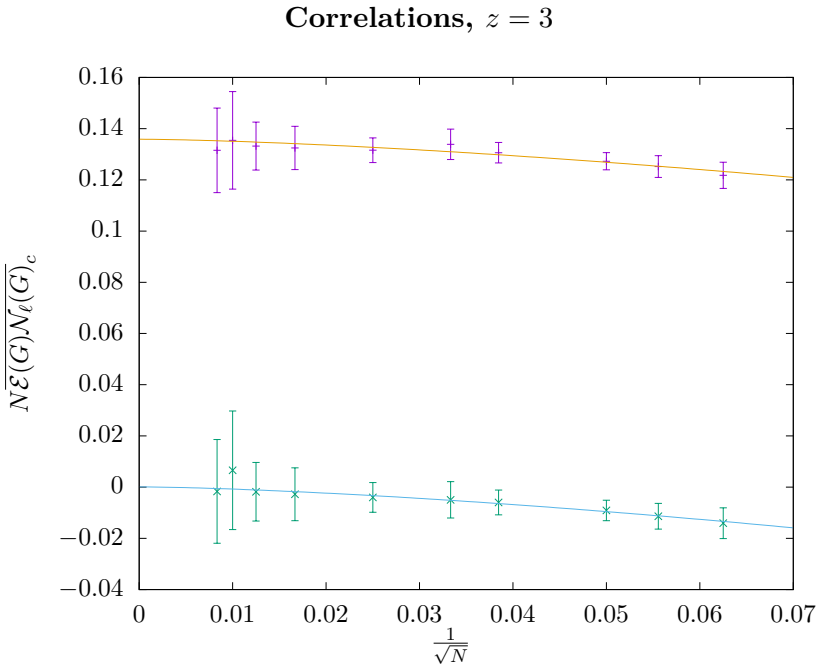


Figure 5.7. From the bottom to the top $\ell = 4, 3$. The form of the fitting function is that of Equation 5.43.

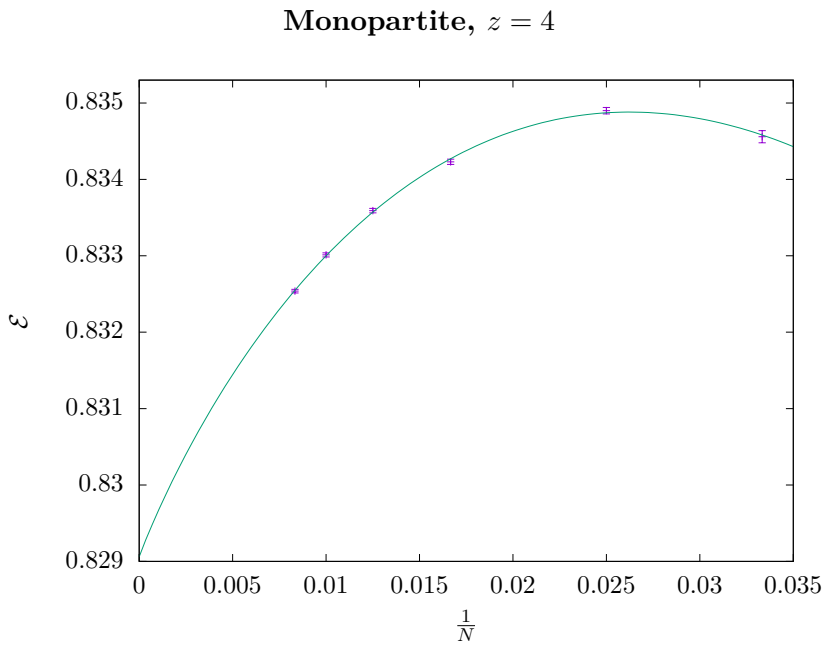


Figure 5.8. Average cost density $\mathcal{E}$ as a function of $1/N$ in the case of RRGs with valence $z = 4$ and link weights exponentially distributed (5.2). The solid line corresponds to the fit obtained using (5.45).

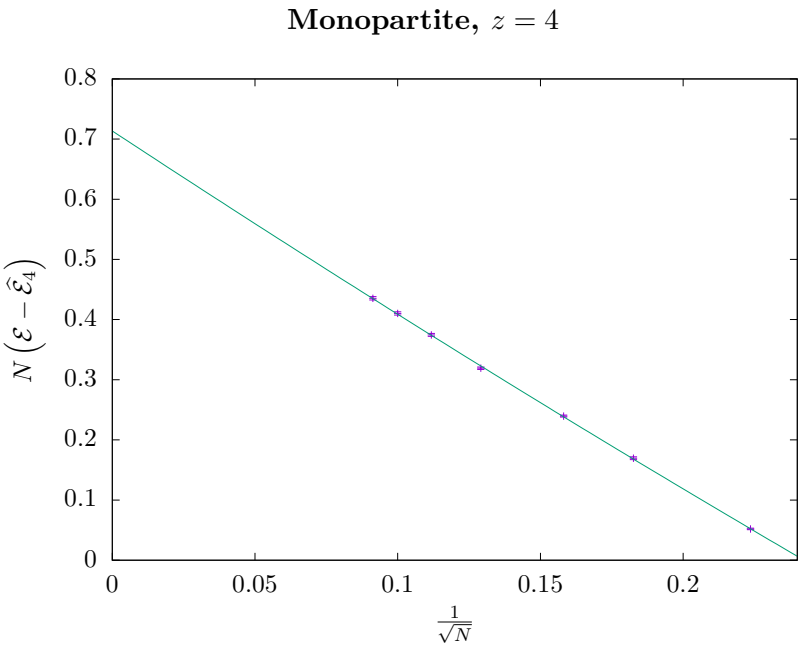


Figure 5.9. Behavior of $N\left(\mathcal{E} - \widehat{\mathcal{E}}_4\right)$ as a function of $1/\sqrt{N}$. The solid line is a quadratic fit in $1/\sqrt{N}$.

By fitting the data with:

$$\mathcal{E} = \mathcal{E}_4 + \frac{c_1}{N} + \frac{\widetilde{c}}{N^{\frac{3}{2}}} + \frac{c_2}{N^2} \tag{5.45}$$

it is found:

$$\begin{aligned} \mathcal{E}_4 &= 0.8291 \pm 0.0005 \\ c_1 &= 0.7 \pm 0.2 \\ \widetilde{c} &= -3 \pm 2 \\ c_2 &= 0.6 \pm 0.5 \end{aligned} \tag{5.46}$$

One more time there is good agreement between the numerical estimate of the asymptotic cost density, $\mathcal{E}_4$, and the cavity calculation $\widehat{\mathcal{E}_4}$ (Table 5.1):

$$\widehat{\mathcal{E}}_4 = 0.82894 \pm 0.00002. \tag{5.47}$$

The analogous of Figure 5.6 in the case of $z = 4$ is given in Figure 5.9.

By fitting with:

$$f(N) = c_1 + \frac{\widetilde{c}}{\sqrt{N}} + \frac{c_2}{N} \tag{5.48}$$

it follows:

$$\begin{aligned} c_1 &= 0.71 \pm 0.01 \\ \widetilde{c} &= -3.2 \pm 0.2 \\ c_2 &= 0.7 \pm 0.5 \end{aligned} \tag{5.49}$$

Also in this case, as can be seen in Figure 5.9, the data trend is linearized for large N by the scaling of the x-axis.

The observation of an $N^{-3/2}$-scaling correction at finite z, by virtue of the interpretation given in subsection 4.1.3, suggests the possibility that the $1/N$-correction receives a contribution due to the graph cycles of all lengths.

5.4.2 Cycles

In this section we present and discuss the measures of the average costs of triangles, squares, and pentagons on random regular graphs with valences $z = 3, 4$. The costs of these polygons have been computed by the method discussed below, that is inspired by the edge-swapping algorithm of section 1.1.2. To be more precise, suppose to generate a sequence of regular graphs $\{G_t\}_{t=0}^{L}$ in which if t is not a multiple of 50, G_{t+1} is obtained from G_t by a single swap of two edges; otherwise G_{t+1} is produced from G_t by $100|E|$ edge swaps. We call *blocks* the 50-graph subsequences composed by graphs that are connected by a single edge swap. For each t, an average energy density $\mathcal{E}(G_t)$ has been computed, by redrawing 10^3 times the weights associated with the edges of G_t according to the distribution (5.2).

