Università degli Studi di Ferrara

Dottorato di Ricerca in Matematica

in convenzione con Università degli Studi di Parma e

Università degli Studi di Modena e Reggio Emilia

Ciclo XXXI, Coordinatore Prof. Massimiliano Mella

A Monte Carlo method for large deviations applied to Erdös-Rényi random graphs

Settore Scientifico Disciplinare MAT/07

Dottoranda	**Tutore**	**Co-Tutore**
Dr. Elena Magnanini	Prof. Claudio Giberti	Prof. Cristian Giardinà

Anni 2015/2018

Contents

Acknowledgements

This thesis is the result of the doctoral research I have undertaken during the last three years. Mathematics made me discover places of my mind and of the world I would never thought to visit. There are so many people who deserve my gratitude for having made this journey possible. I would try to thank all of them for their constant support along this path.

First of all, my special thanks go to my two supervisors for their valuable guidance throughout my doctoral studies and for having taught me how to tackle a problem in research. Thank you for having introduced me to the world of large deviations on random graphs, a topic which is as beautiful as challenging to study. I want to thank Prof. Giberti for having helped me to pursue my research with strong mathematical rigor and a clever point of view on each and every aspect of our research. Thank you for your patience and for the care you put in improving my work with helpful suggestions. For all of that, I am very grateful to you. Likewise, I would like to thank my co-advisor Prof. Giardinà for the fruitful discussions on the algorithmic and theoretical issues related to our work, which gave me a clear view on what we were doing and pushed me to do my best. I admire the resolution and the perseverance you put in tackling the research.

I would like to thank you Marco Prato for the kindness you have always shown me, both on a personal and professional level. Your help has been crucial in achieving some numerical results of my problem.

A special thank goes to Prof. Remco van der Hofstad for having dedicated part of his time to discuss some critical points of my research project and for his precious suggestions. Your brilliant mind together with your kindness and your warm hospitality really inspired me. I am very grateful for having had the possibility to meet you along my path.

None of the work included in this thesis would have been possible without my family, who has always dreamed of me bigger than I made: every seed of love and trust has generated great things. You are my inspiration and the cornerstone of my life. I want to thank my mother Paola for the enormous amount of love, patience and continuous support she has been giving me over the years. Thank you for having escorted me throughout this beautiful life with constant care. To my father Roberto, thank you for your cheerful living, for your inexhaustible patience and for being so proud of me. I would also to express my gratitude to my aunt Lucia and my uncle Ivano for their affection and for our joyful and carefree moments. Finally, I want to express

my gratitude to my grandparents, for their powerful love and for being such a beautiful role model to me. To all of you, thank you for shaping me into the person I am today.

I want to thank my fellows at the PhD office in Modena, i.e. in order of appearance Simone, Vanna, Federica, Chiara, Elena, Carla, Lorenzo, Matteo, Franco, Salvatore, Giorgia, Mathilde, Micaela, Davide, Jean Paul, Francesca and to all members of the HiPeRT Lab for making this journey in the research world funnier and less stressful. Thank you for the countless lunches and dinners, the laughs, the jokes and all the amazing moments we had together during the last three years.

I am immensely grateful to my best friends Cristina, Giulia, Chiara, Silvia, Jessica, Veronica and Eleonora. There are no words for expressing how much I owe to each one of you. Let me say thank you for being so close to me with your love, your reassurance and, most of all, your constant presence in my life. I feel so blessed to have such kind of amazing girls by my side.

My last acknowledgment is dedicated to the team Roberto, Alessandro, Davide and Alberto. I am so grateful for each funny, happy, carefree, holiday moment we spent together: I know that I can always count on your support and your friendship.

Finally I want to thank every person that in these three years has been on my way, taking care of me, with a laugh or a supportive talk, sharing questions, answers, point of views. Each one of you has been the stone of the challenging path that has led me to this day.

Introduction

Rare events

Rare events have increasingly become relevant in the comprehension of problems arising from several domains of science [26, 20, 14, 1]. Indeed, many phenomena occurring in nature involve rare trajectories in dynamical systems and usually the most interesting events are not the typical ones. For example, in order to study chaotic planetary systems, one can rely on the observation of present conditions (within a range of observational error) to infer information about the past and the future. Depending on the initial state, trajectories can differentiate a lot and deviate from their standard behavior: in this context, the large deviations theory is an important tool for making realistic predictions, based on a statistical study of the evolution process. Another case of interest of large deviations is provided by molecular dynamics (and, in a similar way, by supercooled liquid and glasses) which is characterized by long periods of fluctuations around metastable configurations. Sometimes (rare) activation events happen: since they deserve attention as essential steps for chemical reactions, it is useful to have numerical methods for simulating the dynamics in a controlled way, without having to wait for their spontaneous appearance. Further fields of study of large deviations concern rogue waves and the transport of energy in a sample, which can be hindered or facilitated by exceptional trajectories.
Those examples could be tackled in principle by simulating for a long time the true dynamics; however this procedure is not feasible for complex large-scale systems. Most of the time, the smartest strategy for the investigation of such phenomena is to use algorithms able to properly change the probability of rare events in order to make them less infrequent: among such class of schemes, we find *Monte Carlo* methods [6, 32, 31] and the *umbrella sampling* technique [52, 42], which is widely used in computational chemistry when standard numerical strategies are unable to provide insight. From the mathematical point of view, the main challenge when we speak of rare events, is to find an analytical expression of the so-called *rate function*, which represents the exponential rate of decay of the examined event probability. Generally, the exact computation of the rate function is hard to perform and, for this reason, one of the possible goals of the aforementioned algorithms is to provide a numerical approximation of the function via an iterative procedure. The theory of large deviations can be a priori applied to any kind of event exhibiting an atypical behavior: this also includes the random graphs framework [7], which is also a central topic of this thesis.

Random Graphs models

In recent years, much attention has been devoted to the study of complex networks [44, 43] for analyzing problems in a wide class of fields such as information networks (e.g., World Wide Web, citation networks between academic papers), technological networks (e.g. Internet), biological networks (e.g. neural networks) and social networks. These last, are very popular nowadays and owe their name to the fact that they are used for modeling interactions between people such as friendships relationships rather than other kind of shared features. The mathematical tool used for modeling networks are graphs, which are objects consisting in a set of nodes (or vertices), which represent the agents of the interactions and a set of edges, which represent the possible connections between the vertices. Often, real world networks share common features, such as the so-called *small world* property, which, roughly speaking, means that given any two nodes, there is a high probability that they are connected by a short chain of edges. Another common property of real world networks is that they are *scale free*, namely the majority of vertices have few connections, whereas the number of vertices with high connections decays slowly. If we aim at modeling a network using random graphs, we have to choose the model of graph which suits the properties we want to study. For example, the simplest random graph one can build is the dense Erdös-Rényi one (originally defined in [25]): in this model, the connections between the nodes are present with constant probability p. Due to its simplicity, it is not suitable for representing real networks, however, despite its trivial formulation, it gives rise to complex scenarios when we start measuring the large deviations of some observable defined on it.

Large deviations on random graphs

The theory of Large Deviations can be applied to random graphs, for example every time we look at the event that a selected observable, defined on the graph, deviates from its expected value, when the number of nodes grows to infinity. The large deviation theory pursues two main goals: finding the probability of observing the atypical event but also investigating the structures which realize the deviation from the expected behavior (see [14]). The observable on which we want to focus our study is the number of triangles of a dense Erdös-Rényi random graph: the problem of deriving a large deviation principle for the upper tails of such quantity has been widely studied in literature and only a few question are still open [15]. For what concerns the lower tails, the state-of-art is less exhaustive and there are more gaps to fill [55]. A possible strategy used for studying both cases is to merge the Erdös-Rényi model into a wider one, the so-called *Exponential Random Graph* model [5]. It is based on the definition of an Hamiltonian which collects some desirable structures of the graph one wants to measure, such as the number of regular subgraphs properly normalized. For the purpose of studying the deviations of triangles, the suitable Hamiltonian should weigh the number of triangles properly normalized and the number of edges of the graph. In this context, it is possible to define "a pressure"and, according to the choice of the parameter which tunes the density of triangles,

to explore the so-called replica symmetric phase or replica breaking one. The first region has been fully characterized and the expression of the pressure is known in this regime; the second one is hard to explore and it's here that our investigation inserts.

Contribution of this thesis

In this thesis, a Monte Carlo method for the approximation of the *scaled cumulant generating function* of an additive observable is presented. Such function is strictly connected to the theory of large deviations since, when it is possible to apply the Gärtner-Ellis theorem, it turns out to be the Legendre transform of the rate function. Hence, it is possible to recover the first from the second and vice-versa. Moreover, it can be shown that the pressure of an Exponential Random Graph with Hamiltonian including only triangles and edges (properly normalized), coincides, up to an additive constant, with the scaled cumulant generating function of the number of triangles of a dense Erdös-Rényi random graph. In other words, the knowledge of the scaled cumulant generating function allows to provide answers on the replica breaking region mentioned in the paragraph above. The Monte Carlo method we focus on, which is called *Cloning algorithm* and was originally introduced in [29] and further in [28], uses an approach relying on population dynamics: it is based on the evolution of a family of copies of the system which are replicated or killed in such a way that atypical trajectories are favored.
The aim of this thesis is twofold: on one hand, we devise an extended version of the algorithm which works on growing-size graphs and keeps trace of the additive observable *number of triangles* of an Erdös-Rényi random graph; on the other hand, we explore the replica breaking region using three main strategies, among which the extension of the method appears.

Our first contribution consists in providing a formal analysis of the standard version of the method, stressing its strict connection with branching processes, together with numerical simulations on some simple models, which highlight the robustness of the algorithm.

In our second contribution we develop a modified version of the standard Cloning algorithm, which implements a dynamics over growing-size graphs. This strategy consists in looking at an Erdös-Rényi graph of size n as the result of an additive process in which, starting from a small and fixed size, a node is added at each step and connected (or not) to the previous ones with probability p. Since the connections are done with independence this process effectively builds a dense Erdös-Rényi graph. This formulation allows to write in an incremental way the observable one is interested in: it suffices to keep trace of its variations as long as the dynamics goes on. Firstly, we have tested the method on the simplest case of the edge observables, thus showing that the algorithm perfectly converges to the expected curve (which, in this context, is easy to derive). Secondly, we have moved on the triangle observable, tuning, as a starting approach, the parameter of the scaled cumulant generating function in such a way as to work in the so-called *replica symmetric* regime, where its analytical expression is known and, hence,

we have a comparison term for simulations. The major difficulty of the implementation relies on the fact that the method, being devised for running on spaces of graphs which become bigger and bigger, is time consuming. Despite this obstacle, it is still possible to appreciate the convergence of the method to the analytic result even if with less accuracy with respect to the edge case. Finally, still maintaining the setting on the replica symmetric region, we have formulated a heuristic argument, based on a mean-field approximation, for motivating the convergence of the method to the limiting curve.

In our third contribution we push the analysis towards the unresolved region, called *replica breaking*: concretely, this coincides with properly tuning the parameter related to triangles. In this part, the goal of our research is to numerically investigate the structure of the scaled cumulant generating function, whose analytical expression is not known here. Moreover, since the problem of recovering it configures as a variational one on a certain space of symmetric, measurable functions, we aim at finding the structure of the optimizers. We have tackled the problem using three different strategies: first of all, via a well-known optimization method, called *Projected Gradient* algorithm applied to a discretization of the problem; secondly, using the extended version of the Cloning method previously introduced and, finally, through an analysis of the functional (namely the objective function to optimize) over a specific class of functions. This third approach is motivated by the fact that we had an initial guess on the structure of one possible optimizer (in this region, it is not known if it is unique or not).

Organization

The thesis is organized as follows. In Chapter 1, we introduce the reader to the main definitions and theorems concerning the large deviations and the Markov chains theories. We have summarized the preliminaries required for the subsequent discussion. In Chapter 2, we present the Cloning algorithm in its standard setup: we make an analysis of the method stressing its connection with branching processes and we provide some numerical applications which highlight its effectiveness. In Chapter 3, we describe the state-of-art related to the large deviations theory of the triangle observable, keeping the focus on the dense Erdös-Rényi model. We characterize the replica symmetric phase and the replica breaking one for both the rate function and the scaled cumulant generating function and we highlight the open points related to this last region. This chapter basically collects results taken from [18, 17, 19, 15]. In Chapter 4 we present the extended version of the Cloning method, devised for working of growing-size spaces of graphs. We provide the results of simulations for the edge observable and for the triangle one, working on the replica symmetric setting. Finally, we give a heuristic argument for motivating the convergence of the Cloning algorithm to the expected curve (which is known, in this regime). In Chapter 5 we go back to the open problems concerning the replica breaking region and, making use of three different strategies, we tackle the challenge of investigating the behavior of the scaled cumulant generating function here.

Notations

- i.i.d: identically distributed random variables.
- $P(\cdot,\cdot)$ the transition probability of a Markov chain.
- $\mathbb{P}_{n,p}$ the dense, Erdös-Rényi distribution of a graph of size n.
- $\mathbb{E}_p^{ER}$ the expectation with respect to the Erdös-Rényi measure.
- $\{G_{n,p}\}_n$ a sequence of Erdös-Rényi random graphs
- $\mathcal{G}_{n,p}$ the space of all Erdös-Rényi random graphs of size n.
- $\{G_n\}_n$ a sequence of exponential random graphs.
- $\mathcal{H}(G_n)$ the Hamiltonian corresponding to the exponential random graphs model.
- H_1 the edge subgraph and H_2 the triangle subgraph.
- $X^{(n)}$ the adjacency matrix of a graph of size n, with elements $\{X_{ij}\}_{i,j=1,\dots,n}$.
- $T(X^{(n)})$ the number of triangles of a graph of size n.
- $E(X^{(n)})$ the number of edges of a graph of size n.
- $\widetilde{W}$ the space of all graphons equipped with the cut distance $\delta_\square(\cdot,\cdot)$.
- $\tilde{f}$ the elements of the space $\widetilde{W}$.
- $\tilde{P}_\alpha(\cdot,\cdot)$ the tilted probability of the Cloning algorithm.
- $P_\alpha(\cdot,\cdot)$ the normalized transition probability of the Cloning algorithm.
- $k(\cdot)$ the reproduction rate of each clone in the Cloning algorithm.
- $\mu(\alpha)$ the scaled cumulant generating function.
- $I(r)$ the large deviations rate function.
- $I_p(r)$ the Bernoulli relative entropy.

- Ω_n the set of all symmetric adjacency matrices of an Erdös-Rényi random graph of size n, with null elements on the diagonal.
- $< x, y >= \sum_{i=1}^{n} x_i y_i$ the scalar product, with $x, y \in \mathbb{R}^n$.

Chapter 1

Introduction to large deviations

1.1 Motivations

The theory of large deviations is strictly related to rare events, since it deals with the exponential rate of decay of their probabilities, seen as a function of some parameter, such as the temperature of a chemical reaction or a critical noise perturbing a dynamical system. There are many examples, ranging from chemistry ([26]) to turbulence in fluids ([20]) and from theory of random networks ([14]) to economics ([1]), which show that the most interesting situations from the scientific point of view, are not the typical ones but those which are in some sense more difficult to observe, predict and control. For instance, molecular dynamics provides an example of large deviations problem: this is the case of the amorphous state of the glass, whose molecules are described as frozen in a disordered state such that, if we wait a time long enough (of the order of the universe age) the rare thermal fluctuation are able to move the atoms and, in the long, to run the system in a regular status. As described in [14] for the probabilistic framework, the theory of large deviations has developed powerful tools for dealing with linear functionals of independent random objects. For what concerns non-linearity, there is still room for strengthening the mathematical background and we mention the work of Chatterjee [16] among the scientific contributions in this direction. Non- linearity arises in a natural way when we analyze real world networks. For example, in the case of social networks, a natural question is the following: if A is a friend of B and C is a friend of B, is there an increased probability that C will be a friend of A? If this is the case, then the system satisfies the clustering property. Thus, clustering deals with the probability of finding triangles, whose number is a nonlinear functional of the variables that define the network. Furthermore, imagine that you are interested in quantifying the probability that a selected substructure of your network (for example the aforementioned number of triangles), exceeds its average value when the number of interacting agents becomes bigger and bigger. This is an example of large deviation problem, since we look at the probability that a certain observable deviates from its typical value. From the mathematical point of view, solving a large deviation problem consists in finding the

exponential rate at which the probability of the atypical event we are interested in, decreases: this coincides with finding a specific function, called *rate function* which owes its name from this fact. A part from some standard and easy examples, in the majority of cases the direct computation of such function is not easy to perform and one has to rely on numeric simulations such as Monte Carlo methods (we quote [29] and [28] among the wide group of papers which treat this topic).

1.2 Large deviations results

This section introduces some useful definitions and results of large deviation analysis that will be fundamental for the subsequent discussion. A more exhaustive overview of these topics can be found in [21],[22],[24]. The crucial problem when we deal with large deviations is to find a function which satisfies the so-called *large deviation principle*: we start providing the general formulation of it.
Let $\mathcal{X}$ be a Polish space, i.e a complete, separable metric space, equipped with the distance $d : \mathcal{X} \times \mathcal{X} \to [0, +\infty)$. As a preliminary step, we recall the definition of *lower semi-continuous* function.

Definition 1.1 ([22], Definition III.1)**.** *[Lower semi-continuity] The function $f : \mathcal{X} \longmapsto [-\infty, +\infty]$ is lower semi-continuous if it satisfies any of the following properties:*

(i) $\liminf_{n\to+\infty} f(x_n) \geq f(x)$ *for all* $(x_n)_n$*,* x *such that* $x_n \to x$ *in* $\mathcal{X}$*.*

(ii) $\lim_{\varepsilon\to 0} \inf_{y\in \mathcal{B}_\varepsilon(x)} f(y) = f(x)$ *with* $\mathcal{B}_\varepsilon(x) = \{y \in \mathcal{X} : d(x, y) < \varepsilon\}$*.*

(iii) f *has closed level sets, i.e,* $f^{-1}([-\infty, c]) = \{x \in \mathcal{X} : f(x) \leq c\}$ *is closed for all* $c \in \mathbb{R}$*.*

We recall the following property of lower semi-continuous functions:

Proposition 1.1 ([22], Lemma III.3)**.** *A lower semi-continuous function attains a minimum on every non-empty, compact set.*

Lower semi-continuity is a necessary property for a function to be a rate function, as we can see from the definition below:

Definition 1.2 ([22], Definition III.5)**.** *[Rate function] The function $I : \mathcal{X} \longmapsto [0, +\infty]$ is called a rate function if:*

a) $I \not\equiv +\infty$*;*

b) I is lower semi-continuous;

c) I has compact level sets.

At this point we are ready to state the large deviation principle (LDP):

Definition 1.3 ([22], Definition III.6)**.** *[LDP] A sequence of probability measures P_n on $\mathcal{X}$ is said to obey a large deviation principle with speed n and rate function I if*

a) I is a rate function in the sense of Definition 1.2;

b) for any closed set $C \subset \mathcal{X}$:

$$\limsup_{n\to+\infty} \frac{1}{n} \ln P_n(C) \leq - \inf_{r\in C} I(r); \tag{1.1}$$

c) for any open set $O \subset \mathcal{X}$:

$$\liminf_{n\to+\infty} \frac{1}{n} \ln P_n(O) \geq - \inf_{r\in O} I(r). \tag{1.2}$$

Remark 1.1. Consider now a non-empty subset $\Gamma \subset \mathcal{X}$ and denote by $\bar{\Gamma}$ the closure of Γ and by Γ^o the interior of Γ: since $\Gamma^o \subseteq \Gamma \subseteq \bar{\Gamma}$, assuming conditions b) and c), the following chain of inequalities holds:

$$\begin{aligned} - \inf_{r\in\Gamma^o} I(r) \leq \liminf_{n\to+\infty} \frac{1}{n} \ln P_n(\Gamma^o) \leq \liminf_{n\to+\infty} \frac{1}{n} \ln P_n(\Gamma) \leq \limsup_{n\to+\infty} \frac{1}{n} \ln P_n(\Gamma) \leq \limsup_{n\to+\infty} \frac{1}{n} \ln P_n(\bar{\Gamma}) \\ \leq - \inf_{r\in\bar{\Gamma}} I(r), \end{aligned}$$

where the first inequality and the last one derive from (1.1) and (1.2).
It follows that if $\Gamma \subset \mathcal{X}$ is such that

$$\inf_{r\in\bar{\Gamma}} I(r) = \inf_{r\in\Gamma^o} I(r)$$

then

$$\lim_{n\to+\infty} \frac{1}{n} \ln P_n(\Gamma) = - \inf_{r\in\Gamma} I(r),$$

that is the standard way we are used to interpreting the large deviation principle.

Theorem 1.1 ([22], Theorem III.8)**.** *Let P_n satisfy the LDP. Then the associated rate function is unique.*

In general, for a given random sequence $(X_n)_n$, there are three different approaches whereby proving that it satisfies a large deviation principle: the direct way consists in showing that the probability distribution of $(X_n)_n$ satisfies a large deviation principle, according to Definition 1.3. The second one, the indirect method, consists in finding certain functions of $(X_n)_n$ which allow to work out a large deviation principle for the starting random sequence. Finally the third approach, known as contraction method, allows to relate the random sequence $(X_n)_n$ to

another one which satisfies a large deviation principle and then infer from this a large deviation principle for $(X_n)_n$. One of the main results of large deviation theory, the so-called Gärtner-Ellis theorem, makes part of the indirect strategies for deriving an LDP without knowing the distribution of the random variable: before stating the theorem, we introduce the underlying setup, strictly following [24]. Let $W = \{W_n; n = 1, 2, \dots\}$ be a sequence of random vectors which are defined on probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ and which take values in $\mathbb{R}^d$. We define a family of so-called, *moment generating functions*

$$\mu_n(\alpha) = \frac{1}{a_n} \ln \mathbb{E}_n \left(e^{<\alpha, W_n>} \right) \quad n = 1, 2, \dots, \quad \alpha \in \mathbb{R}^d, \tag{1.3}$$

where $\{a_n; n = 1, 2 \dots\}$ is a sequence of positive numbers tending to infinity, $\mathbb{E}_n(\cdot)$ denotes the expectation with respect to P_n and $< \cdot, \cdot >$ indicates the Euclidean inner product on $\mathbb{R}^d$. The function

$$\mu(\alpha) := \lim_{n\to\infty} \mu_n(\alpha) = \lim_{n\to+\infty} \frac{1}{a_n} \ln \mathbb{E}_n \left(e^{<\alpha, W_n>} \right) \quad n = 1, 2, \dots, \quad \alpha \in \mathbb{R}^d, \tag{1.4}$$

provided it exists, is called *scaled cumulant generating function.*
The following hypothesis are assumed to hold:

(a) each function $\mu_n(\alpha)$ is finite for all $\alpha \in \mathbb{R}^d$;

(b) $\mu(\alpha)$ exists for all $\alpha \in \mathbb{R}^d$ and is finite.

For example, hypothesis (a) and (b) are satisfied when W_n is the *nth* partial sum of $X_1, \dots, X_n$ i.i.d random vectors taking values in $\mathbb{R}^d$, i.e when $W_n \equiv S_n := \sum_{i=1}^n X_i$, $n = 1, 2, \dots$, and $\mathbb{E}(e^{<\alpha, X_1>})$ is finite for all $\alpha \in \mathbb{R}^d$. In this case, with $a_n = n$ we have

$$\mu_n(\alpha) = \frac{1}{n} \ln \mathbb{E}(< \alpha, W_n >) = \ln \mathbb{E}(< \alpha, X_1 >) = \mu(\alpha).$$

Cramer's theorem states that the distributions of $\frac{W_n}{n}$ on $\mathbb{R}^d$ have a large deviations property with rate function

$$I(r) = \sup_{\alpha\in\mathbb{R}^d} \{< \alpha, r > -\mu(\alpha)\}, \quad r \in \mathbb{R}^d, \tag{1.5}$$

where $I(r)$ is the Legendre transform of $\mu(\alpha)$. We recall the definition below:

Definition 1.4 (Legendre-Fenchel transform)**.** *The Legendre-Fenchel transform of a function $f : \mathbb{R}^d \longmapsto \mathbb{R}$ is a function $g : \mathbb{R}^d \longmapsto \mathbb{R}$ defined by*

$$g(r) = \sup_{x\in\mathbb{R}^d} \{< r, x > -f(x)\}, \quad r \in \mathbb{R}^d. \tag{1.6}$$

This transform can be denoted in a compact form by $g = f^*$.

When hypothesis (a) and (b) hold, it is possible to generalize Cramer's theorem to the non i.i.d setting, thanks to the Gärtner-Ellis theorem, reported below.

Theorem 1.2 ([24], Theorem II.6.1)**.** *[Gärtner-Ellis Theorem] Assume hypothesis (a) and (b). Let Q_n be the distribution of $\frac{W_n}{a_n}$ on $\mathbb{R}^d$ and let*

$$I(r) = \sup_{\alpha \in \mathbb{R}^d} \{ < \alpha, r > -\mu(\alpha) \}, \quad r \in \mathbb{R}^d, \tag{1.7}$$

where $\mu(\alpha)$ is defined in (1.4)*. Then, the following conclusions hold:*

i) $I(r)$ is convex, lower semi-continuous and non-negative. $I(r)$ has compact level sets and $\inf_{r \in \mathbb{R}^d} I(r) = 0$.

ii) The upper large deviation bound is valid:

$$\limsup_{n \to +\infty} \frac{1}{n} \ln Q_n(C) \leq - \inf_{r \in C} I(r) \quad \text{for each closed set } C \in \mathbb{R}^d. \tag{1.8}$$

iii) Assume in addition that $\mu(\alpha)$ is differentiable for all α. Then, the lower deviation bound is valid:

$$\limsup_{n \to +\infty} \frac{1}{n} \ln Q_n(O) \geq - \inf_{r \in O} I(r) \quad \text{for each open set } O \in \mathbb{R}^d. \tag{1.9}$$

Hence, if $\mu(\alpha)$ is differentiable for all α, then (1.9)*, parts (i) and (ii) imply that $\{Q_n; n = 1, 2 \dots\}$ satisfies a large deviation principle with rate function $I(r)$.*

As described in [22], the Legendre transform of the moment generating function naturally arises as lower bound in the large deviation principle using the exponential Chebyshev's inequality and optimizing over the parameter. If that also holds for the upper bound (which is the most difficult part), then the rate function must be the Legendre transform and (1.7) holds.

Remark 1.2. The Legendre transform of the scaled cumulant generating function yields functions that are necessarily convex: this is due to the fact that $\mu(\alpha)$ is always convex and the Legendre transform of a convex function is convex. Moreover, when $\mu(\alpha)$ is differentiable, the Legendre transform yields to strictly convex functions, namely convex functions with no linear parts (we refer to Section 26 of [48] for the proof). However, there is a priori no reason why rate functions should be convex.

Using a shorthand notation we can resume that if $I(r)$ is convex then $I(r) = \mu^*(\alpha)$ whereas if $I(r)$ is not convex $I(r) \neq \mu^*(\alpha)$. The opposite, i.e $\mu(\alpha) = I^*(r)$, is always true. We report some properties of the scaled cumulant generating function, which could be helpful in clarifying how its critical issues reflect on the rate function and vice-versa. We refer to [53] for a detailed analysis:

a) $\mu(0) = 0$ (this follows by the definition);

b) $\mu(\alpha)$ is convex so it can be non-differentiable only at isolated points;

c) if $I(r)$ is strictly convex then $\mu(\alpha)$ is differentiable;

d) if $I(r)$ is not convex (or it is convex with linear parts) then $\mu(\alpha)$ has at least one non-differentiable point;

e) if $\mu(\alpha)$ is differentiable then the value α such that $\mu'(\alpha) = r$ has the property that $\alpha = I'(r)$. This duality property between the rate function and the scaled cumulant generating function can be also expressed saying that the slopes of $\mu(\alpha)$ correspond to the abscissas of $I(r)$ and vice-versa.

Finally, we quote a general statement from [53] that can be helpful in characterizing the Legendre-Fenchel transform in this context of large deviations theory: such result is described in [53] with explicative examples and the proof can be found in [48] (see also Chap VI of [24] and [51]).

Proposition 1.2 ([53], Result 2)**.** *If $I(r)$ is not convex, the Legendre-Fenchel transform of $\mu(\alpha)$ does not yields $I(r)$, rather it yields the convex envelope of $I(r)$.*

As a direct consequence of Remark 1.2 emerges that non-convex rate functions can not derive from the application of the Gärtner-Ellis theorem and the breakdown of the theorem for this case relies on the differentiability of $\mu(\alpha)$. What about the inverse operation? Is it possible to recover the scaled cumulant generating function once the rate function is known? The answer is yes, and the result is provided by the the Varadhan's theorem (firstly introduced in [54]), which is a generalization of the Laplace's method (or Laplace's approximation) and gives the asymptotic behavior of a large class of integrals. The formulation below is taken from [22].

Theorem 1.3 ([22], Theorem III.3)**.** *[Varadhan's Theorem] Let P_n be a sequence of probability measures which satisfy the large deviation principle with rate function I. Let $f : \mathcal{X} \longmapsto \mathbb{R}$ be a continuous function that is bounded from above, where $\mathcal{X}$ is the Polish space introduced at the beginning of the paragraph. Then*

$$\lim_{n\to+\infty} \frac{1}{n} \ln \int_{\mathcal{X}} P_n(x) e^{nf(x)} dx = \sup_{x\in\mathcal{X}} [f(x) - I(x)]. \tag{1.10}$$

Let us apply the result to a generic sequence $(W_n)_n$ of random vectors with values in $\mathbb{R}^d$, for reading the displayed formulation in a more familiar way: equation (1.10) becomes

$$\lim_{n\to+\infty} \frac{1}{n} \ln \mathbb{E}(e^{nf(W_n)}) = \lim_{n\to+\infty} \frac{1}{n} \ln \int_{\mathbb{R}^d} P_n(r) e^{nf(r)} dr = \sup_{r\in\mathbb{R}^d} [f(r) - I(r)], \tag{1.11}$$

where we recall that $P_n(r)$ represents the probability distribution of W_n. Furthermore, for a fixed $\alpha \in \mathbb{R}^d$ if we set $f(\alpha) =< \alpha, r >$, with $r \in \mathbb{R}^d$, in (1.11) we get, recalling definition (1.4),

$$\mu(\alpha) = \sup_{r\in\mathbb{R}^d} [< \alpha, r > -I(r)], \quad \alpha \in \mathbb{R}^d. \tag{1.12}$$

In other words, as a direct consequence of Varadhan's theorem, it is always possible to recover the scaled cumulant generating function once that the rate function is known, whereas

the opposite operation is allowed only under proper differentiability hypothesis on the scaled cumulant generating function. This represents a crucial point one has to take into consideration in making use of the Gärtner-Ellis theorem for the derivation of the large deviation rate function.
Generally the dissertation on large deviations starts from the simplest setting of independent identically distributed (i.i.d) random variables and then generalizes to the non-independent case. We recall that the first derivation of the large deviation principle for non-independent sequences concerns the discrete-time Markov chains. In these processes the probability of each event depends on the state attained at the previous one, so we lose the independence while maintaining the simplicity of the overall dynamics.

1.3 Basic results on Markov chains

In this part, we report a few basic definitions and properties of discrete-time Markov chains which will turn to be useful throughout Chapter 2.

Definition 1.5 (Discrete time Markov chain)**.** *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(X_i)_{i\geq 0}$ a sequence of random variables defined on Ω and mapping into the finite space $S = \{1, \dots, N\}$. If for all positive integers k and all $x_0, x_1, \dots, x_{k+1} \in S$*

$$\mathbb{P}(X_{k+1} = x_{k+1} | X_k = x_k, \dots, X_0 = x_0) = \mathbb{P}(X_{k+1} = x_{k+1} | X_k = x_k), \tag{1.13}$$

then the sequence $(X_i)_{i\geq 0}$ is called a finite state, discrete Markov chain.

The space S of all possible states is called *phase space* or *alphabet*: the matrix of transition probabilities (from state $x \in S$ to state $y \in S$) at time $k \in \mathbb{N}$, is denoted by P and it is a $N \times N$ matrix (called *transition matrix*) with elements $p(x, y)$, where

$$p(x, y) = \mathbb{P}(X_{k+1} = y | X_k = x), \quad x, y \in S.$$

Such elements must satisfy the properties

$$p(x, y) \geq 0$$
$$\sum_{y \in S} p(x, y) = 1.$$

A realization (or path) of the process up to time k is a sequence $(x_0, x_1, \dots, x_k) \in S^{k+1}$: if the process at time 0 starts from the probability distribution ν, namely $\nu(x_0) = \mathbb{P}(X_0 = x_0)$, then the probability associated to a realization is

$$\begin{aligned} \mathbb{P}^X[x_1, \dots, x_k] &:= \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_k = x_k) = \nu(x_0) p(x_0, x_1) \dots p(x_{k-1}, x_k) \\ &= \nu(x_0) \prod_{i=1}^{k-1} p(x_i, x_{i+1}). \end{aligned} \tag{1.14}$$

When we deal with Markov processes, it often happens to make use of the power of the transition matrix in order to describe the evolution of the system, hence it is convenient to fix a notation for it: we denote by $P^{(n)}$ the nth power of the matrix and by $p^{(n)}(x,y)$ its elements. The distribution at time $n+1$ of the chain is the row vector $\nu^{(n+1)}$ with components

$$\nu^{(n+1)}(y) = \mathbb{P}(X_{n+1} = y) = \sum_{x \in S} \nu^{(n)}(x) p(x,y),$$

that is, in matrix form, $\nu^{(n+1)} = \nu^{(n)} \cdot P$, where this last notation denotes the vector-matrix product. Iteration of this equality yields

$$\nu^{(n+1)} = \nu \cdot P^{(n+1)}. \tag{1.15}$$

It is useful to recall the following definitions:

Definition 1.6 (Stationary measure)**.** *A vector $\pi = \{\pi(x)\}_{x \in S}$ is called an invariant measure of the stochastic matrix P, if:*

$$\pi(x) \in (0,1] \quad \forall\, x \in S,$$
$$\sum_{x \in S} \pi(x) = 1$$

and

$$\pi(y) = \sum_{x \in S} \pi(x) p(x,y), \tag{1.16}$$

that is

$$\pi = \pi \cdot P. \tag{1.17}$$

Definition 1.7 (Reversible measure)**.** *Let P be a transition matrix and π a strictly positive probability vector on S. The pair (P,π) is called reversible if the detailed balance condition*

$$\pi(x)p(x,y) = \pi(y)p(y,x) \tag{1.18}$$

holds for all $x, y \in S$.

Remark 1.3. Note that the sum of (1.18) over x yields

$$\sum_{x \in S} \pi(x)p(x,y) = \sum_{x \in S} \pi(y)p(y,x) = \pi(y) \sum_{x \in S} p(x,y) = \pi(y),$$

so we recover the stationarity condition (1.16). This implies that each reversible measure is also stationary.

A special property of non-negative matrices which brings a lot of advantages when we deal with Markov chains, is irreducibility: this condition guarantees that the distribution of the Markov chain reaches a steady state, i.e a limiting distribution as long as the time runs. We recall the definition below:

Definition 1.8 ([21], Section 3.1)**.** *[Irreducibility] A matrix P with non-negative entries $p(x,y)$ is called irreducible if there exists an $n_0 \in \mathbb{N}$ such that*

$$\min_{(x,y)} p^{(n_0)}(x,y) > 0. \tag{1.19}$$

As we mentioned above, the fact that the Markov chain reaches a stationary distribution in the limit, is a crucial turning point for the analysis of dynamical systems which evolve according to this process and it is a fundamental property when we resort to simulations. The theorem which guarantees the achievement of a steady state of the Markov chain, is the Ergodic theorem: among its hypothesis, the irreducibility condition of the transition matrix appears. We report the statement below, since we will exploit ergodicity for all the models proposed in Chapter 2.

Theorem 1.4 ([50], Theorem 1)**.** *[Ergodic Theorem for the Markov chains] Let $P = \{p(x,y)\}_{(x,y)\in S^2}$ be the transition matrix of a chain with a finite state space $S = \{1,\dots,N\}$.*

(a) if the matrix is irreducible, then there are numbers $\pi(1),\dots,\pi(N)$ such that

$$\pi(y) > 0 \quad \forall y \in \{1,\dots,N\}, \quad \sum_{y\in S} \pi(y) = 1 \tag{1.20}$$

and

$$p^{(n)}(x,y) \to \pi(y) \quad as \quad n \to +\infty \tag{1.21}$$

for every $x \in S$.

(b) Conversely, if there are numbers $\pi(1),\dots,\pi(N)$ which satisfy (1.20) *and* (1.21)*, then the matrix is irreducible.*

(c) The numbers $(\pi(1),\dots,\pi(N))$ satisfy the equations

$$\pi(y) = \sum_{x\in S} \pi(x)p(x,y), \quad y \in \{1,\dots,N\}. \tag{1.22}$$

Note that equations (1.22) represent the stationary condition introduced in Definition 1.6. The reason of this terminology is that the distribution π is unchanged as the time passes. In fact, putting as starting distribution in (1.15) the stationary one, namely $\nu = \pi$, and observing that $\pi = \pi \cdot P^{(n+1)}$, we get

$$\nu^{(n+1)}(y) = \pi(y) = \nu(y) \quad \forall\, y \in S.$$

Furthermore, the stationary distribution is unique. In fact, suppose that another stationary distribution $\tilde{\pi} = (\tilde{\pi}(1),\dots,\tilde{\pi}(N))$ does exist: the stationarity condition implies

$$\tilde{\pi} = \tilde{\pi} \cdot P^{(n)}$$

or, equivalently

$$\tilde{\pi}(y) = \sum_{x\in S} \tilde{\pi}(x)p^{(n)}(x,y), \quad y \in \{1,\dots,N\}.$$

Since $p^{(n)}(x,y) \to \pi(y)$ as $n \to +\infty$ (see (1.21)), in the limit we get

$$\tilde{\pi}(y) = \sum_x \tilde{\pi}(x)\pi(y) = \pi(y),$$

thus showing the uniqueness. As a final consideration, from

$$\lim_{n\to\infty} \nu^{(n+1)}(y) = \lim_{n\to\infty} \sum_x \nu(x)p^{(n)}(x,y) = \sum_x \nu(x) \lim_{n\to\infty} p^{(n)}(x,y) = \pi(y),$$

one can observe that the starting distribution ν does not affect the limiting behavior of the Markov chain, hence it is convenient to fix as starting distribution the stationary one.
As we have seen, the irreducibility of the transition matrix, reflects on the Markov chain different properties: it is convenient to mention another theorem which highlights some other features of irreducible matrices, the Perron- Frobenius theorem. To be precise, the theorem requires that, given a matrix P with non-negative entries $p(x,y)$, for any couple of indices (x,y) there exists $n_0 = n_0(x,y) \in \mathbb{N}$ such that $p^{(n_0)}(x,y) > 0$. Note that such condition is weaker than the one expressed by Definition 1.8, in fact it is implied by it.

Theorem 1.5 ([21], Theorem 3.1.1)**.** *[Perron-Frobenius] Let $P = \{p(x,y)\}_{(x,y)\in S^2}$ be an irreducible matrix. Then P possesses an eigenvalue ρ such that:*

a) $\rho > 0$ is real;

b) for any eigenvalue λ of P, $|\lambda| \leq \rho$;

c) there exist left and right eigenvectors corresponding to the eigenvalue ρ that have strictly positive coordinates;

d) the left and right eigenvectors Λ_ρ^L and Λ_ρ^R corresponding to the eigenvalue ρ are unique up to a constant multiple;

e) for every $x \in S$ and every $\phi = (\phi(1), \ldots, \phi(|S|))$ such that $\phi(y) > 0$ for all $y \in \{1, 2, \ldots, |S|\}$,

$$\lim_{n\to+\infty} \frac{1}{n} \ln \left[\sum_{y=1}^{|S|} p^{(n)}(x,y)\phi(y) \right] = \lim_{n\to+\infty} \frac{1}{n} \ln \left[\sum_{x=1}^{|S|} \phi(x)p^{(n)}(x,y) \right] = \ln(\rho). \tag{1.23}$$

Proof. For the complete proof we refer to [49]. Here we are especially interested in property e). Let $\alpha := \sup_x \Lambda_\rho^R(x)$, $\beta := \inf_x \Lambda_\rho^R(x) > 0$, $\gamma := \sup_y \phi(y)$, $\delta := \inf_y \phi(y) > 0$, where notation $\Lambda_\rho^{(\cdot)}(x)$ denotes that we are considering the component x of the vector. Then, for all $(x,y) \in S^2$,

$$\frac{\gamma}{\beta} p^{(n)}(x,y)\Lambda_\rho^R(y) \geq p^{(n)}(x,y)\phi(y) \geq \frac{\delta}{\alpha} p^{(n)}(x,y)\Lambda_\rho^R(y). \tag{1.24}$$

The right inequality in (1.24) yields

$$\lim_{n\to+\infty}\frac{1}{n}\ln\left[\sum_{y=1}^{|S|}p^{(n)}(x,y)\phi(y)\right]\geq\frac{1}{n}\lim_{n\to+\infty}\ln\left[\frac{\delta}{\alpha}\sum_{y=1}^{|S|}p^{(n)}(x,y)\Lambda_\rho^R(y)\right]$$
$$=\lim_{n\to+\infty}\frac{1}{n}\ln\left[\frac{\delta}{\alpha}\right]+\lim_{n\to+\infty}\frac{1}{n}\ln\left[\sum_{y=1}^{|S|}p^{(n)}(x,y)\Lambda_\rho^R(y)\right]$$
$$=\lim_{n\to+\infty}\frac{1}{n}\ln\left[P^{(n)}\cdot\Lambda_\rho^R\right](x)=\lim_{n\to+\infty}\frac{1}{n}\ln\left(\rho^n\Lambda_\rho^R(x)\right)=\ln(\rho).$$

Similarly, using the left inequality in (1.24), it follows that:

$$\lim_{n\to+\infty}\frac{1}{n}\ln\left[\sum_{y=1}^{|S|}p^{(n)}(x,y)\phi(y)\right]\leq\ln(\rho),$$

and finally

$$\lim_{n\to+\infty}\frac{1}{n}\ln\left[\sum_{y=1}^{|S|}p^{(n)}(x,y)\phi(y)\right]=\ln(\rho).$$

A similar argument leads to prove the second equality in (1.23):

$$\lim_{n\to+\infty}\frac{1}{n}\ln\left[\sum_{x=1}^{|S|}\phi(x)p^{(n)}(x,y)\right]=\ln(\rho).$$

□

We conclude this introductory part reconnecting to large deviations theory, more precisely, as we mentioned at the beginning of the chapter, to large deviations on discrete Markov chains, which is the simplest example of non i.i.d sequence.