We denote by $\mathcal{N}_\ell(G_t)$ the number of cycles of G_t that have length ℓ. For each t, $\mathcal{N}_\ell(G_t)$ has been computed in the case of $\ell = 3, 4, 5$, implementing the formula in Equation 1.5 (chapter 1). At this point consider a given block B; suppose that $\{k\}$, with $G_{t_k}, G_{t_k+1} \in B$, is the sequence of indexes that identifies all the transitions occurring in B of the form:

$$\mathcal{N}_\ell(G_{t_k+1}) - \mathcal{N}_\ell(G_{t_k}) = 1. \tag{5.50}$$

Suppose that T_B^ℓ is the number of transitions of $\mathcal{N}_\ell$ occurring in B. We computed, for each block B, the quantity γ_B^ℓ:

$$\gamma_B^\ell = \frac{1}{T_B^\ell} \sum_{k=1}^{T_B^\ell} \{\mathcal{E}(G_{t_k+1}) - \mathcal{E}(G_{t_k})\}, \tag{5.51}$$

and estimated the cost of a polygon of length ℓ by averaging γ_B^ℓ over all the blocks contained in the sequence.

Each point of the figures 5.10, 5.11, 5.12, 5.13, 5.14, 5.15 is the average of over $\approx 10^6 - 10^7$ different blocks B of γ_B^ℓ, with $\ell = 3, 4, 5$. The first graph of the chain has been drawn uniformly with the configurational model.

In 5.10, 5.11, 5.12, 5.13, 5.14, 5.15 the costs of the cycles per unit length are represented, as a function of $1/N$. We fit the data with a function of the form:

$$g_\ell = \frac{\Gamma_\ell}{\ell} + \frac{c_1}{N} + \frac{c_2}{N^2} + \frac{c_3}{N^3}, \qquad \ell = 3, 4, 5. \tag{5.52}$$

In Table 5.2 all the fit parameters are given, together with the cavity estimates of the asymptotic costs per unit length of the polygons.

Note that the costs of the odd-cycles are all positive: this is a consequence of the fact that the presence of odd-cycles produces frustration, and therefore it raises the

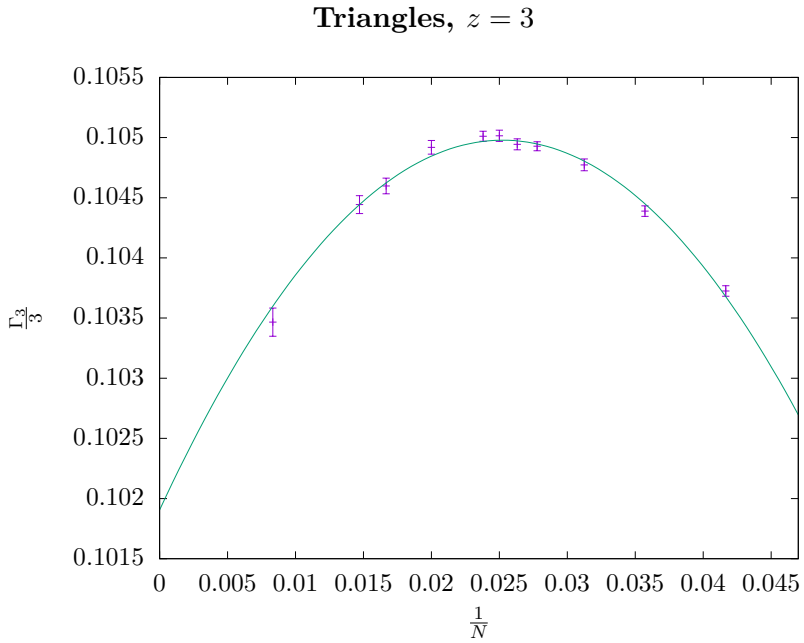


Figure 5.10. Average cost per length unit of a triangle on random regular graphs with nodes having degree $z = 3$. The fitting function has the form Equation 5.52.

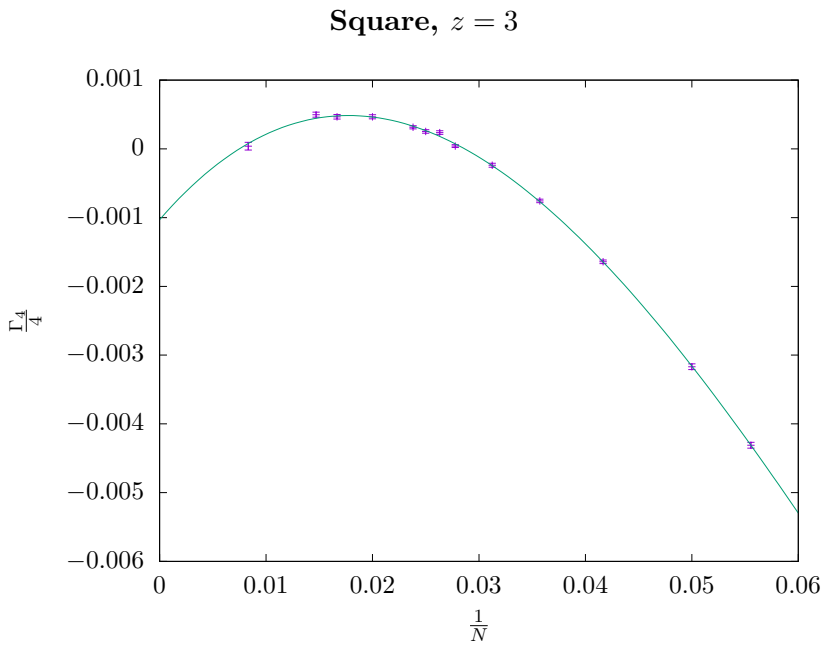


Figure 5.11. Each point represent the average cost per length unit of a square on random regular graphs with nodes having degree $z = 3$. The fitting function has the form Equation 5.52.

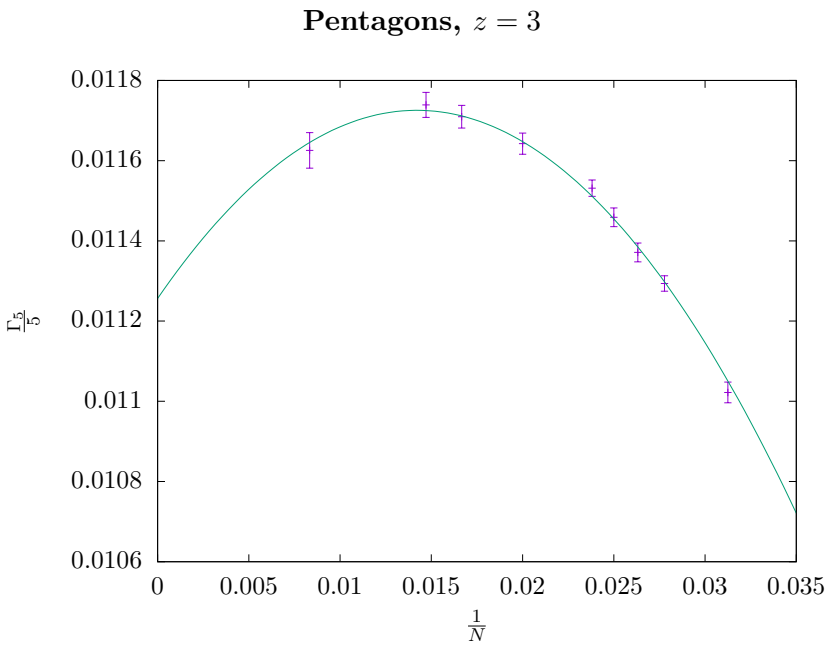


Figure 5.12. Each point represent the average cost per length unit of a pentagon on random regular graphs with nodes having degree $z = 3$. The fitting function has the form Equation 5.52.