1.3.1 Large deviations on Markov chains

Consider a finite state Markov chain $(X_i)_{i\geq 1}$ taking values in a finite space $S=\{1,2,\dots,N\}$; according to the notation introduced in (1.14), we denote by $\mathbb{P}[x_1,\dots,x_n]$ the probability of a realization. The hypothesis under which we work are

$$\begin{aligned}&X_i\in S\quad\text{with}\quad|S|=N\\&X_1,X_2\dots\text{is Markov with transition matrix}\qquad P=\{p(x,y)\}_{(x,y)\in S^2}\\&p(x,y)>0\quad\forall x,y\in S.\end{aligned}\tag{1.25}$$

The ergodic theorem for the Markov chains guarantees the existence of a unique stationary distribution (see Theorem 1.4) which we indicate by $\pi=(\pi(1),\dots,\pi(N))$: we recall that

$\pi(y) > 0 \quad \forall y \in S$. Moreover, we choose π as starting distribution of the Markov chain $(X_i)_{i\geq 1}$.
In order to study the large deviation principle for finite state Markov chains, it is convenient to introduce another sequence of i.i.d random variables $(Y_i)_{i\geq 1}$ with distribution equal to the stationary one, π. First of all, let us define the following object, called *pair empirical measure*

$$L_n^2 = \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_i, X_{i+1})},$$

where δ denotes the Dirac function, with periodic boundary conditions $X_{n+1} = X_1$: it takes values in the space

$$M = \left\{ \rho = \{\rho(x,y)\}_{(x,y)\in S^2} \in [0,1]^{N^2} : \sum_{x,y=1}^{N} \rho(x,y) = 1, \sum_{y=1}^{N} \rho(x,y) = \sum_{y=1}^{N} \rho(y,x) \ \forall x \in S \right\}$$

which turns into a Polish space equipped with the total variation distance

$$d(\mu,\rho) = \frac{1}{2}\sum_{x,y=1}^{N} |\mu(x,y) - \rho(x,y)| \quad \mu,\rho \in M.$$

The key observation is now that the probability of a realization $\mathbb{P}^X[x_1,\dots,x_n]$ of the Markov chain, is a functional of L_n^2 modulo boundary terms. More precisely:

$$\begin{aligned}\mathbb{P}^X[x_1,\dots,x_n] &= \pi(x_1)p(x_1,x_2)\dots p(x_{n-1},x_n) = \frac{\pi(x_1)}{p(x_n,x_{n+1})} e^{\sum_{i=1}^{n} \ln p(x_i,x_{i+1})} \\ &= \frac{\pi(x_1)}{p(x_n,x_{n+1})} e^{\sum_{(x,y)\in S^2}\sum_{i=1}^{n} \delta_{(X_i,X_{i+1})}(x,y) \ln p(x,y)} = \frac{\pi(x_1)}{p(x_n,x_{n+1})} e^{n\sum_{(x,y)\in S^2} L_n^2[x_1,\dots,x_n](x,y)\ln p(x,y)}\end{aligned} \tag{1.26}$$

where $L_n^2[x_1,\dots,x_n]$ is the pair empirical measure associated to the realization $(x_1,\dots,x_n)$. We can repeat the procedure for the random variables $(Y_i)_{i\geq,1}$ thus getting:

$$\mathbb{P}^Y[x_1,\dots,x_n] = \prod_{i=1}^{n}\pi(x_i) = e^{n\sum_{(x,y)\in S^2} L_n^2[x_1,\dots,x_n](x,y)\ln \pi(y)} \tag{1.27}$$

Combining (1.26) with (1.27) we get:

$$\frac{d\mathbb{P}^X}{d\mathbb{P}^Y}[\cdot] = O(1)e^{n\sum_{(x,y)\in S^2} L_n^2[\cdot](x,y)\ln\frac{p(x,y)}{\pi(y)}} = O(1)e^{nF(L_n^2[\cdot])} \tag{1.28}$$

where

$$F(\rho) := \sum_{(x,y)\in S^2} \rho(x,y)\ln\frac{p(x,y)}{\pi(y)}$$

and $\rho \in M$. Equation (1.28), giving the *Radon-Nikodym* derivative of $\mathbb{P}^X[\cdot]$ with respect to $\mathbb{P}^Y[\cdot]$, allows to turn questions about $(X_i)_{i\geq 1}$ into questions about $(Y_i)_{i\geq 1}$, thus simplifying the setup, being $(Y_i)_{i\geq,1}$ i.i.d random variables.
Now we are ready to state the LDP for discrete-time Markov chains:

Theorem 1.6 ([22], Theorem IV.3)**.** *Let $(X_i)_{i\geq 1}$ be a Markov chain and assume that hypothesis* (1.25) *hold. Then the family defined by*

$$P_n^X := \mathbb{P}^X(L_n^2 \in \cdot)$$

satisfies the LDP on M with speed n and rate function

$$I_P(\rho) = \sum_{(x,y)\in S^2} \rho(x,y) \ln \frac{\rho(x,y)}{\sum_z \rho(x,z)p(x,y)}.$$

Proof. For any Borel set $B \subset M$, from (1.28) follows that

$$\frac{1}{n} \ln P_n^X(B) = \frac{1}{n} \ln \int_B \mathbb{P}^X(L_n^2 \in d\rho) = O\left(\frac{1}{n}\right) + \frac{1}{n} \ln \int_B e^{nF(\rho)} \mathbb{P}^Y(L_n^2 \in d\rho).$$

The proof can be deduced by combining two facts:

(a) from the theory of large deviations for i.i.d sequences (in particular from Theorem II.8 in [22]), follows that $P_n^Y := \mathbb{P}^Y(L_n^2 \in \cdot)$ satisfies an LDP on M with speed n and rate function

$$I_\pi(\rho) = \sum_{(x,y)\in S^2} \rho(x,y) \ln \frac{\rho(x,y)}{\sum_z \rho(x,z)\pi(x,y)}; \tag{1.29}$$

(b) as a direct consequence of Varadhan's Theorem 1.3 we know that P_n^X satisfies an LDP on M with speed n and rate function

$$I_P(\rho) = I_\pi(\rho) - F(\rho) = \sum_{(x,y)\in S^2} \rho(x,y) \ln \frac{\rho(x,y)}{\sum_z \rho(x,z)p(x,y)},$$

where the computation directly follows from expressions (1.29) and (1.3.1).

□

Chapter 2

A Monte Carlo method for large deviations: the Cloning algorithm

In this chapter we present the core of our research, i.e a new computation strategy, *The Cloning algorithm*, to evaluate the probability of rare events: the method has been introduced in [29] and further developed in [28]. Cloning is a Monte Carlo method which allows to numerically approximate the scaled cumulant generating function defined in (1.4) using an approach relying on population dynamics. When Gärtner-Ellis theorem holds, the derivation of the rate function is just one step away. The algorithm is based on the evolution of a family of copies of the system which are replicated or killed in such a way as to favor the realization of the atypical trajectories. Our first contribution, provided in this chapter, is developing a formal analysis of the method, with a specific focus on its strict connection with branching processes, together with some simple and original applications which clarify the implementation of the scheme and highlight its effectiveness. The outline of the chapter is the following: Section 2.1 is divided in two parts, since we preliminarily introduce the setup we are interested in, namely large deviations on additive functionals, and after we present the method, with a specific focus on its connection with branching processes. In Section 2.2 we present some illustrative applications of the algorithm and, as a further validation of the method, the related simulations results, thus showing agreement between numerical and analytical computation.

2.1 The Cloning Algorithm

2.1.1 Large deviations for additive functionals

In order to introduce the algorithm, we consider the general setup of a system which evolves according to a Markovian dynamics, as introduced in Section 1.3. Given a discrete Markov chain $(X_n)_{n\geq 0}$ taking values in a finite space S with cardinality $|S| < +\infty$, we are interested in studying the large deviation properties of physical quantities F_T that are additive in time,

namely such that they can be written as

$$F_T = \sum_{n=0}^{T-1} f(X_n, X_{n+1}), \tag{2.1}$$

where f is a function defined on $S \times S$ and taking real values.

Example 2.1. For example, consider a one dimensional lattice with N sites. Imagine that a particle, occupying one site, can jump to the right, to the left or stay still on its position. For a specific transition $x_0 \longrightarrow x_1$, where $x_0, x_1 \in \{1, \dots, N\}$ are the nearest neighbors sites, we can define

$$f(x_0, x_1) = \begin{cases} 1 & \text{a particle jumps to the right} \\ 0 & \text{nothing happens} \\ -1 & \text{a particle jumps to the left} \end{cases}, \tag{2.2}$$

which can be interpreted as a flux of particles.

In particular we are interested in computing the probability of having a current per unit of time. From the theory of large deviations, more precisely rearranging the content of Theorem 1.3 in an informal way, we expect the following exponential behavior:

$$\mathbb{P}\left(\frac{F_T}{T} \approx r\right) = e^{-T[I(r)+o(1)]}.$$

When the Gärtner-Ellis theorem holds, the rate function $I(r)$ can be expressed in terms of Legendre transform of the scaled cumulant generating function $\mu(\alpha)$ defined as:

$$\mu(\alpha) = \lim_{T\to+\infty} \frac{1}{T} \ln \mathbb{E}(e^{\alpha F_T}).$$

We are going to state Theorem 2.1, which shows that the scaled cumulant generating function of the observable F_T is differentiable, hence it is possible to apply the Gärtner-Ellis theorem, thus deriving the large deviations rate function of F_T. Moreover, the theorem shows that the scaled cumulant generating function can be written in a closed form as the logarithm of a spectral radius. Before stating the theorem, it is convenient to introduce a non-negative matrix $\tilde{P}_\alpha$, $\alpha \in \mathbb{R}$ whose elements are:

$$\tilde{p}_\alpha(x, y) := p(x, y) e^{\alpha f(x,y)}, \quad (x, y) \in S^2, \tag{2.3}$$

where P is the transition matrix of the Markov chain. Because the quantities $e^{\alpha f(x,y)}$ are always positive, $\tilde{P}_\alpha$ is irreducible as soon as P is. Let $\rho(\alpha)$ denote the Perron-Frobenius eigenvalue of the matrix $\tilde{P}_\alpha$. The next theorem, shows that $\ln \rho(\alpha)$ coincides with the scaled cumulant generating function $\mu(\alpha)$.

Theorem 2.1 ([21], Theorem 3.1.2)**.** *Let $(X_n)_{n\geq 0}$ be a finite state Markov chain possessing an irriducible transition matrix P and starting distribution ν. For every $r \in \mathbb{R}$, define:*

$$I(r) := \sup_{\alpha\in\mathbb{R}}\{\alpha\, r - \ln \rho(\alpha)\},$$

where $\rho(\alpha)$ is the spectral radius of the matrix $\tilde{P}_\alpha$ defined in (2.3)*. Then F_T defined in (2.1) satisfies the large deviation principle with the convex rate function $I(\cdot)$.*

Proof. Define

$$\mu_T(\alpha) := \frac{1}{T}\ln \mathbb{E}(e^{\alpha\, F_T}).$$

In order to prove the statement and derive the large deviations rate function of F_T, we want to use the Gärtner-Ellis theorem, so we have to check that the limit

$$\mu(\alpha) = \lim_{T\longrightarrow+\infty} \mu_T(\alpha)$$

exists for every $\alpha \in \mathbb{R}$ and that the limit function $\mu(\cdot)$ is finite and differentiable everywhere in $\mathbb{R}$. Developing computations and recalling definition (2.3), we get:

$$\begin{aligned}\mu_T(\alpha) = \frac{1}{T}\ln \mathbb{E}(e^{\alpha\, F_T}) &= \frac{1}{T}\ln \sum_{x_0\dots\, x_T} \nu(x_0)p(x_0,x_1)\dots\, p(x_{T-1},x_T)e^{\alpha\, f(x_0,x_1)}\dots\, e^{\alpha\, f(x_{T-1},x_T)}\\ &= \frac{1}{T}\ln \sum_{x_0,x_T} \nu(x_0)\tilde{p}_\alpha^{(T)}(x_0,x_T) = \frac{1}{T}\ln \sum_{x_0}\nu(x_0)\sum_{x_T}\tilde{p}_\alpha^{(T)}(x_0,x_T). \qquad (2.4)\end{aligned}$$

Introducing the unitary vector $\phi = (1,\dots,1)$, equation (2.4) can be equivalently written as

$$\frac{1}{T}\ln \sum_{x_0}\nu(x_0)\sum_{x_T}\tilde{p}_\alpha^{(T)}(x_0,x_T) = \frac{1}{T}\ln \sum_{x_0}\nu(x_0)\sum_{x_T}\tilde{p}_\alpha^{(T)}(x_0,x_T)\phi(x_T). \qquad (2.5)$$

We now set

$$\begin{aligned}x_{min} &:= \operatorname*{argmin}_x[\nu(x)\sum_{x_T}\tilde{p}_\alpha^{(T)}(x,x_T)\phi(x_T)]\\ x_{max} &:= \operatorname*{argmax}_x[\nu(x)\sum_{x_T}\tilde{p}_\alpha^{(T)}(x,x_T)\phi(x_T)],\end{aligned}$$

thus getting the following inequalities:

$$\begin{aligned}\frac{1}{T}\ln\left(|S|\nu(x_{min})\sum_{x_T\in S}\tilde{p}_\alpha^{(T)}(x_{min},x_T)\phi(x_T)\right) &\leq \mu_T(\alpha)\\ &\leq \frac{1}{T}\ln\left(|S|\nu(x_{max})\sum_{x_T\in S}\tilde{p}_\alpha^{(T)}(x_{max},x_T)\phi(x_T)\right)\end{aligned}$$

and

$$\frac{1}{T}\ln(|S|\nu(x_{min})) + \frac{1}{T}\ln\left(\sum_{x_T\in S}\tilde{p}_\alpha^{(T)}(x_{min},x_T)\phi(x_T)\right) \leq \mu_T(\alpha)$$
$$\leq \frac{1}{T}\ln(|S|\nu(x_{max})) + \frac{1}{T}\ln\left(\sum_{x_T\in S}\tilde{p}_\alpha^{(T)}(x_{max},x_T)\phi(x_T)\right).$$

Since $\tilde{P}_\alpha$ is irreducible, part e) of Theorem 1.5 holds (with $\phi = (1,\dots,1)$). Therefore, computing the limit for $T \to +\infty$, we get:

$$\mu(\alpha) = \ln\rho(\alpha). \tag{2.6}$$

Finally, we remark that, being $\tilde{P}_\alpha$ a positive matrix, as direct consequence of Perron-Frobenius theorem, we have that $\rho(\alpha)$ is a simple root of the characteristic equation of the matrix $\tilde{P}_\alpha$, hence it is differentiable with respect to α (and this reflects on $\mu(\alpha)$). □

Theorem 2.1 shows that the scaled cumulant generating function of the observable (2.1) reduces to the logarithm of the spectral radius of $\tilde{P}_\alpha$.

2.1.2 The method

The computation of $\mu(\alpha) = \ln\rho(\alpha)$ in a closed form is often difficult, especially if the cardinality of the sample space is very large. To overcome this difficulty, it is useful to have numerical methods which allow to compute (at least approximately) the scaled cumulant generating function. In the following, we present a Monte Carlo method, called *Cloning* algorithm, which pursues this aim. The method was firstly introduced in [29] and further developed in [28]. We start from the definition of $\mu(\alpha)$:

$$\mu(\alpha) = \lim_{T\to+\infty}\frac{1}{T}\ln\mathbb{E}(e^{\alpha F_T}) = \lim_{T\to+\infty}\frac{1}{T}\ln\sum_{(x_0,x_1,\dots,x_T)\in S^{T+1}}\nu(x_0)\tilde{p}_\alpha(x_0,x_1)\dots\tilde{p}_\alpha(x_{T-1},x_T),$$

where we have used definition (2.3) for the tilted dynamics. Since the matrix $\tilde{P}_\alpha$ is not stochastic, i.e $\sum_{y\in S}\tilde{p}_\alpha(x,y) \neq 1$, we define a new dynamics described by the matrix P_α, whose elements are:

$$p_\alpha(x,y) = \frac{\tilde{p}_\alpha(x,y)}{k(x)}, \tag{2.7}$$

where $k(x) := \sum_{y\in S}\tilde{p}_\alpha(x,y) = \sum_{y\in S}p(x,y)e^{\alpha f(x,y)}$ is a normalization factor. Using definition (2.7) we can write $\mu(\alpha)$ as follows:

$$\mu(\alpha) = \lim_{T\to+\infty}\frac{1}{T}\ln\sum_{(x_0,\dots,x_T)\in S^{T+1}}\nu(x_0)k(x_0)p_\alpha(x_0,x_1)\dots k(x_{T-1})p_\alpha(x_{T-1},x_T) \tag{2.8}$$
$$= \lim_{T\to+\infty}\frac{1}{T}\ln\sum_{(x_0,x_T)\in S^2}\nu(x_0)R_\alpha{}^{(T)}(x_0,x_T) = \lim_{T\to+\infty}\frac{1}{T}\ln||\nu\cdot R_\alpha^{(T)}||_1, \tag{2.9}$$

where we have defined:

$$R_\alpha(x,y) := k(x)p_\alpha(x,y). \tag{2.10}$$

Note that, as a consequence of the Ergodic Theorem for Markov chains 1.4, the choice of the starting distribution $\nu(x)$ does not affect limit (2.8).
The aforementioned formulation leads us to a break point: in the next paragraph we introduce some results that will turn to be useful to reconnect expression (2.9) with the branching processes framework. For this reason we insert here a brief digression on such kind of processes.

2.1.3 Overview on branching processes

The branching process is a model of population growth which owes its name to the fact that members can reproduce and die causally thus creating a diagram of offspring which realizes a family tree. In this paragraph, we describe the process following the dissertation presented in [27]. Consider a starting population of size N_0 whose individuals belong to the same type. The ith member of the population (with $i \leq N_0$) gives birth, independently from the others, to a random number of direct successors, say Z_i, where Z_i are i.i.d random variables with distribution

$$\mathbb{P}(Z=k) = p(k) \quad \text{with} \quad p(k) \geq 0, \quad k \geq 0, \quad \sum_{k=0}^{+\infty} p(k) = 1$$

and with expectation

$$r := \mathbb{E}(Z) = \sum_k k\, p(k).$$

After this reproduction phase, we get a first generation of descendants of size N_1 and each of them bears a progeny whose size distribution is governed by p. Iterating the scheme, we get that the nth generation, of size N_n, is composed of individuals of the (n-1)st, each of whose gives birth to k descendants with probability $p(k)$: this procedure realizes an integer valued Markov chain $(N_i)_{i=1,\dots,n}$, since the size of each population only depends on the previous step. At time n it is described by the random variable

$$N_n = \sum_{i=1}^{N_{n-1}} Z_i. \tag{2.11}$$

The following proposition, concerning the expectation of N_n, is taken from [27] (Section 3).

Proposition 2.1. *The conditional expectation of N_n, given N_0, is:*

$$\mathbb{E}(N_n|N_0) = N_0\, r^n. \tag{2.12}$$

Proof. In order to prove the statement we proceed by induction, so we first consider the case $n = 1$. We have:

$$\mathbb{E}(N_1|N_0) = \mathbb{E}\left(\sum_{i=1}^{N_0} Z_i \middle| N_0\right) = N_0\mathbb{E}(Z_1) = N_0\, r.$$

We now assume, as induction hypothesis, that (2.12) holds for n. The following equalities hold:

$$\mathbb{E}(N_{n+1}|N_0) = \mathbb{E}\left(\mathbb{E}(N_{n+1}|N_n N_{n-1}\dots N_0)|N_0\right) = \mathbb{E}(\mathbb{E}(N_{n+1}|N_n)|N_0) = \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{N_n} Z_i \middle| N_n\right)\middle| N_0\right)$$

where we have used, respectively, the tower property, the Markov property and the definition of the process. Combining the fact that $\mathbb{E}\left(\sum_{i=1}^{N_n} Z_i|N_n\right) = N_n r$ since, as effect of the conditioning, the size of the population N_n is known, with the induction hypothesis (2.12), the following chain of equalities holds:

$$\mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{N_n} Z_i \middle| N_n\right)\middle| N_0\right) = \mathbb{E}(N_n\, r|N_0) = r\mathbb{E}(N_n|N_0) = r^{n+1}N_0.$$

□

We can generalize this development to multi-type branching processes, so to the case in which each member of the population can give birth to $0 < |S| < +\infty$ different mutant forms, independently with respect to the other. In a parallel way with the one-type process, we denote by $p_i(k_1,\dots,k_{|S|})$ the probability that the individual of type i gives birth to k_1 children of type 1, k_2 children of type 1, ..., $k_{|S|}$ children of type $|S|$. We have:

$$p_i(k_1,\dots,k_{|S|}) \geq 0, \qquad \sum_{k_1,\dots,k_{|S|}=1}^{+\infty} p_i(k_1,\dots,k_{|S|}) = 1 \quad \forall i = 1,\dots,|S|.$$

The size of the population at time n is a random vector

$$N_n = (N_n(1), N_n(2), \dots, N_n(|S|)) \tag{2.13}$$

where $N_n(j)$, $1 \leq j \leq |S|$, is the number of individuals of type j at time n. The evolution of the size of each family is described by

$$N_{n+1}(j) = \sum_{i=1}^{|S|} \sum_{l=1}^{N_n(i)} {Z_l}^{(i,j)} \tag{2.14}$$

where $Z^{(i,j)}$ represents the number of individuals of type j which are offspring of an individual of type i and its distribution is

$$\mathbb{P}(Z^{(i,j)} = k_j) = p_i(k_j)$$

whereas $Z_l^{(i,j)}$ with $l = 1,\dots,N_n(i)$ are $N_n(i)$ i.i.d copies of $Z^{(i,j)}$ with distribution $p_i(k_j)$. At the same way, $(Z_l^{(i,1)},\dots,Z_l^{(i,|S|)})$, with $l = 1,\dots,N_n(i)$ are $N_n(i)$ i.i.d copies of the vector $(Z^{(i,1)},\dots,Z^{(i,|S|)})$ with distribution $p_i(k_1,\dots,k_{|S|})$. More precisely:

$$\mathbb{P}(Z^{(i,1)} = k_1,\dots,Z^{(i,|S|)} = k_{|S|}) = p_i(k_1,\dots,k_{|S|}) \quad \text{with} \quad k_i \geq 0, \quad \forall\, i = 1,\dots,|S|.$$

We collect in the matrix R the average offspring:

$$R(i,j) := \mathbb{E}(Z^{(i,j)}). \tag{2.15}$$

The following proposition, taken from [27] (Section 5), generalizes Proposition 2.1 to the multi-type case.

Proposition 2.2. *Let N_n be the random vector defined in* (2.13)*. The expectation of N_n, conditioned to N_0, is given by:*

$$\mathbb{E}(N_n|N_0) = N_0\, R^{(n)},$$

where $R^{(n)}$ denotes the nth power of the matrix.

Proof. We prove that for each $j \in \{1, \dots, |S|\}$

$$\mathbb{E}(N_n(j)|N_0) = \sum_{k=1}^{|S|} N_0(k)\, r^{(n)}(k,j) \tag{2.16}$$

by induction. We first consider the case $n = 1$: recalling that $N_1(j)$ can be expressed through (2.14), we have

$$\mathbb{E}(N_1(j)|N_0) = \mathbb{E}\left(\sum_{i=1}^{|S|}\sum_{l=1}^{N_0(i)} {Z_l}^{(i,j)}\middle| N_0\right) = \sum_{i=1}^{|S|}\mathbb{E}\left(\sum_{l=1}^{N_0(i)} {Z_l}^{(i,j)}\middle| N_0\right) = \sum_{i=1}^{|S|} N_0(i)\, r(i,j).$$

We now assume that (2.16) holds for n and we prove it for $n+1$:

$$\mathbb{E}(N_{n+1}(j)|N_0) = \mathbb{E}(\mathbb{E}(N_{n+1}(j)|N_0 N_1 \dots\ N_n)|N_0) = \mathbb{E}(\mathbb{E}(N_{n+1}(j)|N_n)|N_0) \tag{2.17}$$

where we have used, respectively, the tower property and the Markov one. Moreover, recalling (2.16), we can proceed from (2.17) as follows:

$$\begin{aligned}\mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{|S|}\sum_{l=1}^{N_n(i)} {Z_l}^{(i,j)}\middle| N_n\right)\middle| N_0\right) &= \mathbb{E}\left(\sum_{i=1}^{|S|}\mathbb{E}\left(\sum_{l=1}^{N_n(i)} {Z_l}^{(i,j)}\middle| N_n\right)\middle| N_0\right) = \mathbb{E}\left(\sum_{i=1}^{|S|} N_n(i) r(i,j)\middle| N_0\right)\\ &= \sum_{i=1}^{|S|} r(i,j)\mathbb{E}(N_n(i)|N_0)\end{aligned} \tag{2.18}$$

where we have exploited the fact that $\mathbb{E}\left(\sum_{l=1}^{N_n(i)} {Z_l}^{(i,j)}\middle| N_n\right) = N_n(i)\mathbb{E}(Z_1^{(i,j)})$, since the conditioning on N_n provides information on $N_n(i)$ and, in addiction, $\mathbb{E}(Z_1^{(i,j)}) = \mathbb{E}(Z^{(i,j)})$, having

$Z_1^{(i,j)}$ the same distribution of $Z^{(i,j)}$. At this point, we can use the induction hypothesis, thus getting from (2.18):

$$\sum_{i=1}^{|S|} r(i,j)\mathbb{E}(N_n(i)|N_0) = \sum_{i=1}^{|S|} r(i,j) \sum_{k=1}^{|S|} N_0(k) r^{(n)}(k,i) = \sum_{k=1}^{|S|} N_0(k) \sum_{i=1}^{|S|} r^{(n)}(k,i) r(i,j)$$
$$= \sum_{k=1}^{|S|} N_0(k) r^{(n+1)}(k,j).$$

□

Having reached the end of this digression on branching processes, we can switch again to the computation of the scaled cumulant generating function introduced at the beginning of the section, since we left as open point its connection with such processes. In order to recap, we recall the expression of the scaled cumulant generating function:

$$\mu(\alpha) = \lim_{T\to+\infty} \frac{1}{T} \ln \sum_{(x_0,\dots,x_T)\in S^{T+1}} \nu(x_0)k(x_0)p_\alpha(x_0,x_1)\dots k(x_{T-1})p_\alpha(x_{T-1},x_T) \tag{2.19}$$
$$= \lim_{T\to+\infty} \frac{1}{T} \ln \sum_{(x_0,x_T)\in S^2} \nu(x_0) {R_\alpha}^{(T)}(x_0,x_T) = \lim_{T\to+\infty} \frac{1}{T} \ln ||\nu \cdot R_\alpha^{(T)}||_1.$$

The following proposition holds:

Proposition 2.3. *Consider a multi-type branching process with $|S|$ species which evolve according to the dynamics described by P_α. Let $N_0(x_0)$ be the number of individuals at time 0 in configuration x_0 and $M_0 := \sum_{x_0=1}^{|S|} N_0(x_0)$ the initial size of the population. Then*

$$\mu(\alpha) = \lim_{T\to+\infty} \frac{1}{T} \ln \frac{M_T}{M_0}, \tag{2.20}$$

where $M_T := \mathbb{E}(\sum_{x_T\in S} N_T(x_T)|N_0)$ is the average size of the population at time T.

Proof. We denote by $N_0(x_0)$ the number of individuals at time 0 in configuration x_0 and by $M_0 := \sum_{x_0=1}^{|S|} N_0(x_0)$ the starting number of individuals.
As a direct consequence of the Ergodic theorem, the starting distribution of the Markov chain does not affect the limiting behavior so that we can choose $\nu(x_0) = \frac{N_0(x_0)}{M_0}$. The expression of $\mu(\alpha)$ becomes:

$$\mu(\alpha) = \lim_{T\to+\infty} \frac{1}{T} \ln \sum_{(x_0,x_T)\in S^2} \frac{N_0(x_0)}{M_0} {R_\alpha}^{(T)}(x_0,x_T). \tag{2.21}$$

In order to conclude, we identify $R_\alpha(x,y) = k(x)p_\alpha(x,y)$ defined in (2.10) with the average number $\mathbb{E}(Z^{(x,y)})$ of individuals of type y coming from an individual of type x (see (2.15)) and we apply Proposition 2.2, thus getting

$$\mu(\alpha) = \lim_{T\to+\infty} \frac{1}{T} \ln \sum_{x_T\in S} \frac{\mathbb{E}(N_T(x_T)|N_0)}{M_0} = \lim_{T\to+\infty} \frac{1}{T} \ln \frac{\mathbb{E}(\sum_{x_T\in S} N_T(x_T)|N_0)}{M_0} = \lim_{T\to+\infty} \frac{1}{T} \ln \frac{M_T}{M_0}.$$

□

The idea of the algorithm is to simulate the average involved in the definition of the scaled cumulant generating function using a family of initial conditions belonging to the phase space, interpreted as a population of individuals, over which a dynamics is implemented. Sometimes we will refer to these individuals with the term *clones*. The dynamics can be interpreted as the natural evolution of the member of a population into a mutant form, as it happens in multi-type branching processes, where the probability of each transition is described by the matrix P_α (2.7). This step, whereby all individuals change their state, is called *evolution*. After having evolved, each individual can reproduce giving life to offspring or die, thus vanishing from the population: this corresponds to the so-called *cloning* phase and, for each clone x, the rate of reproduction is regulated by the normalization factor $k(x)$. The two steps described up to now allow to reproduce formula (2.19), indeed note that each factor of the product $\prod_{t=0}^{T-1} k(x_t)p_\alpha(x_t, x_{t+1})$ can be interpreted as a sequence of steps in which the individual reproduces (according to $k(x_t)$) and evolves (according to $p_\alpha(x_t, x_{t+1})$), following an atypical trajectory which derives from the tilted dynamics.
In order to avoid an explosion of the population, a renormalization phase is also added. It consists in uniformly sampling from the family, after the cloning phase, a number of individuals equal to the starting size of the population, thus selecting the configurations favored by the dynamics and avoiding numerical difficulties (due to possible large sizes of the population).
Summarizing, formula (2.20) is implemented via the following steps:

a) We start with M_0 conditions of the space S which represent a family of initial systems: such family of possible states can be read as a bunch of starting conditions over which we implement the sum (2.19).

b) Each clone evolves according to probability $P_\alpha(\cdot, \cdot)$.

c) Each clone in configuration x is then replicated and gives birth to a number of offspring equal to

$$\begin{cases} \lfloor k(x) \rfloor + 1 & \text{with probability} \quad k(x) - \lfloor k(x) \rfloor \\ k(x) = \lfloor k(x) \rfloor & \text{otherwise} \end{cases} \tag{2.22}$$

where $\lfloor k(x) \rfloor$ represents the greatest integer less than or equal to $k(x)$. In case that $k(x) = 0$ the clone is killed and it doesn't leave offspring.

Since $M_T \approx M_0 e^{T(\ln \rho(\alpha) + o(1))}$ would explode ($\rho(\alpha) > 1$) or vanish ($0 < \rho(\alpha) < 1$) exponentially fast, a renormalization step is also added.

d) Once that the clones have evolved and reproduced, the total number of copies is brought back to M_0, uniformly choosing M_0 clones among those present after the evolution and

reproduction step. At each step τ, after the cloning phase, it is convenient to keep trace of the rescaling factor $S_\tau := \frac{M_\tau}{M_{\tau-1}}$, so that

$$\mu_T(\alpha) = \frac{1}{T}\ln\frac{M_T}{M_0} = \frac{1}{T}\ln\left(\frac{M_T}{M_{T-1}}\frac{M_{T-1}}{M_{T-2}}\dots\frac{M_2}{M_1}\frac{M_1}{M_0}\right) = \frac{1}{T}\ln\left(S_T\cdot S_{T-1}\dots S_1\right)$$
$$= \frac{1}{T}\sum_{\tau=1}^{T}\ln(S_\tau).$$

The method, resumed in Algorithm 1, proceeds by repeating iteratively the four steps above: $\mu(\alpha)$ is estimated by computing

$$\mu_T(\alpha) = \frac{1}{T}\sum_{\tau=1}^{T}\ln(S_\tau).$$

Remark 2.1. A considerable simplification of the scheme occurs when the cloning rates have all the same expression, say $\bar{k}$, i.e they are independent from the state of the system. In such case from (2.8) we get:

$$\mu(\alpha) = \lim_{T\to+\infty}\frac{1}{T}\ln(\bar{k})^T + \lim_{T\to+\infty}\frac{1}{T}\ln\sum_{(x_0,\dots,x_T)\in S^{T+1}}\nu(x_0)p_\alpha(x_0,x_1)\dots p_\alpha(x_{T-1},x_T) = \ln(\bar{k}). \tag{2.23}$$

Equation (2.23) puts the accent on the strict connection between the partition function of the model and the scaled cumulant generating function; such relation can be rephrased saying that the core of the method is actually the numerical approximation of a partition function.

2.2 Applications and numerical results

In order to provide a further analysis of the algorithm, we now focus our attention on some simple applications, thus showing the performances of the method. We set two main scenarios basing on the chosen dynamics, a random walk over the integers and the Ehrenfest's diffusion, and we specialize the additive observable F_T to them. The dynamics is described by birth and death Markov chains on a finite space $S = \{0, 1, 2, \dots, N\}$ and we refer to the setup defined in Section 1.3.

2.2.1 Birth and death chains

Birth and death chains are a particular case of discrete-time Markov chains where transitions allowed consist in only two types: *births* which increase the state of the system by one and *deaths* which make the opposite operation. Let

$$p_x = \mathbb{P}(X_{n+1} = x+1 | X_n = x)$$
$$q_x = \mathbb{P}(X_{n+1} = x-1 | X_n = x),$$

Algorithm 1 Cloning method

Choose a family of initial conditions (*clones*) M_0 according to a starting distribution ν and fix $\alpha \in \mathbb{R}$ and a positive integer T.
FOR $\tau = 1, 2, \ldots, T$

EVOLUTION. Make each clone evolves according to the dynamics described by (2.7).

CLONING. Reproduce each member of the population according to the integer rate (2.22).

SAMPLING. Bring back to M_0 the total number of copies of the system and keep trace of $S_\tau = \frac{M_\tau}{M_{\tau-1}}$.

END
Approximate the scaled cumulant generating function with

$$\mu_T(\alpha) = \frac{1}{T}\sum_{\tau=1}^{T} \ln(S_\tau).$$

be, respectively, the probability that the state of the system turns from x into $x+1$ (p_x) and the probability that it changes from x to $x-1$ (q_x). Furthermore, let $1-q_x-p_x$ the likelihood that no move is made. In the following we analyze three types of birth and death chains which differ in the choice of both the configurations space S and the dynamics.

General case: $S = \mathbb{N}_0$

We first consider the case of a birth and death process with probabilities depending on the state of the system and we impose the detailed balance to find the reversible stationary distribution:

$$\pi(x)p_x = \pi(x+1)q_{x+1} \Longrightarrow \pi(x) = \pi(x-1)\frac{p_{x-1}}{q_x}.$$

Thus

$$\pi(x) = \prod_{k=0}^{x-1} \frac{p_k}{q_{k+1}}\pi(0) \quad \text{with} \quad \pi(0) = \left(\sum_{x=0}^{+\infty}\prod_{k=0}^{x-1}\frac{p_k}{q_{k+1}}\right)^{-1}, \tag{2.24}$$

assuming that the series converges.

Reflecting boundary conditions on $S = [0, 1, \ldots, N]$

An example of birth an death chain is obtained by considering a particle performing a random walk with reflecting conditions on $N+1$ sites: the barrier, located at the edges of the interval $[0, N]$, makes each particle bounce on the previous position. In this case probabilities do not depend on the particular configuration, so we have probability p of moving to the right (one step forward), probability q of moving to the left (one step backward) and probability $1 - p_x - q_x = 1 - q - p$ of staying still on the configuration. We impose again the detailed balance in order to find the stationary distribution:

$$\pi(x)p = \pi(x+1)q \Longrightarrow \pi(x+1) = \frac{p}{q}\pi(x), \quad x \in \{0, \dots, N\}$$
$$\pi(x) = \left(\frac{p}{q}\right)^x \pi(0)$$

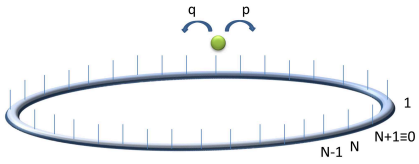
with $\pi(0) = \left(\sum_{x=0}^{N} \left(\frac{q}{p}\right)^x\right)^{-1} = \left(\frac{1-\left(\frac{q}{p}\right)^{N+1}}{1-\frac{q}{p}}\right)^{-1}$; if we specialize to the case $p = q$, from the detailed balance we get:

$$\pi(x) = \frac{1}{N+1}.$$

Therefore the invariant and reversible measure of this model is:

$$\pi(x) = \begin{cases} \pi(0)\left(\frac{p}{q}\right)^x & \text{if } p \neq q \\ \frac{1}{N+1} & \text{if } p = q \end{cases}. \tag{2.25}$$

Periodic boundary conditions on $S = [0, 1, \dots, N]$



A variation on the theme of the reflective random walk is the cyclic random walk. A particle jumps, according to constant probabilities, on $N+1$ sites spread over a Torus, $\mathbb{T}_{[0,N]}$, numbered from 0 to N: with probability p the particle moves on the right, with probability q it moves on the left and with probability $1 - p - q$ it remains on the site where it is located. In this case we note that, in addition to the condition

$$\pi(x)p = \pi(x+1)q, \quad x = 0, \dots, N-1$$

which implies $\pi(x) = \left(\frac{p}{q}\right)^x \pi(0)$, we have the closing bond $\pi(N)p = \pi(0)q$, that is

$$\pi(N) = \left(\frac{p}{q}\right)^N \pi(0) = \pi(0)\frac{q}{p}. \tag{2.26}$$

Equation (2.26) is satisfied only if $p = q$, thus, detailed balance is possible only in this case. If $p \neq q$, in order to find the stationary distribution, we exploit definition $\pi = \pi \cdot P$ that reads as:

$$\pi(x) = \pi(x-1)q_x + \pi(x)(1 - p_x - q_x) + \pi(x+1)p_x$$

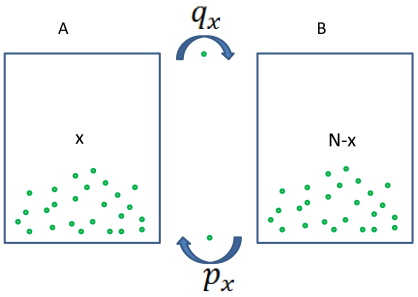
which specialized to $p_x = p$ and $q_x = q$ yields

$$\pi(x) = \pi(x-1)p + \pi(x)(1 - p - q) + \pi(x+1)q. \tag{2.27}$$

From (2.27) it is immediate to check that the uniform measure is stationary. In case $p \neq q$, the detailed balance is not satisfied and we don't have a reversible measure. Summarizing, the uniform distribution is invariant for any $p \neq q$ and it is reversible only in the case $p = q$:

$$\pi(x) = \begin{cases} \frac{1}{N+1} & \text{if } p = q, \text{ reversible} \\ \frac{1}{N+1} & \text{if } p \neq q, \text{ not reversible} \end{cases}.$$

Ehrenfest's diffusion model



A stochastic model of diffusion was proposed by P. and T.Ehrenfest in [23]: N particles numbered from 1 to N are spread over two urns, denoted by A and B. The dynamics evolves as follows: a random number is extracted from the set $S = \{0, 1, 2, \ldots, N\}$ and the corresponding ball is moved from the urn to which it belongs, to the other one, except for the case where 0 is extracted, in which the configuration of the system does not change. The state of the system can be described at a macroscopic level by the number of balls that are in A (or B). Denoting by x the state in which x balls are located in urn A (and $N-x$ in urn B), the possible transitions of the system at each instant of time are: $x \to x+1$, $x \to x-1$ and $x \to x$. The dynamics evolves according to the transition probabilities $p(x, x)$, $p(x, x+1)$, $p(x, x-1)$ which correspond, respectively, to the moves: "the system is not changed" , "the number of particles in urn A is incremented by one", "the number of particles in urn A is decremented by one." Since $p(x, y)$ represents the conditional probability that the event y occurs given that the event x occurred at the previous step, we can precisely establish the likelihood of each transition. The passage $x \to x+1$ only happens when a particle positioned in B is extracted: the probability of this transition is $p(x, x+1) = \frac{N-x}{N+1}$, since we have $N - x$ balls lying in B. The transition

$x \to x-1$ occurs when the number returned by the experiment labels a particle belonging to A, thus we have $p(x, x-1) = \frac{x}{N+1}$. Finally, it remains the case that, being extracted 0, the system does not change the configuration: this happens with probability $p(x,x) = \frac{1}{N+1}$. Such probabilities can be written in a $(N+1 \times N+1)$ square stochastic transition matrix that we report below:

$$P = \begin{bmatrix} \frac{1}{N+1} & \frac{N}{N+1} & 0 & \cdots & \cdots & \cdots & 0 \\ \frac{1}{N+1} & \frac{1}{N+1} & \frac{N-1}{N+1} & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \frac{2}{N+1} & \frac{1}{N+1} & \frac{N-2}{N+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \frac{N-1}{N+1} & \frac{1}{N+1} & \frac{1}{N+1} \\ 0 & 0 & \cdots & \cdots & 0 & \frac{N}{N+1} & \frac{1}{N+1} \end{bmatrix}. \tag{2.28}$$

Remark 2.2. The setup described above, corresponds to a birth and death chain with transition probabilities depending on the state of the system:

$$p_x := \mathbb{P}(X_{n+1} = x+1 | X_n = x) = \frac{N-x}{N+1} \tag{2.29}$$

$$q_x := \mathbb{P}(X_{n+1} = x-1 | X_n = x) = \frac{x}{N+1}. \tag{2.30}$$

Proposition 2.4. *The stationary measure of the Ehrenfest's model is the binomial distribution:*

$$\pi(x) \sim \ Bin\left(N, \frac{1}{2}\right).$$

Proof. Using (2.24) with p_x (2.29) and q_x (2.30) we find:

$$\pi(x) = \frac{N \cdot (N-1) \cdots N-(x-1)}{x \cdot (x-1) \cdots 1} \pi(0) = \frac{N!}{x!(N-x)!} \pi(0) = \binom{N}{x} \pi(0)$$

where $\pi(0)$ can be derived by condition: $1 = \sum_{x=0}^{N} \pi(x) = \pi(0) \sum_{x=0}^{N} \binom{N}{x} = 2^N \pi(0)$, which returns $\pi(0) = 2^{-N}$. Hence, the resulting distribution is $\pi(x) = \binom{N}{x} 2^{-N}$ which is the binomial distribution with N proofs and success probability equals to $\frac{1}{2}$. □

The Gallavotti-Cohen formula

Before turning to numerical tests, we present a meaningful property of the scaled cumulant generating function which arises from the application of the Gallavotti-Cohen fluctuation theorem to additive path functionals. When the observable (2.1) is chosen in a specific way, the theorem highlights a symmetry property on the scaled cumulant generating function (see [36]). The following proposition makes this statement precise.

Proposition 2.5. *Let $(X_n)_{n\geq 0}$ be a Markov chain with irreducible transition matrix P. Choosing as particular observable $F_T^{GC} = \sum_{n=0}^{T-1} f^{GC}(X_n, X_{n+1})$ where $f^{GC}(x,y) = \ln \frac{p(x,y)}{p(y,x)}$, with $x, y \in S$. Then:*

$$\mu^{GC}(\alpha) = \mu^{GC}(-1-\alpha),$$

where we denote by μ^{GC} the cumulant generating function corresponding to F_T^{GC}.

Proof. In order to prove the symmetry, being $\mu^{GC}(\alpha)$ connected to the spectral radius of $\tilde{P}_\alpha$ (as we can see from (2.6)), we recall the definition of the matrix:

$$\tilde{p}_\alpha(x,y) = p(x,y)\left(\frac{p(x,y)}{p(y,x)}\right)^\alpha.$$

First we consider $\tilde{P}_\alpha^T$, i.e the transposed matrix of $\tilde{P}_\alpha$, with elements $\tilde{p}_\alpha^T(x,y)$:

$$\tilde{p}_\alpha^T(x,y) = \tilde{p}_\alpha(y,x) = p(y,x)\left(\frac{p(y,x)}{p(x,y)}\right)^\alpha,$$

then we consider $\tilde{P}_{-1-\alpha}$:

$$\tilde{p}_{-1-\alpha}(x,y) = p(x,y)\frac{p(y,x)}{p(x,y)}\left(\frac{p(y,x)}{p(x,y)}\right)^\alpha = p(y,x)\left(\frac{p(y,x)}{p(x,y)}\right)^\alpha$$

showing the symmetry property:

$$\tilde{P}_\alpha^T = \tilde{P}_{-1-\alpha}.$$

Denoting by $\rho(\alpha)$ the spectral radius of $\tilde{P}_\alpha$ and and remarking that a matrix and its transposed have the same spectrum (so the same spectral radius), we have that $\rho(\alpha) = \rho(-1-\alpha)$. Thus, recalling that $\mu(\alpha) = \ln \rho(\alpha)$, the theorem is proved. □

Proposition 2.6. *Proposition 2.5 becomes empty, i.e $\mu^{GC}(\alpha) = 0$, when the stationary measure is reversible.*

Proof. We start considering the definition

$$\mu(\alpha) = \lim_{T\longrightarrow+\infty} \frac{1}{T}\mathbb{E}(e^{\alpha F_T})$$

with $F_T = \sum_{n=0}^{T-1} f(x_n, x_{n+1})$. If we set $f(x,y) = \ln \frac{p(x,y)}{p(y,x)}$ we get:

$$\mu(\alpha) = \lim_{T\longrightarrow+\infty} \frac{1}{T}\mathbb{E}\left(e^{\alpha \sum_{n=0}^{T-1} \ln \frac{p(x_n,x_{n+1})}{p(x_{n+1},x_n)}}\right).$$

We can now exploit the hypothesis of reversibility, writing the detailed balance:

$$\frac{p(x_n, x_{n+1})}{p(x_{n+1}, x_n)} = \frac{\pi(x_{n+1})}{\pi(x_n)},$$

so we can observe that

$$\sum_{n=0}^{T-1} \ln \frac{p(x_n, x_{n+1})}{p(x_{n+1}, x_n)} = \sum_{n=0}^{T-1} \ln \frac{\pi(x_{n+1})}{\pi(x_n)}$$
$$= -\ln \pi(x_0) + \ln \pi(x_1) + \ln \pi(x_2) - \ln \pi(x_1) \cdots - \ln \pi(x_{T-1}) + \ln \pi(x_T)$$

which is a telescopic sum in which all terms cancel except for the first and the last one. We finally get:

$$F_T = -\ln \pi(x_0) + \ln \pi(x_T).$$

Since $\mathbb{E}(e^{\alpha[-\ln \pi(x_0)+\ln \pi(x_T)]})$ is bounded, we can conclude that:

$$\mu(\alpha) = \lim_{T \longrightarrow +\infty} \frac{1}{T} \mathbb{E}(e^{\alpha[-\ln \pi(x_0)+\ln \pi(x_T)]}) = 0.$$

□

2.2.2 Numerical tests on the periodic random walk

This paragraph aims at applying the algorithm to the random walk with periodic boundary conditions. The tests performed include as observable the function f defined in (2.2). For a clearer application of the algorithm we report the setting we use for the tests. The space of the states consists of the set $S = \{0, \ldots, N\}$ whereas the transition matrix is

$$P = \begin{bmatrix} 1-p-q & p & 0 & \ldots & q \\ q & 1-p-q & p & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p & 0 & \ldots & q & 1-p-q \end{bmatrix}.$$

Furthermore we consider:

i) the observable $F_T = \sum_{n=0}^{T-1} f(X_n, X_{n+1})$ with

$$f(x,y) = \begin{cases} +1 & \text{if } y = x+1, \text{ i.e the particle jumps to the right} \\ -1 & \text{if } y = x-1, \text{ i.e the particle jumps to the left} \\ 0 & \text{otherwise.} \end{cases} \tag{2.31}$$

Note that $f(x_n, x_{n+1})$, with $x_n \in S$ and $x_{n+1} \in \{x_n, x_n - 1, x_n + 1\}$, is a sequence of i.i.d random variables taking values in the set $\{+1, -1, 0\}$.

ii) The cloning rates k that, in this case, do not depend on the specific state of the clone: $k = 1 - p - q + p\, e^{\alpha} + q\, e^{-\alpha}$.

iii) The matrix P_α which contains the evolution of the dynamics:

$$P_\alpha = \begin{bmatrix} \frac{1-p-q}{k} & \frac{p\,e^{\alpha}}{k} & 0 & \dots & \dots & \dots & \frac{q\,e^{-\alpha}}{k} \\ \frac{q\,e^{-\alpha}}{k} & \frac{1-p-q}{k} & \frac{p\,e^{\alpha}}{k} & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \frac{q\,e^{-\alpha}}{k} & \frac{1-p-q}{k} & \frac{p\,e^{\alpha}}{k} \\ \frac{p\,e^{\alpha}}{k} & 0 & \dots & \dots & 0 & \frac{q\,e^{-\alpha}}{k} & \frac{1-p-q}{k} \end{bmatrix}.$$

In case of i.i.d random variables, the computation of $\mu(\alpha)$ is straightforward:

$$\begin{aligned} \mu(\alpha) &= \lim_{T\longrightarrow+\infty} \frac{1}{T} \ln \mathbb{E}\left(e^{\alpha \sum_{n=0}^{T-1} f(X_n, X_{n+1})}\right) \\ &= \lim_{T\longrightarrow+\infty} \frac{1}{T} \ln \mathbb{E}\left(\prod_{n=0}^{T-1} e^{\alpha\, f(X_n, X_{n+1})}\right) = \lim_{T\longrightarrow+\infty} \frac{1}{T} \ln \prod_{n=0}^{T-1} \mathbb{E}\left(e^{\alpha\, f(X_n, X_{n+1})}\right) \\ &= \lim_{T\longrightarrow+\infty} \frac{1}{T} \ln \left[\mathbb{E}(e^{\alpha\, f(X_0, X_1)})\right]^T = \ln \mathbb{E}(e^{\alpha\, f(X_0, X_1)}) = \ln\left(p\, e^{\alpha} + q\, e^{-\alpha} + 1 - p - q\right). \end{aligned} \tag{2.32}$$

Figure 2.1 shows the result of simulations on the model when $p = q$ and when $p \neq q$; the total number of sites considered is $N = 50$ whereas the starting size of the population is $M_0 = 500$. The performances improve as long as the number of initial conditions and iterations grows. However, it is not known if there is an optimal choice of T and M_0 which maximizes the accuracy of the numerical approximation while minimizing the computational effort. In order to get a perfect match as in Figure 2.1 we fixed $T = 1000$.

Preserving the same dynamics, parameters can be adjusted in order to appreciate the symmetry property of the scaled cumulant generating function described in Proposition 2.5. According to its hypothesis we have

$$f^{GC}(x, y) = \ln\left(\frac{p(x, y)}{p(y, x)}\right) = \begin{cases} \ln\left(\frac{p}{q}\right) & \text{if } y = x + 1 \\ \ln\left(\frac{q}{p}\right) & \text{if } y = x - 1\,, \\ 0 & \text{otherwise} \end{cases} \tag{2.33}$$

the matrix $\tilde{P}_\alpha$ turns into:

$$\tilde{P}_\alpha = \begin{bmatrix} 1-p-q & p\left(\frac{p}{q}\right)^{\alpha} & 0 & \dots & q\left(\frac{q}{p}\right)^{\alpha} \\ q\left(\frac{q}{p}\right)^{\alpha} & 1-p-q & p\left(\frac{p}{q}\right)^{\alpha} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p\left(\frac{p}{q}\right)^{\alpha} & 0 & \dots & q\left(\frac{q}{p}\right)^{\alpha} & 1-p-q \end{bmatrix},$$

and the cloning rates take the form:

$$k = p\left(\frac{p}{q}\right)^{\alpha} + q\left(\frac{q}{p}\right)^{\alpha} + 1 - p - q.$$

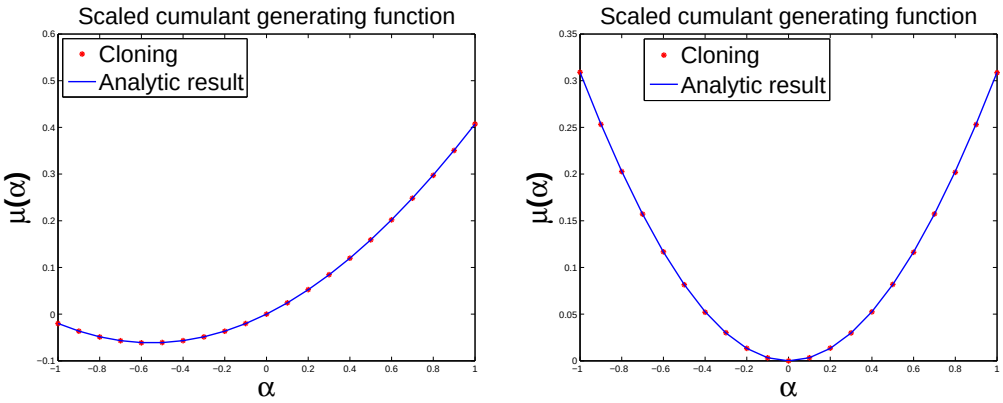


Figure 2.1: Approximation of the scaled cumulant generating function (2.32) via the Cloning algorithm: the dotted line represents the output of the simulation whereas the continuous one is the analytic result (2.32). The tests have been made using $p = \frac{1}{3}$, $q = \frac{1}{9}$ (left) and $p = q = \frac{1}{3}$ (right).

Finally the expected curve is given by the proposition below:

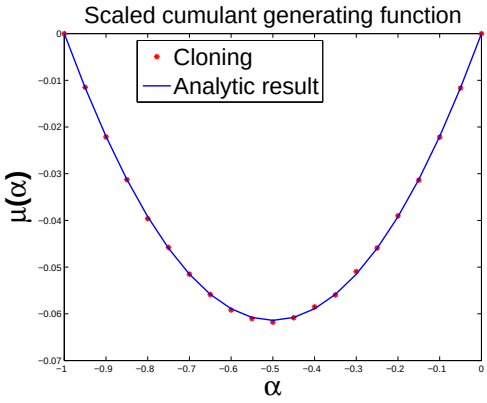


Figure 2.2: Approximation of the scaled cumulant generating function $\mu^{GC}(\alpha)$ with $p = \frac{1}{3}$ and $q = \frac{1}{9}$: the dotted line, which represents the algorithm approximation, perfectly matches the expected curve.

Proposition 2.7. *Let $F_T^{GC} = \sum_{n=0}^{T-1} f^{GC}(X_n, X_{n+1})$ be the Gallavotti-Cohen osservable (2.33). The resulting scaled cumulant generating function is:*

$$\mu^{GC}(\alpha) = \mu\left(\alpha \ln\left(\frac{p}{q}\right)\right). \tag{2.34}$$

Proof.

$$\mu^{GC}(\alpha) = \lim_{T\longrightarrow+\infty} \frac{1}{T} \ln \mathbb{E}(e^{\alpha \sum_{n=0}^{T-1} f^{GC}(X_n, X_{n+1})}) = \lim_{T\longrightarrow+\infty} \frac{1}{T} \ln \mathbb{E}(e^{\alpha \ln \frac{p}{q} \sum_{n=0}^{T-1} f(X_n, X_{n+1})})$$
$$= \mu\left(\alpha \ln \frac{p}{q}\right),$$

where f is defined in (2.31). □

Remark 2.3. Note that the random walk with reflecting boundary conditions is trivial, in the sense that $\mu(\alpha)$ and $\mu^{GC}(\alpha)$ are null. This is a direct consequence of the reversibility of the stationary measure (see (2.25)), combined with Propositions 2.6 and 2.7.

Figure 2.2 shows the result of the simulations when the method is applied to the observable (2.34) with $p \neq q$. When $p = q$ the function vanishes, as a direct consequence of its definition. The algorithm has been run using $T = 1000$, $M_0 = 500$ and $N = 50$.
An analogous analysis can be drawn by considering the Ehrenfest's diffusion: the principal difference with the periodic random walk consists in the fact that the flux F_T is not made of i.i.d random variables, hence we use a different strategy, which leans on Theorem 2.1, in order to find the scaled cumulant generating function.

2.2.3 Numerical tests on the Ehrenfest's diffusion

In this paragraph we apply the Cloning algorithm to the Ehrenfest's diffusion using three different additive observables; we give a synthetic overview of the setting in each case and we compare the expected results with the numerical ones, thus giving a further evidence of the robustness of the algorithm. The phase space $S = \{0, 1, \dots, N\}$ is the same for all tests whereas differences between models are due to the choice of F_T. Note that, in this setup, $f(x_n, x_{n+1})$ is not a sequence of i.i.d random variables because the probability at the present step depends on the previous one.
In order to find an analytical expression of $\mu(\alpha)$ to compare with, we exploit relation 2.6, thus moving the goal on computing the spectral radius of $\tilde{P}_\alpha$. The procedure followed for pursuing this aim is described below:

a) build the matrix for lower cases, namely $N = 1, 2, 3$, identifying the maximum eigenvalue and generalize its structure for the N-dimensional case, thus getting the candidate $\rho(\alpha)$.

b) Compute the eigenvectors associated to the spectral rays of the lower cases ($N = 1, 2, 3$) and repeat the strategy generalizing to the N dimensional case, thus getting u, i.e the eigenvector corresponding to $\rho(\alpha)$.

c) As a final step check that relation $\tilde{P}_\alpha u = \rho(\alpha)\, u$ is verified: this assures that the guessed candidate is effectively an eigenvalue and, if the components of its eigenvector u are all positive, the Perron-Frobenius theorem guarantees that it is the spectral radius.

We start proposing the first flux we dealt with by analogy with the periodic random walk, i.e the symmetric flux.

Symmetric flux

Consider the following parameters:

i) the observable $F_{T,1} = \sum_{n=0}^{T-1} f(X_n, X_{n+1})$ with

$$f(x,y) = \begin{cases} +1 & \text{if } y = x+1, \text{ i.e particles in urn A increase} \\ -1 & \text{if } y = x-1, \text{ i.e particles in urn A decrease} \\ 0 & \text{if } y = x, \text{ i.e nothing happens;} \end{cases} \tag{2.35}$$

ii) the cloning rates $k(x)$ which, in this case, depend on the particular state of the clone:

$$k(x) = \frac{N-x}{N+1}e^{\alpha} + \frac{x}{N+1}e^{-\alpha} + \frac{1}{N+1};$$

iii) the matrix P_α which contains the normalized dynamics, with elements:

$$p_\alpha(x,x) = \frac{1}{(N+1)k(x)}, \quad p_\alpha(x,x+1) = \frac{(N-x)e^{\alpha}}{(N+1)k(x)}, \quad p_\alpha(x,x-1) = \frac{x\,e^{-\alpha}}{(N+1)k(x)}$$
$$\text{with} \quad x = 0, \ldots, N.$$

The scheme described above leads to a spectral radius equal to one. Remembering relation (2.6), we can conclude that:

Proposition 2.8. *The rate function of the flux (2.35) and its Legendre transform for the Ehrenfest's model are:*

$$I(r) = \sup_{\alpha} \alpha\, r = +\infty$$
$$\mu(\alpha) = \ln(1) = 0.$$

We now propose a second flux, which exhibits an asymmetry devised for avoiding the trivial result implied by the symmetric one.

Asymmetric flux

Consider the following quantities:

i) the observable $F_{T,2} = \sum_{n=0}^{T-1} f(X_n, X_{n+1})$ with

$$f(x,y) = \begin{cases} +2 & \text{if } y = x+1, \text{ i.e the number of particles in urn A increase} \\ -1 & \text{if } y = x-1, \text{ i.e the number of particles in urn A decrease} \\ 0 & \text{if } y = x, \text{ i.e nothing happens;} \end{cases} \tag{2.36}$$

ii) the cloning rates $k(x)$ which still depend on the state of the clone:

$$k(x) = \frac{N-x}{N+1}e^{2\alpha} + \frac{x}{N+1}e^{-\alpha} + \frac{1}{N+1};$$

iii) the matrix P_α containing the normalized dynamics, with elements:

$$p_\alpha(x,x) = \frac{1}{(N+1)k(x)}, \quad p_\alpha(x,x+1) = \frac{(N-x)e^{2\alpha}}{(N+1)k(x)} \quad p_\alpha(x,x-1) = \frac{x\,e^{-\alpha}}{(N+1)k(x)}$$
with $\quad x = 0,\dots,N.$

The model leads to a spectral radius equal to $\rho(\alpha) = \frac{1+Ne^{\frac{\alpha}{2}}}{N+1}$. In conclusion:

Proposition 2.9. *The scaled cumulant generating function and the rate function for the flux (2.36) in Ehrenfest's model are respectively:*

$$\mu(\alpha) = \ln\left(\frac{1+N\,e^{\frac{\alpha}{2}}}{N+1}\right) \tag{2.37}$$
$$I(r) = 2r\ln\left(\frac{2r}{N(1-2r)}\right) + \ln[(1-2r)(N+1)]$$

where $I(r)$ is the Legendre transform of $\mu(\alpha)$.

Flux dependent on the state

As a conclusive variation of the observable, we propose a flux which contains the dependence on the state of the system. Consider:

i) the observable $F_{T,3} = \sum_{n=0}^{T-1} f(X_n, X_{n+1})$ with

$$f(x,y) = \begin{cases} \frac{-2x}{c} & \text{if } y = x+1, \text{ i.e particles in urn A increase} \\ \frac{2x}{c} & \text{if } y = x-1, \text{ i.e particles in urn B decrease} \\ 1 & \text{if } y = x, \text{ i.e nothing happens} \end{cases} \tag{2.38}$$

where $c \in \mathbb{R}$;

ii) the cloning rates $k(x)$:

$$k(x) = \frac{N-x}{N+1}e^{\frac{-2\alpha\,x}{c}} + \frac{x}{N+1}e^{\frac{2\alpha\,x}{c}} + \frac{e^{\alpha}}{N+1};$$

iii) the matrix P_α:

$$p_\alpha(x,x) = \frac{e^{\alpha}}{(N+1)k(x)}, \quad p_\alpha(x,x+1) = \frac{(N-x)e^{\frac{-2\alpha\,x}{c}}}{(N+1)k(x)} \quad p_\alpha(x,x-1) = \frac{x\,e^{\frac{2\alpha\,x}{c}}}{(N+1)k(x)}$$
with $\quad$ x=0,. . . ,N.

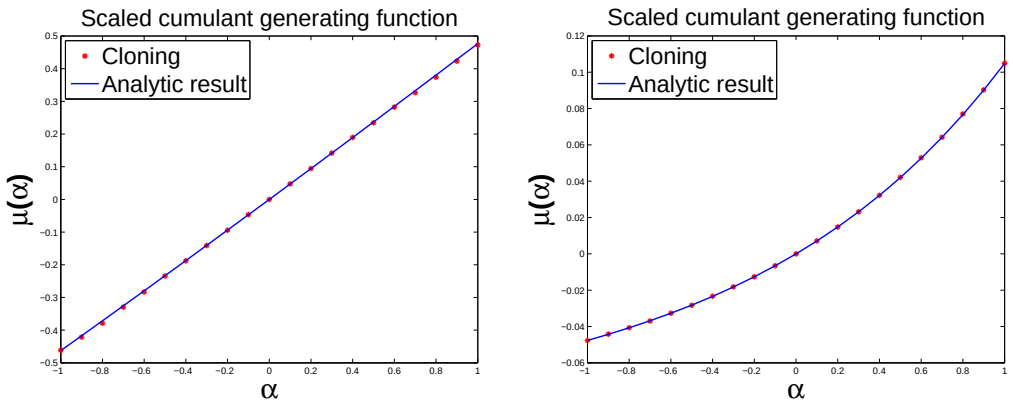


Figure 2.3: Approximation of the scaled cumulant generating function (2.37) (left) and (2.39) (right): the pictures correspond to parameters $N = 16$, $T = 1000$, $M_0 = 500$.

Once again, we can repeat the scheme stated at the beginning of the paragraph which leads to the spectral radius $\rho(\alpha) = \frac{e^{\alpha} + N\, e^{\frac{\alpha}{c}}}{N+1}$. In analogy with the previous cases, we can drawn the same conclusions on the scaled cumulant generating function, resumed by the proposition below:

Proposition 2.10. *The scaled cumulant generating function and the rate function for the flux (2.38) in Ehrenfest's model are respectively:*

$$\mu(\alpha) = \ln\left(\frac{e^{\alpha} + N\, e^{\frac{\alpha}{c}}}{N+1}\right) \tag{2.39}$$

$$I(r) = \frac{cr}{c-1}\ln\left(\frac{N(-1+cr)}{c(1-r)}\right) - \ln\left(\frac{1}{N+1}\left\{\left[\frac{N(-1+cr)}{c(1-r)}\right]^{\frac{c}{c-1}} + N\left[\frac{N(-1+cr)}{c(1-r)}\right]^{\frac{1}{c-1}}\right\}\right)$$

Simulations showed in Figure 2.3 confirm the agreement with the expected result; once again, in order to approach with a good accuracy the the analytic curve we fix the number of iterations to $T = 1000$ with a starting population of $M_0 = 500$ clones. The total number of particles among the two urns is $N = 16$.

The purpose of the applications described in this last section was to illustrate how the algorithm works and to give an idea of how to recover and set the parameters. The method presented in this chapter inspired our research challenge, that is extending the strategy to a more complex scenario, the random graphs framework, in order to infer information on the scaled cumulant generating function. The next chapter aims at giving an overview of this specific context, merging the large deviation theory into the landscape of random graphs and highlighting the open problem we want to tackle: as we will see, the additive observables we deal with, keep trace of some selected structures as long as the graph evolves.

Chapter 3

Large Deviations on random graphs

The purpose of this chapter is to provide an insight into the main results about large deviations of dense Erdös-Rényi random graphs, when we consider as observable the number of triangles. The aforementioned model represents the simplest example of random graph one can build and is characterized by the fact that each edge is present in an independent way with probability p. Originally defined in [25], it has been the subject of several investigations and one of the most studied class of graphs in literature, even if, due to its simplicity, it is not a proper tool for modeling real networks.
On the other side, the theory of Large Deviations studies the probabilities of rare events and the conditional probabilities of events, given that some rare event occurred: when applied to random graphs framework it gives rise to complex and surprising scenarios, despite of the basic structure of the starting model. For example, if we choose as observable the number of triangles of a graph, the two typical questions one could aim at answer concern

a) the probability of observing an atypical number of triangles

b) the structure of the graph which realizes such deviation.

The first problem was firstly investigated in the work of Chatterjee and Dey ([17]) but a crucial turning point was given by Chatterjee and Varadhan ([19]), who made use of the theory of graph limits for merging the problem in a more general framework, in which the rate function can be obtained as the solution of a variational problem. Using this general setup, it is also possible to tackle the second question and investigate how the graph looks like using a proper notion of distance.
This chapter aims at giving an excursus on the known results concerning the free energy and the rate function of the Erdös-Rényi model when the observable considered is the number of triangles. Furthermore we aim at giving a clear picture of the unresolved region of the free energy which we want to investigate. The outline of the chapter is the following: Section 3.1 is devoted to graph limit theory (which is the preliminary basis for studying large deviations via the variational approach) and main results on large deviations of dense Erdös-Rényi random

graphs. Section 3.2 focuses on Exponential Random Graphs (which include the Erdös-Rényi as a particular case) and gives an overview of the state-of-art concerning the free energy of the resulting model. In the final part of this last section we provide an argument to connect the scaled cumulant generating function of the number of triangles of a dense Erdös-Rényi graph, with the free energy resulting from an Exponential Random Graph model (with parameters properly tuned). We conclude by highlighting the purpose of our research with the related open points.

3.1 Graph limit theory and large deviations

3.1.1 Random graphs

This paragraph is devoted to recall some notions that will be used throughout the chapter. A graph G is an object defined by a couple $(\mathcal{V}(G), \mathcal{E}(G))$, where $\mathcal{V}(G)$ denotes the set of vertices (otherwise called nodes) and $\mathcal{E}(G)$ represents the set of edges, i.e the possible connections between the nodes. Given two vertices $\{i, j\} \in \mathcal{V}(G)$ we say that the couple (i, j) belongs to $\mathcal{E}(G)$ if a link is present between them. The set of vertices is typically denoted by its cardinality, for example $\mathcal{V}(G) = [n]$, if the number of vertices runs from 1 to n. The total amount of edges of G will be denoted by $E(G)$, namely $E(G) := |\mathcal{E}(G)|$ and, in a parallel way, the total number of vertices by $|\mathcal{V}(G)|$. One of the characterizing quantities of the vertices is their degree, namely the number of edges incident to them: if we denote by d_i the degree of vertex i, the one of the whole graph will be $l_n := \sum_{i \in [n]} d_i$. We recall that when we consider the degree of a vertex, self loops (namely edges which start and end on the same vertex) are counted twice. When each vertex has the same degree the graph is called *regular*.
We can move forward with the nomenclature saying that a graph is called *directed* when the edges point at a precise direction, meaning that the underlying relation can be read in one verse, otherwise it is called undirected: typically, in directed graphs, the edges are drawn with arrows in order to make explicit the sense of connection. Finally, by the term *simple graph*, we intend an undirected graph without self loops and no multiple edges, i.e edges incident to the same two vertices. It is also convenient to introduce the definition of *bipartite graph*, since we will frequently refer to this terminology throughout the thesis. First, we recall that, in graph theory, an independent set is a set of vertices of the graph, no two of which are adjacent. This means that for any pair of vertices belonging to the set, no edge is present between the two nodes.

Definition 3.1 (Bipartite graph)**.** *A graph G is called bipartite if it is possible to split the set of vertices $\mathcal{V}(G)$ in two independent sets, $\mathcal{A}(G)$ and $\mathcal{B}(G)$, such that every edge belonging to $\mathcal{E}(G)$ connects a vertex of $\mathcal{A}(G)$ to a vertex of $\mathcal{B}(G)$.*

An important characterization of bipartite graphs involves the notion of *cycle*: given an alternating sequence of vertices and edges, say $x_0, e_1, x_1, e_2, \ldots, e_l, x_l$ where $e_i = x_{i-1}x_i, 0 < i \leq l$,

we say that $W := x_0 x_1 \cdots, x_l$ is a *walk* of length l. A cycle is a walk W such that $l \geq 3$, $x_0 = x_l$ and the vertices x_i, $0 < i < l$ are distinct from each other and from x_0 [8].

Theorem 3.1 ([8],Theorem 4)**.** *A graph is bipartite if and only if it has no odd cycles, namely cycles of odd length.*

From the mathematical point of view, it is convenient to represent graphs using the so-called *adjacency matrices*, which consist in $0-1$ elements which denote, respectively, the absence or the presence of a connection between two nodes. In this thesis, adjacency matrices will be denoted by $\{X_{ij}\}_{i,j=1\ldots,n}$ and, according to the model of graph they represent, they embody specific features: for example, when the graph is undirected and with no self loops they are symmetric and with null elements on the diagonal. There exists a disparate variety of models of graphs and some of them turn to be useful for describing real networks.
As we mentioned on the introduction, one of the simplest random graph model, is the *Erdös-Rényi* one ([25],[30]): we recall that in this graph, each pair of nodes is connected with probability p. If p stays fixed as long as the graph grows its dimension (n), we speak of *dense* case whereas if we choose $p = \frac{c}{n}$, where c is a constant, we get the *sparse* case. In this last model, the degree of a random vertex of the graph has Binomial distribution $Bin\left(n-1, \frac{c}{n}\right)$, that converges to a Poisson, $Poiss(c)$, when $n \to +\infty$. Another famous model is the *preferential attachment* one, which describes a graph growing in time ([2],[9]): vertices, connected to a fixed number of edges, are sequentially added to the graph and edges are connected to the receiving vertex with probability proportional to its degree. Another famous model is the *generalized random graph* in which each vertex $i \in [n]$ receives a weight ω_i: given the weights, edges are present in an independent way but the occupation probabilities, say p_{ij}, are not the same for all the edges but depend on the weights of the vertices involved. As described in [5], since this model is essentially tree-like, it does not consider the reciprocity observed in real social networks. One network model that attempts to incorporate reciprocity and follows the statistical mechanics approach, is the *Exponential Random Graph model*. It is based on the definition of an Hamiltonian for weighting the probability measure on the space of graphs. Such measure, which is the Gibbs one, assigns higher probability to graphs which correspond to desiderable properties.
Finally, we quote another model which can be used for describing the behavior of social networks, that is the *random intersection graph* ([34]): in this model a random subset S_v of a given set S is assigned to vertex v and two vertices v and w are connected only if the set they belong to, intersect.

3.1.2 Preliminary results on graph limit theory

The theory of graph limits was firstly introduced by Laszlo Lovász and coauthors in [39], [38], [11], [12], [13] : in this section we give a brief review of such theory, with the main definitions and basic results. We strictly refer to the formalism used by Chatterjee in [15] and [17]. To

start with, we consider a family of graphs $\{G_n\}_n$ of size n, where n goes to infinity: among all possible topics that can be studied on the evolution of such family, one can be interested in identifying specific patterns and counting some selected structures, such as the density of a given subgraph, as long as n grows. In order to provide an answer to this need, we first introduce the notion of homomorphisms between graphs: it is a proper tool for counting patterns since, given two graphs, it maps the edges of the former into the ones of the latter.

Definition 3.2 ([15], Definition 3.1)**.** *Let H and G be two simple graphs: a map $\varphi : \mathcal{V}(H) \longmapsto \mathcal{V}(G)$ is an homomorphism if and only if $(\varphi(i), \varphi(j)) \in \mathcal{E}(G)$ for all $(i, j) \in \mathcal{E}(V)$.*

Let $|Hom(H, G)|$ denote the number of homomorphisms from H in G: this is not exactly the amount of copies of H in G but it is a multiple of it if H is a complete graph. For example, when H is a triangle, we have $|Hom(H, G)| = 6T(G)$, where $T(G)$ denotes the number of triangles of the graph G (indeed, the three vertices of the triangles can be labeled in six possible ways); when H is an edge, we have $|Hom(H, G)| = 2E(G)$. When we normalize $|Hom(H, G)|$ we get the homomorphism density:

Definition 3.3 ([15], Definition 3.2)**.** *Let H and G be finite simple graphs. The homomorphism density of H in G is defined as*

$$t(H, G) := \frac{|Hom(H, G)|}{|\mathcal{V}(G)|^{|\mathcal{V}(H)|}}. \tag{3.1}$$

We remark that, since the maximum number of homomorphisms from a graph of k vertices into a graph of n vertices is n^k, such density properly belongs to $[0, 1]$.
We now move the focus back to the sequence of growing graphs $\{G_n\}_n$: suppose that we want to analyze the evolution of the density of a given simple graph H over the sequence $\{G_n\}_n$, thus studying the limit behavior of $t(H, G_n)$, that we call $t(H)$. In order to read information on such limit element Lovász and coauthors introduced a proper object: the *graphon* (or graph limit) whose definition is reported below.

Definition 3.4 (Graphon, [15])**.** *Let $\mathcal{W}$ be the space of all bounded measurable functions $[0, 1]^2 \longmapsto [0, 1]$ that satisfy $f(x, y) = f(y, x)$ for all $x, y \in [0, 1]$. We refer to the elements of these space as graphons or graph limits.*

Intuitively the idea is to map the limit graph on the unitary square: the coordinates (x, y) on $[0, 1]^2$ represent the equivalent of a couple of vertices whereas $f(x, y)$ denotes the probability of connecting x with y.
For example, the graph limit associated to the dense Erdös-Rényi random graph with parameter p, is the constant function identically equal to p.

Remark 3.1 ([18])**.** Observe that a simple graph G of size n can be read as a graphon by defining

$$f^G(x, y) = \begin{cases} 1 & \text{if } \quad (\lceil nx \rceil, \lceil ny \rceil) \quad \text{is an edge of} \quad G \\ 0 & \text{otherwise} \end{cases}, \tag{3.2}$$

where the symbol $\lceil\cdot\rceil$ denotes the ceiling function. Moreover, we observe that f^G is a Borel measurable function and $f^G(x,y) = f^G(y,x)$.

As a consequence of Remark 3.1, the space of graphs turns out to be a subset of the space of graphons. It is now possible to define a distance between the elements of this last space, the so-called *cut distance*.

Definition 3.5 (Cut Distance, [17]). *For any $f, g \in \mathcal{W}$ we define the distance*

$$\delta_\square(f,g) := \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} [f(x,y) - g(x,y)]\, dx\, dy \right|. \tag{3.3}$$

At this point a remark is in order: note that, thinking in terms of graphs, it is enough to relabel the vertices for getting an object which, a priori is different, but has substantially the same properties of the starting one. As the graphs can be identified up to permutations, the same consideration can be applied to graphons, which can be classified in a parallel way up to bijections, thus grouping them in equivalence classes.

Definition 3.6 ([15], Section 3.3). *A measure-preserving bijection of $[0,1]$ is a map $\sigma : [0,1] \longmapsto [0,1]$ such that σ is a bijection and, for any Borel set S, the Lebesgue measures $\sigma(S)$ and $\sigma^{-1}(S)$ are both equal to that of S.*

Definition 3.7 (Equivalence relation). *Let Σ the space of all measure-preserving bijections $\sigma : [0,1] \longmapsto [0,1]$: we say that $f(x,y) \sim g(x,y)$ (namely f is equivalent to g) if it exists $\bar\sigma \in \Sigma$ such that*

$$g(x,y) = f(\bar\sigma(x), \bar\sigma(y)) =: f_{\bar\sigma}(x,y). \tag{3.4}$$

The quotient space with respect to this equivalence relation is denoted by $\widetilde{\mathcal{W}}$ and we indicate by $\tilde f$ the equivalence class of f. The quotient space has some interesting properties, for example, it can be equipped with the natural extension of the cut distance:

$$\delta_\square(\tilde f, \tilde g) := \inf_{\bar\sigma_1,\bar\sigma_2\in\Sigma} (f_{\bar\sigma_1}(x,y), g_{\bar\sigma_2}(x,y)),$$

for any $\tilde f, \tilde g \in \widetilde{\mathcal{W}}$, thus becoming a metric space. Furthermore:

Theorem 3.2 ([40], Theorem 5.1). *The metric space $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact.*

The framework defined up to now outlines the starting problem of properly devising a limit for $t(H, G_n)$ but the definition of convergence still lacks: before switching to it, we need to introduce the notion of subgraph density, namely the density of a fixed subgraph on a graphon. Such definition naturally emerges by noting that, thanks to expression (3.2) of f^G, it is possible to write the homomorphism density (3.1) in terms of the graphon ([15], Exercise 3.1):

$$t(H,G) = \int_{[0,1]^k} \prod_{(i,j)\in\mathcal{E}(H)} f^G(x_i,x_j) dx_1 \cdots dx_k. \tag{3.5}$$

As a direct consequence of (3.5), we introduce the following definition:

Definition 3.8 (Subgraph density, [17]). *Take a simple graph H with $\mathcal{V}(H) = [k]$ and $f \in \mathcal{W}$; the subgraph density of H in f is defined as*

$$t(H, f) := \int_{[0,1]^k} \prod_{(i,j)\in\mathcal{E}(H)} f(x_i, x_j) dx_1 \cdots dx_k. \tag{3.6}$$

For example if H is a triangle, (3.6) reads as follows:

$$t(H, f) := \int_0^1 \int_0^1 \int_0^1 f(x, y) f(y, z) f(z, x)\, dx\, dy\, dz. \tag{3.7}$$

Proposition 3.1. *The subgraph density $t(H, \cdot)$ is continuous for any finite simple graph H.*

It is now possible to precise the notion of convergence:

Definition 3.9 ([17]). *A sequence of graphs $\{G_n\}_n$ is said to converge to f in $\mathcal{W}$, if for every finite, simple graph H,*

$$\lim_{n\to\infty} t(H, G_n) = t(H, f).$$

Given a graphon f and a measure-preserving bijection $\bar{\sigma} \in \Sigma$, it is possible to show that

$$t(H, f) = t(H, f_{\bar{\sigma}})$$

and, from the equality displayed above and the fact that $\tilde{f} = \{f_{\bar{\sigma}} | \bar{\sigma} \in \Sigma\}$ immediately follows that

$$t(H, f) = t(H, \tilde{f}), \quad f \in \tilde{f}. \tag{3.8}$$

We stress that it is crucial to keep in mind relation (3.8) in order to avoid ambiguities in the notation during the dissertation of the following part of this thesis.

3.1.3 Large deviations

Throughout all this section we will refer to large deviations of dense Erdös-Rényi random graphs sequences $\{G_{n,p}\}_n$, where each graph $G_{n,p}$ belongs to the space $\mathcal{G}_{n,p}$. We recall that, according to this model, each node of the graph is connected or not, with independence, respectively with probability p and $1 - p$ and this probability does not depend on the size n of the graph. Let $X^{(n)} = \{X_{ij}\}_{i,j=1,\dots,n}$ be the symmetric adjacency matrix of an Erdös-Rényi random graph: each random variable $X_{ij} \in \{0, 1\}$ has a Bernoulli distribution of parameter p and denotes the presence ($X_{ij} = 1$) or absence ($X_{ij} = 0$) of the edge. We precise that the diagonal of the adjacency matrix is null since loops are not allowed and that each matrix $X^{(n)}$ lives in a space Ω_n with cardinality $|\Omega_n| = 2^{\binom{n}{2}}$. We denote by $\mathbb{P}_{n,p}$ the Erdös-Rényi measure: given a graph with adjacency matrix $X^{(n)}$, it is defined as follows

$$\mathbb{P}_{n,p}(X^{(n)}) := p^{\sum_{i<j\le n} X_{ij}} (1-p)^{\binom{n}{2} - \sum_{i<j\le n} X_{ij}} = (1-p)^{\binom{n}{2}} e^{h_p E(X^{(n)})}, \tag{3.9}$$

having settled $h_p := \ln \frac{p}{1-p}$ and recalling that by $E(X^{(n)}) = \sum_{i<j\leq n} X_{ij}$ we indicate the total number of edges of the graph. In the following, the expectation with respect to this measure will be denoted by $\mathbb{E}_p^{ER}(\cdot)$.
For a fixed $p \in (0,1)$, let $I_p(x) : [0,1] \mapsto [0,1]$ be the Bernoulli relative entropy

$$I_p(x) := x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}.$$

It is possible to extend $I_p(x)$ to $\mathcal{W}$ for any $f \in \mathcal{W}$ as follows:

$$I_p(f) = \int_{[0,1]^2} I_p(f(x,y))dxdy = \int_{[0,1]^2} \left[f(x,y) \ln \left(\frac{f(x,y)}{p} \right) + (1-f(x,y)) \ln \left(\frac{1-f(x,y)}{1-p} \right) \right] dxdy. \tag{3.10}$$

We recall that the function is defined with continuity over the whole interval $[0,1]$ by setting $I_p(0) = I_p(1) = 0$ and this convention also extends to $I_p(f)$.

Remark 3.2. The Bernoulli relative entropy (3.10) which takes as argument a graphon $f \in \mathcal{W}$, can be extended to the space $\widetilde{\mathcal{W}}$ by defining $I_p(\tilde{f}) := I_p(f)$: in other words, with notation $I_p(\tilde{f})$ we simply mean that the function is computed in a representative element f of the equivalence class $\tilde{f}$.

Another important property of $I_p(f)$ is that it is a convex function and, as a consequence, its extension $I_p(\tilde{f})$ on the space $\widetilde{\mathcal{W}}$ is lower-semicontinuous with respect to the cut metric ([19], Lemma 2.1).
Since the main task of this paragraph is to provide an overview on the main large deviations results for the sequence $\{G_{n,p}\}_n$, we start reporting the definition of large deviation principle, however a preliminary remark is in order.

Remark 3.3 ([18])**.** The Erdös-Rényi random graph $G_{n,p}$ induces, respectively, a probability distribution $\mathbb{P}_{n,p}$ on the space $\mathcal{W}$ through the map $G_{n,p} \longmapsto f^{G_{n,p}}$ and a probability distribution $\widetilde{\mathbb{P}}_{n,p}$ on $\widetilde{\mathcal{W}}$ through the map $G_{n,p} \longmapsto f^{G_{n,p}} \longmapsto \tilde{f}^{G_{n,p}} = \widetilde{G}_{n,p}$.

Large Deviation Principle (LDP) for dense ER random graphs [19]. The following theorem gives a Large Deviation Principle for the measure $\widetilde{\mathbb{P}}_{n,p}$ induced by the Erdös-Rényi random graph on the quotient space $\widetilde{\mathcal{W}}$ of graphons equipped with the cut metric.

Theorem 3.3. *([19], Theorem 2.3) For any fixed $p \in (0,1)$, the sequence $\widetilde{\mathbb{P}}_{n,p}$ obeys a large deviation principle on the space $\widetilde{W}$ with rate function defined in* (3.10)*.*
More precisely, for any closed set $\widetilde{F} \subseteq \widetilde{W}$:

$$\limsup_{n\to+\infty} \frac{2}{n^2} \ln \widetilde{\mathbb{P}}_{n,p}(\widetilde{F}) \leq - \inf_{\tilde{f}\in\widetilde{F}} I_p(\tilde{f}) \tag{3.11}$$

and for any open set $\widetilde{U} \subseteq \widetilde{W}$:

$$\liminf_{n\to+\infty} \frac{2}{n^2} \ln \widetilde{\mathbb{P}}_{n,p}(\widetilde{U}) \geq - \inf_{\tilde{f}\in\widetilde{U}} I_p(\tilde{f}). \tag{3.12}$$

As we have done before for a generic sequence $\{G_n\}_n$, we can consider the subgraph density of a simple graph H on the sequence of dense Erdös-Rényi graphs $\{G_{n,p}\}_n$ and look for the probability that such density exceeds a given threshold: what can be said about the exponential rate of decay of such probability? A first general answer is provided by Theorem 3.4.

Large deviation principle for subgraphs densities [19]. Consider the subgraphs density (3.6) and define

$$\phi^+(H,p,r) = \inf\{I_p(\tilde{f}) : \tilde{f} \in \widetilde{\mathcal{W}}, t(H,\tilde{f}) \geq r, \quad r \in [0,1]\}, \tag{3.13}$$

$$\phi^-(H,p,r) = \inf\{I_p(\tilde{f}) : \tilde{f} \in \widetilde{\mathcal{W}}, t(H,\tilde{f}) \leq r, \quad r \in [0,1]\}, \tag{3.14}$$

where $\widetilde{\mathcal{W}}$ is the space of graphons and H is a fixed subgraph. Since $I_p(\tilde{f})$ is lower semicontinuous in the space $\widetilde{\mathcal{W}}$ under the metric $\delta_\square$, the infimum is always attained. The following theorem, specialized to the Erdös-Rényi model, shows how the search of the rate function turns into a variational problem. In case of dense Erdös-Rényi graphs, due to the independent, identical distribution of the edges, it is possible to show that $\mathbb{E}(t(H,G_{n,p})) = p^{E(H)}$. Such consideration explains why $p^{E(H)}$ turns out to be the candidate threshold to compare with, when we speak of large deviations.

Theorem 3.4 ([15], Theorem 6.3)**.** *Let $\{G_{n,p}\}_n$ a sequence of random graphs belonging to the Erdös-Rényi model with $p \in (0,1)$. Then*

$$\lim_{n\to\infty} \frac{2}{n^2} \ln \mathbb{P}(t(H,G_{n,p}) \geq r^{E(H)}) = -\phi^+(H,p,r), \quad r > p, \tag{3.15}$$

$$\lim_{n\to\infty} \frac{2}{n^2} \ln \mathbb{P}(t(H,G_{n,p}) \leq r^{E(H)}) = -\phi^-(H,p,r), \quad r < p. \tag{3.16}$$

Let

a) $\widetilde{F}^+(H,p,r)$ be the set of minimizers of $I_p(\tilde{f})$ subject to the constraint $t(H,\tilde{f}) \geq r^{E(H)}$,

b) $\widetilde{F}^-(H,p,r)$ be the set of minimizers of $I_p(\tilde{f})$ subject to the constraint $t(H,\tilde{f}) \leq r^{E(H)}$.

Depending on the structure of minimizers we can distinguish two different regions:

The symmetric phase. We say that r is in the *symmetric phase* if $\widetilde{F}^+(H,p,r)$ contains only constant functions. Similarly, for $r < p^{E(H)}$ we say that r is in the *symmetric phase* if $\widetilde{F}^-(H,p,r)$ consists of constant functions.

The symmetry breaking phase. We say that r is in the *symmetry broken phase* if $\widetilde{F}^+(H,p,r)$ contains non-constant graphons. Similarly, for $r < p^{E(H)}$ we say that r is in the *symmetry broken phase* if $\widetilde{F}^-(H,p,r)$ consists of non-constant graphons.

The next large deviation result we need to introduce, requires the notion of *convex minorant*. For this reason, we make a brief digression on the geometric properties of $I_p(\cdot)$: this is the content of the following lemma.

Lemma 3.1 ([41], Lemma A.1)**.** *The function $x \mapsto I_p(\sqrt[\gamma]{x})$ with domain $(0,1)$ is convex for $0 < \gamma \leq 1$. If $\gamma > 1$ and*

$$p \geq p_0(\gamma) := \frac{\gamma - 1}{\gamma - 1 + e^{\frac{\gamma}{\gamma - 1}}} \tag{3.17}$$

then the function is also convex. If $\gamma > 1$ and $0 < p < p_0(\gamma)$, then the function has exactly two inflection points, with a region of concavity in the middle. Finally the function has infinite derivatives at both endpoints of $(0,1)$.

If we specialize to the case $\gamma = 2$ we can work out from Lemma 3.1 that the function $I_p(\sqrt{x})$ is convex when $p \geq p_0 = \frac{1}{1+e^2}$, whereas it has two inflection points when $p < p_0$. The *convex minorant* is get replacing the concave part with a linear part which consists of a segment between the two inflection points, as shown in Figure 3.1.

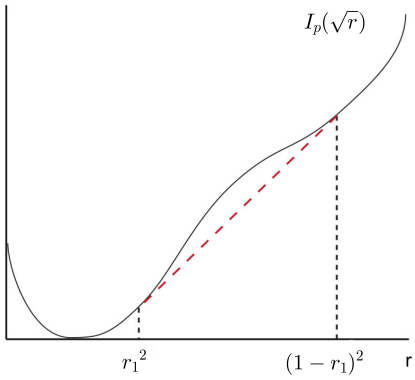


Figure 3.1: Convex minorant of $I_p(\sqrt{r})$: the concave part between the two inflection points r_1^2 and $(1-r_1)^2$ is replaced by the red, dotted line.

Keeping in mind the definition of convex minorant, we can characterize the large deviations of upper tails of triangles (for the Erdös-Rényi, dense model) through the following theorem.