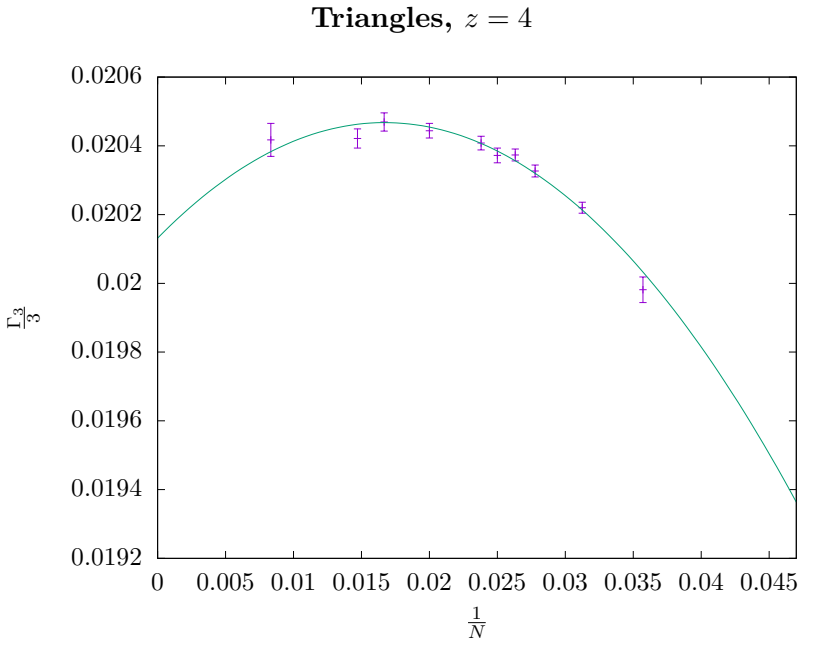


Figure 5.13. Each point represent the average cost per length unit of a triangle on random regular graphs with nodes having degree $z = 4$. The fitting function has the form Equation 5.52.

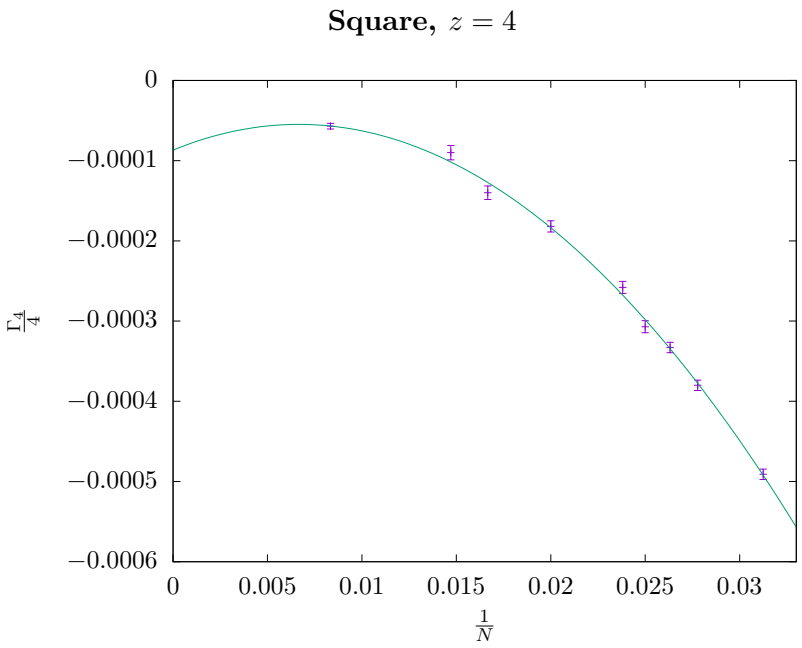


Figure 5.14. Each point represent the average cost per length unit of a square on random regular graphs with nodes having degree $z = 4$. The fitting function has the form Equation 5.52.

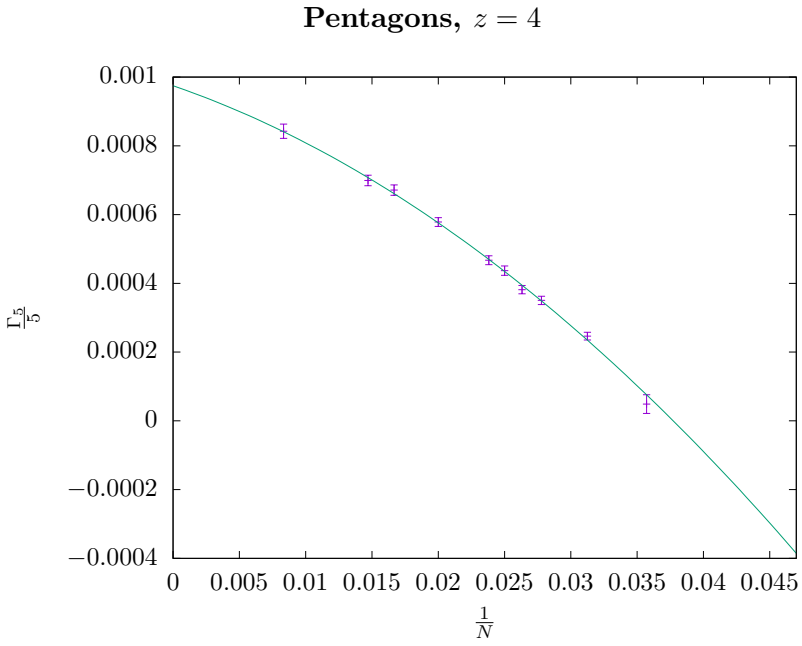


Figure 5.15. Each point represent the average cost per length unit of a pentagon on random regular graphs with nodes having degree $z = 4$. The fitting function has the form Equation 5.52.

	$z=3$		$z=4$	
	Simulations	Cavity	Simulations	Cavity
$\ell=3$				
$\Gamma_3/3$	0.1019 ± 0.0002	0.10167 ± 0.00001	0.02013 ± 0.00009	0.020118 ± 0.000005
c_1	-0.24 ± 0.01	–	0.040 ± 0.008	–
c_2	-4.8 ± 0.02	–	-1.2 ± 0.2	–
$\ell=4$				
$\Gamma_4/4$	-0.0010 ± 0.0001	-0.001144 ± 0.000005	-0.00009 ± 0.00001	-0.000098 ± 0.000004
c_1	0.18 ± 0.01	–	0.010 ± 0.002	–
c_2	-5.6 ± 0.5	–	0.72 ± 0.05	–
c_3	26 ± 0.5	–	–	–
$\ell=5$				
$\Gamma_5/5$	0.01126 ± 0.00006	0.011202 ± 0.000006	0.00097 ± 0.00003	0.000965 ± 0.000004
c_1	0.066 ± 0.005	–	-0.013 ± 0.003	–
c_2	-2.3 ± 0.1	–	-0.33 ± 0.07	–

Table 5.2. Summary table of the fit parameters and of the cavity predictions obtained with population dynamics for triangles, square and pentagons, and $z = 3, 4$.

cost on average. The numerical estimates of the costs of the polygons are in good agreement with the cavity predictions.

At this point we make the following conjecture about the form of the expansion of the average cost $\mathcal{E}$ in powers of $1/N$:

$$\mathcal{E} = \mathcal{E}_\infty + \frac{1}{N}\left(C + \sum_{\ell=3}^{\infty} \frac{(z-1)^\ell}{2\ell}\Gamma_\ell\right) + o\left(\frac{1}{N}\right), \tag{5.53}$$

where $\mathcal{E}_\infty$ is the asymptotic cost, C is a constant that is supposed not to depend on the presence of cycles of length $\ell \geq 3$, and $\frac{(z-1)^\ell}{2\ell} \equiv \langle \mathcal{N}_\ell \rangle$ is the average number of cycles of length ℓ on RRGs for $N \to \infty$. If the conjecture is true we expect that the connected correlations defined in Section 5.4.1 verify:

$$N\overline{\mathcal{E}(G)\mathcal{N}_\ell(G)}_c \xrightarrow{N\to\infty} \Gamma_\ell \overline{\mathcal{N}_\ell^2(G)}_c, \tag{5.54}$$

Since for large N, $\mathcal{N}_\ell$ has a Poisson distribution (chapter 1), and then $\overline{\mathcal{N}_\ell^2(G)}_c = \frac{(z-1)^\ell}{2\ell}$, we expect that:

$$N\overline{\mathcal{E}(G)\mathcal{N}_\ell(G)}_c \xrightarrow{N\to\infty} \Gamma_\ell \frac{(z-1)^\ell}{2\ell}. \tag{5.55}$$

This prevision is verified for $\ell = 3$, indeed, from the measure of the connected correlation functions, and Equation 5.44 it follows that:

$$\frac{N\overline{\mathcal{E}(G)\mathcal{N}_3(G)}_c}{3} \equiv a_3 = 0.136 \pm 0.001 \text{ is compatible with } \frac{4}{3}\frac{\Gamma_3}{3} = 0.1359 \pm 0.0003 \tag{5.56}$$

5.4.3 Fractional Matching

In this section we discuss some numerical results about the fractional matching problem on random regular graphs with valence $z = 3$. For the fractional matching

case a consideration allows to argue that the asymptotic ($N \to \infty$) cost density should be equal to the corresponding cost density of the matching, for every fixed valence z. The fundamental ingredient is that we are working on a locally tree-like ensemble of graphs in the limit $N \to \infty$. The argument is that on a tree with finite but arbitrarily large size, the space of feasible solutions of the matching problem is always equal to that of the fractional matching. Indeed the only edges that can be added to the matching with a non-integer occupation number are those that belong to a cycle, and on the trees there are no cycles.

In Figure 5.16 the average optimal cost densities for $30 \leq N \leq 120$ are reported, in the case of exponential distribution of the weights (5.2). Each point corresponds to the average of the optimal cost of 10^6 independent realizations of the system. Each realization has been constructed by generating a regular graph with the configurational model (1.1.2), and then by assigning to each edge a random weight, exactly as in the matching case.

As discussed in chapter 4, in the fully connected case the difference between the optimal cost density in the matching and in the fractional matching cases is equal, up to $o(1/N)$, to the term that is thought to be due to the odd-cycles of length $\ell \geq 3$:

$$\sum_{\substack{k \geq 3, \\ k \text{ odd}}} \frac{I_k}{k}, \tag{5.57}$$

where the I_k's are defined in chapter 4. If this fact still holds at z finite, one may try to see if the $1/N$-correction to the asymptotic cost density is equal to the term C/N of Equation 5.53. Moreover, if there are no terms that depend on cycles contributing to the $1/N$ correction of the fractional, one may expect that there are no terms proportional to $N^{-\frac{3}{2}}$, as a consequence of the interpretation given in Section 4.1.3.

This fact motivated to choose to fit the data represented in Figure 5.16 with a function of the form:

$$\mathcal{E} = \mathcal{E}_F + \frac{c_1}{N} + \frac{c_2}{N^2} + \frac{c_3}{N^3}. \tag{5.58}$$

One obtains the following parameters for the best fit:

$$\begin{aligned} \mathcal{E}_F &= 0.837905 \pm 0.000009 \\ c_1 &= -0.032 \pm 0.002 \\ c_2 &= -5.6 \pm 0.1 \\ c_3 &= 24 \pm 2 \end{aligned} \tag{5.59}$$

As previously argued, the asymptotic cost density is compatible with that of the matching problem in the case $z = 3$:

$$\widehat{\mathcal{E}}_3 = 0.83789 \pm 0.00002 \tag{5.60}$$

At this point, in order to test the hypothesis about the anomalous scaling above discussed, a possibility is to study the behavior of $N(\mathcal{E} - \mathcal{E}_F)$ as a function of $1/N$. This behavior is reported in Figure 5.17.

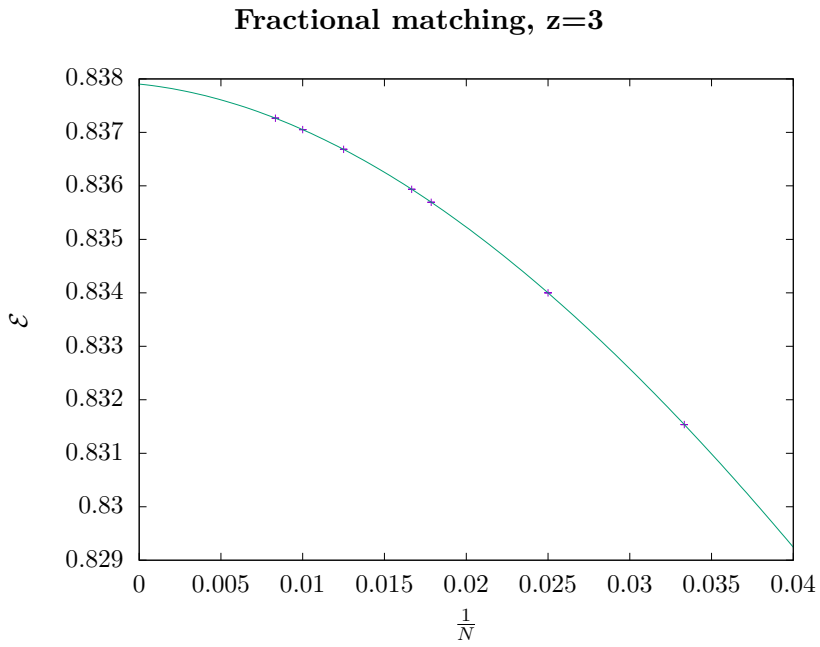


Figure 5.16. Average optimal cost density in the random link fractional matching problem on random regular graphs with connectivity $z = 3$, as a function of $1/N$.

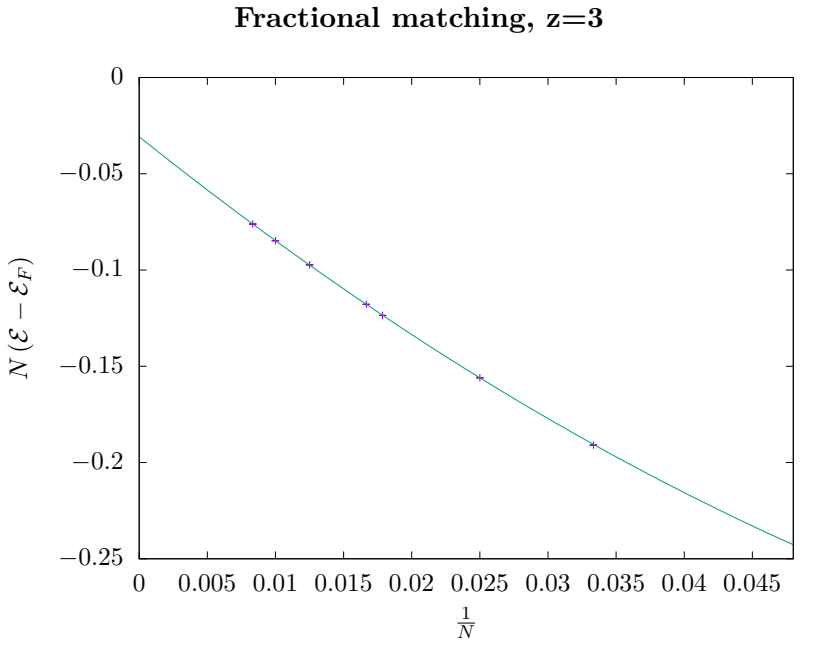


Figure 5.17. Behavior of $N(\mathcal{E} - \mathcal{E}_F)$ as a function of $1/N$. The solid line is a quadratic fit in $1/N$.

The fitting function has the form:

$$f(N) = c_1 + \frac{c_2}{N} + \frac{c_3}{N^2}, \tag{5.61}$$

z	$\mathcal{E}_z \pm \Delta\mathcal{E}_z$
20	0.78310 ± 0.00006
40	0.8026 ± 0.0002
60	0.80899 ± 0.00009
70	0.8109 ± 0.0002
80	0.8123 ± 0.0001
90	0.8133 ± 0.0001
100	0.8146 ± 0.0001

Table 5.3. Energy densities $\mathcal{E}_z$ estimated with a fitting function of the form Equation 5.64 for different values of the valence z.

with:

$$\begin{aligned} c_1 &= -0.0308 \pm 0.0003 \\ c_2 &= -5.65 \pm 0.03 \\ c_3 &= 26 \pm 1 \end{aligned} \tag{5.62}$$

In Figure 5.17 the fact that the data points are linearized for large N by the chosen scaling of the axes is an evidence of the absence of an anomalous correction proportional to $N^{-3/2}$. At this point we can discuss the idea that c_1 is equal to C, the first $1/N$-term of Equation 5.53. To avoid ambiguities we will denote by $c^f \equiv c_1$, the estimate of the $1/N$-correction of the fractional matching, and by $c^s = 1.096 \pm 0.006$ the estimate of the $1/N$-correction of the standard matching (Section 5.4.1). If we suppose that c^f is a measure of C, then, we expect that:

$$c^f + \sum_{\ell=3}^{\infty} \frac{(z-1)^\ell}{2\ell} \Gamma_\ell \quad \text{has to be compatible with} \quad c^s. \tag{5.63}$$

For $z = 3$, the left side of Equation 5.63, truncated at $\ell = 5$, is equal to *left*= 0.549 ± 0.002, that is not compatible with $c^s = 1.096 \pm 0.006$. This can be due to different reasons. First of all we must take into account the fact that we computed only the first three terms of the series; in order to test Equation 5.63 one has in principle to study the behavior of the contributions of higher length polygons. Moreover, regardless of the cancellation of the cycle-term, it is not guaranteed that the $1/N$-correction of the fractional matching is equal to C/N. This problem could be deepen by studying the costs of the cycles of the fractional matching.