Theorem 3.5 ([41], Theorem 1.1)**.** *Fix $0 < p \leq r < 1$ and let H be a finite, simple, regular graph of degree $d \geq 2$. Let $\{G_{n,p}\}_n$ a sequence of random graphs drawn from the Erdös-Rényi distribution.*

(i) If the point $(r^d, I_p(r))$ lies on the convex minorant of the function $I_p(x^{1/d})$ then

$$\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{P}(t(H, G_{n,p}) \geq r^{E(H)}) = -I_p(r)$$

(ii) If the point $(r^d, I_p(r))$ does not lie on the convex minorant of the function $I_p(x^{1/d})$ then

$$\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{P}(t(H, G_{n,p}) \geq r^{E(H)}) > -I_p(r)$$

Proposition 3.2. *When H is a triangle ($d = 2$), the symmetric phase corresponds to the following regions:*

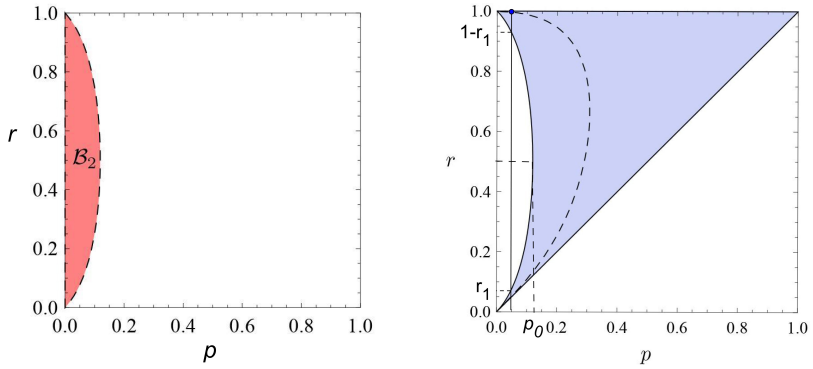
a) if $p \geq \frac{1}{1+e^2} \approx 0.1192$ then $r \in [p, 1)$ (this immediately follows from Lemma 3.1 and item (i) of Theorem 3.5);

b) if $p < \frac{1}{1+e^2} \approx 0.1192$ then $r \in [p, r_1] \cup [1 - r_1, 1)$ where r_1 is the unique solution of

$$\ln \frac{r}{1-r} - (1-2r) \ln \frac{p}{1-p} = 0 \qquad r \in (0, 1/2).$$

This last condition is recoverable from [41, Lemma A.2].

The breaking phase consists of all points (p, r) such that $(r^2, I_p(r))$ does not lie on the convex minorant: sticking to the notation given in [41], we denote such region by $\mathcal{B}_2$. The set is represented in Figure 3.2(a); Figure 3.2(b) puts the accent on the replica symmetric region (the blue one) and on its the profile which was guessed to be the optimal one before the work of Lubetzky and Zhao [41] (dotted line). Furthermore, the figure reports the critical values introduced in Proposition 3.2 and the threshold $p_0 = \frac{1}{1+e^2}$.
Rearranging statement (i) of Theorem 3.5 we can say that if the point $(r^2, I_p(r))$ lies on the



(a) (b)

Figure 3.2: Replica breaking region (3.2(a)) and phase diagram for the upper tails of triangle density (3.2(b)). The dotted line on the phase diagram represents the first, not optimal profile splitting the two regions (replica breaking and replica symmetric) given in the first work of Chatterjee and Dey ([17]). Both pictures are taken from [41].

.

convex minorant of $I_p(\sqrt{x})$, then

$$\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{P}(t(H, G_{n,p}) \geq r^{E(H)}) = \lim_{n\to\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{P}\left(\frac{T(G_{n,p})}{\binom{n}{3}} \geq r^3\right) \tag{3.18}$$

$$= \lim_{n\to\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{P}\left(\frac{T(G_{n,p})}{n} \geq \frac{r^3}{6}\right) = -I_p(r) \tag{3.19}$$

where $\frac{T(G_{n,p})}{n}$ denotes the normalized number of triangles of the graph or, equivalently (using formulation given in [19] where the correction terms of the limit are computed)

$$\mathbb{P}\left(\frac{T(G_{n,p})}{n} \geq \frac{r^3}{6}\right) = \exp\left(-n^2 \frac{I_p(r)}{2}\left(1 + O(n^{-\frac{1}{2}})\right)\right). \tag{3.20}$$

The problem of deriving a large deviation principle for the upper tails of triangles of a dense Erdös-Rényi model was initially faced by Chatterjee and Dey in [17]. The rate function was proved to be the relative Bernoulli entropy in a precise region which they called *replica symmetric*: roughly speaking the required condition for things to work was to choose $p > \frac{2}{2+e^{\frac{3}{2}}}$. Such interval of p, however, was not the optimal one: the precise characterization, summed up in Proposition 3.2, is due to the work of Lubetzky and Zhao [41] and it's also extended to a generic regular subgraph.
For what concerns the rate function of the lower tails of triangles of the dense Erdös-Rényi graph, we refer to [55] (Theorem 2.1). The main result basically states that there are two curves depending on p that we denote by $\bar{r}(p)$ and $\underline{r}(p)$ (such curves are found in [55], at least partially, through a numerical approach) which mark three regions: the one which lies above $\bar{r}(p)$ characterizes the replica symmetric, the one which lies below $\underline{r}(p)$ characterizes the replica breaking whereas the region included between $\underline{r}(p)$ and $\bar{r}(p)$ is unresolved. Again, for the replica breaking region, the analytic expression of the rate function is unknown whereas it coincides with the Bernoulli relative entropy above $\bar{r}(p)$. Figure 3.3 portrays the overview given in this section.

3.2 Exponential Random graphs and limiting free energy

The purpose of this paragraph is to define the Exponential Random Graph model providing the main results on the limiting free energy: in order to present this summary, we strictly follow [18]. This section is, at a first glance, disjoint from the large deviations context but, as we will argue at the end of the paragraph, such topics are deeply related.
Let $\mathcal{G}_n$ the space of simple graphs on n vertices: for a fixed graph $G_n \in \mathcal{G}_n$ one can define a probability measure, more precisely the Gibbs one, starting from the Hamiltonian

$$\mathcal{H}(G_n) := \sum_{i=1}^{k} \beta_i t(H_i, G_n), \tag{3.21}$$

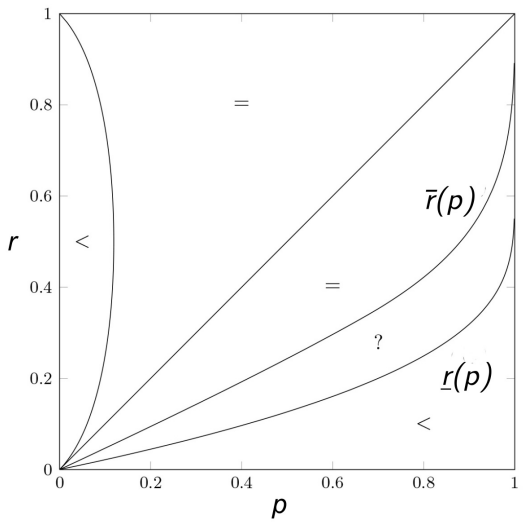


Figure 3.3: Phase diagram for triangle density: the symbol $<$ denotes the replica breaking phase, the symbol $=$ indicates the replica symmetric one, whereas the region denoted by ? is unresolved (more precisely it is not known if replica symmetric, replica breaking or both occur in it). The picture is taken from [55].

where β_i are real parameters, H_i are simple graphs (for example H_1 is the edge subgraph and H_2 is the triangle subgraph) and $t(H_i, G_n)$ is the homomorphism density defined in (3.1). The function $\mathcal{H}(G_n)$ is sometimes called *sufficient statistic*. On this model, we define the probability distribution

$$\mathbb{P}(G_n) = \frac{e^{\mathcal{H}(G_n)}}{Z_n}, \tag{3.22}$$

where Z_n is the normalizing partition function.
Even if in the standard setting the Hamiltonian weighs the regular subgraphs of the graph G_n (e.g. triangles and edges), it is possible to introduce a more general approach in which the Hamiltonian (3.21) is a continuous, bounded function on the space $(\widetilde{\mathcal{W}}, \delta_\square)$ which takes as argument the equivalence class of a graphon. In fact, we know that it is always possible to map elements of $\mathcal{G}_n$ into elements of $\widetilde{\mathcal{W}}$. In this context we can define the pressure

$$\psi_n(\beta) = \frac{1}{n^2} \ln \sum_{\widetilde{G}_n \in \widetilde{\mathcal{W}}} e^{n^2 \mathcal{H}(\widetilde{G}_n)} \tag{3.23}$$

and give the following theorem:

Theorem 3.6 ([18], Theorem 3.1)**.** *Let $\mathcal{H} : \widetilde{\mathcal{W}} \mapsto \mathbb{R}$ be a bounded continuous function and let*

$$I(\tilde{f}) := \int_{[0,1]^2} I(f(x,y))dxdy = \int_{[0,1]^2} f(x,y)\ln\left(f(x,y)\right) + (1-f(x,y))\ln\left(1-f(x,y)\right)\, dxdy \tag{3.24}$$

be the natural extension of the entropy to the space $\widetilde{\mathcal{W}}$. Then,

$$\psi(\beta) := \lim_{n\to\infty} \psi_n(\beta) = \sup_{\tilde{f}\in\widetilde{\mathcal{W}}} \left(\mathcal{H}(\tilde{f}) - \frac{1}{2}I(\tilde{f})\right). \tag{3.25}$$

Furthermore, the set of maximizers $\mathcal{F}^* \subseteq \widetilde{\mathcal{W}}$ is non-empty and compact.

In the same way as we did for problem (3.13), we can distinguish two regions basing on the set maximizers $\mathcal{F}^*$: we say that the Exponential Random Graph is in the *replica symmetric* phase if $\mathcal{F}^*$ only consists of constant functions. If the elements of the set are non-constant functions we say that the graph is in the *replica breaking* phase. The two next paragraphs are devoted to separately analyze these two regimes.
Even if the statement of Theorem 3.6 involves a general bounded, continuous $\mathcal{H}(\widetilde{G}_n)$, for the subsequent discussion we will focus on the model given by Hamiltonian (3.21). More precisely, we denote by

$$\mathcal{H}(\widetilde{G}_n) = \sum_{i=1}^{k} \beta_i t(H_i, G_n), \tag{3.26}$$

the natural extension of $\mathcal{H}(G_n)$ defined in (3.21) to the space of graphons and we consider $\psi_n(\beta)$ as given in (3.23).

Remark 3.4. We stress that the mapping between $\mathcal{H}(G_n)$ and $\mathcal{H}(\widetilde{G}_n)$ can be made more explicit by the following considerations. First, note that G_n can be seen as a graph limit f^{G_n} as specified in Remark 3.1: at this point $\mathcal{H}(G_n)$ reads as

$$\mathcal{H}(f^{G_n}) = \sum_{i=1}^{k} \beta_i t(H_i, f^{G_n}),$$

where $t(H_i, f^{G_n})$ is the subgraph density defined in (3.8). Finally, recalling relation (3.8), we have $t(H_i, f^{G_n}) = t(H_i, \tilde{f}^{G_n}) = t(H_i, \tilde{G}_n)$, being $\tilde{f}^{G_n} = \tilde{G}_n$.

Under the hypothesis that the coefficients $\{\beta_i\}_{i=2,\dots,k}$ in (3.26) are non-negative, we will see in the next paragraph that the graph is in the *replica symmetric* region and the variational problem (3.25) becomes a scalar problem.

3.2.1 Replica symmetric phase

The core results concerning the free energy when the replica symmetric occurs involve a simplified form of the variational problem which turns into a scalar one and the discussion of its

possible solutions as long as the parameters change. This is the content of the statements below.

Theorem 3.7 ([18], Theorem 4.1)**.** *Suppose that* $\{\beta_i\}_{i=2,\dots,k}$ *are non-negative and* $\mathcal{H}(\widetilde{G}_n)$ *is chosen as in* (3.26)*: then*

$$\lim_{n\to\infty} \psi_n(\beta) = \sup_{0\le u\le 1} \left[\sum_{i=1}^{k} \beta_i u^{E(H_i)} - \frac{1}{2} I(u) \right] \tag{3.27}$$

where

$$I(u) := u\ln(u) + (1-u)\ln(1-u). \tag{3.28}$$

Proof. The proof can be found in [18, Theorem 4.1]. □

The possible scenarios which emerge solving the scalar problem (3.27) are collected in the following theorem:

Theorem 3.8 ([18], Theorem 4.2)**.** *Let* G_n *be an Exponential Random Graph with sufficient statistic* $\mathcal{H}$ *defined in* (3.26)*. Furthermore we assume that* $\{\beta_i\}_{i=2,\dots,k}$ *are non-negative. Then*

a) If the maximization problem (3.27) *is solved at a unique value* u^**, then* G_n *is indistinguishable from the Erdös-Rényi random graph* G_{n,u^*} *in the large limit, in the sense that* $\widetilde{G}_n$ *converges to the constant function* u^* *in probability as* $n \to +\infty$*;*

b) Even if the maximizer is not unique, the set U *of maximizers is a finite subset of* $[0,1]$ *and*

$$\min_u \delta_\square(\widetilde{G}_n, \tilde{u}) \to 0 \qquad \text{in probability as} \quad n \to +\infty,$$

where $\tilde{u}$ *is the image of the constant function* u *in* $\widetilde{\mathcal{W}}$ *and* u *is randomly picked from the set of maximizers* U *according to some probability distribution.*

In both cases the Exponential Random Graph G_n*, when* n *is sufficiently large, looks like an Erdös-Rényi random graph.*

Item (a) of theorem (3.8) was firstly proved by Bhamidi et al. in specific window of parameters called *high temperature regime* (see [5], Theorem 7.1); in such region the scalar problem has a unique solution and the resulting free energy is smooth. It is worth while reporting the statement, since it will turn to be useful in Chapter 4 and its formulation could help in focusing the limiting behavior of the Exponential Random Graph from another angle. We define

$$\varphi(u) := \sum_{i=1}^{k} 2E(H_i)\beta_i u^{E(H_i)-1}, \tag{3.29}$$

noticing that this function is involved in the first derivative of (3.27) with respect to u. The following theorem holds:

Theorem 3.9 ([5], Theorem 7)**.** *Let G_n be an Exponential Random Graph with sufficient statistic* (3.26)*. Assume that parameters $\{\beta_i\}_{i=2,\dots,n}$ are non-negative and such that equation $\frac{e^{\varphi(u)}}{1+e^{\varphi(u)}} = u$ has a unique solution. Moreover let $\{X_{1,\xi}, X_{2,\xi}, \dots, X_{\xi-1,\xi}\}$ be an arbitrary collection of $\xi - 1$ edges ($\xi - 1 \in \mathbb{N}$). Then for all $(x_1, \dots, x_{\xi-1}) \in \{0,1\}^{\xi-1}$*

$$|\mathbb{P}(X_{1,\xi} = x_1, X_{2,\xi} = x_2, \dots, X_{\xi-1,\xi} = x_{\xi-1}) - (u^*)^{\sum_{i=1}^{\xi-1} x_i}(1-u^*)^{\xi-1-\sum_{i=1}^{\xi-1} x_i}| \overset{n\to+\infty}{\to} 0,$$

where u^* solves $\frac{e^{\varphi(u)}}{1+e^{\varphi(u)}} = u$.

As a direct consequence of Theorem 3.8 follows that the conclusion of Theorem 3.9 can be drawn for the more general setting in which the parameters $\{\beta_i\}_{i=2,\dots,n}$ are simply non-negative: the role of the unique solution of equation $\frac{e^{\varphi(u)}}{1+e^{\varphi(u)}} = u$ is played by the optimizer, u^*, of the scalar problem (3.27).

Remark 3.5. If we specialize Hamiltonian (3.26) restricting to the case $k = 2$, so we only consider triangles and edges density, we get:

$$\mathcal{H}(\widetilde{G}_n) = \frac{2\beta_1 E(G_n)}{n^2} + \frac{6\beta_2 T(G_n)}{n^3}, \tag{3.30}$$

where we recall that $E(G_n)$ and $T(G_n)$ respectively denote the number of edges and triangles of G_n and the coefficients multiplied by β_1 and β_2 appear since, when H is complete, the homomorphism density counts subgraphs multiple times (more precisely two times when H_i is an edge and six times when it's a triangle). Setting

$$6\beta_2 := \alpha, \quad 2\beta_1 := h_p = \ln\frac{p}{1-p} \tag{3.31}$$

with $p \in (0,1)$, statistic (3.30) reads

$$\mathcal{H}(\widetilde{G}_n) = \frac{h_p E(G_n)}{2n^2} + \frac{\alpha\, T(G_n)}{6n^3}. \tag{3.32}$$

Remark 3.6. Note that the free energy (3.25), when $\mathcal{H}(\widetilde{G}_n)$ is chosen as in (3.30), can be written as:

$$\psi(\alpha) = \frac{1}{2}\sup_{\tilde{f}\in\widetilde{\mathcal{W}}}\left[\alpha\frac{t(H_2,\tilde{f})}{3} + h_p\, t(H_1,\tilde{f}) - I(\tilde{f})\right],$$

or, equivalently (absorbing $\frac{h_p}{2}t(H_1,\tilde{f})$ into $\frac{I(\tilde{f})}{2}$)

$$\psi(\alpha) = \frac{1}{2}\sup_{\tilde{f}\in\widetilde{\mathcal{W}}}\left[\alpha\frac{t(H_2,\tilde{f})}{3} - I_p(\tilde{f})\right] - \frac{\ln(1-p)}{2}, \tag{3.33}$$

where $I_p(\tilde{f})$ is the Bernoulli relative entropy defined in (3.10).

When we are in replica symmetric regime, thanks to Theorem 3.7 the variational problem (3.33) turns into:

$$\psi(\alpha) = \frac{1}{2} \sup_{0 \le u \le 1} \left[\alpha \frac{u^3}{3} - I_p(u) \right] - \frac{\ln(1-p)}{2} \quad \alpha \ge 0. \tag{3.34}$$

In the following theorem, we consider the scalar problem present in equation (3.34).

Theorem 3.10. *Consider the scalar problem*

$$\sup_{0 \le u \le 1} \left(\alpha \frac{u^3}{3} - I_p(u) \right), \quad \alpha \ge 0. \tag{3.35}$$

The following statements hold:

a) The stationary points of (3.35) *are found solving the fixed-point equation*

$$\frac{e^{\alpha\, u^2 + h_p}}{e^{\alpha\, u^2 + h_p} + 1} = u. \tag{3.36}$$

b) If $p \ge \frac{2}{2+e^{3/2}} \approx 0.31$ *then* (3.36) *has a unique solution,* $u^*(h_p, \alpha)$*, which represents the maximum of* (3.35)*. Furthermore* $u^*(h_p, \alpha)$ *is continuous in* α *and* $\sup_{0 \le u \le 1} \left(\alpha \frac{u^3}{3} - I_p(u) \right)$ *is a differentiable function of* α*.*

c) If $p < \frac{2}{2+e^{3/2}} \approx 0.31$ *then there exists an interval* $[\alpha_1, \alpha_2]$ *such that* (3.36) *has more than one solution when* $\alpha \in [\alpha_1, \alpha_2]$*.*

d) When $p < \frac{2}{2+e^{3/2}} \approx 0.31$*, the maximum of* (3.35) *is reached at a unique* $u^*(h_p, \alpha)$*, except for a single value of* α*, that we call* $\bar{\alpha}$*, where the maximum is attained at two distinct* $u^*(h_p, \bar{\alpha})$*. Furthermore* $u^*(h_p, \alpha)$ *is discontinuous function of* α *(jump at* $\bar{\alpha} \in (\alpha_1, \alpha_2)$*) and, as a consequence,* $\sup_{0 \le u \le 1} \left(\alpha \frac{u^3}{3} - I_p(u) \right)$ *is not differentiable at* $\bar{\alpha}$*.*

We discuss the statements point by point:

a) Equation (3.36) is simply get deriving (3.35) with respect to u and rearranging terms in order to isolate the unknown.

b) The full study of equation (3.36) is made in [17]: in particular here it's proved that above the threshold $\bar{p} = \frac{2}{2+e^{3/2}}$ equation (3.36) has a unique solution. In order to show that $u^*(h_p, \alpha)$ is continuous in α, we display the geometric argument proposed by Lubetzky and Zhao in [41]. By the change of variable $u := \sqrt[3]{x}$, problem (3.35) can be written as follows:

$$\sup_{0 \le x \le 1} \left(\alpha \frac{x}{3} - I_p(\sqrt[3]{x}) \right), \quad \alpha \ge 0. \tag{3.37}$$

The following proposition gives a geometric characterization of the maxima and it is drawn from a discussion on the discontinuity of the solution $u^*(h_p, \alpha)$ provided in [41].

Proposition 3.3. *x^* maximizes (3.37) if and only if the tangent to the function $x \mapsto I_p(\sqrt[3]{x})$ at x^* has slope $\frac{\alpha}{3}$ and lies below the curve.*

Proof. If the tangent to the function $x \mapsto I_p(\sqrt[3]{x})$ at x^* has slope $\frac{\alpha}{3}$ and lies below the curve then the following condition holds

$$I_p(\sqrt[3]{x}) \geq \frac{\alpha}{3}(x - x^*) + I_p(\sqrt[3]{x^*}) \tag{3.38}$$

and, rearranging terms, it follows that

$$\frac{\alpha}{3}x - I_p(\sqrt[3]{x}) \leq \frac{\alpha}{3}x^* - I_p(\sqrt[3]{x^*}), \tag{3.39}$$

so x^* is a maximum.
Vice-versa, if x^* is a maximizer, then it is a stationary point and hence satisfies $\frac{\alpha}{3} = I_p'(\sqrt[3]{x^*})$, so the slope of the tangent line in x^* must be $\frac{\alpha}{3}$. Furthermore, since x^* is a maximum, it still holds inequality (3.39) (and consequently (3.38)), so the tangent lies below the curve $I_p(\sqrt[3]{x})$. □

For any $\alpha \in \mathbb{R}$, there is a unique lower tangent $y = \frac{\alpha}{3}(x-x^*)+I_p(\sqrt[3]{x^*})$ to the curve $I_p(\sqrt[3]{x})$ in x^*: furthermore, from Lemma 3.1 we know that if $p \geq \frac{2}{2+e^{\frac{3}{2}}}$, as in our hypothesis, $I_p(\sqrt[3]{x})$ is convex and the maximizer x^* of (3.37) is unique.
As a consequence, as α grows, the related tangent point $x^* = u^{*3}$ does the same, moving on the smooth, convex curve $I_p(\sqrt[3]{x})$.

c) if $p < \frac{2}{2+e^{\frac{3}{2}}}$, it exists an interval $[\alpha_1, \alpha_2]$ in which one can find more than one solution of (3.36): the proof, with the precise characterization of these two values, is given in [17, Lemma 12].
Here, we just report how to recover α_1 and α_2: for a fixed $h_p < \bar{h} = \ln(2) - \frac{3}{2}$ (which corresponds to $\bar{p} = \frac{2}{2+e^{\frac{3}{2}}}$) one can find two distinct solutions, $0 < a_1(h_p) < \frac{1}{2} < a_2(h_p)$, of equation

$$\ln(x) + \frac{1+x}{2x} + h_p = 0$$

which lead to

$$\alpha_1 = \frac{(1+a_1)^3}{2a_1} \tag{3.40}$$

$$\alpha_2 = \frac{(1+a_2)^3}{2a_2}. \tag{3.41}$$

The possible cases are the following:

- If $\alpha \in (\alpha_1, \alpha_2)$ there are three solutions of (3.36)

- if $\alpha = \alpha_1$ or $\alpha = \alpha_2$ there are two solutions of (3.36).
- $\alpha \notin [\alpha_1, \alpha_2]$ there is one solution of (3.36).

In general, among all the solutions of (3.36), there is a unique stationary point which realizes the maximum except for a special value of α, say $\bar{\alpha}(h_p) \in (\alpha_1, \alpha_2)$, which corresponds to two maxima (there is not a closed form for $\bar{\alpha}(h_p)$; in order to find out such value one must resort to simulations). Figure 3.2(a) shows how the critical parameter $\bar{\alpha}(h_p)$ changes as h_p varies (blue curve). The picture also represents $\alpha_1(h_p)$ and $\alpha_2(h_p)$ given by equations (3.40) and (3.41) (red curves). We recall that the inner region delimited by such red curves corresponds to three solutions of the fixed-point equation, among which we can select one maximizer except for the case in which the parameters of the optimization problem (3.35) lie on $\bar{\alpha}(h_p)$ and we have two maximizers.

d) The proof of point (d) is given in [47, Proposition 3.2] starting from a maximization problem for the more general function

$$l(u, \beta_1, \beta_2) = \beta_1 u + \beta_2 u^k - \frac{I_p(u)}{2}, \quad \beta_1 \in \mathbb{R}, \beta_2 \geq 0, \tag{3.42}$$

where k is an integer greater or equal to two.
Note that choosing $(\beta_1, \beta_2) = (\frac{h_p}{2}, \frac{\alpha}{6})$ and $k = 3$, one recovers (up to the constant $\frac{1}{2}$) the scalar problem (3.35). Keeping the focus on this setting, Radin and Yin proved that, for a fixed p, the limiting free energy, has a first order phase transition in correspondence of the critical value $\bar{\alpha}(h_p)$, introduced in item c). Figure 3.2(a) represents the curve $\bar{\alpha}(h_p)$ as a function of h_p: it lies between $\alpha_1(h_p)$ and $\alpha_2(h_p)$ expressed by equations (3.40), (3.41) (the so-called V-shaped region) and ends in a critical point $(h_p^c, \alpha^c) = (\ln(2) - \frac{3}{2}, \frac{27}{8})$ which realizes a second order phase transition. Figure (3.2(a)) represents the V-shaped region with the phase transition $\bar{\alpha}(h_p)$ whereas Figure (3.2(b)) shows the discontinuity of the solution $u^* \left(\frac{h_p}{2}, \frac{\alpha}{6}\right)$ crossing the critical value $\bar{\alpha}\left(\frac{h_p}{2}\right)$.

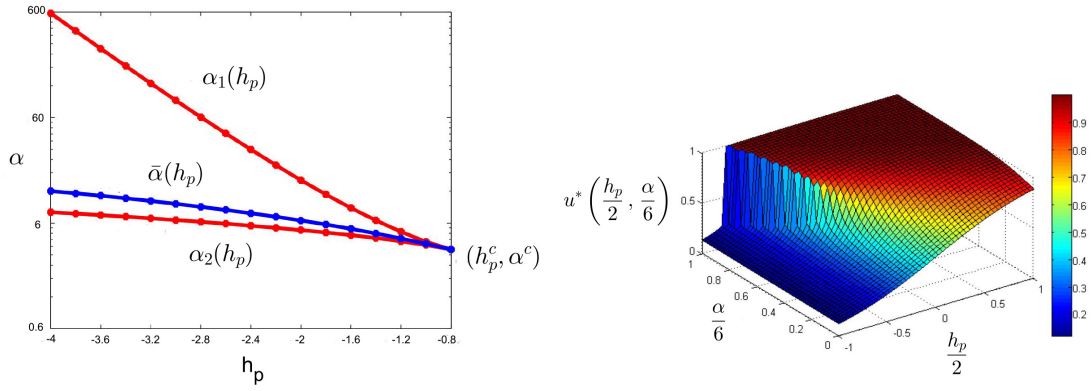
Remark 3.7. From items b) and d) in Theorem 3.10, it follows that the scalar problem (3.35) has a unique solution for any p and any $\alpha \geq 0$: in fact, even when the critical α realizes two maxima, we can pick at random one of them according to item b) of Theorem 3.8. More precisely from (3.34) follows

$$\psi(\alpha) = \frac{1}{2}\left\{\sup_{0 \leq u \leq 1}\left[\alpha\frac{u^3}{3} - I_p(u)\right] - \ln(1-p)\right\} = \alpha\frac{u^{*3}}{6} - \frac{I_p(u^*)}{2} - \frac{\ln(1-p)}{2}, \quad \alpha \geq 0 \tag{3.43}$$

where u^* solves (3.36).

Geometrical interpretation of the Phase transition

An accurate description of the phenomenon of the phase transition in terms of triangles density is given by Chatterjee and Diaconis and formalized in [18](Theorem 5.1): we briefly summarize



(a) (b)

Figure 3.4: V-shaped region (3.2(a)) (picture from [47]) and discontinuity of the solution $u^*\left(\frac{h_p}{2}, \frac{\alpha}{6}\right)$ crossing the critical value $\bar{\alpha}\left(\frac{h_p}{2}\right)$. (3.2(b)) (picture from [18])

.

the picture. The expected number of triangles is the first derivative of the free energy $\psi(\alpha)$, given in (3.43), with respect to α: this means that if $u^*(h_p, \alpha)$ shows a jump, the average number of triangles behaves the same. More precisely, crossing the critical value $\bar{\alpha}(h_p)$, which corresponds to two maximizers $u_1^*(h_p, \bar{\alpha})$ and $u_2^*(h_p, \bar{\alpha})$ of (3.35), the model switches from a lower density of triangles to an higher one, skipping intermediate structures. A geometric interpretation of the phenomenon is provided below and Figure 3.5 gives an intuition of it.
First of all, looking at the scalar problem (3.35), we recall that, using the change of variable $u := \sqrt[3]{x}$, the phase transition happens when $I_p(\sqrt[3]{x})$ is not convex (this corresponds to $p < p_0(3)$, whose equation is given in (3.17)). Furthermore from Proposition 3.3 we know that, in this specific setting, the tangent line in both maxima x_1^* and x_2^* has the same critical slope $\frac{\bar{\alpha}}{3}$ and lies below $I_p(\sqrt[3]{x})$. The storyline of the optimal solution passes through three steps: first, the maximizer is unique and grows with α as long as α reaches the critical value $\bar{\alpha}(h_p)$; at this special value the scalar problem has two maximizers and the tangent line touches $I_p(\sqrt[3]{x})$ at x_1^* and x_2^*, as shown in Figure 3.5. After the critical value the maximizer turns to be unique again but it jumps over the interval (x_1^*, x_2^*) and goes on growing continuously with respect to α. Figure 3.6 gives another evidence of the phase transition for $p = 0.2$: the maximizer (alias the density of triangles) passes from being close to 0.1 to being close to 0.9. In this case the critical value is $\bar{\alpha} \approx 4.75$. Note that in Figure 3.6 the three values of α chosen for the pictures, lie inside the interval (α_1, α_2), where α_1 and α_2 can be recovered through equations (3.40) and (3.41). In fact, in both cases we can observe that there are three solutions of the fixed-point equation (3.36) (right column of Figure 3.6).

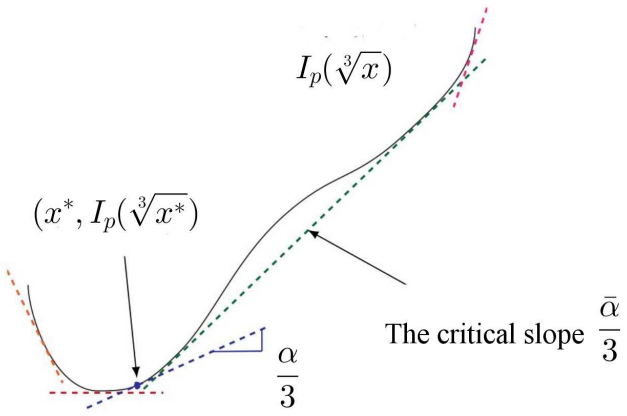


Figure 3.5: Geometrical argument for the phase transition (picture from [41]).

All the conclusions drawn up to now, concern the replica symmetric region and lean on the hypothesis of Theorems 3.7 and 3.8 of having an Hamiltonian $\mathcal{H}(\widetilde{G}_n) = \sum_{i=1}^{k} \beta_i t(H_i, G_n)$ with non- negative coefficients $\{\beta_i\}_{i=2,\dots,k}$: Chatterjee and Diaconis proved that the replica symmetric phase also occurs when such coefficients are negative but small enough in absolute value. The next theorem makes this statement precise.

Theorem 3.11 ([18], Theorem 6.1)**.** *Consider the Exponential Random Graph with sufficient statistic* $\mathcal{H}(\widetilde{G}_n) = \sum_{i=1}^{k} \beta_i t(H_i, G_n)$*. Suppose that* $\{\beta_i\}_{i=1,\dots,k}$ *are such that*

$$\sum_{i=2}^{k} |\beta_i| E(H_i)(E(H_i) - 1) < 2, \tag{3.44}$$

where $E(H_i)$ *is the number of edges of* H_i*. Then the conclusions of Theorems 3.7 and 3.8 hold for this vector of parameters* $(\beta_1, \dots, \beta_k)$*.*

In particular, when we consider H_2, i.e triangles, recalling that $\beta_2 = \frac{\alpha}{6}$ (from (3.31)) and $E(H_2) = 3$, condition (3.44) reads

$$|\alpha| < 2. \tag{3.45}$$

Remark 3.8. An important remark is in order: when we start considering negative values of α, the solution of the fixed-point equation (3.36) which realizes the maximum is lower than p. Note that such solution represents the average density of triangles. Thinking in terms of large deviations, this means that we are looking at the lower tails of triangles. Moreover the solution of (3.36) is unique when α is negative and this can be worked out observing that the function $\varphi(u) := \frac{e^{\alpha u^2 + h_p}}{e^{\alpha u^2 + h_p} + 1} - u$, with $u \in (0, 1)$, crosses the abscissas axis (in fact, note that $\varphi(0) > 0$ and $\varphi(1) < 0$) and that it has a negative first derivative: $\frac{d\varphi(u)}{du} = 2\alpha\varphi(u)(1 - \varphi(u)) - 1 < 0$, if $\alpha < 0$.

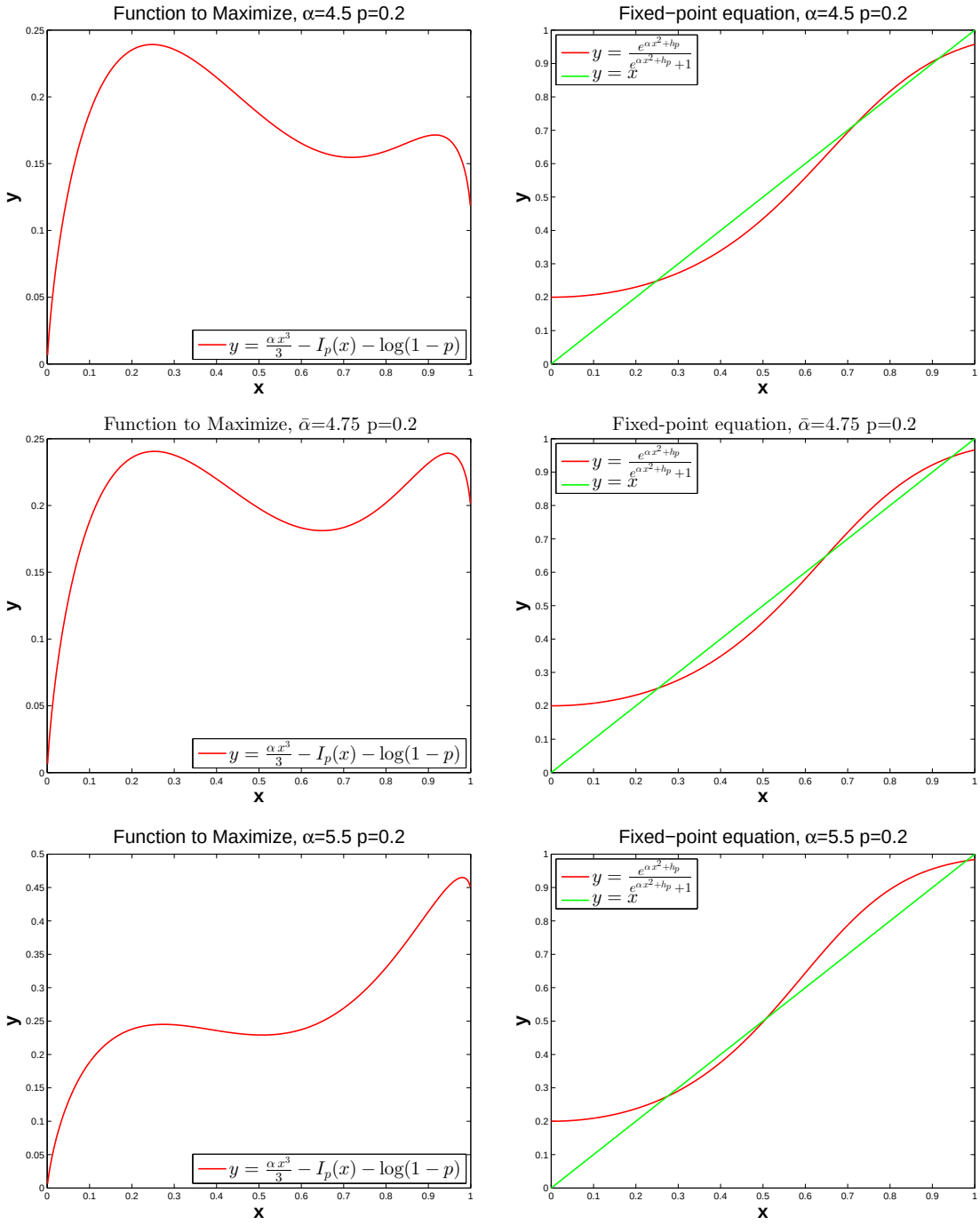


Figure 3.6: Representation of the function $\alpha\frac{u^3}{3} - I_p(u) - \log(1-p)$ involved in the optimization problem (3.43) (left) and graphical representation of the fixed-point equation which arises from the problem (right). We can observe that the maximum is close to 0.1 (first row) up to $\bar{\alpha} \approx 4.75$, at which the function is optimized at two values (central row). Above $\bar{\alpha}$, which corresponds to a phase transition, the maximum is close to 0.9 (bottom row).

3.2.2 Replica breaking

Up to now we have investigated the behavior of the free energy when the vector of parameters β lies in the replica symmetric region but what happens when the coefficients $\{\beta_i\}_{i=2,\dots,k}$ are not positive neither sufficiently negative and small in absolute value? We quote the two following theorems, proved in [18], which describe the behavior of the Exponential Random Graph when the statistic $\mathcal{H}$ only consists of triangles and edges and parameters are chosen as in (3.31).
To be precise, we work in the following setting:

$$\mathcal{H}(\widetilde{G}_n) = \frac{1}{2}\left[\alpha\frac{t(H_2,f)}{3} + h_p\, t(H_1,f)\right] \tag{3.46}$$

and we want to maximize $\mathcal{H}(f) - \frac{I(f)}{2}$.

Theorem 3.12 ([18], Theorem 6.2)**.** *Consider the Exponential Random Graph with sufficient statistic* (3.46)*: then for any given value of p there exists a positive constant $C(h_p)$, sufficiently large, so that whenever $\alpha \le -C(h_p)$, $\mathcal{H}(f) - \frac{I(f)}{2}$ is not maximized at any constant function. Consequently, if G_n is an Exponential Random Graph with this sufficient statistic, then there exists $\varepsilon > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}(\delta_\square(\widetilde{G}_n, \widetilde{C}) > \varepsilon) = 1,$$

where $\widetilde{C}$ is the set of constant functions. In other words, G_n does not behave like and Erdös-Rényi random graph, in the large limit.

Proof. We follow the proof given by Chatterjee and Diaconis in [18]. To start with, fix p and define $\gamma := -\frac{\alpha}{6}$. We recall that, with a few algebraic steps, one can write the statistic $\mathcal{H}$ using the relative entropy $I_p(f)$ in the following way:

$$\mathcal{H}(f) - \frac{1}{2}I(f) = -\gamma\, t(H_2,f) - \frac{1}{2}I_p(f) - \frac{1}{2}\ln(1-p), \tag{3.47}$$

so that the variational problem reads

$$\psi(\gamma) = \sup_{f\in\mathcal{W}}\left[\mathcal{H}(f) - \frac{1}{2}I(f)\right] = \sup_{f\in\mathcal{W}}\left[-\gamma\, t(H_2,f) - \frac{1}{2}I_p(f) - \frac{1}{2}\ln(1-p)\right]. \tag{3.48}$$

The proof is made by contradiction: we start supposing that α is negative (so that γ is positive) and that the constant function $f(x,y) = u$ maximizes $\mathcal{H}(f) - \frac{1}{2}I(f)$. Then, we turn the maximization problem into a minimization one, thus getting:

$$\min_{f\in\mathcal{W}}\left[\gamma\, t(H_2,f) + \frac{1}{2}I_p(f)\right] = \gamma\, u^3 + \frac{1}{2}I_p(u).$$

The considerations that follow aim at showing that it is possible to find out a best optimizer, thus leading to an absurd.
From the optimality condition of u it follows that $\gamma\, u^3 + \frac{1}{2}I_p(u) \le \gamma\, x^3 + \frac{1}{2}I_p(x) \quad \forall x \in [0,1]$ and

this implies $u \in (0,1)$ (indeed, we can exclude the extremal values because the first derivative in such points is infinite). Furthermore, the stationarity condition yields

$$0 = \frac{d}{dx}\left(\gamma\, x^3 + \frac{1}{2} I_p(x)\right)\Bigg|_{x=u} = 3\gamma u^2 + \frac{1}{2}\ln\left(\frac{u(1-p)}{p(1-u)}\right) \tag{3.49}$$

and, rearranging terms,

$$u^2 = -\frac{1}{6\gamma}\ln\left(\frac{u(1-p)}{p(1-u)}\right). \tag{3.50}$$

From (3.50) follows that, since $u \in (0,1)$ and $p \in (0,1)$, $\lim_{\gamma\to+\infty} u(\gamma) = 0$. As a consequence

$$\lim_{\gamma\to+\infty} \min_{0\le x\le 1}\left(\gamma\, x^3 + \frac{1}{2} I_p(x)\right) = \frac{I_p(0)}{2} = \frac{1}{2}\ln\frac{1}{1-p}. \tag{3.51}$$

We now separately consider the so-called *bipartite graphon*

$$g_{0,p}(x,y) := \begin{cases} 0 & \text{if} \quad (x,y) \in \left[0,\frac{1}{2}\right]^2 \cup \left[\frac{1}{2},1\right]^2 \\ p & \text{if} \quad (x,y) \in \left[0,\frac{1}{2}\right] \times \left[\frac{1}{2},1\right] \cup \left[\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right] \end{cases}; \tag{3.52}$$

for almost every (x,y,z) it realizes a null density of triangles, indeed $g_{0,p}(x,y)g_{0,p}(y,z)g_{0,p}(z,x) = 0$ and consequently $t(H_2, g_{0,p}) = 0$. Recalling that $I_p(p) = 0$, so restricting to the subsets of $[0,1]$ where $g_{0,p}(x,y) = 0$, the integral of the Bernoulli relative entropy immediately follows:

$$\int_0^1\int_0^1 \frac{I_p(g_{0,p}(x,y))}{2} dx\, dy = \frac{1}{2}\left(\frac{1}{4} I_p(0) + \frac{1}{4} I_p(0)\right) = \frac{1}{4}\ln\frac{1}{1-p}. \tag{3.53}$$

Since $\frac{1}{4}\ln\frac{1}{1-p} < \frac{1}{2}\ln\frac{1}{1-p}$ the conclusion is that the constant function u can not be the minimizer of $\gamma\, t(H_2, f) + \frac{1}{2} I_p(f)$.
The final part of the statement follows from Theorem 3.2 in [18] and the compactness of $\widetilde{W}$. □

In order to complete the scenario we also report the following theorem which describes the limiting behavior of the free energy when $\alpha \to -\infty$. It is a special case of Theorem 7.1 in [18] when we consider the sufficient statistic (3.46), namely when we deal with parameters h_p and α (which respectively tune the density of edges and triangles) settled in (3.31).

Theorem 3.13 ([18], Theorem 7.1)**.** *For a fixed p, consider the sufficient statistic* (3.46) *with edges and triangles and let $\mathcal{F}^*(\alpha)$ be the set of maximizers of the variational problem*

$$\sup_{\tilde f \in \widetilde{\mathcal{W}}}\left[\mathcal{H}(\tilde f) - \frac{I(\tilde f)}{2}\right].$$

Then

$$\lim_{\alpha\to-\infty} \sup_{\tilde f\in\mathcal{F}^*(\alpha)} \delta_\square(\tilde f, \tilde g_{0,p}) = 0$$

and

$$\lim_{\alpha\to-\infty} \psi(\alpha) = \frac{1}{4}\ln\frac{1}{1-p}. \tag{3.54}$$

Roughly speaking, the unique maximizer of the variational problem, in the negative limit for α, is the bipartite graphon. Furthermore the bipartite graphon is the function which minimizes the relative entropy $I_p(f)$ among all graphons with a null density of triangles. This is formalized by the following theorem.

Remark 3.9. The bipartite graphon (3.52) preserves the same properties of the bipartite graph of Definition 3.1. Since in bipartite graphs the number of odd cycles is null, also bipartite graphons must be characterized by a null density of odd regular subgraphs. As a consequence the density triangles $t(H_2, g_{0,p})$ is null.