These considerations suggest to test at finite z another cancellation that holds in the fully connected case: that one of the whole $1/N$-correction in the case of loopy matching, for weights distributed exponentially.

5.4.4 Loopy Fractional Matching

In this section we want to discuss some numerical simulations about the loopy fractional matching. The starting idea was to test if, even at finite z, for exponentially

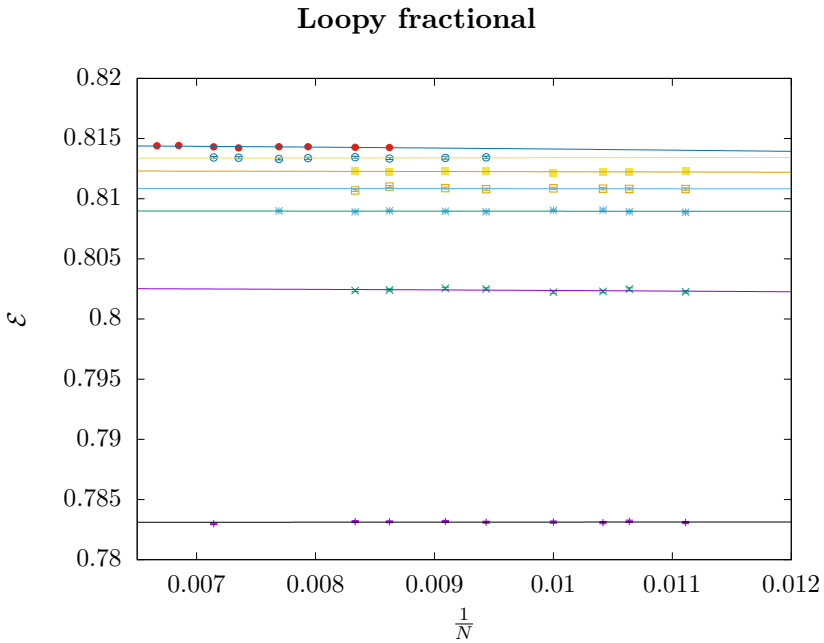


Figure 5.18. From the bottom to the top the data represent measure of the average cost density as a function of $1/N$ for $z = 20, 40, 60, 70, 80, 90, 100$.

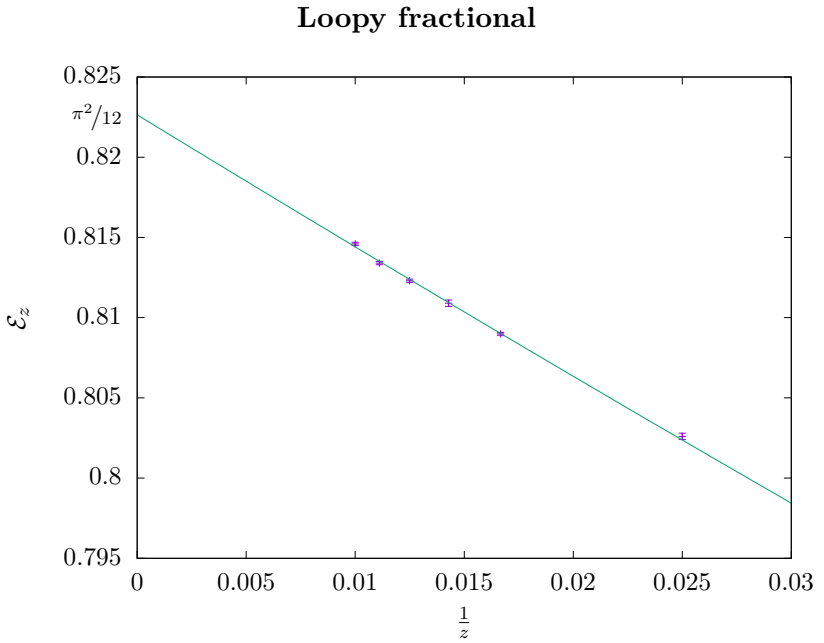


Figure 5.19. Asymptotic cost of the loopy fractional matching on the ensembles of RRG with connectivity z.

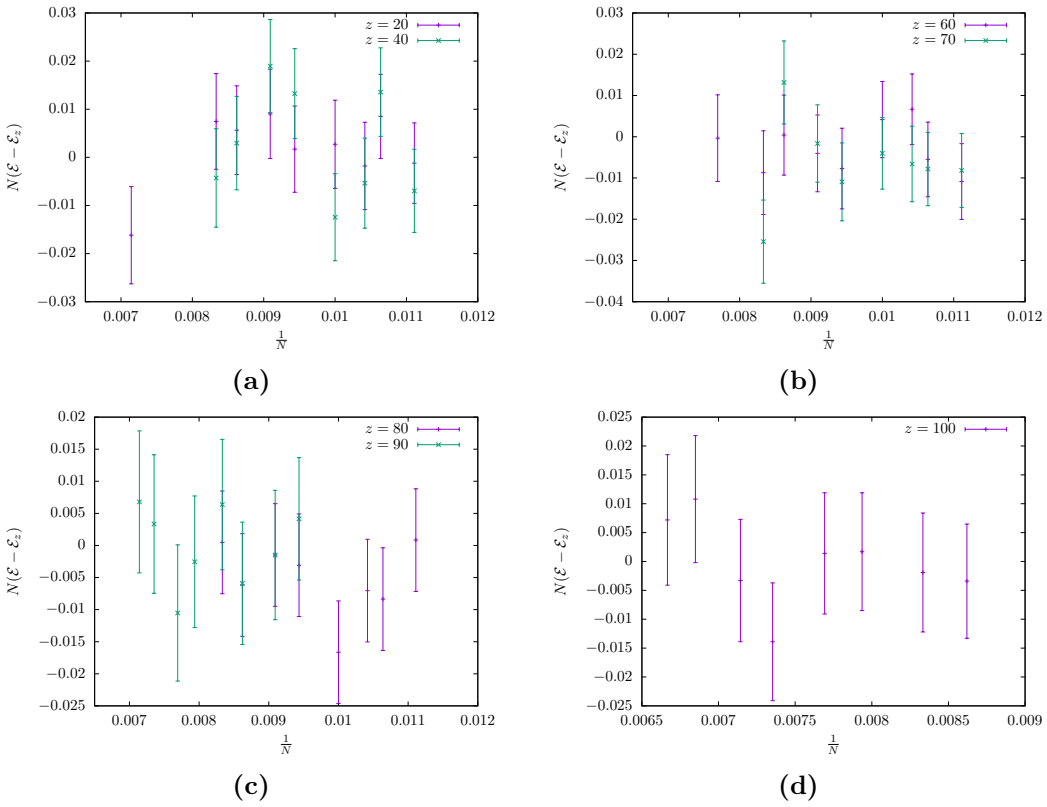


Figure 5.20. Measures of $N(\mathcal{E}-\mathcal{E}_z)$ as function of $1/N$, for different values of the connectivity z in the loopy fractional matching.

distributed weights there is a $1/N$-correction to the asymptotic cost density or not, as in the fully connected case (section 4.2). Nevertheless the estimates turned out to be noisy. More accurate considerations require more statistics.

The algorithm that has been used is analogous to that discussed in Section 5.4.2 for the costs of the cycles. Unlike Section 5.4.2 here we focus only on the ground state energy of the system.

The procedure was the following: we constructed a Markov chain of regular graphs, at each step we computed the average optimum energy by redrawing 10^3 times the weights associated with the edges, and finally, we averaged over the whole chain. As in the case of the cycles (Section 5.3) the chain was divided into blocks of length 50.

In addition to the corrections, we are also interested in checking if the limit of the asymptotic costs densities for $z \to \infty$ reproduces the expected $\pi^2/12$ (chapter 4). For large values of z, the generation of the first graph of the chain with the configurational method may be inefficient. For this reason we used the Havel-Hakimi theorem (chapter 1) to generate the first graph. However this generation is not uniform, and then, in order to "forget" the initial condition, we started to use the chain after $10^3|E|$ swaps of edges, starting from the first graph.

In Figure 5.18 are represented the cost densities $\mathcal{E}$ as functions of $1/N$ for $z = 20, 40, 60, 70, 80, 90, 100$, along with fitting functions of the form:

$$f(N) = \mathcal{E}_z + \frac{b_z}{N}. \tag{5.64}$$

Each point of Figure 5.18 corresponds to a chain of length 10^3. In the best-fitting functions of the form (5.64), the energy densities are those reported in Table 5.3, while all the b_z are compatible with 0. To study in greater detail the $1/N$ correction, in the figures 5.20 we report the behavior of the quantities $N(\mathcal{E} - \mathcal{E}_z)$ as a function of $1/N$ for different values of z. At this level of precision the data show no trend, and all the points are compatible with 0. This fact suggests that, even at finite z, there is not a $1/N$-correction to the asymptotic cost of the loopy fractional matching.

In Figure 5.19 we represent the estimates of the asymptotic costs $\mathcal{E}_z$ as a function of $1/z$. The data of Figure 5.19 are fitted with a function:

$$f(z) = c_1 + \frac{c_2}{z} + \frac{c_3}{z^2}, \tag{5.65}$$

and we obtained:

$$\begin{aligned} c_1 &= 0.82235 \pm 0.00006 \\ c_2 &= -0.812 \pm 0.006 \\ c_3 &= 0.55 \pm 0.09. \end{aligned} \tag{5.66}$$

Note that, as expected, c_1 is compatible with $\pi^2/12 = 0.822467....$

5.4.5 Assignment Problem

In this section we discuss some numerical results about the assignment problem on random regular graphs with valence $z = 3$. This problem is equal to that of section

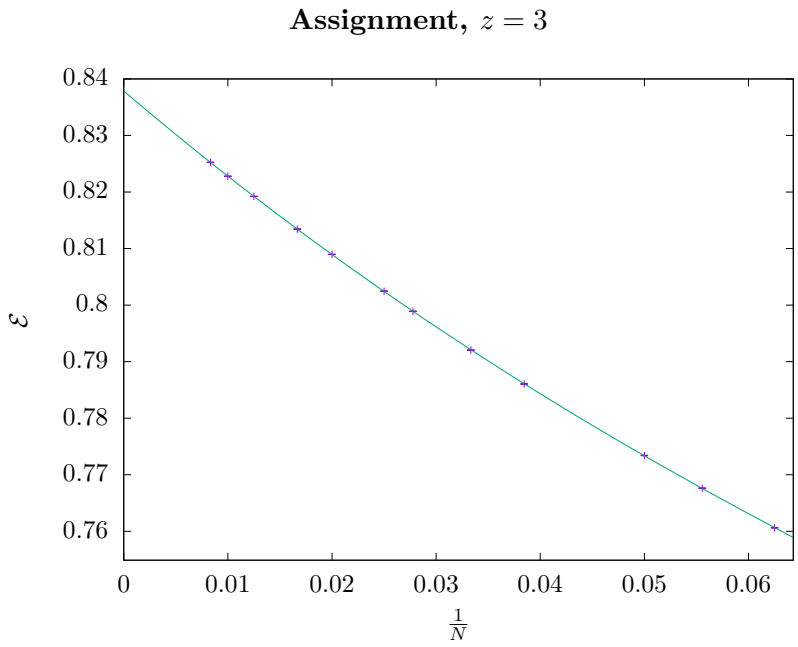


Figure 5.21. Density of cost of the assignment matching as a function of $1/N$ in the case of random bipartite regular graph, with bipartite set of equal cardinality, and $z = 3$. The costs of the edges are drawn independently from the exponential distribution (5.2).

5.4.1, except for the fact that now we have to consider only a subset of the whole set of regular graphs; indeed the assignment problem is defined on bipartite graphs $G = (W, Z; E)$, with $|W| = |Z|$. We may expect the average optimal cost in the limit $N \to \infty$ to be equal to that of the standard matching problem, as the RRG ensemble is locally tree like in the large graph limit, and it is immediate to see that all the trees can be thought as bipartite[2].

In Figure 5.21 the average optimal cost densities are reported for $16 \leq N \leq 120$, in the case of exponential distribution of the weights (5.2). Each point corresponds to the average of the optimal costs of 10^7 independent realizations of the system. Each realization has been constructed by generating a bipartite regular graph with bipartite sets of the same cardinality, and then by assigning to each edge of the random graph a random weight. The generation of the graphs has been done with the configurational model (subsection 1.1.2). As in the matching case, the uncertainties of the estimates are improved by appropriately summing and subtracting an heuristic cost.

Since in the fully connected case there are no anomalous scaling corrections to the optimal average thermodynamic cost at finite sizes, we try to fit the data with a function of the form:

$$\mathcal{E} = \mathcal{E}_A + \frac{c_1}{N} + \frac{c_2}{N^2} + \frac{c_3}{N^3}. \tag{5.67}$$

The fit gives:

[2] A graph G is bipartite if and only if it does not contain odd-cycles.

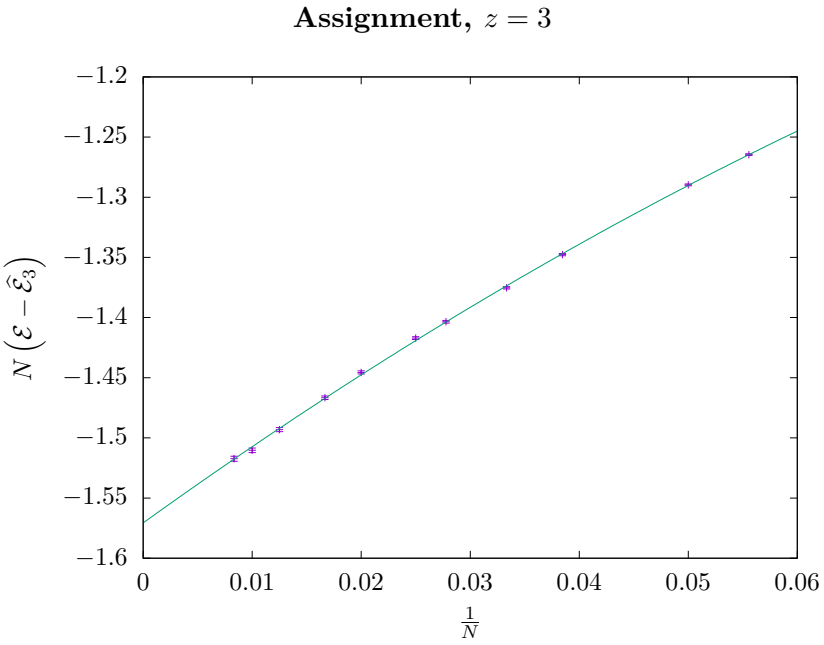


Figure 5.22. Behavior of $N(\mathcal{E} - \widehat{\mathcal{E}}_3)$ as a function of $1/N$. Numerical data points are shown along a quadratic fit in $1/N$. At this level of accuracy and number of acquisitions the data trend appears linearized for large N by the choice of the axis scaling.

$$\begin{aligned} \mathcal{E}_A &= 0.83786 \pm 0.00007 \\ c_1 &= -1.569 \pm 0.009 \\ c_2 &= 6.5 \pm 0.3 \\ c_3 &= -19 \pm 3, \end{aligned} \tag{5.68}$$

and then, as anticipated, $\mathcal{E}_A$ is in agreement with the estimate $\mathcal{E}_3$, and the cavity prediction $\widehat{\mathcal{E}}_3$. Now we can repeat the fit (5.67) by taking fixed and equal to $\widehat{\mathcal{E}}_3$ the zero-order term of the $1/N$-expansion of $\mathcal{E}$. This allows to improve the estimate of c_1, c_2, c_3. The new estimate can be done by fitting $N(\mathcal{E} - \widehat{\mathcal{E}})$, as in Figure 5.22. Choosing a function of the form:

$$f(N) = c_1 + \frac{c_2}{N} + \frac{c_3}{N^2}, \tag{5.69}$$

the best fit is defined by:

$$\begin{aligned} c_1 &= -1.57068 \pm 0.002361 \\ c_2 &= 6.5144 \pm 0.1495 \\ c_3 &= -18.0995 \pm 2.097 \end{aligned} \tag{5.70}$$

In this case one can observe that the $1/N$-correction, c_1, has opposite sign with respect to the $1/N$-correction of the standard matching. This difference, once again, could be due to the absence of odd-cycles on bipartite graphs.