Theorem 3.14 ([18], Theorem 7.3)**.** *Let $g_{0,p}(x,y)$ be the function defined in* (3.52)*, take $p \in (0,1)$ and let H_2 be a triangle: if f is any element of $\mathcal{W}$ which minimizes $I_p(f)$ among all f such that $t(H_2, f) = 0$, then $\tilde{f} = \tilde{g}_{0,p}$.*

The last proposition we report, highlights that a necessary condition for a representative element $f \in \tilde{f}$ for being a maximizer of the variational problem

$$\sup_{\tilde{f}\in\widetilde{\mathcal{W}}} \frac{1}{2}\left[\alpha\frac{t(H_2,\tilde{f})}{3} + h_p\, t(H_1,\tilde{f})\right] - \frac{I(\tilde{f})}{2} = \sup_{\tilde{f}\in\widetilde{\mathcal{W}}} \frac{1}{2}\left[\alpha\frac{t(H_2,\tilde{f})}{3} - I_p(\tilde{f}) - \log(1-p)\right], \quad (3.55)$$

is to solve equation (3.56).

Theorem 3.15 ([18], Theorem 6.3)**.** *If $\tilde{f} \in \widetilde{W}$ maximizes $\frac{\alpha}{3}t(H_2,\tilde{f}) - I_p(\tilde{f})$ then any representative element $f \in \tilde{f}$ must satisfy for almost all $(x,y) \in [0,1]^2$,*

$$f(x,y) = \frac{e^{\alpha\int_0^1 f(x,z)f(z,y)dz + h_p}}{e^{\alpha\int_0^1 f(x,z)f(z,y)dz + h_p} + 1} \quad \alpha \in \mathbb{R}. \quad (3.56)$$

Moreover, any maximizing function must be bounded away from 0 and 1.

We remark that (3.36) is the scalar version of (3.56): moreover, if we choose $f(x,y)$ identically equal to any of the solutions of (3.36) when $\alpha \in \mathbb{R}$, we have that (3.56) is verified. In conclusion, such constant functions belong to the stationary points of the functional $\frac{\alpha}{3}t(H_2,\tilde{f}) - I_p(\tilde{f})$.

Proof. Let h be a bounded symmetric measurable function from $[0,1]^2$ into $\mathbb{R}$. For each $u \in \mathbb{R}$ define

$$f_u(x,y) := f(x,y) + uh(x,y). \quad (3.57)$$

First, suppose that the maximizer f is bounded away from 0 and 1. Then, if u is sufficiently small, f_u belongs to the space of graphons $\mathcal{W}$ and, since f is an optimizer, the following inequality holds:

$$\frac{\alpha}{3}t(H_2,f_u) - I_p(f_u) \leq \frac{\alpha}{3}t(H_2,f) - I_p(f).$$

In particular,

$$\frac{\partial}{\partial u}\left(\frac{\alpha\, t(H_2,f_u)}{3} - I_p(f_u)\right)\Bigg|_{u=0} = 0. \quad (3.58)$$

The derivative is given by

$$\frac{\partial}{\partial u}\frac{\alpha}{3}t(H_2,f_u)-\frac{\partial}{\partial u}I_p(f_u),$$

where

$$\begin{aligned}
&\frac{\partial}{\partial u}t(H_2,f_u)=\frac{\partial}{\partial u}\int_0^1\int_0^1\int_0^1 f_u(x,y)f_u(y,z)f_u(z,x)dxdydz\\
&=\int_0^1\int_0^1\int_0^1\frac{\partial f_u(x,y)}{\partial u}f_u(y,z)f_u(z,x)+f_u(x,y)\frac{\partial f_u(y,z)}{\partial u}f_u(z,x)+f_u(x,y)f_u(y,z)\frac{\partial f_u(z,x)}{\partial u}dxdydz\\
&=3\int_0^1\int_0^1\frac{\partial f_u(x,y)}{\partial u}\int_0^1 f_u(y,z)f_u(z,x)dxdydz\\
&=3\int_0^1\int_0^1 h(x,y)\int_0^1 f_u(y,z)f_u(z,x)dxdydz,
\end{aligned}$$

and

$$\frac{\partial I_p(f_u)}{\partial u}=\int_0^1\int_0^1\frac{\partial}{\partial u}I_p\left(f(x,y)+uh(x,y)\right)dxdy=\int_0^1\int_0^1 h(x,y)\ln\left(\frac{f_u(x,y)(1-p)}{p(1-f_u(x,y))}\right)dxdy.$$

Imposing condition (3.58), it can be concluded that for any measurable function $h(x,y):[0,1]\to\mathbb{R}$ must hold:

$$\int_0^1\int_0^1 h(x,y)\left(\alpha\int_0^1 f(y,z)f(z,x)dz-\ln\left(\frac{f(x,y)(1-p)}{p(1-f(x,y))}\right)\right)dxdy=0. \tag{3.59}$$

In particular, choosing $h(x,y)=\left(\alpha\int_0^1 f(y,z)f(z,x)dz-\ln\left(\frac{f(x,y)(1-p)}{p(1-f(x,y))}\right)\right)$ (which is a bounded function since f is bounded by hypothesis) the conclusion of the theorem follows. As a final step, we remove the assumption that f is bounded away from 0 and 1 and we prove the theorem under the hypothesis that $f\in\mathcal{W}$ is a generic maximizer.
For each $u\in[0,1]$ and a fixed $\omega\in(0,1)$ we define

$$f_{u,\omega}(x,y):=(1-u)f(x,y)+u\max\{f(x,y),\omega\}. \tag{3.60}$$

We observe that $f_{u,\omega}(x,y)$ is a symmetric bounded measurable function from $[0,1]^2\to[0,1]$ and

$$\frac{\partial f_{u,\omega}(x,y)}{\partial u}=\max\{f(x,y),\omega\}-f(x,y)=(\omega-f(x,y))_+, \tag{3.61}$$

where notation $(\cdot)_+$ denotes the positive part of the argument. Using the same strategy of the first part of the proof, we get:

$$\begin{aligned}
\frac{\partial}{\partial u}t(H_2,f_{u,\omega})&=3\int_0^1\int_0^1\frac{\partial f_{u,\omega}(x,y)}{\partial u}\int_0^1 f_{u,\omega}(y,z)f_{u,\omega}(z,x)dxdydz\\
\frac{\partial I_p(f_{u,\omega})}{\partial u}&=\int_0^1\int_0^1\frac{\partial f_{u,\omega}(x,y)}{\partial u}\ln\left(\frac{f_{u,\omega}(x,y)(1-p)}{p(1-f_{u,\omega}(x,y))}\right)dxdy.
\end{aligned}$$

Finally, using (3.61), we obtain:

$$\begin{aligned}
&\frac{\partial}{\partial u}\left(\frac{\alpha}{3}t(H_2, f_{u,\omega}) - I_p(f_{u,\omega})\right)\Big|_{u=0} \\
&= \int_0^1\int_0^1 \left(\alpha\int_0^1 f(y,z)f(z,x)dz - \ln\left(\frac{f(x,y)(1-p)}{p(1-f(x,y))}\right)\right)(\omega - f(x,y))_+ dxdy \\
&= \int_0^1\int_0^1 \left(\alpha\int_0^1 f(y,z)f(z,x)dz + h_p - \ln\left(\frac{f(x,y)}{1-f(x,y)}\right)\right)(\omega - f(x,y))_+ dxdy \\
&\geq \int_0^1\int_0^1 \left(-C - \ln\left(\frac{f(x,y)}{1-f(x,y)}\right)\right)(\omega - f(x,y))_+ dxdy,
\end{aligned}$$

where C is a positive constant depending on h_p (we recall that $h_p = \ln\frac{p}{1-p}$), α, H_1 and H_2. If ω is so small that

$$-C - \ln\frac{\omega}{1-\omega} > 0 \tag{3.62}$$

and $f(x,y) < \omega$ then

$$\frac{\partial}{\partial u}\left(\frac{\alpha}{3}t(H_2, f_{u,\omega}) - I_p(f_{u,\omega})\right)\Big|_{u=0} > 0$$

on a set of positive Lebesgue measure. In conclusion $f(x,y) \geq \omega$ almost everywhere and this prove that the maximizer must be bounded away from zero. In order to prove that the maximizer is bounded away from one, the same argument can be used with $f_{u,\omega}(x,y) := (1-u)f(x,y) - u(f(x,y)-\omega)_+$. □

3.2.3 From the Exponential Random graph to Erdös-Rényi

Up to now, we have analyzed the Exponential Random Graph model quoting the main results concerning the free energy. In this paragraph we want to bring the focus back to the scaled cumulant generating function defined in (1.4), when the observable is the normalized number of triangles of a dense Erdös-Rényi random graph, underlining the connection between these two functions. To begin with, we stress the link between the Exponential Random Graph and the Erdös-Rényi model. The parallelism is suggested by the comparison between the Erdös-Rényi measure

$$\mathbb{P}_{n,p}(X^{(n)}) = (1-p)^{\binom{n}{2}} e^{h_p E(X^{(n)})} \tag{3.63}$$

and the Exponential Random Graph probability distribution

$$\mathbb{P}(X^{(n)}) = \frac{e^{\mathcal{H}(X^{(n)})}}{Z_n} : \tag{3.64}$$

setting $\mathcal{H}(X^{(n)}) := h_p E(X^{(n)})$ and checking that $Z_n = \frac{1}{(1-p)^{\binom{n}{2}}}$ it is possible to map (3.63) into (3.64). The control can be easily done using the Newton binomial, $(a+b)^M = \sum_{k=0}^M \binom{M}{k} a^k b^{M-k}$,

with $a = e^{h_p}$ and $b = 1$:

$$Z_n = \sum_{X^{(n)} \in \Omega_n} e^{h_p \mathcal{H}(X^{(n)})} = \sum_{k=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{k} e^{h_p k} = (1 + e^{h_p})^{\binom{n}{2}} = \left(1 + \frac{p}{1-p}\right)^{\binom{n}{2}} = \left(\frac{1}{1-p}\right)^{\binom{n}{2}}.$$

Speaking in terms of the sufficient statistic (3.21), which we report here

$$\mathcal{H}(G_n) = \sum_{i=1}^{k} \beta_i t(H_i, G_n),$$

we get the Erdös-Rényi model when we include only edges in the Hamiltonian and we set the related parameter to h_p, in other words when $\beta = (h_p, 0, \ldots, 0)$.
Exploiting this connection we want to highlight the link between the free energy defined in (3.25) and the scaled cumulant generating function. We recall that, when the Gärtner Ellis theorem holds, it is the Legendre transform of the rate function. We report the definition below, using the notation $\mathbb{E}_p^{ER}(\cdot)$ for denoting the expectation with respect to the measure (3.63) (according to the terminology fixed at the beginning of Subsection 3.1.3):

$$\mu(\alpha) = \lim_{n \to +\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{E}_p^{ER}(e^{\alpha \frac{T(X^{(n)})}{n}}) = \lim_{n \to +\infty} \frac{1}{\binom{n}{2}} \ln \sum_{X^{(n)} \in \Omega_n} e^{\alpha \frac{T(X^{(n)})}{n}} \mathbb{P}_{n,p}(X^{(n)}) \tag{3.65}$$

$$= \lim_{n \to +\infty} \frac{1}{\binom{n}{2}} \ln \sum_{X^{(n)} \in \Omega_n} e^{\alpha \frac{T(X^{(n)})}{n}} e^{h_p E(X^{(n)})} (1-p)^{\binom{n}{2}} = \ln(1-p) + 2\psi(\alpha) \tag{3.66}$$

$$= \sup_{\tilde{f} \in \widetilde{\mathcal{W}}} \left(\frac{\alpha\, t(H_2, \tilde{f})}{3} - I_p(\tilde{f}) \right), \tag{3.67}$$

where we have used expression (3.33) for $\psi(\alpha)$. Moreover, we recall that $\mathbb{P}_{n,p}(X^{(n)})$ is the Erdös measure defined in (3.9), $h_p = \ln \frac{p}{1-p}$ and $\alpha \in \mathbb{R}$. Note that, thanks to Theorem 3.7, when α is such that replica symmetric occurs, i.e when $\alpha > -2$, (3.67) reads

$$\mu(\alpha) = \alpha \frac{u^{*3}}{3} - I_p(u^*), \quad \alpha > -2 \tag{3.68}$$

where u^* is the solution of (3.36).
When the Gärtner Ellis theorem holds, it is possible to switch from the rate function to the scaled cumulant generating function (alias the free energy) simply computing the Legendre transform. What happens when Gärtner Ellis theorem does not hold?
In Subsection 3.1.3 we quoted a large deviation result for the upper tails of triangles of the Erdös-Rényi model (Theorem 3.5), distinguishing between the replica symmetric region and the replica breaking one, according to the domain of the rate function argument. Such result refers to the variational problem (3.13). At the same time the Erdös-Rényi random graph can be seen as a particular case of a wider class of graphs, the exponential ones, where once

again it is possible to read these two regions solving another variational problem, (3.67), a priori different from (3.13). Maintaining the focus on the Erdös-Rényi model and keeping in mind the strict connection between the scaled cumulant generating function and the free energy stressed in (3.66), it comes natural to think that, known the window of parameters for which the replica breaking occurs in the large deviations setting, it is possible to switch to the corresponding replica breaking region for the scaled cumulant generating function of triangles, properly mapping the parameters. Unfortunately this is not always true and the possibility of reading the same information on the rate function and on the free energy does not follow straightforward. In order to give an intuition of this statement, we recover to the argument proposed by Lubetzky and Zhao in [41]. We recall that, as seen in Subsection 3.1.3, the replica breaking phase when we speak in terms of large deviations principle, consists of all points (p, r) such that $(r^2, I_p(r))$ does not lie on the convex minorant of $x \mapsto I_p(\sqrt{x})$ (recall from Lemma 3.1 that the function $I_p(\sqrt[\gamma]{x})$ is not convex if $\gamma > 1$ and $p < p_0(\gamma)$): we called this region $\mathcal{B}_2$. On the other side we know from the geometrical interpretation given in Subsection 3.2.1 that the solution of the scalar variational problem (3.35) jumps over the concave part of $I_p(\sqrt[3]{x})$ as long as α overcomes the critical value $\bar{\alpha}$. By analogy with $\mathcal{B}_2$, we denote by $\mathcal{B}_3$ the set of points (p, r) such that $(r^3, I_p(r))$ does not lie on the convex minorant of $x \mapsto I_p(\sqrt[3]{x})$. It can be proved that $\mathcal{B}_2 \subset \mathcal{B}_3$ (Lemma A.5 of [41]): in other words, the replica breaking phase entirely lies inside the region $\mathcal{B}_3$ which is never visited by the solution of the scalar problem, or equivalently by the solution of the fixed-point equation (3.36) which returns the maximum. In conclusion, when α is positive and we are consequently looking at deviations of upper tails, we can precisely identify a broken symmetry region for the parameter r of the rate function whereas, looking at the sufficient statistic (3.32), there is replica symmetry everywhere and the free energy can be obtained solving the scalar problem (3.35).

Final overview and open problems

In order to conclude, we report a brief summary of Chapter 3 in order to give an overview of the scenario:

a) The graph limit theory allows us to describe and study the behavior of a graph when its size grows to infinity. It makes use of symmetric measurable functions called graphons used to represent the limit of a graph: such functions live in a space equipped with a metric, where it is possible to define a distance and the equivalence relation (3.4).

b) Using the space of graphons it is possible to derive a large deviation principle for the upper tails of triangles of an Erdös-Rényi random graph (dense case) and identify two regions, the symmetric phase and the breaking one: the first is characterized by a rate function which coincides with the Bernoulli relative entropy $I_p(r)$ and this holds in a precise window of the parameter r. For what concerns the replica breaking regime, only a lower bound of the rate function is known.

Turning to lower tails it is still possible to identify such two regions and we refer to [55] for an insightful overview.

c) In a parallel way one can discuss the limiting behavior of the free energy of an Exponential Random Graph: this class of graph includes as a specific case the Erdös-Rényi model. Keeping the focus on this context, the scaled cumulant generating function of the normalized number of triangles, coincides with the free energy of an Exponential Random Graph (with parameters properly tuned and statistic made of triangles and edges) up to an additive constant (see (3.66)).

d) The analytic expression of the free energy (and consequently of the scaled cumulant generating function) is known when the parameter α which tunes the triangle density is positive, slightly negative or when it goes to minus infinity. Furthermore when α lies below a negative constant $-C(h_p)$ which depends on p (recall Theorem 3.12), the solution of the variational problem (3.67) is no more the constant function. This coincides with the unresolved region. At this point it is important to highlight that, speaking in terms of large deviations, when α is negative we are looking at the lower tails of triangles: there is always this parallelism between the rate function and the free energy but, when we are in replica breaking regime, how to switch from the characterization of the former into the one of the second, still rests an open problem.

From the summary above a few open questions arise:

a) Looking at the variational problem (3.67), is it possible to find out an expression of the scaled cumulant generating function when $\alpha \leq -C(h_p)$? What is the structure of the maximizer below such threshold of α?

b) When we look at both the upper and lower tails of triangles, what is the analytic form of the rate function in the replica breaking region?

c) In replica breaking regime, is it possible to translate the properties of the rate function in terms of the free energy and vice-versa?

Chapter 4

Extended version of the Cloning algorithm

4.1 Presentation of the extended method

In this section, we introduce an extended version of the Cloning algorithm, described in Chapter 2 in its original setting. To recap, the standard approach of the method consists in computing the scaled cumulant generating function of an additive observable $F_T = \sum_{n=0}^{T-1} f(X_n, X_{n+1})$ via the formula

$$\mu(\alpha) = \lim_{T\to+\infty} \frac{1}{T} \ln \sum_{(x_0,\dots,x_T)\in S^{T+1}} \nu(x_0)k(x_0)p_\alpha(x_0,x_1)\dots k(x_{T-1})p_\alpha(x_{T-1},x_T),$$

where the sum $\sum_{n=0}^{T-1} f(X_n, X_{n+1})$ is split and hidden into the transition probabilities $p_\alpha(\cdot,\cdot)$ and $(X_n)_{n\geq 0}$ is a finite state Markov chain taking values on a space S. Such formula is implemented through a population dynamics scheme which starts from a bunch of initial conditions with distribution $\nu(\cdot)$ and makes them evolve with probability $p_\alpha(\cdot,\cdot)$ and clone according to a rate $k(\cdot)$. We recall that the scheme is resumed in Algorithm 1. The core idea of this chapter is that a priori the method can be applied to a generic additive observable, in particular to the number of subgraphs of a dense Erdös-Rényi random graph. In this context, this section aims at

a) providing a modified version of the Cloning algorithm based on a dynamics which evolves on Erdös-Rényi, growing-size matrices;

b) showing a simple application of this extended version of the method to the observable *number of edges*.

4.1.1 A dynamics between growing-size matrices

We start introducing a Markovian dynamics between adjacency matrices which grow their size of one unit per step. More precisely, the graph at volume n is thought as obtained by a sequence of steps, each of them corresponding to the addition of a vertex. The new vertex is connected to the other ones independently with probability p. Along this growing-size process, one can keep trace of one selected additive observable which plays the role of F_T, defined in (2.1) for the standard approach. For example, one can choose as observable the number of a fixed subgraph of the evolving graph, such as the number of edges or triangles. We consider a sequence $(X^{(n)})_n$, with $X^{(n)} \in \mathcal{G}_{n,p}$ and transition probabilities $P(\cdot,\cdot)$ on $\mathcal{G}_{n-1,p} \times \mathcal{G}_{n,p}$ and we introduce the set of vectors with whom we fringe the matrix $X^{(n-1)}$:

$$\mathcal{L}_n = \{Y_n = (X_{1,n}, X_{2,n}, \dots, X_{n-1,n}, 0) |\, X_{i,n} \in \{0,1\}, i = 1 \dots, n-1\}. \tag{4.1}$$

If $Y_n \in \mathcal{L}_n$, let $X^{(n)} = X^{(n-1)} \vee Y_n \in \mathcal{G}_{n,p}$ be the matrix fringed with the vector Y_n, i.e a new graph obtained from $X^{(n-1)}$ adding a vertex (n) with new connections expressed by Y_n.
On $\mathcal{L}_n$, we have the probability measure ρ_n defined as:

$$\rho_n(Y_n) = p^{|Y_n|}(1-p)^{(n-1)-|Y_n|} \tag{4.2}$$

$$= (1-p)^{(n-1)} e^{h_p |Y_n|} \tag{4.3}$$

with $h_p = \ln \frac{p}{1-p}$ and

$$|Y_n| = \sum_{i=1}^{n-1} X_{i,n}. \tag{4.4}$$

Then, the transition probability between two graphs is defined as follows:

$$P(X^{(n-1)}, X^{(n)}) = \begin{cases} \rho_n(Y_n), & \text{if exists } Y_n \in \mathcal{L}_n \text{ such that } X^{(n)} = X^{(n-1)} \vee Y_n \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

4.1.2 Compatibility condition for Erdös-Rényi measures

In general, we say that the measures γ_n defined on the space $\mathcal{G}_{n,p}$, $n = 2, 3, \dots$, are *compatible* if

$$\sum_{Y_n \in \mathcal{L}_n} \gamma_n(X^{(n-1)} \vee Y_n) = \gamma_{n-1}(X^{(n-1)}), \quad n = 3, 4 \dots. \tag{4.6}$$

It is possible to show that this condition is satisfied from the Erdös-Rényi measure with parameter p that we recall below:

$$\mathbb{P}_{n,p}(X^{(n)}) = (1-p)^{\binom{n}{2}} e^{h_p E(X^{(n)})}$$

with $h_p = \ln \frac{p}{1-p}$. The first member of (4.6) can be written as follows:

$$\sum_{Y_n \in \mathcal{L}_n} \mathbb{P}_{n,p}(X^{(n-1)} \vee Y_n) = (1-p)^{\binom{n}{2}} \sum_{Y_n \in \mathcal{L}_n} e^{h_p E(X^{(n-1)} \vee Y_n)}. \tag{4.7}$$

Observing that

$$E(X^{(n-1)} \vee Y_n) = E(X^{(n-1)}) + |Y_n|$$

the sum (4.7) can be written as:

$$\sum_{Y_n \in \mathcal{L}_n} e^{h_p E(X^{(n-1)} \vee Y_n)} = \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} e^{h_p E(X^{(n-1)})+\ell} = e^{h_p E(X^{(n-1)})} \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} e^{h_p \ell}, \quad (4.8)$$

since there are $\binom{n-1}{\ell}$ vectors of $\mathcal{L}_n$ of length ℓ. Recalling the definition of h_p and using the Newton binomial, (4.8) becomes

$$\sum_{Y_n \in \mathcal{L}_n} e^{h_p E(X^{(n-1)} \vee Y_n)} = e^{h_p E(X^{(n-1)})} \frac{1}{(1-p)^{n-1}}.$$

Finally, substituting in (4.7) we get

$$\sum_{Y_n \in \mathcal{L}_n} \mathbb{P}_{n,p}(X^{(n-1)} \vee Y_n) = (1-p)^{\binom{n}{2}-(n-1)} e^{h_p E(X^{(n-1)})} = (1-p)^{\binom{n-1}{2}} e^{h_p E(X^{(n-1)})} = \mathbb{P}_{n-1,p}(X^{(n-1)}),$$

which shows, according to definition (4.6), that the Erdös-Rényi probabilities are compatible. Note that the compatibility can be also inferred noting that, thanks to independence, the Erdös-Rényi probability satisfies the condition:

$$\mathbb{P}_{n,p}(X^{(n-1)} \vee Y_n) = p^{|Y_n|}(1-p)^{(n-1)-|Y_n|} \mathbb{P}_{n-1,p}(X^{(n-1)}),$$

where $p^{|Y_n|}(1-p)^{(n-1)-|Y_n|}$ is the connection probability between the vertex n and the graph $X^{(n-1)}$. Recalling definition (4.2) we get:

$$\mathbb{P}_{n,p}(X^{(n-1)} \vee Y_n) = \rho_n(Y_n) \mathbb{P}_{n-1,p}(X^{(n-1)})$$

and this condition implies the compatibility:

$$\sum_{Y_n \in \mathcal{L}_n} \mathbb{P}_{n,p}(X^{(n-1)} \vee Y_n) = \sum_{Y_n \in \mathcal{L}_n} \rho_n(Y_n) \mathbb{P}_{n-1,p}(X^{(n-1)}) = \mathbb{P}_{n-1,p}(X^{(n-1)}),$$

being $\sum_{Y_n \in \mathcal{L}_n} \rho_n(Y_n) = 1$.

Remark 4.1. Note that given a sequence of vectors $(Y_\xi)_{\xi \geq 3}$ with $Y_\xi \in \mathcal{L}_\xi$ and $X^{(2)} \in \mathcal{G}_{2,p}$ it is always possible to recursively build $X^{(n)} = X^{(2)} \vee Y_3 \vee \cdots \vee Y_n$. Moreover, applying (4.2), we get another representation of the Erdös-Rényi probability:

$$\mathbb{P}_{n,p}(X^{(n)}) = \mathbb{P}_{2,p}(X^{(2)}) \prod_{\xi=3}^{n} \rho_\xi(Y_\xi) \quad (4.9)$$

$$= \mathbb{P}_{2,p}(X^{(2)}) P(X^{(2)}, X^{(3)}) \cdots P(X^{(n-1)}, X^{(n)}). \quad (4.10)$$

Vice-versa, given $X^{(n)} \in \mathcal{G}_{n,p}$, one can determine (not in an unique way) a sequence of vectors which generates the graph $X^{(n)}$, $\mathbb{P}_{n,p}(X^{(n)})$ can be computed as in (4.9) and the product is independent on the choice of the sequence.

4.1.3 An easy case: the Cloning algorithm applied to the edges observable

Having defined the additive dynamics, we can now apply the algorithm to the edges observable of a dense Erdös-Rényi random graph, for reproducing the scaled cumulant generating function. First, we recall that the graph at size n can be recursively built in the following way:

$$X^{(n)} = X^{(2)} \vee \{X_{13}, X_{23}\} \vee \{X_{14}, X_{24}, X_{34}\} \vee \cdots \vee \{X_{1n}, X_{2n}, \ldots, X_{n-1,n}\}. \tag{4.11}$$

Note that each term of decomposition (4.11) is a vector $Y_\xi \in \mathcal{L}_\xi$ introduced in (4.1) (with $\xi = 3 \ldots, n$), where the last element is omitted, since it is null. The quantity we are interested in, can be written in an incremental way as reported below:

$$E(X^{(n)}) = E(X^{(2)}) + \sum_{\xi=3}^{n} |Y_\xi|.$$

We recall that the Cloning algorithm implements a discrete time dynamics described by a Markov chain: in the present case the trajectory is given by the decomposition (4.11) and the transition probabilities $P(X^{(\xi-1)}, X^{(\xi)})$ by the Erdös Rényi distribution (then with independence). According to (4.9), (4.3) and decomposition (4.11), we have:

$$\mathbb{P}_{n,p}(X^{(n)}) = e^{h_p X_{12}}(1-p)e^{h_p(X_{13}+X_{23})}(1-p)^2 \cdots e^{h_p(X_{1n}+\cdots+X_{n-1n})}(1-p)^{n-1}. \tag{4.12}$$

It is now important to explicitly recall the scaled cumulant generating function which we want to reproduce:

$$\begin{aligned}\mu(\alpha) &= \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{E}_p^{ER}(e^{\alpha\, E(X^{(n)})}) = \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \sum_{X^{(n)}\in\Omega_n} e^{\alpha\, E(X^{(n)})} \mathbb{P}_{n,p}(X^{(n)}) \\ &= \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \sum_{X^{(n)}\in\Omega_n} e^{\alpha\, E(X^{(n)})} e^{h_p E(X^{(n)})} (1-p)^{\binom{n}{2}},\end{aligned}$$

Using decomposition (4.11) and the Erdös-Rényi distribution written in form (4.12), we get:

$$\begin{aligned}&\mu_n(\alpha) = \\ &\frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} \sum_{X_{13},X_{23}} \sum_{X_{14},X_{24},X_{34}} \cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{\alpha\, X_{12}} e^{\alpha(X_{13}+X_{23})} e^{\alpha(X_{14}+X_{24}+X_{34})} \cdots e^{\alpha(X_{1n}+\cdots+X_{n-1n})} \cdots \\ &\cdots e^{h_p X_{12}}(1-p) e^{h_p(X_{13}+X_{23})}(1-p)^2 e^{h_p(X_{14}+X_{24}+X_{34})}(1-p)^3 \cdots e^{h_p(X_{1n}+\cdots+X_{n-1n})}(1-p)^{n-1}.\end{aligned} \tag{4.13}$$

Note that in (4.13), we can recognize the transition probabilities of the original dynamics, given by the Erdös Rényi measure:

$$P(X^{(\xi-1)}, X^{(\xi)}) = p^{|Y_\xi|}(1-p)^{(\xi-1)-|Y_\xi|} = (1-p)^{(\xi-1)} e^{h_p|Y_\xi|}, \quad \xi \in \{3, \ldots, n\}. \tag{4.14}$$

Rearranging expression (4.13) of $\mu_n(\alpha)$ we get:

$$\mu_n(\alpha) = \frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} e^{h_p X_{12}}(1-p) \sum_{X_{13},X_{23}} e^{\alpha(X_{13}+X_{23})} e^{h_p(X_{13}+X_{23})}(1-p)^2 \sum_{X_{14},X_{24},X_{34}} e^{\alpha(X_{14}+X_{24}+X_{34})} e^{h_p(X_{14}+X_{24}+X_{34})}(1-p)^3 \cdots$$
$$\cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{\alpha(X_{1n}+\cdots+X_{n-1n})} e^{h_p(X_{1n}+\cdots+X_{n-1n})}(1-p)^{n-1} \tag{4.15}$$

and expression (4.15) suggests how to define the tilted measure of the dynamics, i.e:

$$\tilde{P}_\alpha(X^{(\xi-1)}, X^{(\xi)}) := e^{\alpha\,|Y_\xi|} p^{|Y_\xi|}(1-p)^{(\xi-1)-|Y_\xi|}, \quad \xi \in \{3,\ldots,n\}. \tag{4.16}$$

At this point, in order to get the transition probability of the Markov chain, we have to normalize, thus we define:

$$P_\alpha(X^{(\xi-1)}, X^{(\xi)}) := \frac{\tilde{P}_\alpha(X^{(\xi-1)}, X^{(\xi)})}{k(X^{(\xi-1)})}, \tag{4.17}$$

where $k(X^{(\xi-1)})$ is the partition function:

$$k(X^{(\xi-1)}) := \sum_{X_{1,\xi}X_{2,\xi}\ldots X_{\xi-1,\xi}} e^{\alpha(X_{1,\xi}+\cdots+X_{\xi-1,\xi})} e^{h_p(X_{1,\xi}+\cdots+X_{\xi-1,\xi})}(1-p)^{\xi-1} \tag{4.18}$$

$$= \sum_{|Y_\xi|=0}^{\xi-1} \binom{\xi-1}{|Y_\xi|} e^{\alpha|Y_\xi|} p^{|Y_\xi|}(1-p)^{(\xi-1)-|Y_\xi|} = (p\,e^\alpha + 1 - p)^{\xi-1}. \tag{4.19}$$

Finally, using all the quantities defined up to now, namely $\tilde{P}_\alpha(\cdot,\cdot)$, $P_\alpha(\cdot,\cdot)$ and the normalization factor $k(\cdot)$, we can write the moment generating function as follows:

$$\mu_n(\alpha) = \frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} \sum_{X_{13},X_{23}} \sum_{X_{14},X_{24},X_{34}} \cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{h_p X_{12}}(1-p) \frac{\tilde{P}_\alpha(X^{(2)}, X^{(3)})}{k(X^{(2)})} k(X^{(2)}) \frac{\tilde{P}_\alpha(X^{(3)}, X^{(4)})}{k(X^{(3)})} k(X^{(3)}) \cdots$$
$$\cdots \frac{\tilde{P}_\alpha(X^{(n-1)}, X^{(n)})}{k(X^{(n-1)})} k(X^{(n-1)})$$
$$= \frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} \sum_{X_{13},X_{23}} \sum_{X_{14},X_{24},X_{34}} \cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{h_p X_{12}}(1-p) P_\alpha(X^{(2)}, X^{(3)}) k(X^{(2)}) P_\alpha(X^{(3)}, X^{(4)}) k(X^{(3)}) \cdots \tag{4.20}$$
$$\cdots P_\alpha(X^{(n-1)}, X^{(n)}) k(X^{(n-1)}).$$

Note that the expression of the rates (4.19) can be put into (4.20) yielding the product $\prod_{k=2}^{n-1} k(X^{(k)}) = (p\,e^\alpha + 1 - p)^{\frac{(n-2)(n-1)}{2}}$ which can be brought out from the sum, since it does not depend on the adjacency matrices. Moreover, $\ln\left(\prod_{k=2}^{n} k(X^{(k)})\right) = \frac{(n-2)(n-1)}{2} \ln(p\,e^\alpha + 1 - p)$

and, in the limit, this is the only part which does not vanishes, thanks to the term $\frac{(n-2)(n-1)}{2}$ which has the same speed of $\binom{n}{2}$. In conclusion we get

$$\mu(\alpha) = \ln(p\, e^{\alpha} + 1 - p),$$

which is the function the method should reproduce.
In a parallel way as we have done in Subsection 2.1.2, we describe how the scheme performs the numerical approximation of the objective function, mapping the standard approach into this new one. The observable over the time F_T, defined in (2.1), is replaced by the additive observable which takes into account the increment of the selected subgraph, in this specific case the number of edges. For the following description, let us denote by $X^{(\xi-1)}$ a generic matrix of size $\xi - 1$ involved in the algorithm. The starting population consists in a family of adjacency matrices of size two which grow their size by one vertex at each step. This makes part of the evolution phase, in which two events happen: each matrix evolves in a wider space (and this concretely corresponds to fringing the starting matrix $X^{(\xi-1)}$ with a vector $Y_\xi \in \mathcal{L}_\xi$) and the new vertex is connected (or not) to the others according to probability (4.17). After the evolution step, a cloning phase occurs: each adjacency matrix leaves a number of offspring expressed by its rate $k(X^{(\xi-1)})$ (properly turned into integer according to the rule of the standard method). Finally, the normalization step rests the same: we uniformly pick a number of matrices equal to the starting population and the cycle restarts. Summarizing:

a) A family of M_2 adjacency matrices of size two, drawn from the Bernoulli distribution of parameter p constitutes the starting population of the method. Such matrices represent the so-called clones.

b) Each clone evolves in a wider space and the new vertex is connected (or not) to the others according to probability $P_\alpha(\cdot,\cdot)$ given in (4.17), but remains to be eventually connected to future vertices.

c) Each clone in configuration $X^{(\xi-1)}$ is then replicated and gives birth to a number of offspring equal to its rate

$$k(X^{(\xi-1)}) = \begin{cases} \lfloor k(X^{(\xi-1)}) \rfloor + 1 & \text{with probability} \quad k(X^{(\xi-1)}) - \lfloor k(X^{(\xi-1)}) \rfloor \\ \lfloor k(X^{(\xi-1)}) \rfloor & \text{otherwise} \end{cases}$$

where $\lfloor k(X^{(\xi-1)}) \rfloor$ represents the integer part of $k(X^{(\xi-1)})$. In case that $k(X^{(\xi-1)}) = 0$ the clone is killed and it does not leave offspring.

d) Once that the clones have evolved and reproduced, the total number of copies is brought back to M_2, uniformly choosing M_2 clones among those present after the evolution and reproduction step. Recall that M_2 represents the size of the family, not the size of the

graphs involved, which is equal to two only at the first step. At each step $\xi - 1$, after the cloning phase, it is convenient to keep trace of the rescaling factor $S_{\xi-1} := \frac{M_{\xi-1}}{M_{\xi-2}}$, so that

$$\mu_n(\alpha) = \frac{1}{\binom{n}{2}} \ln \frac{M_n}{M_2} = \frac{1}{\binom{n}{2}} \ln \left(\frac{M_n}{M_{n-1}} \frac{M_{n-1}}{M_{n-2}} \dots \frac{M_4}{M_3} \frac{M_3}{M_2} \right) = \frac{1}{\binom{n}{2}} \ln \left(S_n \cdot S_{n-1} \dots S_3 \right)$$
$$= \frac{1}{\binom{n}{2}} \sum_{\xi=3}^{n} \ln(S_\xi).$$

The algorithm proceeds by repeating iteratively the three steps above: $\mu(\alpha)$ is estimated by computing

$$\mu_n(\alpha) = \frac{1}{\binom{n}{2}} \sum_{\xi=3}^{n} \ln(S_\xi).$$

The whole scheme is resumed in Algorithm 2 for a fixed $\alpha \in \mathbb{R}$ and $p \in (0, 1)$.

Algorithm 2 Extended Cloning algorithm

Consider a family of M_2 adjacency matrices of size two drawn from a Bernoulli distribution of parameter p. Choose some $n \in \mathbb{N}$.
For $\xi = \{3 \dots, n\}$ compute $\mu_n(\alpha)$ as follows:

1. Evolve each matrix of the population according to the transition probability $P_\alpha(X^{(\xi-1)}, X^{(\xi)})$;

2. Reproduce each matrix according to the average rate $k(X^{(\xi-1)})$ (if the rate is null, the clone is killed, otherwise, it leaves offspring), where

$$k(X^{(\xi-1)}) = \begin{cases} \lfloor k(X^{(\xi-1)}) \rfloor + 1 & \text{with probability} \quad k(X^{(\xi-1)}) - \lfloor k(X^{(\xi-1)}) \rfloor \\ \lfloor k(X^{(\xi-1)}) \rfloor & \text{with probability} \quad 1 - \left(k(X^{(\xi-1)}) - \lfloor k(X^{(\xi-1)}) \right) \rfloor \end{cases}.$$

3. Extract a number of individuals equal to the starting one, M_2, with uniform probability from those survived after the phase 2;

4. Compute $S_\xi = \frac{M_\xi}{M_{\xi-1}}$ so that $S_n \cdots S_1 = \frac{M_n}{M_2}$.

End
At the end of the cycle, approximate $\mu(\alpha)$ with

$$\mu_n(\alpha) = \frac{1}{\binom{n}{2}} \sum_{\xi=3}^{n} \ln(S_\xi).$$

Remark 4.2. Note that there are two parameters which determine the accuracy of the method: the number of iterations, which, in this case, coincides with the volume of the graphs involved

in the process and the initial size of the population M_2, which represents a sample of the whole space of states. Clearly, the more the number of iterations grows, the more the limiting behavior of the graph is accurate; at the same time, the bigger the initial population is, the better the average involved in the definition of $\mu(\alpha)$ is approximated. At the best of my knowledge, no study on how the choice of the parameters reflects on the error of the algorithm is present in the state-of-art.

Remark 4.3. The description of the implementation scheme leads to a practical remark: we can notice that, for a fixed $\xi \in \{3, \dots, n\}$, the sum involved in the reproduction rate (4.18) runs over $\xi - 1$ possible elements which correspond to the components of the fringing vector. To be precise, the generic reproduction rate $k(X^{(\xi-1)})$, in this case does not depend on the whole adjacency matrix but only on the new possible connections. As a further simplification, for the edge observable, the rate can be written in the closed form (4.19). For this reason, it is unnecessary to carry on through simulations a population of matrices: the only information we need is to know how many connections are possible at each step and to assign a probability to each case. In other words all simulations can be run replacing the adjacency matrices with integers (namely we perform a Markov process on $\mathbb{N}$) and this considerably simplifies the implementation.

At this point, all the quantities we need for running the algorithm are defined and the returned output is shown in Figure 4.1: the curve reproduced by the method fits the expected one with a very good accuracy. When the structure of the observable gets more difficult and we lose the

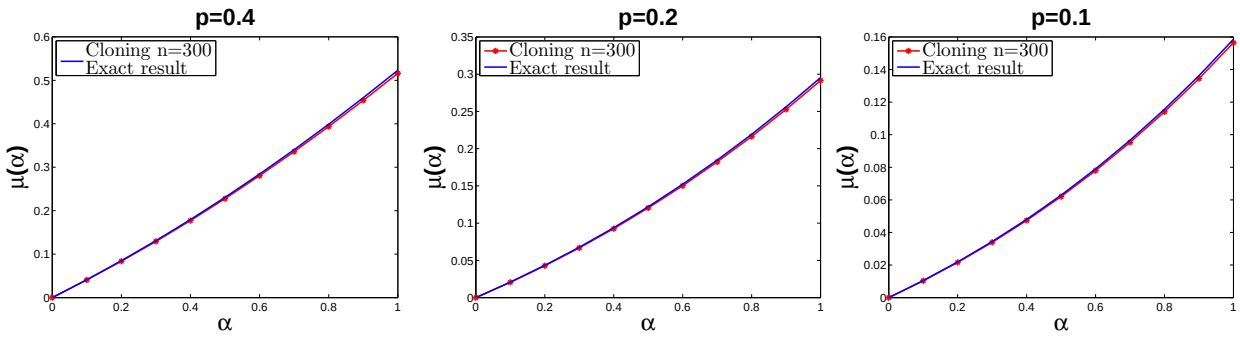


Figure 4.1: Numerical approximation of the scaled cumulant generatig function for the edge observable: pictures show different values of p. The dotted line represents the algorithm output, whereas the blue continuous one, the expected curve. For a final size of the graph $n = 300$ we get a very good approximation of the objective function.

independence of its additive components, the computation of the scaled cumulant generating function is not easy to perform. It is the case of the triangle observable, which despite its apparent simplicity, gives rise to different and complex scenarios in such estimation.

4.2 Investigation of the replica symmetric phase

4.2.1 The Cloning Algorithm applied to the triangle observable

The main purpose of this section is to apply the extended version of the Cloning algorithm to the dense Erdös-Rényi model for the approximation of the scaled cumulant generating function of the triangle observable (properly normalized). As we have seen in Chapter 3, such function is strictly related to the free energy of an Exponential Random Graph with a statistic including only triangles and edges. According to the choice of the parameter which tunes the triangles density, we explore the so-called *replica symmetric* phase or the *replica breaking* one. In this section, we implement the extended Cloning method introduced in Subsection 4.1.1 working in replica symmetric regime and we provide a heuristic argument concerning its convergence to the analytic result.
We start addressing the problem

Problem 4.1. Solve

$$\mu(\alpha) = \sup_{f \in \mathcal{W}} \left[\frac{\alpha\, t(H_2, f)}{3} - I_p(f) \right], \quad \alpha > -2 \tag{4.21}$$

where $f(x, y) : [0, 1]^2 \to [0, 1]$ is measurable and symmetric.