Chapter 6

Conclusions

In this thesis work, we studied some random link matching problems on the ensemble of random regular graphs $\mathbb{G}_{\mathrm{RRG}}(N, z)$. For all cases, we assumed that the weights associated with the edges are independent and identically distributed according to an exponential law.

Of the standard matching, we studied the sample-averaged energy density $\mathcal{E}$ for $N \to \infty$ and the $1/N$-correction to $\mathcal{E}$ using the cavity method. The cavity estimates of $\mathcal{E}$, performed at the replica-symmetric level of approximation, turned to be in good agreement with the numerical simulations for $z = 3, 4$. The study of the $1/N$-correction was addressed by making a conjecture, motivated by some already known results about the matching problem on fully connected topologies. The fundamental idea is that such correction can be written as the sum of two contributions: one due to the presence of cycles on the graph, and one that is cycle-independent. The presence of a cycle-dependent contribution is corroborated by the numerical observation of an anomalous correction proportional to $N^{-3/2}$. The cavity estimations of the costs of the cycles of length $\ell = 3, 4, 5$ turns to be in good agreement with the numerical simulations for $z = 3, 4$. Assuming at finite z the validity of a property of the fully connected case, we tried to estimate the cycle-independent term, using the $1/N$-correction of the fractional matching. For $z = 3$ the validity of such property seems not to be consistent with our conjecture. About this, a more in-depth analysis is required.

We performed some preliminary numerical studies about the fractional matching, the loopy fractional matching, and the assignment problem. In the fractional case we did not observe for $z = 3$ an anomalous correction, as it happens on the fully connected topology. This suggests the absence of a cycle-dependent $1/N$-correction in this case.

For the loopy fractional we verified that for large z, $\mathcal{E}$ tends to $\pi^2/12$, as expected. Moreover, as it happens to the same problem on the fully connected topology, even at finite z there seems not to be present $1/N$-corrections to the asymptotic costs.

For the assignment problem we verified that the asymptotic cost for $z = 3$ is equal to the the asymptotic cost for $z = 3$ of the standard matching. The finite size corrections instead are different between the two cases. As the only difference between the two problems is that the assignment is defined on bipartite graphs, i.e. graphs without odd-cycles, this can be seen as a further confirmation of the presence

of cycle-dependent contributions in finite size corrections.

The programs written in this thesis for the numerical simulations can be reused for different purposes: it is possible to compute higher length polygon contributions, to study the costs of the polygons for large z for a comparison with the fully connected case, and to study in more detail the finite size corrections of the fractional and the loopy fractional matching. Moreover, the technique used to estimate the costs of the cycles has proved to be a powerful tool and can be readapted to other problems defined on random graph ensembles.

Bibliography

[AH89] K. Appel and W. Haken. "Every Planar Map is Four Colorable". American Mathematical Society, 1989. doi:10.1090/conm/098. [Cited on page 10.]

[Ald01] D. J. Aldous. "The $\zeta(2)$ limit in the random assignment problem". *Random Structures and Algorithms*, 18(4):381–418, 2001. doi:10.1002/rsa.1015. [Cited on page 40.]

[AM00] S. M. Aji and R. J. McEliece. "The generalized distributive law". *IEEE transactions on Information Theory*, 46(2):325–343, 2000. doi:10.1109/18.825794. [Cited on page 25.]

[AS04] D. Aldous and J. M. Steele. "The objective method: probabilistic combinatorial optimization and local weak convergence". In *Probability on discrete structures*, pages 1–72. Springer, 2004. [Cited on page 7.]

[BBB93] N. Biggs, N. L. Biggs, and E. N. Biggs. "Algebraic graph theory", volume 67. Cambridge university press, 1993. [Cited on page 2.]

[BFK+01] B. Bollobás, W. Fulton, A. Katok, F. Kirwan, and P. Sarnak. "Cambridge studies in advanced mathematics". In *Random graphs*, volume 73. Cambridge University Press New York, 2001. [Cited on page 3.]

[Bol80] B. Bollobás. "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs". *European Journal of Combinatorics*, 1(4):311–316, 1980. [Cited on page 8.]

[Bré10] E. Brézin. "Introduction to statistical field theory". Cambridge University Press, 2010. [Cited on pages 16 and 17.]

[Bru67] S. G. Brush. "History of the Lenz-Ising model". *Reviews of modern physics*, 39(4):883, 1967. [Cited on page 16.]

[BSS05] M. Bayati, D. Shah, and M. Sharma. "Maximum weight matching via max-product belief propagation". *arXiv preprint cs/0508101*, 2005. doi:10.1109/TIT.2007.915695. [Cited on page 28.]

[Cal07] G. Caldarelli. "Scale-free networks: complex webs in nature and technology". Oxford University Press, 2007. [Cited on page 3.]

[CC05] T. Castellani and A. Cavagna. "Spin-glass theory for pedestrians". *Journal of Statistical Mechanics: Theory and Experiment*, 2005(05):P05012, 2005. doi:10.1088/1742-5468/2005/05/P05012. [Cited on page 32.]

[CDMS17] S. Caracciolo, M. P. D'Achille, E. M. Malatesta, and G. Sicuro. "Finite-size corrections in the random assignment problem". *Physical Review E*, 95(5), may 2017. doi:10.1103/physreve.95.052129. [Cited on page 40.]

[Cla07] D. M. Clark. "Graphs and Optimization". *Journal of Inquiry Based Learning in Mathematics*, 2, 2007. [Cited on page 9.]

[CLRS09] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. "Introduction to algorithms". MIT press, 2009. [Cited on page 13.]

[DJK11] B. Dezső, A. Jüttner, and P. Kovács. "LEMON – an Open Source C++ Graph Template Library". *Electronic Notes in Theoretical Computer Science*, 264(5):23–45, jul 2011. doi:10.1016/j.entcs.2011.06.003. [Cited on page 55.]

[DMS13] A. Dembo, A. Montanari, and N. Sun. "Factor models on locally tree-like graphs". *The Annals of Probability*, 41(6):4162–4213, nov 2013. doi:10.1214/12-aop828. [Cited on page 29.]

[Edm09] J. Edmonds. "Paths, Trees, and Flowers". In *Classic Papers in Combinatorics*, pages 361–379. Birkhäuser Boston, 2009. doi:10.1007/978-0-8176-4842-8$_2$6. [Cited on page 55.]

[Ell11] D. Ellis. "The expansion of random regular graphs". *Lecture Notes, Lent*, 2011. [Cited on page 4.]

[FA86] Y. Fu and P. W. Anderson. "Application of statistical mechanics to NP-complete problems in combinatorial optimisation". *Journal of Physics A: Mathematical and General*, 19(9):1605–1620, jun 1986. doi:10.1088/0305-4470/19/9/033. [Cited on page 37.]

[FH93] K. H. Fischer and J. A. Hertz. "Spin glasses", volume 1. Cambridge university press, 1993. [Cited on page 32.]

[Gne17] B. V. Gnedenko. "Theory of probability". Routledge, 2017. [Cited on page 15.]

[Hua09] K. Huang. "Introduction to statistical physics". CRC press, 2009. [Cited on page 14.]

[Kad09] L. P. Kadanoff. "More is the same; phase transitions and mean field theories". *Journal of Statistical Physics*, 137(5-6):777, 2009. doi:10.1007/s10955-009-9814-1. [Cited on page 16.]

[KGV87] S. Kirkpatrick, C. Gelatt, and M. Vecchi. "Optimization by Simulated Annealing". In *Readings in Computer Vision*, pages 606–615. Elsevier, 1987. doi:10.1016/b978-0-08-051581-6.50059-3. [Cited on page 37.]

[KS78] S. Kirkpatrick and D. Sherrington. "Infinite-ranged models of spin-glasses". *Physical Review B*, 17(11):4384–4403, jun 1978. doi:10.1103/physrevb.17.4384. [Cited on page 33.]

[Kuh05] H. W. Kuhn. "The Hungarian method for the assignment problem". *Naval Research Logistics*, 52(1):7–21, feb 2005. doi:10.1002/nav.20053. [Cited on page 13.]

[KV18] B. Korte and J. Vygen. "Combinatorial Optimization". Springer Berlin Heidelberg, 2018. doi:10.1007/978-3-662-56039-6. [Cited on page 9.]

[LL60] L. Landau and E. Lifshitz. Classical mechanics. *ed: Pergamon Press, Oxford*, 1960. [Cited on page 14.]