Thanks to Theorems 3.8 and 3.11, Problem 4.1 turns into the scalar one

$$\mu(\alpha) = \sup_{0 \leq u \leq 1} \left[\alpha \frac{u^3}{3} - I_p(u) \right], \quad \alpha > -2 \tag{4.22}$$

returning as solution

$$\mu(\alpha) = \alpha \frac{u^{*3}}{3} - I_p(u^*),$$

where u^* solves the fixed-point equation (3.36). As we mentioned at the beginning of the chapter, despite the extended version of the method can be applied to a generic additive observable, we want to focus our attention on triangles, for reproducing the related scaled cumulant generating function in the replica symmetric region. For a graph of size n, the number of triangles has the form $T(X^{(n)}) = \frac{1}{6} \sum_{i,j,k=1}^{n} X_{ij} X_{jk} X_{ki}$: note that the triplets involved in this last sum are not independent since they can potentially share common factors. We recall that it is possible to see the graph at size n as the union of random vectors according to decomposition (4.11). With the same spirit used for edges, we can write the number of triangles of a generic graph $X^{(\xi)}$, leaning on the information of the previous step:

$$T(X^{(\xi)}) = T(X^{(\xi-1)}) + \theta_{\xi-1,\xi}, \tag{4.23}$$

where the term $\theta_{\xi-1,\xi}$ represents the partial increment of triangles in the passage from the graph $X^{(\xi-1)}$ to the graph $X^{(\xi)}$. It has the following expression:

$$\theta_{\xi-1,\xi} = \frac{1}{2}\sum_{i,j=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}, \tag{4.24}$$

where the factor $X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}$ can be interpreted as a wedge with base in i and j and common vertex in ξ. The derivation of expression (4.24) is reached through the computations below:

$$\begin{aligned}
T(X^{(\xi)}) &= \frac{1}{6}\sum_{i,j,l=1}^{\xi} X_{ij}^{(\xi)} X_{jl}^{(\xi)} X_{li}^{(\xi)} = \frac{1}{6}\sum_{i=1}^{\xi}\sum_{j=1}^{\xi}\left[\sum_{l=1}^{\xi-1} X_{ij}^{(\xi)} X_{jl}^{(\xi)} X_{li}^{(\xi)} + X_{ij}^{(\xi)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}\right] \\
&= \frac{1}{6}\sum_{i=1}^{\xi}\left[\sum_{j=1}^{\xi}\sum_{l=1}^{\xi-1} X_{ij}^{(\xi)} X_{jl}^{(\xi)} X_{li}^{(\xi)} + \sum_{j=1}^{\xi} X_{ij}^{(\xi)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}\right] \\
&= \frac{1}{6}\sum_{i=1}^{\xi}\left[\sum_{j=1}^{\xi-1}\sum_{l=1}^{\xi-1} X_{ij}^{(\xi)} X_{jl}^{(\xi-1)} X_{li}^{(\xi)} + \sum_{l=1}^{\xi-1} X_{i\xi}^{(\xi)} X_{\xi l}^{(\xi)} X_{li}^{(\xi)} + \sum_{j=1}^{\xi-1} X_{ij}^{(\xi)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)} + X_{i\xi}^{(\xi)} X_{\xi\xi}^{(\xi)} X_{\xi i}^{(\xi)}\right] \\
&= \frac{1}{6}\Bigg\{\sum_{i=1}^{\xi-1}\sum_{j=1}^{\xi-1}\sum_{l=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{jl}^{(\xi-1)} X_{li}^{(\xi-1)} + \sum_{j=1}^{\xi-1}\sum_{l=1}^{\xi-1} X_{\xi j}^{(\xi)} X_{jl}^{(\xi-1)} X_{l\xi}^{(\xi)} + \sum_{i=1}^{\xi-1}\sum_{l=1}^{\xi-1} X_{i\xi}^{(\xi)} X_{\xi l}^{(\xi)} X_{li}^{(\xi-1)} \\
&+ \sum_{l=1}^{\xi-1} X_{\xi\xi}^{(\xi)} X_{\xi l}^{(\xi)} X_{l\xi}^{(\xi)} + \sum_{i=1}^{\xi-1}\sum_{j=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)} + \sum_{j=1}^{\xi-1} X_{\xi j}^{(\xi)} X_{j\xi}^{(\xi)} X_{\xi\xi}^{(\xi)} + \sum_{i=1}^{\xi-1} X_{i\xi}^{(\xi)} X_{\xi\xi}^{(\xi)} X_{\xi i}^{(\xi)} \\
&+ X_{\xi\xi}^{(\xi)} X_{\xi\xi}^{(\xi)} X_{\xi\xi}^{(\xi)}\Bigg\}.
\end{aligned}$$

In conclusion, recalling that the elements on the diagonal are null (since self loops are not allowed), the expression displayed above reduces to:

$$\begin{aligned}
T(X^{(\xi)}) &= \frac{1}{6}\sum_{i,j,l=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{jl}^{(\xi-1)} X_{li}^{(\xi-1)} + \frac{1}{6}\sum_{i,l=1}^{\xi-1} X_{i\xi}^{(\xi)} X_{\xi l}^{(\xi)} X_{li}^{(\xi-1)} \\
&+ \frac{1}{6}\sum_{j,l=1}^{\xi-1} X_{\xi j}^{(\xi)} X_{jl}^{(\xi-1)} X_{l\xi}^{(\xi)} + \frac{1}{6}\sum_{i,j=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)} \\
&= T(X^{(\xi-1)}) + \frac{1}{2}\sum_{i,j=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}.
\end{aligned}$$

Remark 4.4. For simplifying the notation, in the following the triplet $X_{ij}X_{j\xi}X_{\xi i}$ will replace $X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}$.

Thanks to the incremental formulation, the total number of triangles up to size n can be expressed as

$$T(X^{(n)}) = T(X^{(2)}) + \sum_{\xi=2}^{n-1} \theta_{\xi,\xi+1} = \frac{1}{2} \sum_{\xi=2}^{n-1} \sum_{i,j=1}^{\xi} X_{ij} X_{j,\xi+1} X_{i,\xi+1} \tag{4.25}$$

(observe that $T(X^{(2)}) = 0$), whereas the normalized number of triangles can be expressed in two ways.

a) The first way derives from (4.25):

$$\frac{T(X^{(n)})}{n} = \frac{1}{n} \sum_{\xi=2}^{n-1} \theta_{\xi,\xi+1} = \frac{1}{2n} \sum_{\xi=2}^{n-1} \sum_{i,j=1}^{\xi} X_{ij} X_{j,\xi+1} X_{i,\xi+1}$$

b) The second way leans on a telescopic sum and relation (4.23):

$$\begin{aligned} \frac{T(X^{(n)})}{n} &= \sum_{\xi=2}^{n-1} \frac{T(X^{(\xi+1)})}{\xi+1} - \frac{T(X^{(\xi)})}{\xi} \qquad = \sum_{\xi=2}^{n-1} \frac{T(X^{(\xi)})}{\xi+1} + \frac{\theta_{\xi,\xi+1}}{\xi+1} - \frac{T(X^{(\xi)})}{\xi} \\ &= \sum_{\xi=2}^{n-1} \frac{\theta_{\xi,\xi+1}}{\xi+1} - \frac{T(X^{(\xi)})}{\xi(\xi+1)}. \end{aligned} \tag{4.26}$$

As we have already done for the edges, it is important to recall the objective function we want to reproduce:

$$\mu(\alpha) = \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \mathbb{E}_p^{ER}(e^{\alpha \frac{T(X^{(n)})}{n}}) = \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \sum_{X^{(n)}\in\Omega_n} e^{\alpha \frac{T(X^{(n)})}{n}} \mathbb{P}_{n,p}(X^{(n)}) \tag{4.27}$$

$$= \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \sum_{X^{(n)}\in\Omega_n} e^{\alpha \frac{T(X^{(n)})}{n}} e^{h_p E(X^{(n)})} (1-p)^{\binom{n}{2}}. \tag{4.28}$$

Using expression (4.26) for the normalized number of triangles and (4.12) for the Erdös-Rényi measure, we can develop (4.27) as follows:

$$\begin{aligned} &\mu_n(\alpha) = \\ &\frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} \sum_{X_{13},X_{23}} \sum_{X_{14},X_{24},X_{34}} \cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{\alpha\left(\frac{\theta_{2,3}}{3} - \frac{T(X^{(2)})}{2\cdot 3}\right)} e^{\alpha\left(\frac{\theta_{3,4}}{4} - \frac{T(X^{(3)})}{3\cdot 4}\right)} \cdots e^{\alpha\left(\frac{\theta_{n-1,n}}{n} - \frac{T(X^{(n-1)})}{n\,(n-1)}\right)} \\ &\cdots e^{h_p X_{12}} (1-p) e^{h_p (X_{13}+X_{23})} (1-p)^2 e^{h_p (X_{14}+X_{24}+X_{34})} (1-p)^3 \cdots e^{h_p (X_{1n}+\cdots+X_{n-1n})} (1-p)^{n-1}. \end{aligned} \tag{4.29}$$

Note that, being $h_p = \ln \frac{p}{1-p}$, we can read in (4.29) the transition probabilities of the original dynamics, described by the Erdös Rényi measure (recall (4.3) and (4.5)):

$$P(X^{(\xi-1)}, X^{(\xi)}) = e^{h_p |Y_\xi|} (1-p)^{(\xi-1)}, \quad \xi \in \{3, \ldots, n\}.$$

Rearranging expression (4.29) of $\mu_n(\alpha)$ we get:

$$\mu_n(\alpha) =$$
$$\frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} e^{h_p X_{12}}(1-p) \sum_{X_{13},X_{23}} e^{\alpha\left(\frac{\theta_{2,3}}{3}-\frac{T(X^{(2)})}{2\cdot 3}\right)} e^{h_p(X_{13}+X_{23})}(1-p)^2 \sum_{X_{14},X_{24},X_{34}} e^{\alpha\left(\frac{\theta_{3,4}}{4}-\frac{T(X^{(3)})}{3\cdot 4}\right)} e^{h_p(X_{14}+X_{24}+X_{34})}(1-p)^3 \cdots$$
$$\cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{\alpha\left(\frac{\theta_{n-1,n}}{n}-\frac{T(X^{(n-1)})}{n\,(n-1)}\right)} e^{h_p(X_{1n}+\cdots+X_{n-1n})}(1-p)^{n-1} \tag{4.30}$$

and we arrive to define the tilted measure of the dynamics:

$$\tilde{P}_\alpha(X^{(\xi-1)}, X^{(\xi)}) := e^{\alpha\left(\frac{\theta_{\xi-1,\xi}}{\xi}-\frac{T(X^{(\xi-1)})}{\xi(\xi-1)}\right)} e^{h_p|Y_\xi|}(1-p)^{(\xi-1)}, \quad \xi \in \{3,\ldots,n\}.$$

Finally, we normalize the tilted measure thus getting the new transition probabilities of the Markov chain:

$$P_\alpha(X^{(\xi-1)}, X^{(\xi)}) := \frac{\tilde{P}_\alpha(X^{(\xi-1)}, X^{(\xi)})}{k(X^{(\xi-1)})},$$

where $k(X^{(\xi-1)})$ is the partition function:

$$k(X^{(\xi-1)}) := \sum_{X_{1,\xi}X_{2,\xi}\ldots X_{\xi-1,\xi}} e^{\alpha\left(\frac{\theta_{\xi-1,\xi}}{\xi}-\frac{T(X^{(\xi-1)})}{\xi(\xi-1)}\right)} e^{h_p(X_{1,\xi}+\cdots+X_{\xi-1,\xi})}(1-p)^{\xi-1}. \tag{4.31}$$

In conclusion the scaled cumulant generating function can be written as follows:

$$\mu_n(\alpha) =$$
$$= \frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} \sum_{X_{13},X_{23}} \sum_{X_{14},X_{24},X_{34}} \cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{h_p X_{12}}(1-p) \frac{\tilde{P}_\alpha(X^{(2)}, X^{(3)})}{k(X^{(2)})} k(X^{(2)}) \frac{\tilde{P}_\alpha(X^{(3)}, X^{(4)})}{k(X^{(3)})} k(X^{(3)}) \cdots$$
$$\cdots \frac{\tilde{P}_\alpha(X^{(n-1)}, X^{(n)})}{k(X^{(n-1)})} k(X^{(n-1)})$$
$$= \frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} \sum_{X_{13},X_{23}} \sum_{X_{14},X_{24},X_{34}} \cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{h_p X_{12}}(1-p) P_\alpha(X^{(2)}, X^{(3)}) k(X^{(2)}) P_\alpha(X^{(3)}, X^{(4)}) k(X^{(3)}) \cdots$$
$$\cdots P_\alpha(X^{(n-1)}, X^{(n)}) k(X^{(n-1)}).$$

Remark 4.5. Observe that the product $\prod_{\xi=2}^{n-1} k(X^{(\xi)})$ returns, up to the constant $(1-p)^{\binom{n}{2}}$, the partition function of the Exponential Random Graph with statistic including edges and triangles and parameters α and h_p: $Z_n = \sum_{X^{(n)}\in\Omega_n} e^{\alpha\frac{T(X^{(n)})}{n}+h_p E(X^{(n)})}$.

4.2.2 A turning point for the implementation

Looking at the reproduction rate (4.31), one can immediately observe that, for a selected $\xi \in \{3,\ldots,n\}$, the sum involved in the computation of the partition function, runs over $2^{\xi-1}$

possible vectors and hence attained configurations in the passage from $X^{(\xi-1)}$ to $X^{(\xi)}$. Note that $2^{\xi-1}$ coincides with the number of dispositions (admitting repetitions), of 0 and 1 over $\xi-1$ sites. As a consequence, the computation of the partition function, i.e the cloning rate, is numerically unaffordable from the numerical point of view when the global size of the graph is pushed too further. Furthermore, note that the partial increment of triangles from size $\xi-1$ to size ξ, namely $\frac{1}{2}\sum_{i,j=1}^{\xi-1} X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{\xi i}^{(\xi)}$, strongly depends on the whole structure of the adjacency matrix up to the step $\xi-1$ and, as a consequence, we need to record in the code all adjacency matrices involved. Since we can not transfer on the integers the information we are interested in (as we have done for the edge observable, see Remark 4.3), the numerical computation of the reproduction rate is hard to perform and simulations are allowed only for small sizes of graphs. A remarkable improvement in the computational complexity would be done by finding the way for counting the adjacency matrices which lead to the same increment, thus performing the sum (4.31) in a most efficient way. Even if this intuition looks promising, the combinatorics underlying this goal is very complex. We provide an example below.

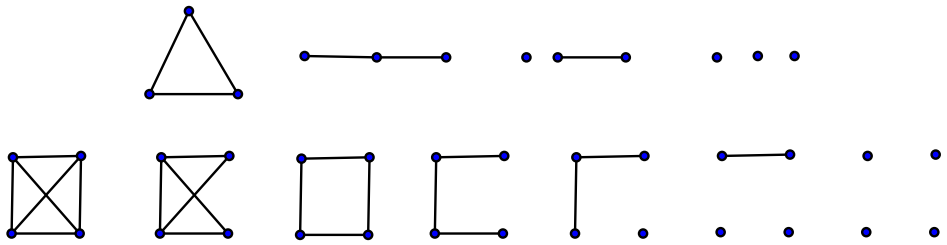


Figure 4.2: Upper row: structures of the subgraphs that we have to count in $X^{(\xi-1)}$ (from left to right) in order to know how many times, adding three edges, we get an increment of respectively 3,2,1,0 triangles. The same information is provided by the picture in the lower row, for the case in which four possible connections are added.

Example 4.1. Assume that, in the passage from the matrix $X^{(\xi-1)}$ to $X^{(\xi)}$, the vector Y_ξ used for fringing $X^{(\xi-1)}$ and containing all possible connections, has three edges in positions $k_1, k_2, k_3 \in \{1,\dots,\xi-1\}$ with $k_1 \neq k_2 \neq k_3 \neq k_1$ (recall that the last element of Y_ξ is null). In other words, make the hypothesis that the elements of the vector $Y_\xi = (X_{1,\xi}, X_{2,\xi}, \dots, X_{\xi-1,\xi}, 0)$ are such that

$$X_{i,\xi}^{(\xi)} = \begin{cases} 1 & \text{if} \quad i = k_1, k_2, k_3 \\ 0 & \text{otherwise} \end{cases}, \quad i \in \{1,\dots,\xi-1\}.$$

As a consequence, the increment of triangles $\theta_{\xi-1,\xi} = \sum_{i,j>i}^{\xi-1} X_{ij}^{(\xi-1)} X_{j\xi}^{(\xi)} X_{i\xi}^{(\xi)}$ can be written as follows:

$$\theta_{\xi-1,\xi} = \sum_{i,j>i}^{\xi-1} X_{ij}^{(\xi-1)} \mathbb{1}_{\{(i,j)|X_{j\xi}^{(\xi)} X_{i\xi}^{(\xi)}=1\}} = X_{k1,k2}^{(\xi-1)} + X_{k1,k3}^{(\xi-1)} + X_{k2,k3}^{(\xi-1)}. \tag{4.32}$$

From (4.32) it is evident that $\theta_{\xi-1,\xi} \in \{0,1,2,3\}$ and the precise value depends on the structure of the matrix $X^{(\xi-1)}$.
Repeating the reasoning for a vector Y_ξ with four edges indexed by different integers $k_1, k_2, k_3, k_4 \in \{1,\dots,\xi-1\}$, we get the increment:

$$\theta_{\xi-1,\xi} = X^{(\xi-1)}_{k1,k2} + X^{(\xi-1)}_{k1,k3} + X^{(\xi-1)}_{k2,k3} + X^{(\xi-1)}_{k1,k4} + X^{(\xi-1)}_{k2,k4} + X^{(\xi-1)}_{k3,k4}$$

which belongs to the set $\{0,\dots,6\}$.
In general, we can say that, if $m \in \{1,\dots,\xi-1\}$ denotes the number of edges present in the vector Y_ξ, the increment $\theta_{\xi-1,\xi}$ belongs to the set $\{0,\dots,\binom{m}{2}\}$. Note that the maximum number of triangles which can be created is reached when $m = \xi - 1$, namely the fringing vector Y_ξ is the unitary one, a part from the last null element: moreover, note that $\binom{\xi-1}{2}$ is the maximum number of edges possible for the starting graph $X^{(\xi-1)}$.
From the geometrical point of view, it turns out that if we want to know how many configurations (among the $2^{\xi-1}$ possible ones) bring the same increment, θ, of triangles, we have to count the number of subgraphs of $X^{(\xi-1)}$ with m vertices and θ edges. Such structures are represented in Figure 4.2 for $m = 3$ (upper row) and $m = 4$ (lower row).

In conclusion, the starting goal of simplifying the computation of the rate function (4.19) by grouping the configurations bringing the same information, is hard to reach, since the problem, even in its geometrical form, still has an high complexity.

4.2.3 Numerical results

We now move on numerical results. We recall from Subsection 3.2.3 that the scaled cumulant generating function of the triangle observable in the Erdös-Rényi dense model, coincides, up to an additive constant, with the free energy of an Exponential Random Graph with statistic including triangles and edges. When we are in replica symmetric regime

$$\mu(\alpha) = \alpha\frac{u^{*3}}{3} - I_p(u^*), \quad \alpha > -2^1, \tag{4.33}$$

where u^* optimizes the scalar problem (4.22) and $I_p(\cdot)$ is the Bernoulli relative entropy. Figures 4.3 and 4.4 show, for different values of p, that the method converges to the expected curve, despite the small size of graphs (note however the scale n^2 in (4.28)). Moreover for smaller sizes, such as $n = 3, 4$ a direct computation of the moment generating function is possible

[1] To be precise, replica symmetric regime extends up to a negative value of α which could be lower than -2, as mentioned by Theorem 3.12. However, in this part we restrict to this region, where replica symmetric occurs for sure.

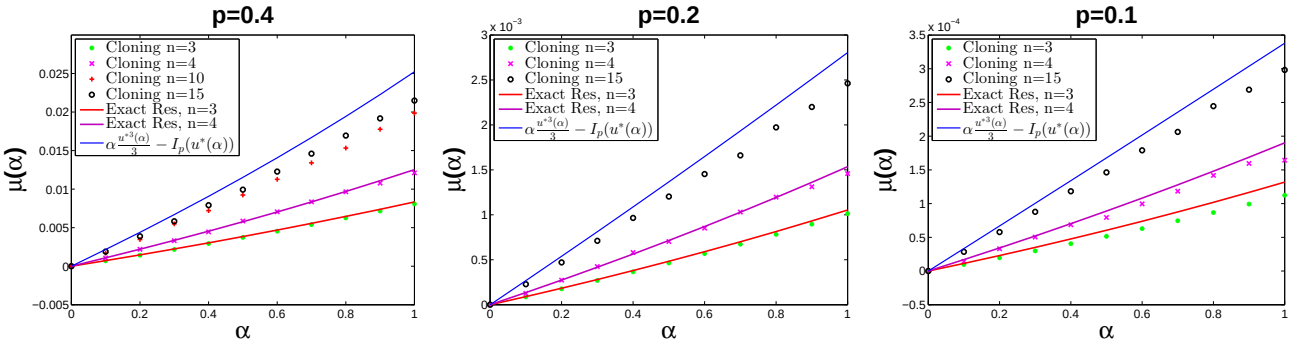


Figure 4.3: Numerical approximation of the scaled cumulant generating function of the triangle observable in the interval $[0, 1]$ (α varies with step 0.1): the pictures show different values of p. The dotted line represents the output of the method for different sizes n of the graphs involved, whereas the continuous blue line represents the analytic result. As the size grows, the dotted line approaches the limiting curve. For $n = 3, 4$ an exact computation of the moment generating function is possible and we use it as a comparison term (such exact functions are represented in the picture through the red and pink continuous lines). The starting size of the family used for reproducing the pictures is $M_2 = 1000$.

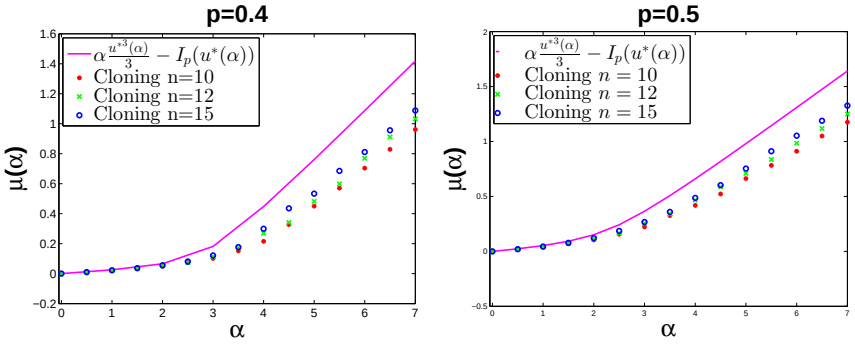


Figure 4.4: Numerical approximation of the scaled cumulant generating function of the triangle observable for $p = 0.4$ and $p = 0.5$ in the interval $\alpha \in [0, 7]$ (α varies with step 0.5). The dotted line represents the output of the method for different sizes n of the graphs involved, whereas the continuous pink line represents the analytic result. As the size grows, the dotted line approaches the limiting curve. The starting size of the family used for reproducing the pictures is $M_2 = 1000$.

($\mu_3(\alpha)$ and $\mu_4(\alpha)$ are drawn in Figure 4.3). For example

$$\begin{aligned}\mu_3(\alpha) &= \frac{1}{\binom{3}{2}} \ln \sum_{X_{12}} e^{h_p X_{12}}(1-p) \sum_{X_{13}X_{23}} e^{\frac{\alpha}{3}(X_{12}X_{13}X_{23})} e^{h_p(X_{12}+X_{23})}(1-p)^2 \\ &= \frac{1}{\binom{3}{2}} \ln \Big(p(1-p^2+p^2 e^{\frac{\alpha}{3}}) + 1 - p \Big)\end{aligned}$$

and, with a more complex scenario, an analogous computation can be performed for deriving $\mu_4(\alpha)$. A part from the numerical evidence shown by Figures 4.3 and 4.4, it is possible to provide a heuristic argument of the convergence based on a mean field approximation which turns out to be exact in the graph large size limit. This strategy has already been explored in [45] and in [17], in order to prove that the quantities $L_{ij} := \sum_{k\neq i,j} \frac{X_{ik}X_{jk}}{n}$, namely the normalized number of wedges, satisfy a set of mean field equations.

4.2.4 A mean field approximation with a cubic interaction

In this paragraph, we argue the convergence of the extended version of the Cloning algorithm (resumed in Algorithm 2) to the scaled cumulant generating function (4.33) of the normalized number of triangles. We stress that the conclusion holds in replica symmetric regime.
As a first step, in order to compute $\lim_{n\to+\infty} \mu_n(\alpha)$, we can write more explicitly the expression of $\mu_n(\alpha)$ reported in (4.30), thus getting:

$$\begin{aligned}&\mu_n(\alpha) = \\ &\frac{1}{\binom{n}{2}} \ln \sum_{X_{12}} e^{h_p X_{12}}(1-p) \sum_{X_{13},X_{23}} e^{\alpha\left(\frac{\sum_{i,j=1}^2 X_{ij}X_{j3}X_{i3}}{2\cdot 3} - \frac{1}{6}\frac{\sum_{i,j,k=1}^2 X_{ij}X_{jk}X_{ki}}{2\cdot 3}\right)} e^{h_p(X_{13}+X_{23})}(1-p)^2 \cdot \\ &\cdot \sum_{X_{14},X_{24},X_{34}} e^{\alpha\left(\frac{\sum_{i,j=1}^3 X_{ij}X_{j4}X_{i4}}{2\cdot 4} - \frac{1}{6}\frac{\sum_{i,j,k=1}^3 X_{ij}X_{jk}X_{ki}}{3\cdot 4}\right)} e^{h_p(X_{14}+X_{24}+X_{34})}(1-p)^3 \cdots \\ &\cdots \sum_{X_{1n},X_{2n},\ldots,X_{n-1n}} e^{\alpha\left(\frac{\sum_{i,j=1}^{n-1} X_{ij}X_{jn}X_{in}}{2\cdot n} - \frac{1}{6}\frac{\sum_{i,j,k=1}^{n-1} X_{ij}X_{jk}X_{ki}}{n\cdot(n-1)}\right)} e^{h_p(X_{1n}+\cdots+X_{n-1n})}(1-p)^{n-1} \qquad (4.34)\end{aligned}$$

and the derivation so far is exact. At this point we recall Theorem 3.9 which basically states that, when the size of the Exponential Random Graph grows to infinity, the distribution of any finite subgraph is the Erdös-Rényi one, namely the connections become asymptotically independent. This consideration should support the following substitution, which consists in a mean-field approximation:

$$X_{ij} \longmapsto \frac{1}{\xi-1} \sum_{m_1=1}^{\xi-1} X_{m_1\xi} \qquad (4.35)$$

Observe that, when n is large enough, the two variables X_{ij} and $\frac{1}{\xi-1}\sum_{m_1=1}^{\xi-1} X_{m_1\xi}$ have the same expectation and $Var(\frac{1}{\xi-1}\sum_{m_1=1}^{\xi-1} X_{m_1\xi}) = \frac{1}{\xi-1} X_{1\xi} \to 0$ when $\xi \to +\infty$. The advantage

of using the mean field approximation (4.35) relies on the possibility to write the exponents in (4.34) in a simplified way, as we show below:

$$\begin{aligned}
&\frac{\sum_{i,j=1}^{\xi-1} X_{ij}X_{j\xi}X_{i\xi}}{2\xi} - \frac{1}{6}\frac{\sum_{i,j,k=1}^{\xi-1} X_{ij}X_{jk}X_{ki}}{\xi(\xi-1)} = \\
&\frac{1}{2\xi(\xi-1)}\sum_{i,j,m_1=1}^{\xi-1} X_{m_1\xi}X_{j\xi}X_{i\xi} - \frac{1}{6\xi(\xi-1)}\sum_{i,j,k=1}^{\xi-1}\left[\sum_{m_1=1}^{\xi-1}\frac{X_{m_1\xi}}{\xi-1}\sum_{m_2=1}^{\xi-1}\frac{X_{m_2\xi}}{\xi-1}\sum_{m_3=1}^{\xi-1}\frac{X_{m_3\xi}}{\xi-1}\right] = \\
&\frac{1}{2\xi(\xi-1)}\sum_{i,j,m_1=1}^{\xi-1} X_{m_1\xi}X_{j\xi}X_{i\xi} - \frac{1}{6\xi(\xi-1)}\frac{1}{(\xi-1)^3}\left[\sum_{m_1=1}^{\xi-1} X_{m_1\xi}\sum_{m_2=1}^{\xi-1} X_{m_2\xi}\sum_{m_3=1}^{\xi-1} X_{m_3\xi}\right]\sum_{i,j,k=1}^{\xi-1} 1,
\end{aligned} \tag{4.36}$$

$\xi \in \{3,\dots,n\}$. At this point, note that the term $\frac{1}{(\xi-1)^3}$ simplifies with the sum $\sum_{i,j,k=1}^{\xi-1} 1$ and we finally get:

$$\frac{1}{2\xi(\xi-1)}\sum_{i,j,m_1=1}^{\xi-1} X_{m_1\xi}X_{j\xi}X_{i\xi} - \frac{1}{6\xi(\xi-1)}\sum_{m_1,m_2,m_3=1}^{\xi-1} X_{m_1\xi}X_{m_2\xi}X_{m_3\xi} = \frac{1}{3\xi(\xi-1)}\sum_{i,j,k=1}^{\xi-1} X_{i\xi}X_{j\xi}X_{k\xi}, \tag{4.37}$$

where, in the last passage, the indices are again called with the familiar notation i, j, k. Using expression (4.37) for the exponents, we reach a simplified version of $\mu_n(\alpha)$, which we call $\tilde{\mu}_n(\alpha)$:

$$\begin{aligned}
\tilde{\mu}_n(\alpha) :=& \frac{1}{\binom{n}{2}}\ln\sum_{X_{12}} e^{h_p X_{12}}(1-p)\sum_{X_{13},X_{23}} e^{\frac{\alpha}{3\cdot 2}\left(\frac{\sum_{i,j,k=1}^{2} X_{i3}X_{j3}X_{k3}}{3}\right)} e^{h_p(X_{13}+X_{23})}(1-p)^2 \cdot \\
&\cdot \sum_{X_{14},X_{24},X_{34}} e^{\frac{\alpha}{4\cdot 3}\left(\frac{\sum_{i,j,k=1}^{3} X_{i4}X_{j4}X_{k4}}{3}\right)} e^{h_p(X_{14}+X_{24}+X_{34})}(1-p)^3 \cdots \\
&\cdots \sum_{X_{1n},X_{2n},\dots,X_{n-1n}} e^{\frac{\alpha}{n(n-1)}\left(\frac{\sum_{i,j,k=1}^{n-1} X_{in}X_{jn}X_{kn}}{3}\right)} e^{h_p(X_{1n}+\cdots+X_{n-1n})}(1-p)^{n-1}.
\end{aligned}$$

Finally we pass to the limit, thus getting:

$$\begin{aligned}
\lim_{n\to+\infty}\tilde{\mu}_n(\alpha) &= \lim_{n\to+\infty}\frac{1}{\binom{n}{2}}\ln\sum_{X_{12}} e^{h_p X_{12}}(1-p)\prod_{\xi=2}^{n-1}\sum_{\{X_{i\xi+1}\}_{i=1,\dots,\xi}\in\{0,1\}^\xi} e^{\alpha\frac{\left(\sum_{i=1}^{\xi} X_{i\xi+1}\right)^3}{3\xi^2}+h_p\sum_{i=1}^{\xi} X_{i\xi+1}}(1-p)^\xi \\
&= \lim_{n\to+\infty}\frac{1}{\binom{n}{2}}\ln\prod_{\xi=2}^{n-1}\sum_{\{X_{i\xi+1}\}_{i=1,\dots,\xi}\in\{0,1\}^\xi} e^{\frac{\alpha\left(\sum_{i=1}^{\xi} X_{i\xi+1}\right)^3}{3\xi^2}+h_p\sum_{i=1}^{\xi} X_{i\xi+1}}(1-p)^\xi.
\end{aligned} \tag{4.38}$$

Making use of the change of variable $m := \frac{\sum_{i=1}^{\xi} X_{i\xi+1}}{\xi}$ we can turn the sum in (4.38) into $\sum_m \mathcal{N}_\xi(m) e^{\xi[\frac{\alpha}{3}m^3+h_p m]}(1-p)^\xi$, where $\mathcal{N}_\xi(m)$ can be interpreted as the number of configurations having a given instantaneous magnetization m. Such number can be expressed through a binomial coefficient whose behavior for large ξ is given in terms of the entropy function of a Bernoulli process: $\mathcal{N}_\xi(m) = \binom{\xi}{\xi m} = e^{\xi[-I(m)+o(1)]}$, where $I(m) = m\ln(m) + (1-m)\ln(1-m)$. To leading exponential order in ξ, the partition function can be written as displayed below:

$$\lim_{n\to+\infty} \tilde{\mu}_n(\alpha) = \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \prod_{\xi=2}^{n-1} \int_0^1 e^{\xi\left\{\frac{\alpha}{3}m^3 - I_p(m)+o(1)\right\}} dm, \tag{4.39}$$

where we recall that $I_p(m) = m\ln\frac{m}{p} + (1-m)\ln\frac{1-m}{1-p}$. The integral involved in (4.39) can be evaluated using the Laplace's method, to get as final result the function

$$\begin{aligned}\lim_{n\to+\infty} \tilde{\mu}_n(\alpha) &= \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln \prod_{\xi=2}^{n-1} e^{\xi \sup_m\left\{\frac{\alpha}{3}m^3 - I_p(m)+o(1)\right\}} = \lim_{n\to+\infty} \frac{1}{\binom{n}{2}} \ln\, e^{\binom{n}{2}\sup_m\left\{\frac{\alpha}{3}m^3 - I_p(m)\right\}} \\ &= \sup_m \left[\frac{\alpha}{3}m^3 - I_p(m)\right] \end{aligned} \tag{4.40}$$

which coincides with the scalar problem related to the triangles observable in replica symmetric regime. As a further check observe that deriving (4.40) with respect to m we get

$$\frac{\partial}{\partial m}\left\{\frac{\alpha}{3}m^3 - I_p(m)\right\} = 0 \Rightarrow \alpha\, m^2 = \frac{\partial I_p(m)}{\partial m}, \tag{4.41}$$

where $\frac{\partial I_p(m)}{\partial m} = \ln\frac{m(1-p)}{p(1-m)}$. Rearranging (4.41) we recover the well-known fixed-point equation

$$\frac{e^{\alpha\, m^2 + h_p}}{e^{\alpha\, m^2+h_p}+1} = m, \tag{4.42}$$

which yields the scaled cumulant generating function

$$\lim_{n\to+\infty} \tilde{\mu}_n(\alpha) = \frac{\alpha}{3}m^{*3} - I_p(m^*),$$

where m^* satisfies (4.42). In conclusion, thanks to the substitution (4.35), we have recovered the scaled cumulant generating function of the normalized number of triangles in the replica symmetric regime (see (4.33)).

Chapter 5

Investigation of the replica breaking phase of the Exponential Random graph

Throughout Chapter 4, we focused our attention on the replica symmetric regime: now we want to switch to the replica breaking one, namely we want to numerically investigate the following problem:

Problem 5.1. Solve

$$\sup_{f \in \mathcal{W}} \left[\frac{\alpha\, t(H_2, f)}{3} - I_p(f) \right] \tag{5.1}$$

where $f(x, y) : [0, 1]^2 \to [0, 1]$ is symmetric and $\alpha \leq -2$.

The interest in such investigation is motivated by the fact that the state-of-art (presented in Chapter 3) provides a solution to this problem only for positive or slightly negative values of α (Theorems 3.8, 3.11) and for the limit $\alpha \to -\infty$ (Theorem 3.13), but very little is known when $\alpha \leq -2$. To be precise, it has been proved in Theorem 3.12 that there exists a negative threshold, called $-C(h_p)$, below which the solution of Problem 5.1 stops being the constant one, hence, strictly speaking, the replica breaking regime starts below $-C(h_p)$. Focusing on this behavior, we recall question a) we asked at the end of Chapter 3:
"Is it possible to find out an expression of the scaled cumulant generating function of the triangle observable when $\alpha \leq -C(h_p)$? What is the structure of the maximizer below such threshold of α?"

In order to tackle the question, we investigate Problem 5.1 via three approaches:

a) using a numerical optimization method (Projected Gradient);

b) using the extended version of the Cloning algorithm;

c) investigating the behavior of the variational problem when the solution is supposed to be the generalized bipartite graphon:

$$g^{(a)}_{p_i,p_b}(x,y) = \begin{cases} p_i & \text{if} \quad (x,y) \in [0,a]^2 \cup [a,1]^2 \\ p_b & \text{if} \quad (x,y) \in [0,a] \times [a,1] \cup [a,1] \times [0,a] \end{cases}, \tag{5.2}$$

where $p_i, p_b, a \in [0,1]$.

We start from the first strategy.

5.1 Projected Gradient method

5.1.1 Introduction to the Projected Gradient method

Gradient methods are by far the most standard and popular iterative schemes aimed at solving problem

$$\min_{x \in \Gamma} h(x) \tag{5.3}$$

where $\Gamma \subset \mathbb{R}^n$ is a non-empty, closed and convex set and $h : \Gamma \longmapsto \mathbb{R}$ is a continuously differentiable function over Γ. When $\Gamma = \mathbb{R}^n$ and therefore there are no restrictions on the unknown variable x, one speaks of *unconstrained optimization*, otherwise of *constrained optimization*. Many of these methods lean on the so-called *iterative descent idea*, which consists in generating a sequence of iterates $\{x^{(k)}\}_{k \in \mathbb{N}}$ in such a way that h is decreased at each step. The legitimate hope is that this iterative procedure will eventually lead to approach a global minimum or, at least, a stationary point of h. The decrease in the objective function is imposed by moving along a *descent direction* with a sufficiently small positive steplength, which can be kept fixed or may vary at each iteration: there exist different strategies aiming at adapting the choice of the steplength in order to improve the convergence rate of gradient methods while maintaining an affordable computational demand.

Definition 5.1. *A vector $x^* \in \Gamma$ is a stationary point of h over Γ if*

$$\nabla h(x^*)^T (y - x^*) \geq 0, \quad \forall\ y \in \Gamma. \tag{5.4}$$

Definition 5.2. *A vector $d \in \mathbb{R}^n$ is a descent direction for h at the point $x \in \mathbb{R}^n$ if*

$$< \nabla h(x), d > < 0.$$

Algorithm 2 Gradient Projection (GP) method

Choose the starting point $x^{(0)} \in \Gamma$, set the parameters $\beta, \delta \in (0,1)$, $0 < \tau_{min} < \tau_{max}$.
FOR $k = 0, 1, 2, \ldots$

STEP 1. Choose $\tau_k \in [\tau_{min}, \tau_{max}]$.

STEP 2. Compute the projection $y^{(k)} = P_\Gamma(x^{(k)} - \tau_k \nabla h(x^{(k)}))$;
if $y^{(k)} = x^{(k)}$, then $x^{(k)}$ is a stationary point and GP stops.

STEP 3. Define the descent direction $d^{(k)} = y^{(k)} - x^{(k)}$.

STEP 4. Set $\lambda_k = 1$.

STEP 5. Backtracking loop:
IF $h(x^{(k)} + \lambda_k d^{(k)}) \leq h(x^{(k)}) + \beta\lambda_k < \nabla h(x^{(k)}), d^{(k)} >$ THEN
go to STEP 6
ELSE
set $\lambda_k = \delta\lambda_k$ and go to STEP 5.
ENDIF

STEP 6. Set $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$.

END

Generally speaking, a gradient method is an iterative algorithm which, starting from an initial guess $x^{(0)} \in \mathbb{R}^n$, generates a sequence of the form

$$x^{(k+1)} = x^{(k)} + \tau_k d^{(k)}, \quad k = 0, 1, 2, \ldots \tag{5.5}$$

where $d^{(k)}$ is a descent direction at $x^{(k)}$ and τ_k is a positive parameter, denominated steplength in the aforementioned discussion. A simple and well studied algorithm for the solution of the constrained optimization problem (5.3) is the Gradient Projection (GP) method, whose general iteration is given by

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + \lambda_k d^{(k)} = \\ &= x^{(k)} + \lambda_k \left(P_\Gamma(x^{(k)} - \tau_k \nabla h(x^{(k)})) - x^{(k)} \right), \end{aligned} \tag{5.6}$$

where $\lambda_k \in (0, 1]$ is the linesearch parameter, τ_k is a positive steplength and P_Γ is the standard Euclidean projection onto Γ. The linesearch parameter is determined by means of a backtracking loop where the so-called *Armijo rule* is imposed (this approach is called *along the feasible direction*): such rule is expressed by inequality

$$h(x^{(k)} + \lambda_k d^{(k)}) \leq h(x^{(k)}) + \beta\lambda_k < \nabla h(x^{(k)}), d^{(k)} >$$

which guarantees the sufficient decrease of the objective function and is imposed at step 5. of Algorithm 2. The described scheme generates a sequence with stationary limit points, as proved in [4, Proposition 2.3.1, Proposition 2.3.3]. Furthermore, when the objective function is convex and admits at least one minimum point, the convergence of the whole sequence (5.6) to a solution of problem (5.3) is established in [33].

5.1.2 Application of the Projected Gradient method and numerical results

In this subsection we apply the projected gradient method to the discretized variational problem (5.1) turned into a minimization problem. Note that the function h considered in Subsection 5.1.1 is the functional

$$\frac{\alpha\, t(H_2, f)}{3} - I_p(f), \quad f \in \mathcal{W}. \tag{5.7}$$

We introduce a discrete analogous of the space of graphons $\mathcal{W}$: since each element of the space is a symmetric function $f(x,y) : [0,1]^2 \longmapsto [0,1]$, we represent f by a $n \times n$ matrix with elements $\{f_{i,j}\}_{i,j=1,\dots,n} \in [0,1]$, thus discretizing the square $[0,1]^2$. Actually, note that thanks to Theorem 3.15 we can restrict the research of the minimum on functions bounded away from zero and one, so we introduce the bond $f_{ij} \in [\varepsilon, 1-\varepsilon]$ for some $\varepsilon > 0$. Moreover, we add the condition $f_{ij} = f_{ji}$ which imposes the symmetry. Let us denote by $\mathbb{R}^{n\times n}$ the space of $n \times n$ real matrices and define

$$\Gamma_\varepsilon := \{f \in \mathbb{R}^{n\times n} : f_{ij} = f_{ji}, f_{ij} \in [\varepsilon, 1-\varepsilon] \quad \forall i,j \text{ and for some } \varepsilon > 0\}. \tag{5.8}$$

The discretized version of (5.7) reads

$$\mathcal{D}_n^{(\alpha)}(f) := \frac{1}{n^2}\left[\frac{-\alpha}{3n}\sum_{i,j,k=1}^{n} f_{ij}f_{jk}f_{ki} + \sum_{i,j=1}^{n}\left(f_{ij}\ln\frac{f_{ij}}{p} + (1-f_{ij})\ln\frac{1-f_{ij}}{1-p}\right)\right], \quad \alpha \le -2, f \in \Gamma_\varepsilon \tag{5.9}$$

and yields the minimization problem

Problem 5.2. For $\alpha \le -2$ solve

$$\min_{f\in\Gamma_\varepsilon} \frac{1}{n^2}\left[\frac{-\alpha}{3n}\sum_{i,j,k=1}^{n} f_{ij}f_{jk}f_{ki} + \sum_{i,j=1}^{n}\left(f_{ij}\ln\frac{f_{ij}}{p} + (1-f_{ij})\ln\frac{1-f_{ij}}{1-p}\right)\right]. \tag{5.10}$$

The legitimate hope is that Problem 5.2 converges to the continuous one when $n \to +\infty$. The convergence of the GP method to a stationary point, even in absence of convexity of the objective function, is proved in [46, Theorem 1],[10] under the assumption that the objective function satisfies the so-called *Kurdyka-Lojasiewicz* inequality (KL) ([35],[37]) and its gradient

is L-Lipschitz continuous. Such property was originally proved for real analytic functions and turns out to hold true for a large class of functions such as p norms, Kullback-Leibler divergence and indicator functions of box plus equality constraints. Coming back to our application, the objective function $D_n^{(\alpha)}(f)$ is real analytic and, thus, it satisfies the Kurdyka-Lojasiewicz property, whereas the L-Lipschitz continuity is guaranteed by the compactness of Γ_ε: therefore the GP method converges to a stationary point when applied to Problem 5.2. In order to implement the scheme, beside the discretized objective function defined in (5.9), we need the expression of the gradient: for each $m, l \in \{1, \dots, n\}$ its generic element takes the form

$$
\begin{aligned}
\frac{D_n^{(\alpha)}(f)}{\partial f_{ml}} &= -\frac{\alpha}{3n^3} \sum_{i,j,k} \frac{\partial}{\partial f_{ml}} f_{ij} f_{jk} f_{ki} + \frac{1}{n^2} \sum_{i,j} \frac{\partial}{\partial f_{ml}} \left(f_{ij} \ln \frac{f_{ij}}{p} + (1 - f_{ij}) \ln \left(\frac{1 - f_{ij}}{1 - p} \right) \right) \\
&= -\frac{\alpha}{3n^3} \sum_{i,j,k=1}^{n} \left[\frac{\partial f_{ij}}{\partial f_{ml}} f_{jk} f_{ki} + f_{ij} \frac{\partial f_{jk}}{\partial f_{ml}} f_{ki} + f_{ij} f_{jk} \frac{\partial f_{ki}}{\partial f_{ml}} \right] + \frac{1}{n^2} \sum_{i,j=1}^{n} \left[\frac{\partial f_{ij}}{\partial f_{ml}} \ln \frac{f_{ij}}{p} - \frac{\partial f_{ij}}{\partial f_{ml}} \ln \frac{1 - f_{ij}}{1 - p} \right] \\
&= -\frac{\alpha}{3n^3} \sum_{i,j,k=1}^{n} f_{jk} f_{ki} \mathbb{1}_{[i=m, j=l]} + f_{ij} f_{ki} \mathbb{1}_{[j=m, k=l]} + f_{ij} f_{jk} \mathbb{1}_{[k=m, i=l]} + \frac{1}{n^2} \sum_{i,j=1}^{n} \ln \frac{f_{ij}(1-p)}{p(1-f_{ij})} \mathbb{1}_{[i=m, j=l]} \\
&- \frac{\alpha}{3n^3} \left(\sum_{k=1}^{n} f_{lk} f_{km} + \sum_{i=1}^{n} f_{im} f_{li} + \sum_{j=1}^{n} f_{lj} f_{jm} \right) + \frac{1}{n^2} \ln \left(\frac{f_{ml}(1-p)}{p(1-f_{ml})} \right) \\
&= \frac{1}{n^2} \left[-\frac{\alpha}{n} \sum_{k=1}^{n} f_{lk} f_{km} + \ln \left(\frac{f_{ml}(1-p)}{p(1-f_{ml})} \right) \right]. \qquad (5.11)
\end{aligned}
$$

Remark 5.1. From (5.11), note that $-\frac{\alpha}{n} \sum_{k=1}^{n} f_{lk} f_{km} + \ln \left(\frac{f_{ml}(1-p)}{p(1-f_{ml})} \right) = 0$ is the discretized version of (3.56) and, when f_{lm}, for any $l, m \in \{1, \dots, n\}$, is identically equal to the solution of the fixed-point equation (3.36), the gradient vanishes.