[LL68] L. D. Landau and E. M. Lifshitz. "Statistical physics: V. 5: course of theoretical physics". Pergamon press, 1968. [Cited on page 14.]

[LMPS18] C. Lucibello, E. M. Malatesta, G. Parisi, and G. Sicuro. "The random fractional matching problem". *Journal of Statistical Mechanics: Theory and Experiment*, 2018(5):053301, may 2018. doi:10.1088/1742-5468/aabbc8. [Cited on pages 42 and 43.]

[LP09] L. Lovász and M. D. Plummer. "Matching theory", volume 367. American Mathematical Soc., 2009. [Cited on pages 11 and 13.]

[LPS17] C. Lucibello, G. Parisi, and G. Sicuro. "One-loop diagrams in the random Euclidean matching problem". *Physical Review E*, 95(1), jan 2017. doi:10.1103/physreve.95.012302. [Cited on page 41.]

[LW04] S. Linusson and J. Wästlund. "A proof of Parisi's conjecture on the random assignment problem". *Probability Theory and Related Fields*, 128(3):419–440, mar 2004. doi:10.1007/s00440-003-0308-9. [Cited on page 41.]

[Mej10] N. Mej. "Networks: an introduction", 2010. [Cited on page 3.]

[Méz15] M. Mézard. Cavity method: message-passing from a physics perspective. In "Statistical Physics, Optimization, Inference, and Message-Passing Algorithms", pages 95–126. Oxford University Press, dec 2015. doi:10.1093/acprof:oso/9780198743736.003.0004. [Cited on page 33.]

[Mil06] P. Miller. "Applied Asymptotic Analysis". American Mathematical Society, jun 2006. doi:10.1090/gsm/075. [Cited on page 33.]

[MKI+03] R. Milo, N. Kashtan, S. Itzkovitz, M. E. Newman, and U. Alon. "Uniform generation of random graphs with arbitrary degree sequences". *arXiv preprint cond-mat/0312028*, 106:1–4, 2003. [Cited on page 5.]

[MM09] M. Mezard and A. Montanari. "Information, physics, and computation". Oxford University Press, 2009. [Cited on pages 19, 25, 28, 29, 40, and 49.]

[MP85] M. Mézard and G. Parisi. "Replicas and optimization". *Journal de Physique Lettres*, 46(17):771–778, 1985. doi:10.1051/jphyslet:019850046017077100. [Cited on pages 37 and 39.]

[MP87] M. Mézard and G. Parisi. "On the solution of the random link matching problems". *Journal de Physique*, 48(9):1451–1459, 1987. doi:10.1051/jphys:019870048090145100. [Cited on pages 39 and 40.]

[MP01] M. Mézard and G. Parisi. "The Bethe lattice spin glass revisited". *The European Physical Journal B*, 20(2):217–233, mar 2001. doi:10.1007/pl00011099. [Cited on page 44.]

[MPV87] M. Mézard, G. Parisi, and M. Virasoro. "Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications", volume 9. World Scientific Publishing Company, 1987. doi:10.1142/0271. [Cited on pages 31, 32, 33, 37, and 38.]

[Mus10] G. Mussardo. "Statistical field theory: an introduction to exactly solved models in statistical physics". Oxford University Press, 2010. [Cited on pages 15 and 16.]

[MZ03] C. G. M. Mihail and E. Zegura. "The markov chain simulation method for generating connected power law random graphs". In *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments*, volume 111, page 16. SIAM, Philadelphia, 2003. [Cited on pages 5 and 6.]

[New18] M. Newman. "Networks". Oxford university press, 2018. [Cited on pages 3 and 4.]

[Nis01] H. Nishimori. "Statistical Physics of Spin Glasses and Information Processing". Oxford University Press, jul 2001. doi:10.1093/acprof:oso/9780198509417.001.0001. [Cited on page 33.]

[Nis05] M. Niss. "Phenomena, Models and Understanding: The Lenz-Ising Model and Critical Phenomena 1920-1971". PhD thesis, Roskilde Universitet, 2005. [Cited on page 16.]

[NPS05] C. Nair, B. Prabhakar, and M. Sharma. "Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures". *Random Structures and Algorithms*, 27(4):413–444, 2005. doi:10.1002/rsa.20084. [Cited on page 41.]

[Par83] G. Parisi. "Order Parameter for Spin-Glasses". *Physical Review Letters*, 50(24):1946–1948, jun 1983. doi:10.1103/physrevlett.50.1946. [Cited on page 33.]

[Par88] G. Parisi. "Statistical field theory". Addison-Wesley, 1988. [Cited on pages 14, 16, 17, and 20.]

[Par98] G. Parisi. "A conjecture on random bipartite matching". *arXiv preprint cond-mat/9801176*, 1998. [Cited on page 40.]

[PR02] G. Parisi and M. Ratiéville. "On the finite size corrections to some random matching problems". *The European Physical Journal B*, 29(3):457–468, oct 2002. doi:10.1140/epjb/e2002-00326-3. [Cited on pages 39 and 40.]

[PS98] C. H. Papadimitriou and K. Steiglitz. "Combinatorial optimization: algorithms and complexity". Courier Corporation, 1998. [Cited on page 11.]

[Sag13] S. Sagitov. "Weak convergence of probability measures". *Chalmers University of technology and Gothenburg University*, 2013. [Cited on page 7.]

[Sal11] J. Salez. "Some implications of local weak convergence for sparse random graphs". Theses, Université Pierre et Marie Curie - Paris VI ; Ecole Normale Supérieure de Paris - ENS Paris, July 2011. URL https://tel.archives-ouvertes.fr/tel-00637130. [Cited on page 7.]

[Tay81] R. Taylor. "Constrained switchings in graphs". In *Lecture Notes in Mathematics*, pages 314–336. Springer Berlin Heidelberg, 1981. doi:10.1007/bfb0091828. [Cited on page 6.]

[vdH16] R. van der Hofstad. "Random Graphs and Complex Networks". Cambridge University Press, 2016. doi:10.1017/9781316779422. [Cited on pages 3 and 7.]

[VL05] F. Viger and M. Latapy. "Fast generation of random connected graphs with prescribed degrees". *arXiv preprint cs/0502085*, 2005. [Cited on page 6.]

[VM84] J. Vannimenus and M. Mézard. "On the statistical mechanics of optimization problems of the travelling salesman type". *Journal de Physique Lettres*, 45(24):1145–1153, 1984. doi:10.1051/jphyslet:0198400450240114500. [Cited on page 38.]

[WÏ0] J. Wästlund. "The mean field traveling salesman and related problems". *Acta Mathematica*, 204(1):91–150, 2010. doi:10.1007/s11511-010-0046-7. [Cited on pages 42 and 43.]

[Wil99] T. G. Will. "Switching Distance Between Graphs with the Same Degrees". *SIAM Journal on Discrete Mathematics*, 12(3):298–306, jan 1999. doi:10.1137/s0895480197331156. [Cited on page 6.]

[Wol00] W. Wolf. "The Ising model and real magnetic materials". *Brazilian Journal of Physics*, 30(4):794–810, 2000. [Cited on page 16.]

[Wor80] N. C. Wormald. "Some problems in the enumeration of labelled graphs". *Bulletin of the Australian Mathematical Society*, 21(1):159–160, 1980. [Cited on page 8.]

[Wor81] N. C. Wormald. "The asymptotic distribution of short cycles in random regular graphs". *Journal of Combinatorial Theory, Series B*, 31(2):168–182, 1981. [Cited on page 8.]

[Wor99] N. C. Wormald. "Models of random regular graphs". *London Mathematical Society Lecture Note Series*, pages 239–298, 1999. [Cited on page 4.]

[WT01] M. Welling and Y. W. Teh. "Belief optimization for binary networks: A stable alternative to loopy belief propagation". In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 554–561. Morgan Kaufmann Publishers Inc., 2001. [Cited on page 28.]

[YFW03] J. S. Yedidia, W. T. Freeman, and Y. Weiss. "Understanding belief propagation and its generalizations". *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003. [Cited on pages 22, 25, and 28.]

[YFW05] J. S. Yedidia, W. T. Freeman, and Y. Weiss. "Constructing free-energy approximations and generalized belief propagation algorithms". *IEEE Transactions on information theory*, 51(7):2282–2312, 2005. doi:10.1109/TIT.2005.850085. [Cited on page 21.]