5.1.3 Numerical results

The goal of this section is to investigate the structure of the minimizers of Problem 5.2 by applying the GP method to the discretized functional (5.9). We recall that α is negative and precisely $\alpha \leq -2$. At a macroscopic level, we can draw the following conclusions:

a) simulations highlight the presence of two kind of minimizers: the constant ones and those who exhibit a chessboard pattern. Such minima are revealed for all values of α tested and constant functions turn out to be numerically equivalent to the solution, u^*, of the fixed-point equation (3.36) that we recall below:

$$
\frac{e^{\alpha u^2 + h_p}}{e^{\alpha u^2 + h_p} + 1} = u, \qquad (5.12)
$$

where $h_p = \ln \frac{p}{1-p}$.

b) Chessboard structures are different and their patterns strongly depend on the initial condition of the method. However, all the structures revealed by the algorithm for the negative α tested, approximate the bipartite graphon

$$g_{0,p}(x,y) = \begin{cases} 0 & \text{if} \quad (x,y) \in \left[0,\frac{1}{2}\right]^2 \cup \left[\frac{1}{2},1\right]^2 \\ p & \text{if} \quad (x,y) \in \left[0,\frac{1}{2}\right] \times \left[\frac{1}{2},1\right] \cup \left[\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right] \end{cases},$$

as we will argue.

c) There exists a critical value of α below which the optimizer of $\mathcal{D}_n^{(\alpha)}(f)$ stops being the constant solution and starts being a chessboard one.

We provide the details of the analysis below. Being interested in the solution of Problem 5.1 for $\alpha \to -\infty$, here we consider two negative cases, which are significant because they show a change in the structure of the optimizer: $\alpha = -109$ and $\alpha = -110$, for $p = 0.4$.

Description of numerical results

Figures 5.1, 5.2, 5.3 show the stationary solutions found by Algorithm 2 for the variational problem (5.2), starting from 12 different conditions when $p = 0.4$ and $\alpha = -110$. Figures 5.4, 5.5, 5.6 return the same information when $p = 0.4$ and $\alpha = -109$. The method has been implemented using a 40×40 grid and two different choices of τ_k (put side by side in each figure): in the first case the steplength is fixed whereas in the second one it has been settled using the Barzilai-Borwein rules [3]. Such rules allow to update the steplength at each iteration in order to converge faster to a stationary point; the downside of this approach is that the objective function can not be explored as thoroughly as with a fixed (and small) steplength.
The two choices of the steplength together with the initial conditions, bring to light different stationary points: let us denote by f^* such solutions. All constant functions get from the twelve starting conditions and present in Figures 5.1, 5.2, 5.3 (blue frames) are approximately equal to $f_1^* \approx 0.120082$ with $\mathcal{D}_{40}^{(-110)}(f_1^*) \approx 0.255916$. All chessboard patterns present in the same figures numerically realize the same value of $\mathcal{D}_{40}^{(-110)}$, hence we denote by f_2^* this class of solutions and we have $\mathcal{D}_{40}^{(-110)}(f_2^*) \approx 0.255361$. Figures 5.4, 5.5, 5.6 provide the same information but with $\alpha = -109$. In this case, all constant functions are approximately equal to $f_1^* \approx 0.120486$ with $\mathcal{D}_{40}^{(-109)}(f_1^*) \approx 0.255336$ and the chessboard patterns provide a functional value $\mathcal{D}_{40}^{(-109)}(f_2^*) \approx 0.255356$.

Characterization of constant functions

Note that for both $\alpha = -109$ and $\alpha = -110$ the constant solution f_1^* coincides with the numerical solution of the fixed-point equation (5.12) up to the sixth decimal digit.

	$u^*(\alpha)$	Constant solution f_1^*	Chessboard solution f_2^*	$\mathcal{D}_{40}^{(\alpha)}(f_1^*)$	$\mathcal{D}_{40}^{(\alpha)}(f_2^*)$
$\alpha = -109$	0.120486	0.120486	$p_i \approx 0.000114, p_b \approx 0.398829$	0.255336	0.255356
$\alpha = -110$	0.120082	0.120082	$p_i \approx 0.000100, p_b \approx 0.399100$	0.255916	0.255361

Table 5.1: Stationary solutions returned by the GP method for $p = 0.4$ and $\alpha = -109, -110$.

Characterization of chessboard structures

We can observe that chessboard stationary solutions shown in Figures 5.1, 5.2, 5.3, 5.4, 5.5, 5.6 exhibit different patterns. We recall that they are obtained, for $p = 0.4$, starting from twelve initial conditions with, respectively, $\alpha = -109$ and $\alpha = -110$. Table 5.1 resumes the numerical response: we have denoted by p_i and p_b the two values respectively assumed by the light blue and the dark blue frames of the chessboards (this notation is coherent with definition (5.2)). Despite their diversity, chessboard solutions share two main peculiarities: the first one is that they assume two values close to 0 and p, thus reminding the bipartite graphon

$$g_{0,p}(x,y) = \begin{cases} 0 & \text{if} \quad (x,y) \in \left[0,\frac{1}{2}\right]^2 \cup \left[\frac{1}{2},1\right]^2 \\ p & \text{if} \quad (x,y) \in \left[0,\frac{1}{2}\right] \times \left[\frac{1}{2},1\right] \cup \left[\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right] \end{cases} \tag{5.13}$$

(however, we notice from Table 5.1 that, when $\alpha = -110$, the pair $(p_i.p_b)$ is closer to $(0,p)$ than the pair obtained for $\alpha = -109$).
The second one is that, when they are passed as argument to the discretized functional $\mathcal{D}_{40}^{(\alpha)}(f)$, they return a good approximation of the continuous functional:

$$-\frac{\alpha\, t(H_2, g_{0,p})}{3} + I_p(g_{0,p}) = -\frac{\ln(1-p)}{2} \approx 0.255412 \quad \forall\alpha,$$

since $t(H_2, g_{0,p}) = 0$.
It is evident from simulations that a similarity between the chessboard patterns and the bipartite graphon (5.13) arises, if only because the discretized functional $\mathcal{D}_{40}^{(\alpha)}(f_2^*)$ ($\alpha = -109, -110$) approximates the continuous one.
Moving on exact results, thanks to Proposition 5.2, we can work out the membership of all chessboard structures taking values 0 and p to the same equivalence class of the bipartite graphon $g_{0,p}$ with respect to relation (3.4), thus justifying, at least partially, the numerical results.
Before stating Proposition 5.2, we introduce two properties portraying a specific class of patterns. Consider two subsets A and A^c (where A^c denotes the complement of A) of $[0,1]$ with the two following properties:

a) $|A| = |A^c| = \frac{1}{2}$ where $|\cdot|$ denotes the Lebesgue measure;

b) A and A^c cover the whole of $[0,1]$ and they are disjoint up to sets of null Lebesgue measure.

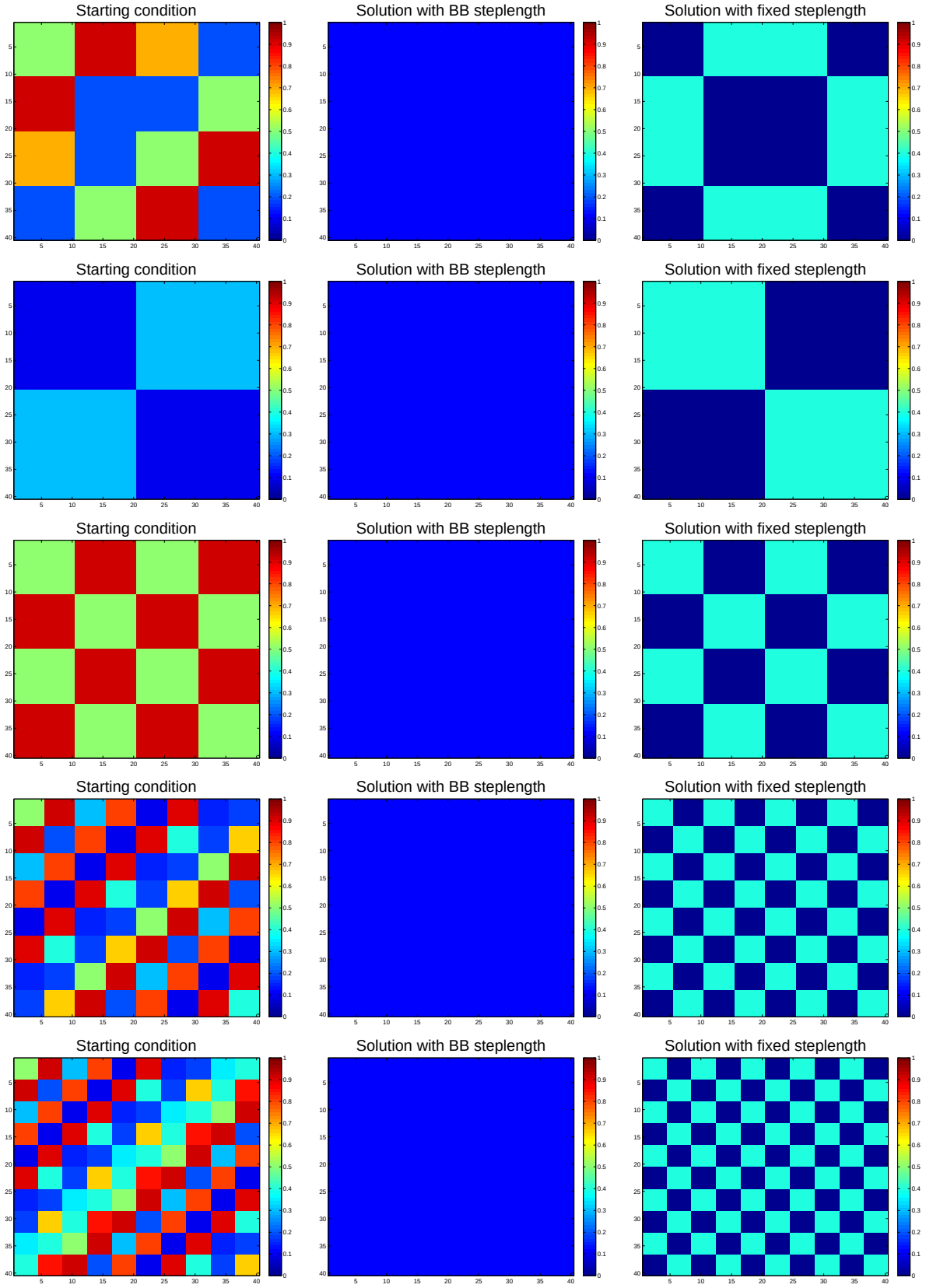


Figure 5.1: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -110$; the picture reproduces, respectively, the starting condition (left column), the solution found by the method GP with the Barzilai-Borwein steplength (central column) and the solution found with the fixed steplength (right column). The blue frame denotes the constant function $f_1^* \approx 0.120082$ whereas the chessboard patterns take values $p_i \approx 0.000100$ (light blue) and $p_b \approx 0.399100$ (dark blue). The chessboard solution is the best optimizer when $\alpha = -110$.

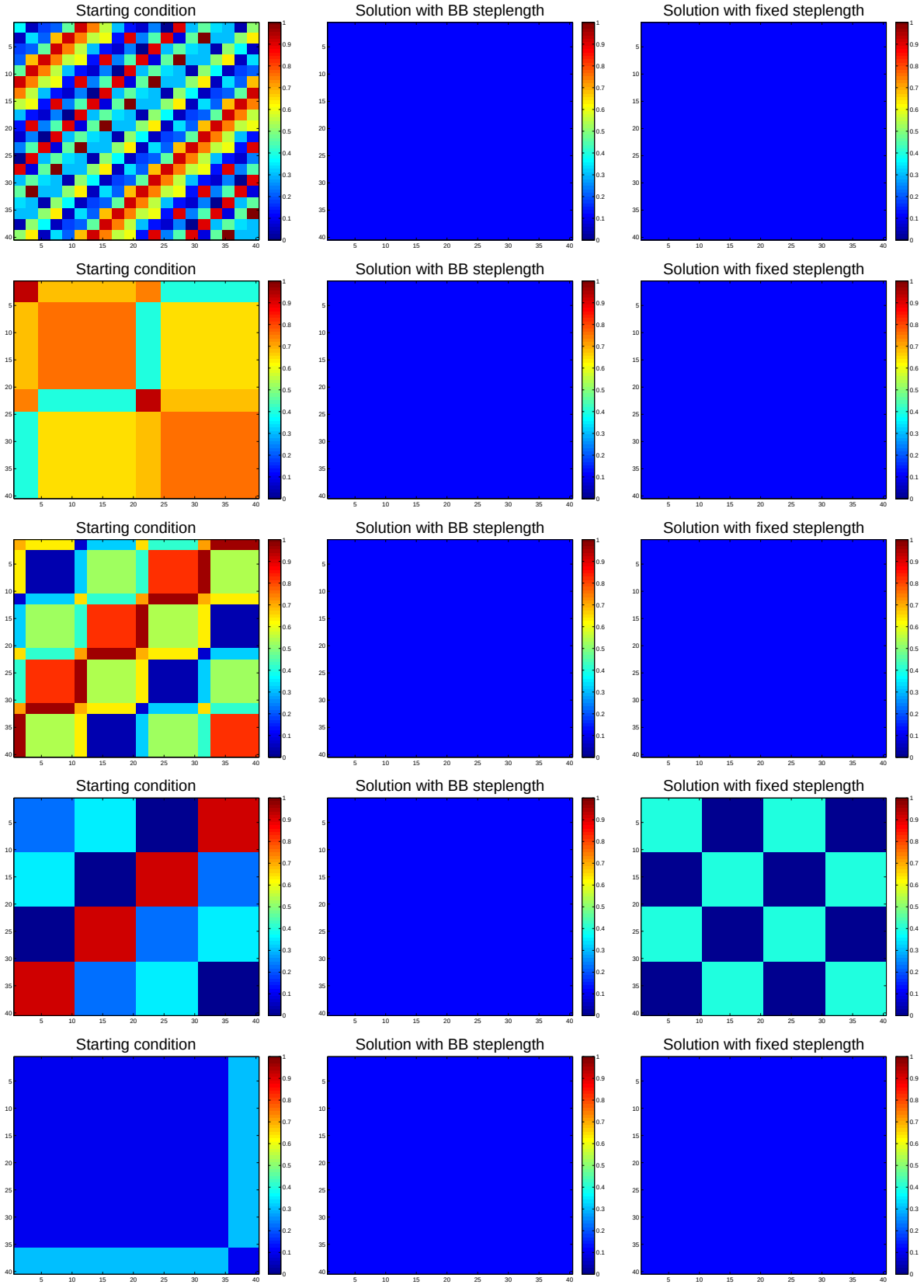


Figure 5.2: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -110$; the picture reproduces, respectively, the starting condition (left column), the solution found by the method GP with the Barzilai-Borwein steplength (central column) and the solution found with the fixed steplength (right column). The blue frame denotes the constant function $f_1^* \approx 0.120082$ whereas the chessboard patterns take values $p_i \approx 0.000100$ (light blue) and $p_b \approx 0.399100$ (dark blue). The chessboard solution is the best optimizer when $\alpha = -110$.

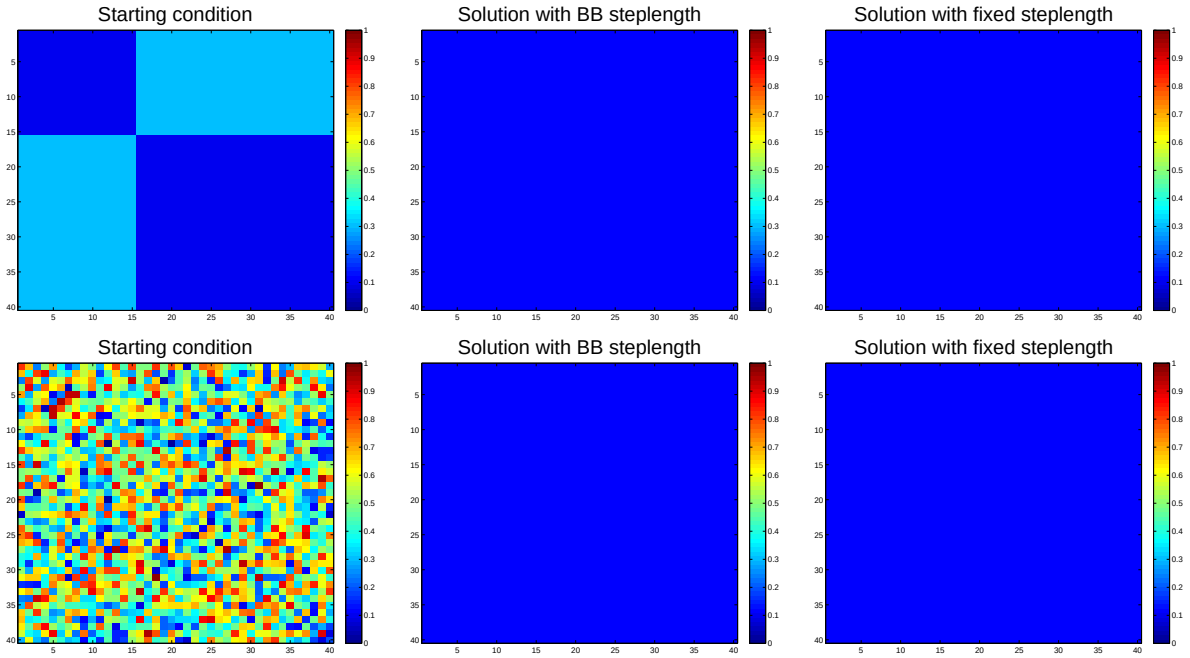


Figure 5.3: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -110$; the picture reproduces, respectively, the starting condition (left column), the solution found by the method GP with the Barzilai-Borwein steplength (central column) and the solution found with the fixed steplength (right column). The blue frame denotes the constant function $f_1^* \approx 0.120082$ whereas the chessboard patterns take values $p_i \approx 0.000100$ (light blue) and $p_b \approx 0.399100$ (dark blue). The chessboard solution is the best optimizer when $\alpha = -110$.

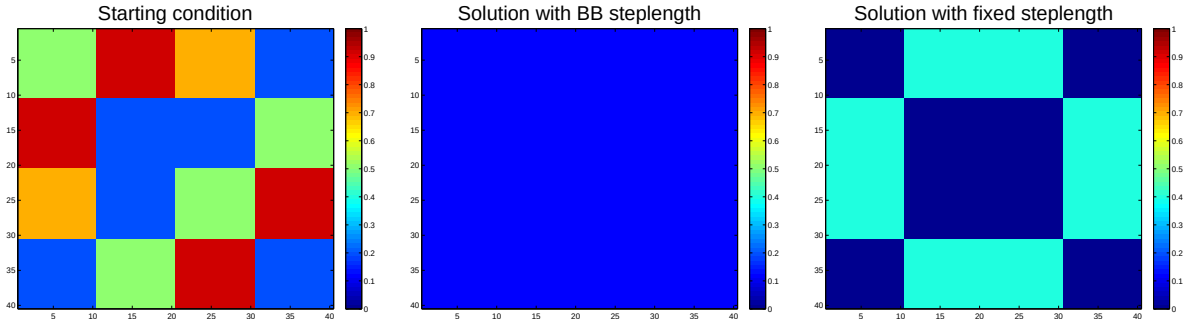


Figure 5.4: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -109$. The blue frame denotes the constant function $f_1^* \approx 0.120486$ whereas the chessboard patterns take values $p_i \approx 0.000114$ (light blue) and $p_b \approx 0.398829$ (dark blue). The panel shows, in the order, the starting condition and the solutions respectively get using the Barzilai-Borwein and the constant steplength. The constant solution is the best optimizer when $\alpha = -109$.

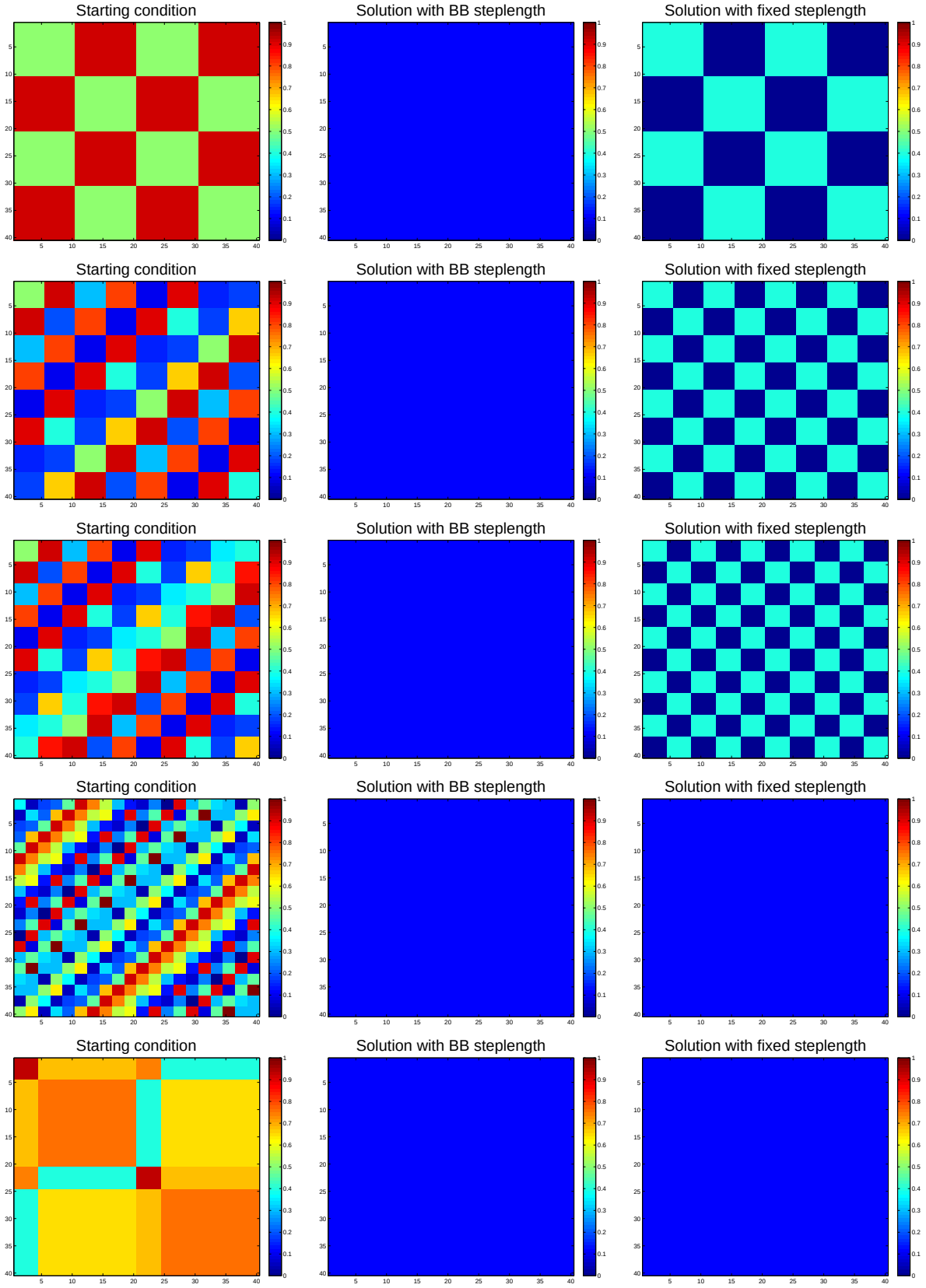


Figure 5.5: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -109$; the picture reproduces, respectively, the starting condition (left), the solution found by the method GP with the Barzilai-Borwein steplength (center) and the solution found with the fixed steplength (right). The blue frame denotes the constant function $f_1^* \approx 0.120486$ whereas the chessboard patterns take values $p_i \approx 0.000114$ (light blue) and $p_b \approx 0.398829$ (dark blue). The constant solution is the best optimizer when $\alpha = -109$.

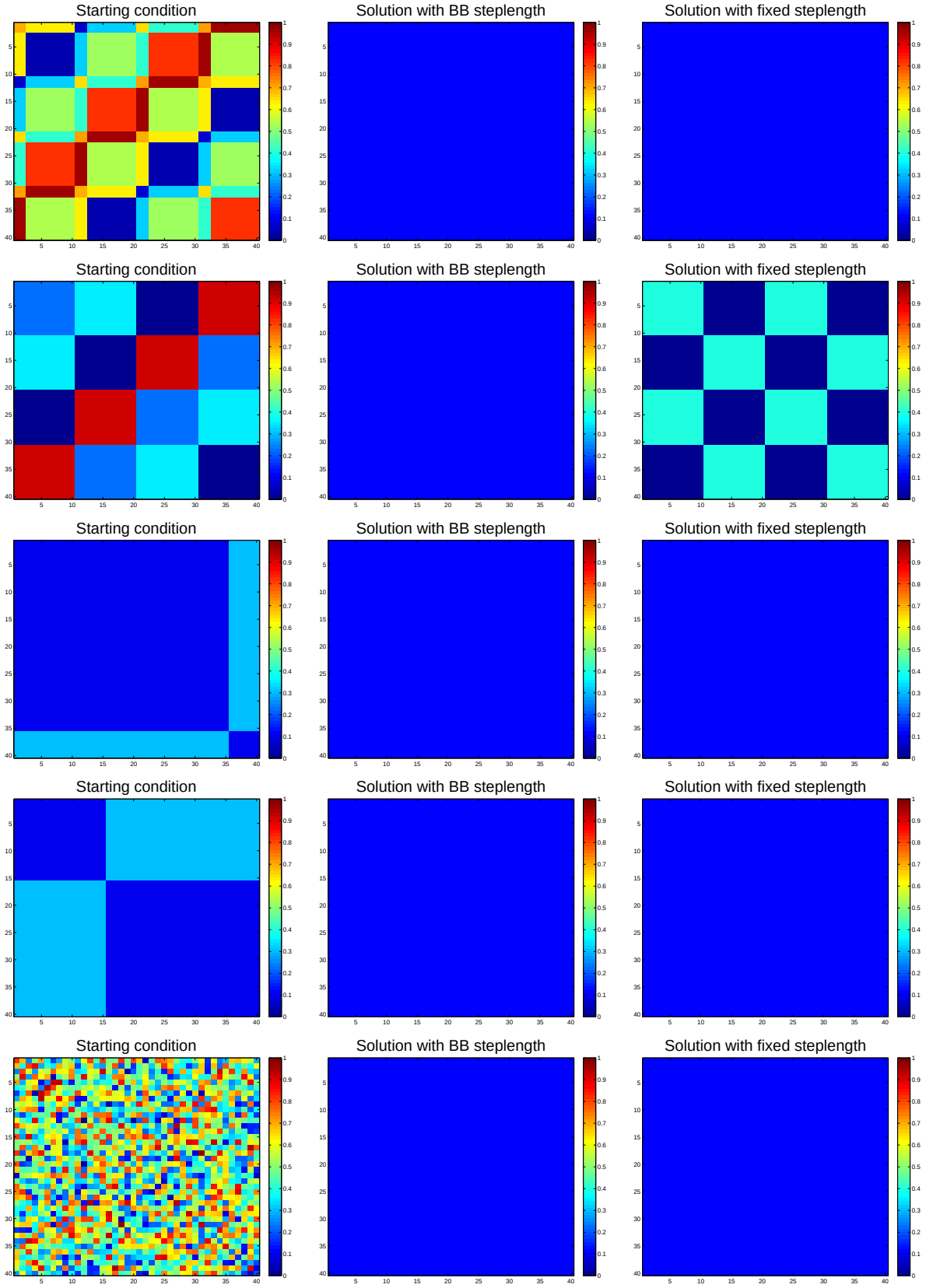


Figure 5.6: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -109$; the picture reproduces, respectively, the starting condition (left), the solution found by the method GP with the Barzilai-Borwein steplength (center) and the solution found with the fixed steplength (right). The blue frame denotes the constant function $f_1^* \approx 0.120486$ whereas the chessboard patterns take values $p_i \approx 0.000114$ (light blue) and $p_b \approx 0.398829$ (dark blue). The constant solution is the best optimizer when $\alpha = -109$.

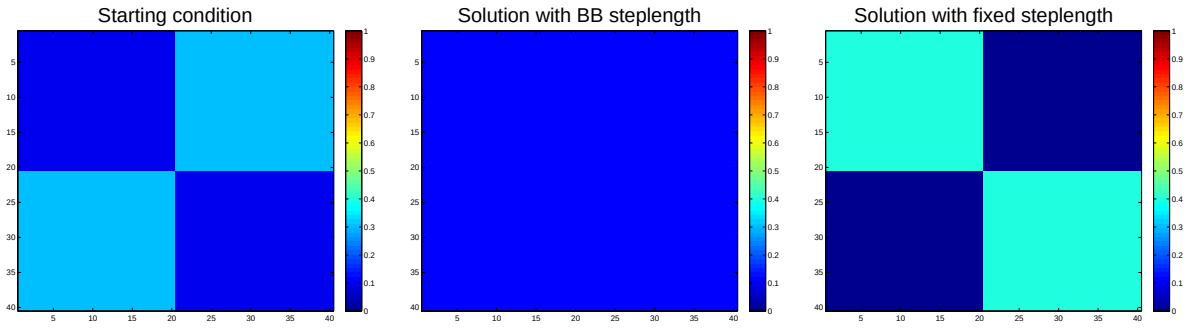


Figure 5.7: Stationary solutions of Problem 5.2 for $p = 0.4$ when $\alpha = -109$; the picture reproduces, respectively, the starting condition (left), the solution found by the method GP with the Barzilai-Borwein steplength (center) and the solution found with the fixed steplength (right). The blue frame denotes the constant function $f_1^* \approx 0.120486$ whereas the chessboard patterns take values $p_i \approx 0.000114$ (light blue) and $p_b \approx 0.398829$ (dark blue). The constant solution is the best optimizer when $\alpha = -109$.

The following proposition holds:

Proposition 5.1. *Take A and A^c with the properties a) and b) listed above. Define the function*

$$g(x,y) := \begin{cases} 0 & \text{if} \quad (x,y) \in [A \times A] \cup [A^c \times A^c] \\ p & \text{if} \quad (x,y) \in [A \times A^c] \cup [A^c \times A]. \end{cases} \tag{5.14}$$

Then

$$g(x,y)g(y,z)g(z,x) = 0 \quad \forall (x,y,z).$$

Moreover

$$\int_0^1 \int_0^1 I_p(g(x,y)) = -\frac{\ln(1-p)}{2}.$$

Proof. If $(x,y) \in A^c \times A^c$ or $(x,y) \in A \times A$ then $g(x,y) = 0$ and the the result follows. Let's start from $(x,y) \in A \times A^c$ so that $g(x,y) = p$.
In order to take a value of $g(y,z)$ different from 0, we have to choose $(y,z) \in A^c \times A$ and necessarily follows, in the end, that $(z,x) \in A \times A$ thus leading to $g(z,x) = 0$. It follows straightforward by the symmetry of $g(x,y)$ that the same conclusion holds starting from $(x,y) \in A^c \times A$.
For what concerns the relative entropy, a direct computation leads to:

$$\begin{aligned} \int_0^1 \int_0^1 I_p(g(x,y)) &= \int\int_{A\times A} I_p(0) + \int\int_{A^c\times A^c} I_p(0) + 2\int\int_{A\times A^c} I_p(p) = \frac{I_p(0)}{2} = \frac{1}{2}\ln\frac{1}{1-p} \\ &= -\frac{\ln(1-p)}{2}. \end{aligned}$$

□

An immediate consequence of Proposition 5.1 is that for any g defined as in (5.14), $\frac{\alpha\, t(H_2,g)}{3} - I_p(g) = -I_p(g) = \frac{\ln(1-p)}{2}$.
The following example shows that a particular configuration for which properties a) and b) hold, is the chessboard. This helps to justify why the class of functions with this pattern returned by the GP method, which we called f_2^*, approximates $\frac{\ln(1-p)}{2}$. In fact, having f_2^* values close to 0 and p, we can lean on Proposition 5.1 for explaining the reason of this numerical evidence.

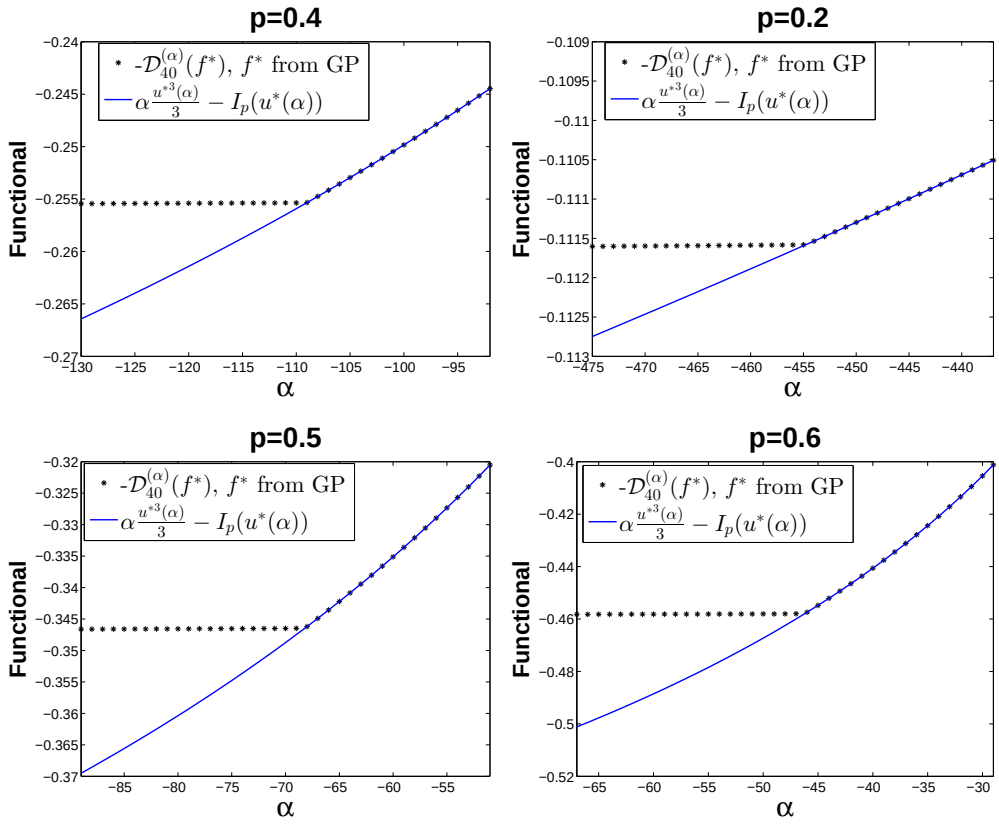


Figure 5.8: Plot of $-\mathcal{D}_{40}^{(\alpha)}(f^*)$ as a function of α, where f^* is the stationary solution returned by the Gradient Projection method (black, dotted line), overlapped with the function $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$, where $u^*(\alpha)$ solves (5.12) (continuous, blue line). The sign of $\mathcal{D}_{40}^{(\alpha)}(f^*)$ is changed to be consistent with the original maximization (and not minimization) Problem 5.1.

Example 5.1 (Chessboard)**.** For a fixed $n \in \mathbb{N} = \{1, 2, \dots\}$, let $V^n = \{0, 2, 4, \dots, 2^n - 4\}$ be the set of even integers of cardinality $|V^n| = \frac{2^n}{2} - 1$. Note that the case $n = 1$ is empty by definition. Set

$$A := \left[0, \frac{1}{2^n}\right] \bigcup_{k \in V^n} \left[\frac{2}{2^n} + k\frac{1}{2^n}, \frac{2}{2^n} + (k+1)\frac{1}{2^n}\right] \tag{5.15}$$

and let A^c be the complement set of A on $[0,1]$.
When $n=1$, since V^1 is empty, we have $A=\left[0,\frac{1}{2}\right]$.
When $n=2$ we have that $|V^2|=1$, consequently $V^2=\{0\}$ and

$$A=\left[0,\frac{1}{4}\right]\cup\left[\frac{2}{4},\frac{2}{4}+\frac{1}{4}\right]=\left[0,\frac{1}{4}\right]\cup\left[\frac{1}{2},\frac{3}{4}\right],\quad (k=0).$$

When $n=3$ we have $|V^3|=3$ and consequently $V^3=\{0,2,4\}$ and

$$A=\left[0,\frac{1}{8}\right]\left[\frac{2}{8},\frac{2}{8}+\frac{1}{8}\right]\cup\left[\frac{2}{8}+\frac{2}{8},\frac{2}{8}+\frac{3}{8}\right]\cup\left[\frac{2}{8}+\frac{4}{8},\frac{2}{8}+\frac{5}{8}\right]=\left[0,\frac{1}{8}\right]\cup\left[\frac{1}{4},\frac{3}{8}\right]\cup\left[\frac{1}{2},\frac{5}{8}\right]\cup\left[\frac{6}{8},\frac{7}{8}\right],\quad (k=0,2,4).$$

One can check that the two properties a) and b) are satisfied: in fact $|A|=\frac{1}{2^n}+\frac{1}{2^n}\left(\frac{2^n}{2}-1\right)=\frac{1}{2}$ and, by construction, A and A^c are disjoint, with $|A^c|=\frac{1}{2}$.
Observe that, setting A as in (5.15), the function $g(x,y)$ defined in (5.14) assumes a chessboard structure, with a number of frames depending on the fixed n.

As a further step, we quote the following proposition, which identifies all the functions defined in (5.14) up to bijections and is extracted from the proof of Theorem 7.3 in [18]:

Proposition 5.2. *All the functions defined under the hypothesis of Proposition 5.1, belong to the same equivalence class of the bipartite graphon* (5.13)*, according to the equivalence relation* (3.4)*.*

Figure 5.8 shows how $-\mathcal{D}_{40}^{(\alpha)}(f^*)$ changes as a function of α when it is computed on the best optimizer found by the Gradient Projection method for $p=0.4$, $p=0.2$, $p=0.5$ and $p=0.6$ (α varies with unitary step). The sign of $\mathcal{D}_{40}^{(\alpha)}(f^*)$ is negative in the plot, since Problem 5.2 is formulated as a minimization one, while we want to be consistent with the original maximization Problem 5.1. Note that respectively around $\alpha=-110$, $\alpha=-455$, $\alpha=-69$ and $\alpha=-47$ the optimal solution f^* of $\mathcal{D}_{40}^{(\alpha)}$ switches from the constant one f_1^* (approximating the fixed-point equation (5.12)) to one of the chessboard functions f_2^* ascribable to the generalized bipartite graphon $g_{p_i,p_b}^{(a)}$ with $p_i\approx 0$, $p_b\approx p$ and $a=\frac{1}{2}$.
We now turn to the second approach of our investigation, namely the numerical response returned by the extended Cloning algorithm introduced in Section 4.1.

5.2 Application of the Cloning algorithm

The second approach we use to investigate the behavior of the scaled cumulant generating function in the large, negative limit of α, is the Cloning algorithm. It is in order to remark that the method is not meant for running through large values of such parameter, since it is a population dynamics scheme whose reproduction rates are exponential in α and then they explode or vanish very fast. For this reason, simulations show criticalities when the probability of the graph p grows above 0.6 because the number of triangles and edges, combined with the

size of α makes the population die very soon. Moreover, as can be worked out from simulations of Subsection 5.1.3 (Figure 5.8) and as we will see in Section 5.3, the negative threshold, brought to light by our numerical investigation, which discriminates between the constant solution and a different one in Problem 5.1, seems to steeply decrease (or increase, in absolute value) as long as p does the same. For the reasons mentioned above, we have selected for simulations the two intermediate values $p = 0.4$ and $p = 0.5$. Figure 5.9 shows the scaled cumulant generating function reproduced by the method for $n = 10$, $n = 12$, $n = 15$ when $p = 0.4$ (top) and $p = 0.5$ (bottom). We recall that the algorithm implements a growing-size dynamics where n coincides with the volume of the graphs belonging to the population of the last iteration. Both pictures reproduce a wider interval, respectively $[-150, 7]$ and $[-82, 7]$ (on the right) and a zoomed one, $[-150, 0]$ and $[-82, 0]$ (on the left).

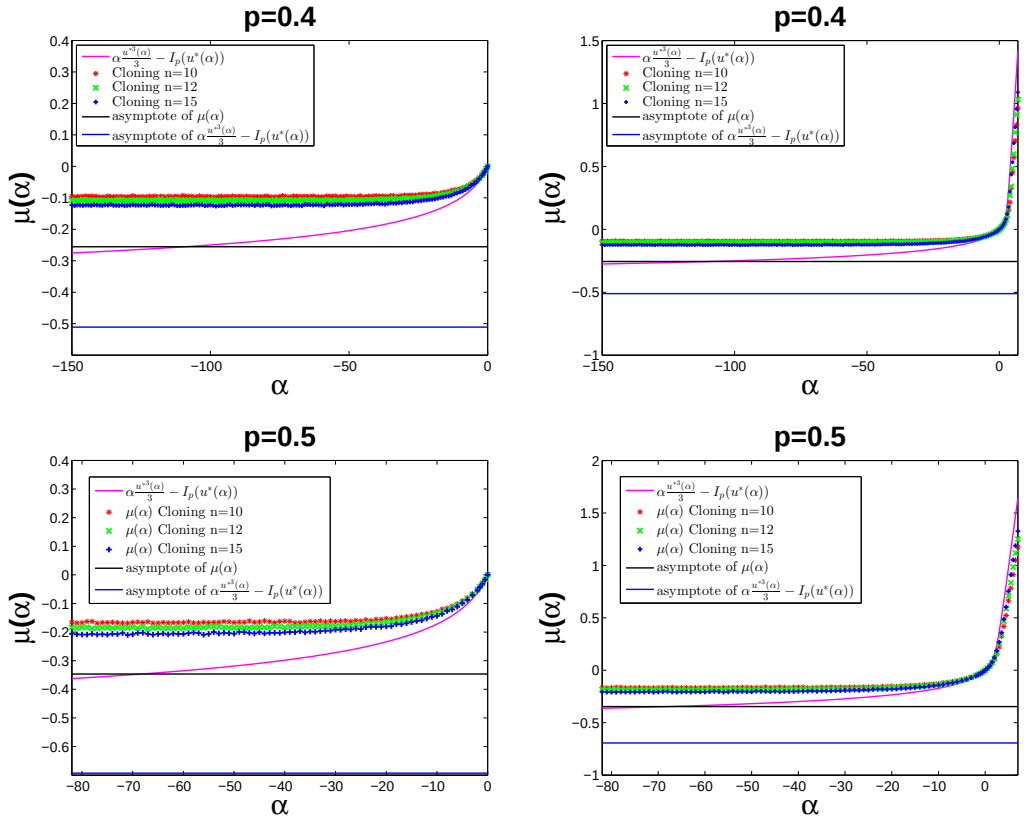


Figure 5.9: Numerical approximation of the scaled cumulant generating function of the triangle observable reproduced by the Cloning algorithm for $p = 0.4$ (top) and $p = 0.5$ (bottom). The picture also shows the function $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$, where $u^*(\alpha)$ solves (5.12) (pink curve) and its asymptote $\ln(1-p)$ (blue curve) together with the known asymptote of the scaled cumulant generating function $\frac{1}{2}\ln(1-p)$ (black curve).

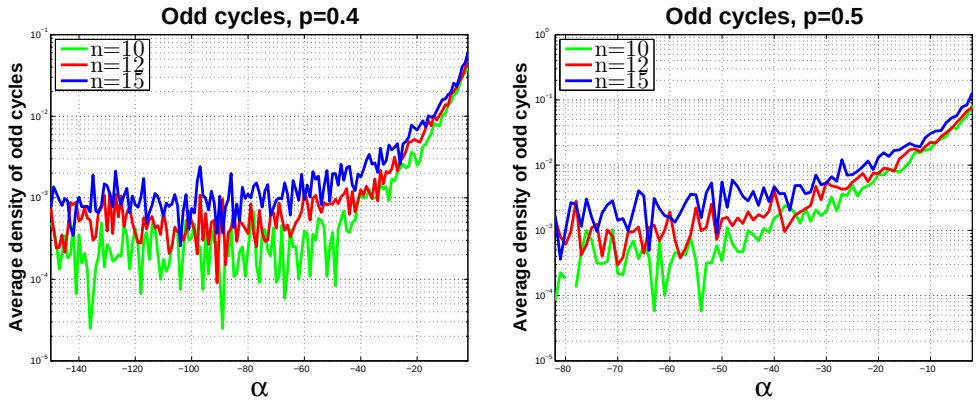


Figure 5.10: Average density of odd cycles per population when $p = 0.4$ (left) and $p = 0.5$ (right).

Note that the figure also includes the function $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$, where $u^*(\alpha)$ solves (5.12) (pink, continuous curve) and its asymptote $\ln(1-p)$ (blue curve); moreover, the picture reports the correct asymptote of the scaled cumulant generating function which is known from Theorem 3.13 and is equal to $\frac{1}{2}\ln(1-p)$. Since the core of the information is carried by the population of graphs, it is also possible to check out some other observables such as the density of odd cycles, as long as α decreases. We recall from Remark 3.9 that, when such number vanishes, the graph is bipartite and then this observable potentially provides information on the kind of graphs belonging to the population. We also remind the reader that each run of the algorithm involves, for a fixed α, a starting family of clones, according to Algorithm 2. Figure 5.10 shows the average density of odd cycles per population for $p = 0.4$ and $p = 0.5$ as long as α decreases: the starting size of the family used for reproducing the picture is $M_2 = 1000$. Note that, for all sizes of graphs $n = 10$, $n = 12$, $n = 15$, such average number flattens to zero when α decreases. Another quantity that it possible to monitor is the average density of triangles per population, as α varies; in fact we know that in replica symmetric such density must be numerically close to the solution u^* of the fixed-point equation (5.12) whereas in replica breaking we have no information. If we expect, below a critical threshold, to go towards a bipartite graphon $g_{0,p}$ (which is reached, however, only in the limit $\alpha \to -\infty$), then, crossing such value, we should observe a decrease in the density of triangles, which is null in $g_{0,p}$. Figure 5.11 shows the average density of triangles per population when $p = 0.4$ and $p = 0.5$: in the first case, for $n = 15$, it is possible to appreciate that, approximately around $\alpha = -110$ the density flattens towards zero (note the coherence with the critical threshold revealed by Figure 5.8 for $p = 0.4$) whereas in the positive, or slightly negative, range of α, the density sticks to the expected one, namely the one known for the replica symmetric regime. When $n = 12$, as natural, the triangle density recovered from the population, approaches the expected one in a worst way. For what

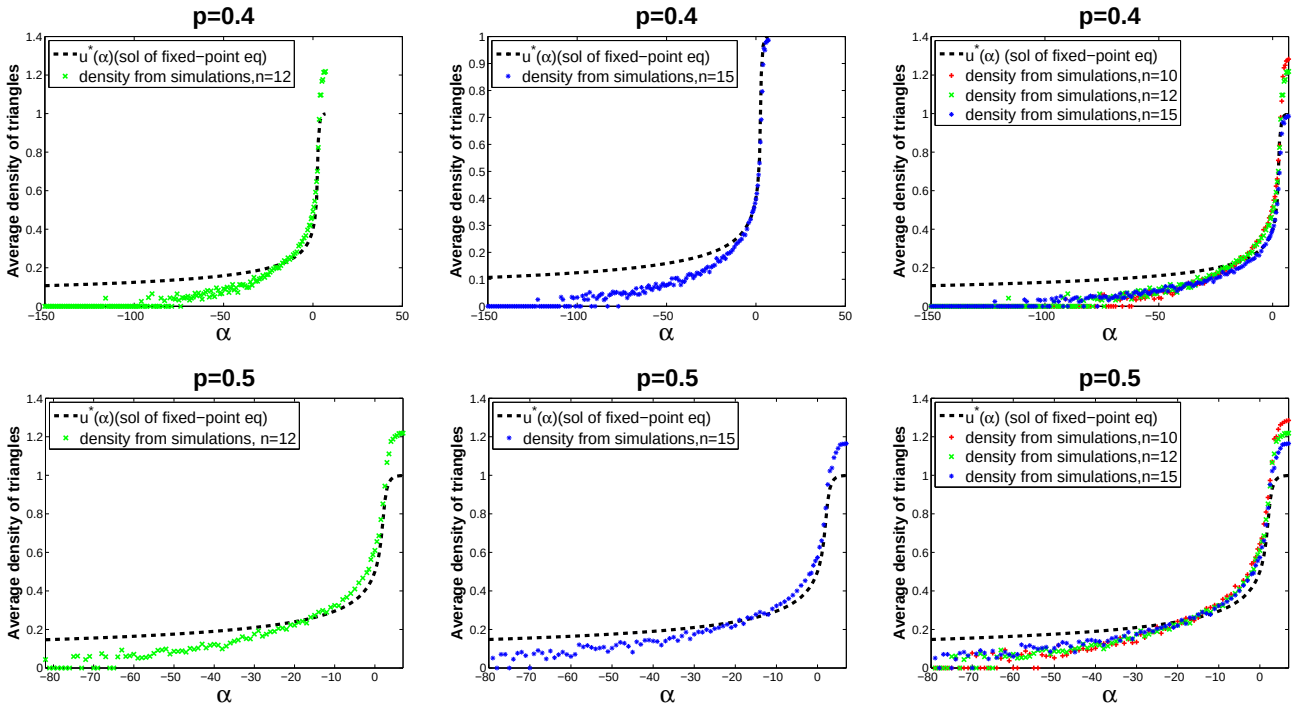


Figure 5.11: Average density of triangles per population when $p = 0.4$ (top) and $p = 0.5$ (bottom). The picture reports the statistic for $n = 12$ (left), $n = 15$ (middle) and $n = 10$, $n = 12$, $n = 15$ together (right)

concerns $p = 0.5$, the statistic is presented in the same way as $p = 0.4$ but the information on the null density, that we expect to observe around $\alpha = -67$, is less legible.

5.3 Optimization of the functional over the class of generalized bipartite graphons

We conclude our analysis via the third approach, i.e starting from the study of functional (5.1) and restricting the research of the solution on the set of graphons $\{g^{(a)}_{p_i,p_b}\}_{p_i,p_b,a}$, where $g^{(a)}_{p_i,p_b}$ is defined in (5.2). To be precise we tackle the problem

Problem 5.3. Solve

$$\sup_{\{g^{(a)}_{p_i,p_b}\}_{p_i,p_b,a}} \left[\frac{\alpha\, t(H_2, g^{(a)}_{p_i,p_b})}{3} - I_p(g^{(a)}_{p_i,p_b}) \right], \quad \alpha \leq -2 \tag{5.16}$$

which turns into

Problem 5.4. Solve

$$\sup_{p_i,p_b,a\in[0,1]} \left[\frac{\alpha\, t(H_2, g^{(a)}_{p_i,p_b})}{3} - I_p(g^{(a)}_{p_i,p_b}) \right], \quad \alpha \leq -2. \tag{5.17}$$

We know from Theorem 3.12 that there exists a sufficiently negative constant below which the solution of Problem 5.1 stops being the constant one (corresponding to the Erdös-Rényi graphon). In this last part of our investigation, by addressing Problem 5.4, we want to analyze the behavior of the solution when we absorb the constant graphon into a more general structure, with a further degree of freedom. Indeed, the subset of graphons $\mathcal{R} = \{g^{(a)}_{p_i,p_b}(x,y) |\, a,\, p_i,\, p_b \in [0,1]\} \subset \mathcal{W}$ contains both the constant graphon (obtained by setting $p_i = p_b$) and the generalized bipartite graphon (5.2). We further remark that any constant graphon is represented by infinitely many elements of $\mathcal{R}$, those with fixed $p_i = p_b$ and arbitrary value of a. In this setting, we aim at answering the following questions:

a) Even if we restrict the investigation on the specific class of graphons $\{g^{(a)}_{p_i,p_b}\}_{p_i,p_b,a}$, is it visible that the optimizer stops being the constant function below a critical value of α? If yes, what is the value of the threshold?

b) What happens to the structure of the graphon (namely at the parameters p_i, p_b, a) below such value of α?

In the following discussion, the (negative) threshold which differentiates among the homogeneous solution and the inhomogeneous one, will be called $\alpha_c(h_p)$. Unfortunately, we do not have any proof that the transition in $\mathcal{R}$ is the real transition occurring in the whole space $\mathcal{W}$ and that $\alpha_c(h_p)$ coincides with the negative constant $-C(h_p)$ mentioned by Theorem 3.12. However, some evidence for the equality of the two critical points will be obtained by intertwining the numerical results of the present section with the ones of Section 5.1.
In order to make $\frac{\alpha\, t(H_2, g^{(a)}_{p_i,p_b})}{3} - I_p(g^{(a)}_{p_i,p_b})$ explicit, it is convenient to write (5.2) as follows:

$$g^{(a)}_{p_i,p_b}(x,y) = p_i \left(\mathbb{1}_{[0,a]\times[0,a]} + \mathbb{1}_{[a,1]\times[a,1]} \right) + p_b \left(\mathbb{1}_{[0,a]\times[a,1]} + \mathbb{1}_{[a,1]\times[0,a]} \right) \tag{5.18}$$

Using expression (5.18) and developing computations we get

$$\begin{aligned} t(g^{(a)}_{p_i,p_b}, H_2) &= \int_0^1 \int_0^1 \int_0^1 g^{(a)}_{p_i,p_b}(x,y) g^{(a)}_{p_i,p_b}(y,z) g^{(a)}_{p_i,p_b}(z,x) dx dy dz \\ &= p_i^3[a^3 + (1-a)^3] + 3p_b^2 p_i a(1-a)^2 + 3p_b^2 p_i a^2(1-a) \end{aligned} \tag{5.19}$$

and, for what concerns the Bernoulli relative entropy,

$$I_p(g^{(a)}_{p_i,p_b}) = \int_0^1 \int_0^1 I_p(g^{(a)}_{p_i,p_b}(x,y))dxdy = 2a(1-a)I_p(p_b) + [a^2 + (1-a)^2]I_p(p_i). \tag{5.20}$$

Combining (5.19) and (5.20) the resulting function reads:

$$\begin{aligned}
&\alpha \frac{t(g^{(a)}_{p_i,p_b}, H_2)}{3} - I_p(g^{(a)}_{p_i,p_b}) = \\
&= \frac{\alpha}{3}\{p_i^3[a^3 + (1-a)^3] + 3p_b^2 p_i a(1-a)^2 + 3p_b^2 p_i a^2(1-a)\} - 2a(1-a)I_p(p_b) - [a^2 + (1-a)^2]I_p(p_i) \\
&=: \mathcal{F}_\alpha(p_i, p_b, a).
\end{aligned} \tag{5.21}$$

Remark 5.2. We observe that, since the Bernoulli relative entropy is defined by continuity over the interval $[0,1]$ by setting $I_p(0) = I_p(1) = 0$, the function $\mathcal{F}_\alpha(p_i, p_b, a)$ is continuous and the supremum over $[0,1]^3$ is always reached. In conclusion there is at least one triplet (p_i^*, p_b^*, a^*) of optimizers.

Remark 5.3. The function $\mathcal{F}_\alpha(p_i, p_b, a)$ enjoys, with respect to the variable a, a reflectional symmetry around $\frac{1}{2}$, being $\mathcal{F}_\alpha(p_i, p_b, a) = \mathcal{F}_\alpha(p_i, p_b, 1-a)$.

In conclusion, the Problem (5.4) turns into

$$\begin{aligned}
&\max_{p_i,p_b,a\in[0,1]} \mathcal{F}_\alpha(p_i, p_b, a) = \\
&\max_{p_i,p_b,a\in[0,1]} \frac{\alpha}{3}\{p_i^3[a^3 + (1-a)^3] + 3p_b^2 p_i a(1-a)^2 + 3p_b^2 p_i a^2(1-a)\} - 2a(1-a)I_p(p_b) - [a^2 + (1-a)^2]I_p(p_i).
\end{aligned} \tag{5.22}$$

In the next paragraph, we will address problem (5.22) looking for

$$\mathcal{F}_\alpha(p_i^*, p_b^*, a^*) = \max_{p_i,p_b,a\in[0,1]} \mathcal{F}_\alpha(p_i, p_b, a). \tag{5.23}$$

5.3.1 Gradient computation and solutions analysis

The computation of the gradient of the function (5.21) with respect to (p_i, p_b, a) leads to the following system, describing the condition $\nabla\mathcal{F}_\alpha(p_i, p_b, a) = 0$:

$$\begin{cases}
\alpha\left[p_i^2(3a(a-1)+1) + ap_b^2(1-a)\right] - (a^2 + (1-a)^2)\ln\left(\frac{p_i(1-p)}{p(1-p_i)}\right) = 0 & (a) \\
2a(1-a)\left[p_i p_b \alpha - \ln\left(\frac{p_b(1-p)}{p(1-p_b)}\right)\right] = 0 & (b) \\
(2a-1)\left[\alpha(p_i^3 - p_b^2 p_i) + 2(I_p(p_b) - I_p(p_i))\right] = 0 & (c)
\end{cases}, \quad \begin{array}{c} \alpha \leq -2. \\ p_i, p_b, a \in [0,1] \end{array} \tag{5.24}$$

Existence condition of the solutions

In order that system (5.24) admits a solution, observe from equation (a) that it must be $(a^2+(1-a)^2)\ln\left(\frac{p_i(1-p)}{p(1-p_i)}\right) \leq 0$, otherwise, since $\alpha \leq -2$, all the terms turn out to be strictly negative. This condition yields $\ln\left(\frac{p_i(1-p)}{p(1-p_i)}\right) \leq 0$ and consequently $\frac{p_i(1-p)}{p(1-p_i)} \leq 1$, which implies

$$p_i \leq p. \tag{5.25}$$

We now separately discuss the possible solutions of system (5.24) according to the choice of a.

Case $a = 1$

If we set $a = 1$, system (5.24) becomes:

$$\begin{cases} \alpha\, p_i^2 - \ln\left(\frac{p_i(1-p)}{p(1-p_i)}\right) = 0 & (a) \\ \alpha(p_i^3 - p_b^2 p_i) + 2(I_p(p_b) - I_p(p_i)) = 0 & (c) \end{cases}, \quad \alpha \leq -2. \tag{5.26}$$

Note that, equation (a) coincides with the fixed-point equation $\frac{e^{\alpha\, u^2+h_p}}{e^{\alpha\, u^2+h_p}+1} = u$, whose solution u^* is unique when α is negative (see Remark 3.8). Substituting $p_i = u^*$ in equation (c) we get:

$$\alpha({u^*}^3 - u^* p_b^2) + 2(I_p(p_b) - I_p(u^*)) = 0. \tag{5.27}$$

Let us define $\mathcal{C}_\alpha(p_b) := 2I_p(p_b) - \alpha\, u^* p_b^2 + \alpha\, {u^*}^3 - 2I_p(u^*)$: in order to find the possible zeros of the function, we observe first of all that it is convex, since it is a combination of $-\alpha\, u^* p_b^2$ (note that $-\alpha > 0$ since α is negative) and $2I_p(p_b)$ which are convex. Moreover, the minimum of $\mathcal{C}_\alpha(p_b)$ is attained at $p_b = u^*$ and this last conclusion is justified by

$$\frac{d\mathcal{C}_\alpha(p_b)}{dp_b} = 0 \Rightarrow \frac{e^{\alpha\, u^* p_b + h_p}}{e^{\alpha\, u^* p_b + h_p} + 1} - p_b = 0, \tag{5.28}$$

being $h_p = \ln\frac{p}{1-p}$. We already know that a possible solution of equation (5.28) is $p_b = u^*$: it is, as expected, the unique one, being the function $\mathcal{D}_\alpha(p_b) := \frac{e^{\alpha\, u^* p_b + h_p}}{e^{\alpha\, u^* p_b + h_p}+1} - p_b$ strictly decreasing and crossing the abscissas axis. Combining the fact that the minimum of $\mathcal{C}_\alpha(p_b)$ is attained at $p_b = u^*$ with the convexity of the function $\mathcal{C}_\alpha(p_b)$ and with the fact that u^* is such that $\mathcal{C}_\alpha(u^*) = 0$, we can conclude that u^* is the unique zero of the function. To summarize, we have found that the unique possible solution of system (5.26) when $a = 1$, is the pair $(p_i^*, p_b^*) = (u^*, u^*)$. Moreover, it is in order to stress that, when $p_i = p_b = u^*$, the generalized bipartite graphon (5.2) coincides with the Erdös-Rényi graphon identically equal to u^*. Finally note that $u^* < p$ so the solution is acceptable.

Case $a = 0$

As we have observed in Remark 5.3, the function shows a symmetry around $\frac{1}{2}$, hence the same conclusions drawn for the case $a = 1$ can be drawn for the case $a = 0$.

Remark 5.4. We observe that when $a = 1$ or $a = 0$ it is natural that the model falls back into the Erdös-Rényi one. In fact, in correspondence of these two extremal values of a, the graphon $g^{(a)}_{p_i,p_b}$ respectively turns into $g^{(1)}_{p_i}$ (when $a = 1$) or $g^{(0)}_{p_b}$ ($a = 0$) and its structure becomes no longer bipartite. Moreover observe that, taking for example $a = 1$, the function $\mathcal{F}_\alpha(p_i, p_b)$ reduces to $\mathcal{F}_\alpha(p_i) = \frac{\alpha}{3}p_i^3 - I_p(p_i)$ and the maximization problem $\max_{p_i\in[0,1]} \frac{\alpha}{3}p_i^3 - I_p(p_i)$ is the standard scalar problem one has to solve in replica symmetric for recovering the scaled cumulant generating function of the normalized number of triangles. The computation of the eigenvalues of the Hessian matrix of $\mathcal{F}_\alpha(p_i, p_b, a)$ allows to check that the pair (u^*, u^*) corresponds, as expected, to a maximum. This turns out to be true for any a and, in particular, for $a = 1$ and $a = 0$, which are the cases discussed up to now.

Case $a \neq \frac{1}{2}, 0, 1$

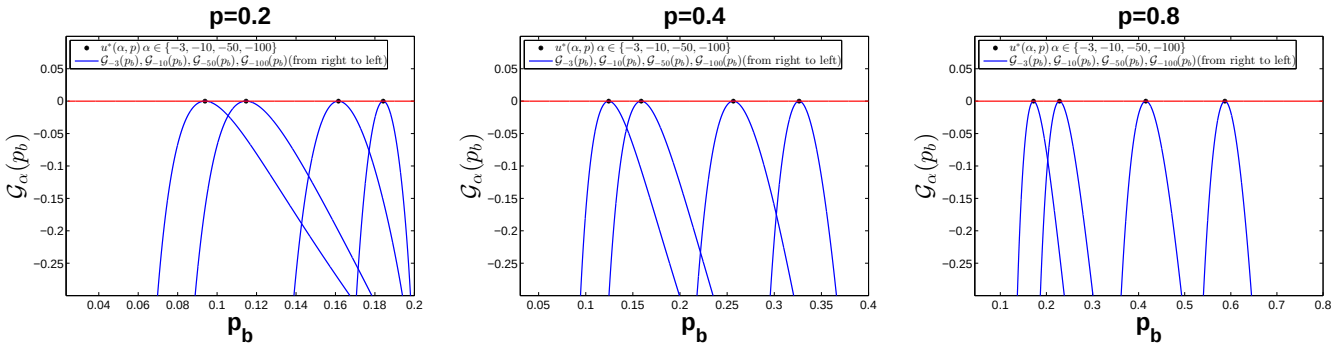


Figure 5.12: Plot of $\mathcal{G}_\alpha(p_b)$ for different values of p: each picture shows that the function is concave and its unique zero, which also coincides with its maximum, is attained at $p_b = u^*$ (black star). In each picture, for a fixed p, we overlap the plot of $\mathcal{G}_\alpha(p_b)$ for different values of α. The plot is defined over the interval of p_b expressed by condition (5.30).

In this case, system (5.24) becomes:

$$\begin{cases} \alpha\left[p_i^2(3a(a-1)+1) + ap_b^2(1-a)\right] - (a^2+(1-a)^2)\ln\left(\frac{p_i(1-p)}{p(1-p_i)}\right) = 0 & (a) \\ p_i p_b \alpha - \ln\left(\frac{p_b(1-p)}{p(1-p_b)}\right) = 0 & (b)\ , \quad \alpha \leq -2. \\ \alpha(p_i^3 - p_b^2 p_i) + 2(I_p(p_b) - I_p(p_i)) = 0 & (c) \end{cases} \tag{5.29}$$

Domain of $\mathbf{p_b}$: From equation (b) we can make $p_i = \frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)}$ explicit and check when the condition $0 \leq p_i \leq 1$ is satisfied. Inequality $\frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \geq 0$ implies $\ln \frac{p_b(1-p)}{p(1-p_b)} \leq 0$ (since α is negative) and hence $p_b \leq p$. Inequality $\frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \leq 1$ implies $p_b \geq \frac{1}{\alpha} \ln \frac{p_b(1-p)}{p(1-p_b)}$. In conclusion we get the bond

$$\frac{1}{\alpha} \ln \frac{p_b(1-p)}{p(1-p_b)} \leq p_b \leq p. \tag{5.30}$$

Substituting the expression of p_i in (c), we recover the following equation:

$$\alpha \left[\left(\frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \right)^3 - \frac{p_b}{\alpha} \ln \frac{p_b(1-p)}{p(1-p_b)} \right] + 2 \left[p_b \ln \frac{p_b}{p} + (1-p_b) \ln \frac{1-p_b}{1-p} - \frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \ln \left(\frac{1}{p\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \right) \right.$$
$$\left. - \left(1 - \frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \right) \ln \left(\frac{1}{1-p} \left[1 - \frac{1}{\alpha\, p_b} \ln \frac{p_b(1-p)}{p(1-p_b)} \right] \right) \right] = 0. \tag{5.31}$$

Let us call $\mathcal{G}_\alpha(p_b)$ the first member of (5.31). Note that, when $\frac{1}{\alpha} \ln \left(\frac{p_b(1-p)}{p(1-p_b)} \right) \leq p_b \leq p$ (as expressed in condition (5.30)) all the logarithms are well defined. Since equation $\mathcal{G}_\alpha(p_b) = 0$ is hard to solve, we have done a numerical investigation, which has led to conclude that $\mathcal{G}_\alpha(p_b)$ turns out to be a concave function whose unique zero coincides with $p_b = u^*$ as shown in Figure 5.12. In conclusion, the pair $(p_i, p_b) = (u^*, u^*)$ is the unique solution of system (5.29) that we have numerically found. Note that such pair of solutions, when substituted in equation (a), leads to $(2a^2 - 2a + 1) \left(\alpha\, u^{*2} - \ln \left(\frac{u^*(1-p)}{p(1-u^*)} \right) \right) = 0$ which is verified $\forall\, a$. In other words, we can again remark that when $p_i = p_b = u^*$, the generalized bipartite graphon $g^{(a)}_{p_i,p_b}$ assumes everywhere the constant value u^* and the parameter a loses its meaning.
Up to now, the optimization over the parameters of $g^{(a)}_{p_i,p_b}$ has uniquely led to the constant function identically equal to u^*, which corresponds to a maximum (see Remark 5.4). It still rests to analyze the case $a = \frac{1}{2}$.

Case $a = \frac{1}{2}$

We finally consider the case $a = \frac{1}{2}$: making this substitution, system (5.24) becomes:

$$\begin{cases} \alpha \left[\frac{p_i^2}{4} + \frac{p_b^2}{4} \right] - \frac{1}{2} \ln \left(\frac{p_i(1-p)}{p(1-p_i)} \right) = 0 & (a) \\ \frac{1}{2} \left(\alpha\, p_b p_i - \ln \left(\frac{p_b(1-p)}{p(1-p_b)} \right) \right) = 0 & (b) \end{cases}, \quad \alpha \leq -2, \tag{5.32}$$

whereas the function (5.21) takes the following form:

$$\mathcal{F}_\alpha \left(p_i, p_b, \frac{1}{2} \right) = \frac{\alpha}{3} \left(\frac{p_i^3}{4} + \frac{3}{4} p_b^2 p_i \right) - \frac{1}{2} I_p(p_b) - \frac{1}{2} I_p(p_i).$$

Description of numerical results

We have performed a numerical investigation of the solutions of (5.32) and, by checking the eigenvalues of the Hessian matrix, we have selected the pairs corresponding to a maximum.

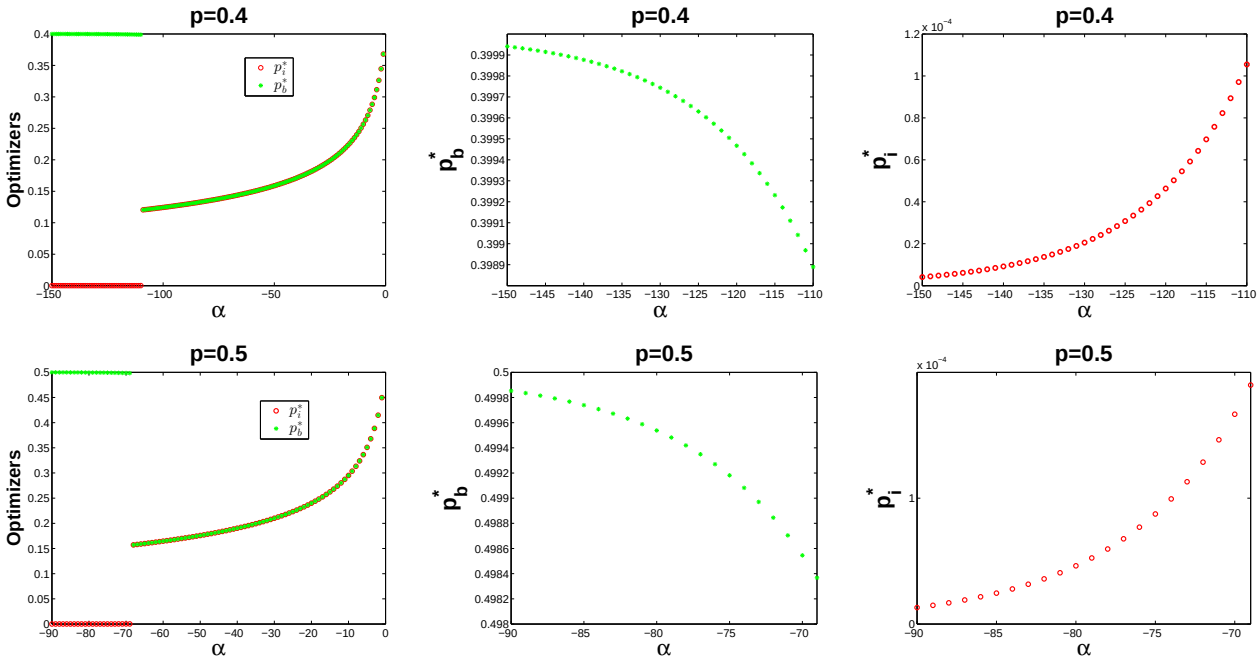


Figure 5.13: Upper panel: the left picture represents the optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ of $\mathcal{F}_\alpha\left(p_i, p_b, \frac{1}{2}\right)$ as functions of $\alpha \in [-150, 0]$, where α varies with unitary step and $p = 0.4$. The pictures in the middle and in the right columns show the behavior of $p_b^*(\alpha)$ and $p_i^*(\alpha)$ in the zoomed interval $[-150, -110]$. Lower panel: the same data of the upper panel for $p = 0.5$, $\alpha \in [-90, 0]$ and $\alpha \in [-90, -69]$ (zoomed pictures).

Among all the maxima found, we have then identified the best optimizer of $\mathcal{F}_\alpha\left(p_i, p_b, \frac{1}{2}\right)$. Figure 5.13 shows the pairs of optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ of $\mathcal{F}_\alpha(p_i, p_b, a)$ when α varies with unitary step. The upper panel shows the case $p = 0.4$ in the interval $[-150, 0]$ (left column) and in the zoomed one $[-150, -110]$ (central and right column); the lower panel shows the same data for $p = 0.5$ in the interval $[-90, 0]$ and $[-90, -69]$ (zoomed visual). The central and right columns of Figure 5.13 show that $p_i^*(\alpha)$ and $p_b^*(\alpha)$ are two slowly varying function that converge to the limits 0 and p whereas the left column of Figure 5.13, gives an evidence of the fact that, for fixed p, the function $\alpha \to (p_i^*(\alpha), p_b^*(\alpha))$ drastically changes its trend, respectively around $\alpha_c(h_p) = -110$ for $p = 0.4$ and $\alpha_c(h_p) = -69$ for $p = 0.5$. More precisely, crossing $\alpha_c(h_p)$ from above, the first two components of the optimizer $(p_i^*(\alpha), p_b^*(\alpha), \frac{1}{2})$ stop being equal to $u^*(\alpha)$ and $(p_i^*(\alpha), p_b^*(\alpha))$ jumps towards the point $(0, p)$. Further evidence of this transition is given in Figure 5.14 where the curves $\alpha \to \mathcal{F}_\alpha\left(p_i^*(\alpha), p_b^*(\alpha), \frac{1}{2}\right)$ and $\alpha \to \alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$ are displayed together with the asymptote of the scaled cumulant generating function of the normalized number of triangles, which is known thanks to Theorem 3.13 (red, continuous line). This figure shows that there is a value of α at which the two curves separate one from the other.

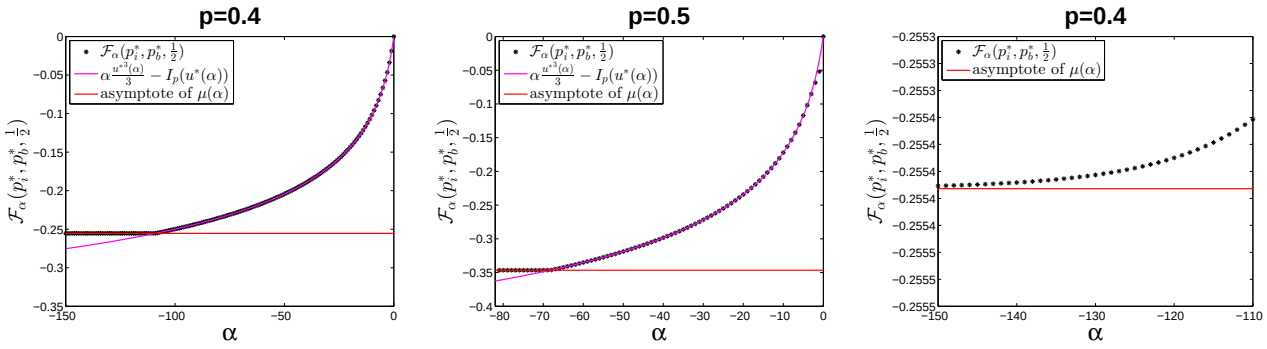


Figure 5.14: Left and central column: plot of $\mathcal{F}_\alpha(p_i^*(\alpha), p_b^*(\alpha), \frac{1}{2})$ as a function of α, together with the function $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$, respectively for $p = 0.4$ and $p = 0.5$ (pink continuous line). The horizontal red line represents the asymptotic value of function $\mu(\alpha)$ as $\alpha \to -\infty$. Right column: zoom on the interval $[-150, -110]$ for the case $p = 0.4$.

This is the critical value $\alpha_c(h_p)$ at which the curve $\alpha \to (p_i^*(\alpha), p_b^*(\alpha))$ makes the jump visible in Figure 5.13 (left column). Note that Figure 5.15 further clarifies this behavior: it shows how the maximum of the function $\mathcal{F}_\alpha(p_i, p_b, \frac{1}{2})$, passes from lying on the interior to being close to the edge, crossing $\alpha_c(h_p)$.
We report a summary of the described results below, in order to shed additional light to the observed behavior. We recall that we refer to the case $a = \frac{1}{2}$.

Critical threshold

The left and central column of Figure 5.14 give an evidence of the fact that there is a value of α at which $\mathcal{F}_\alpha\left(p_i^*, p_b^*, \frac{1}{2}\right)$ departs from $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$. Crossing this critical value, which we have called $\alpha_c(h_p)$, the sequence of optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ makes the jump visible both in Figure 5.13 (left column) and in Figure 5.15. The critical values $\alpha_c(h_p)$ turn out to be approximately -110 and -69, respectively for $p = 0.4$ and $p = 0.5$. Note that other thresholds are reported in Table 5.2 for different values of p whereas Figure 5.16 shows the fitting curve of the data.

Above the critical threshold

Above the critical threshold, $\mathcal{F}_\alpha\left(p_i^*, p_b^*, \frac{1}{2}\right)$ perfectly suits $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$, namely the pair of optimizers coincides with $p_i^*(\alpha) = p_b^*(\alpha) = u^*(\alpha)$. We recall that $\alpha\frac{u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$ is related to the scalar problem characterizing the replica symmetric regime.

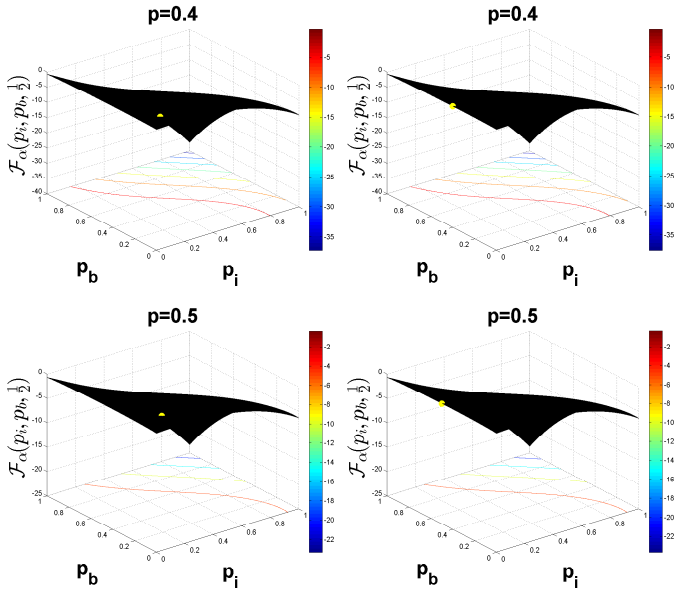


Figure 5.15: Top: jump of the pair of optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ of $\mathcal{F}_\alpha(p_i, p_b, \frac{1}{2})$ when α passes from -109 to -110 and $p = 0.4$ (yellow point). Bottom: jump of the pair of optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ when α passes from -68 to -69 and $p = 0.5$.

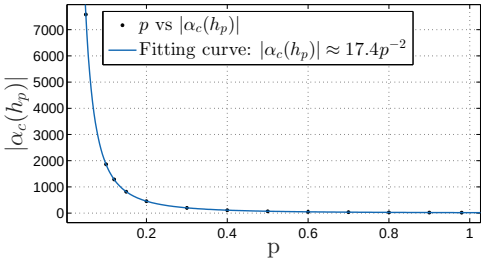


Figure 5.16: fitting curve of data present in Table 5.2.

$\mathbf{p}$	$\alpha_\mathbf{c}(\mathbf{h_p})$
0.05	-7580
0.1	-1861
0.12	-1286
0.15	-817
0.2	-455
0.3	-199
0.4	-110
0.5	-69
0.6	-47
0.7	-34
0.8	-25
0.9	-19
0.98	-17

Table 5.2: critical values of α as a function of p.

Below the critical threshold

Below the critical threshold, the pairs of optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ squeeze towards $(0, p)$ as long as α decreases, without never really reach neither 0 or p (note that this is coherent with Theorem 3.15 which states that the maximizer of Problem 5.1 must be bounded away from zero). As can be inferred from the picture, below $\alpha_c(h_p)$, the function $\mathcal{F}_\alpha(p_i^*, p_b^*, \frac{1}{2})$ seems

to settle in at the asymptote, $\frac{1}{2}\ln(1-p)$, of the scaled cumulant generating function of the normalized number of triangles. Note that $\mathcal{F}_\alpha(0,p,\frac{1}{2}) = -\frac{1}{2}I_p(0) = \frac{1}{2}\ln(1-p)$ which is the equation of the asymptote mentioned above.
However we remark that, below the critical value of α, the asymptote of the scaled cumulant generating function is never really reached by the function $\mathcal{F}_\alpha\left(p_i^*, p_b^*, \frac{1}{2}\right)$, since the optimal pair of parameters assumes values which are close, not equal to 0 and p. This can be also appreciated by looking at the right column of Figure 5.14, which shows a zoomed visual of $\mathcal{F}_\alpha(p_i^*, p_b^*, \frac{1}{2})$ on the interval $[-150, -110]$ together with the asymptote of $\mu(\alpha)$.

Conclusions

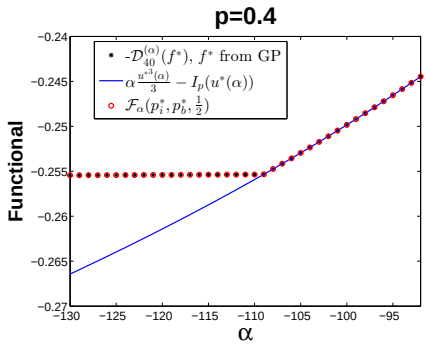


Figure 5.17: Plot of the curves resulting from the investigation with the GP method (black dotted curve) and from the numerical analysis of Problem 5.4 (red dotted curve) overlapped with $\frac{\alpha\, u^{*3}(\alpha)}{3} - I_p(u^*(\alpha))$ (continuous blue curve characterizing the replica symmetric regime).

In order to summarize, we provide an answer to the questions we asked at the beginning of this Subsection. More precisely, the investigation of Problem 5.4 has led to the following conclusions:

a) the optimization over the parameters (p_i, p_b, a) has brought to light that the case $a = \frac{1}{2}$ seems to be the interesting one, if we want to observe a pair of optimizers $(p_i^*(\alpha), p_b^*(\alpha))$ different from the constant solution $u^*(\alpha)$. Setting $a = \frac{1}{2}$, we have observed that, above a critical and negative value of α (which depends on p), the optimal triplet of parameters which maximizes $\mathcal{F}_\alpha(p_i, p_b, a)$ coincides with $\left(u^*(\alpha), u^*(\alpha), \frac{1}{2}\right)$: in other words we recover the constant solution which characterizes the replica symmetric regime. Speaking of critical values, it is possible to give a numerical estimation of them (which depends on p) and the values that we have found are collected in Table 5.2. Note that the critical values of α found in this subsection, perfectly agree with the ones revealed by the Projected Gradient method introduced in Subsection 5.1.1, which had no initial bias concerning the structure of the optimizer (see Figure 5.8, cases $p = 0.4$ and $p = 0.5$). As a further

evidence, we show in Figure 5.17 the perfect match between the numerical curve returned by the GP method ($-\mathcal{D}_{40}^{(\alpha)}(f^*)$) and the one obtained from the analysis of the variational problem over the subclass of generalized bipartite graphons ($\mathcal{F}_\alpha(p_i^*(\alpha), p_b^*(\alpha), \frac{1}{2})$).

b) Below the critical threshold, the maximizers differentiate and $p_i^*(\alpha)$ and $p_b^*(\alpha)$ respectively become closer and closer to 0 and p, as long as α decreases.

As a general consideration, we stress that, for each p, the constant $\alpha_c(h_p)$ that we have found through simulations is for sure less or equal to the critical one, $-C(h_p)$, mentioned by Theorem 3.12: however, we strongly believe that they should coincide, since the simplest way for the constant graphon to break its structure is to split in a bipartite one.

Conclusions

Rare events are widely studied in several scientific fields, since the most interesting aspects of natural phenomena are not the typical ones but those which are in some sense more difficult to observe, predict and control. The theory of large deviations has developed mathematical tools for computing the probability that a certain event deflects from its average behavior. From the practical point of view, the derivation of the rate function is difficult to perform and it could be helpful to resort to simulations. However, the approach of simulating the true dynamics is time consuming, since by definition, rare events are difficult to observe. Helpful numerical tools for investigating rare events are Monte Carlo methods, which are tailored for weighting the probabilities in such a way that the they become less infrequent. This thesis inserts in this context by presenting a Monte Carlo method, called Cloning algorithm, for the approximation of the so-called scaled cumulant generating function: it is strictly related to the theory of large deviations since, when the Gärtner-Ellis theorem holds, it coincides with the Legendre transform of the rate function. Our contribution consists in devising an extended version of the Cloning algorithm, which is specifically tailored for working on random graphs, aiming at approximating the scaled cumulant generating function of some additive observable which we measure on them. In this thesis, the method is applied to the dense Erdös-Rényi model, considering as observable the number of triangles. A very complex large deviations theory is hidden behind the simplicity of this problem and some questions are still open. Keeping the focus on the dense Erdös-Rényi model, this thesis also aims at investigating the unresolved region concerning the scaled cumulant generating function of such observable.

First, we introduced the standard version of the Cloning algorithm, providing a formal analysis of its formulation. In particular, we highlighted its strict connection with branching processes, which can be intuitively forecast from the fact that the method follows a population dynamics approach in which each individual is reproduced or killed according to some rate. We tested the scheme on some simple models, the random walk and the Ehrenfest diffusion, where the dynamics was described by birth and death chains on a finite space of configurations, for reproducing the scaled cumulant generating function of an additive observable. These simple applications brought to light the effectiveness of the method, since simulations return the objective function with a very good accuracy.

Second, we devised an extended version of Cloning algorithm designed for working on the

random graphs setting. This new approach consists in implementing a dynamics on growing-size spaces of graphs where, starting from a chosen size, typically small, a vertex is added and then connected to the previous ones. On the grounds of this process, we specialized to the dense Erdös-Rényi model connecting the edges at each step with independence according to probability p. The main advantage of this strategy consists in the fact that it is possible to keep trace in an incremental way of the observable one wants to measure along the process. First, we tested this new version of the Cloning algorithm on the edge observable, which is the simplest one, since the independence of the random variables involved allows different simplifications. For example, it is possible to implement the dynamics without involving the adjacency matrices in the code, since all the information necessary for running the method can be transferred on integers, thus reducing the computational cost. Simulations results are favored by this strategy, which allows to run the method along several iterations, and they return an optimal approximation of the objective function, which, in this case, is simple to derive. For what concerns the triangle observable, things are more complex: due to the lack of independence it is necessary to involve adjacency matrices in the process; however, the most complex part consists in computing the reproduction rate of each member of the population. This number, coincides with a partition function which, in the triangles case, sums over an exponential number of terms at each step (we recall that matrices grow their size of one unit per iteration). A significant simplification would be obtained by finding a smart way for counting all the matrices related to the same increment of triangles, in order to perform the sum over less terms. The combinatorics underlying this goal is however very complex. As a first step, we run simulations on the so-called replica symmetric region where the analytical expression of the scaled cumulant generating function is known and can be used as a comparison term. Despite the aforementioned difficulty, simulations, possible for small sizes of the graphs, show interesting results and reveal the progressive convergence of the method to the limiting curve. Keeping the setting on the replica symmetric region, we finally provided a heuristic argument on the convergence, which turns out to be exact when the number of nodes of the graphs grows to infinity.

We finally moved the investigation on the replica breaking region, which is unresolved, since the expression of the scaled cumulant generating function is not known here: this regime is revealed when α is negative, where α is the parameter which tunes the triangles density. For both replica symmetric and replica breaking regions, the analytical computation of the scaled cumulant generating function coincides with solving a variational problem over a certain space of measurable, symmetric functions called graphons. The only information we have on replica breaking regime, is the fact that the solution of the problem stops being the constant one (which, instead, characterizes the replica symmetric phase) and changes its structure. The breaking point must correspond to a sufficiently negative value of α, below which starts the unresolved window. Moreover we know that, in the negative limit of α, the unique optimizer is the so-called bipartite graphon with parameters 0 and p. Leaning on the variational formulation, we proposed three strategies for investigating such controversial region. In the first approach, we

studied the problem by means of a well-known optimization method, the Gradient Projected one: the downside of this approach consists in the fact that the output of the algorithm strongly depends on the initial condition and, in case of non-convex functions, it does not guarantee the convergence at a minimum. It is however possible to show that, even in lack of convexity, the method, applied to the discretized optimization problem, converges to a stationary point. Despite the aforementioned criticalities, the Projected Gradient algorithm returns promising results. To recap we can say that:

a) the method brings to light the existence of different stationary points, which, however, can be grouped in two classes: the constant solution, which is always a stationary point for the variational problem, independently from the choice of the regime (replica symmetric or replica breaking one) and a collection of functions with a chessboard structure assuming values $p_i \approx 0$ and $p_b \approx p$ which approximate the bipartite graphon with parameters 0 and p.

b) from numerical results is clearly appreciable that there exists a critical value of α, crossing which the optimizer switches from the constant function to a chessboard one, assuming values $p_i \approx 0$ and $p_b \approx p$.

As a second strategy of investigation, we run the extended version of the Cloning algorithm: the method is not meant for coping with large absolute values of α, since the reproduction rates of the population are exponential in α and then they could explode or vanish very fast. For this reason, simulations do not reveal significant features of the objective curve, however something can be worked out by monitoring both the average density of triangles and of odd cycles per population. More precisely:

a) for a fixed p, we observed that the average density of triangles followed the typical trend of replica symmetric regime, up to a certain negative constant and then it drifted towards zero. Such behavior could let forecast the passage between a homogeneous structure to another one, where triangles are progressively suppressed from the graph. It is in order to point out that the algorithm deals with graphs and not graphons, however, if we expect to go towards a bipartite structure, the progressive disappearance of odd cycles, and hence of triangles, is exactly what we should see. The average density of triangles starts approaching zero from a value of α which is comparable with the critical one revealed by the Gradient Projection method and corresponding to the change of optimizer of the variational problem. In such sense, the response of this investigation is coherent with the one returned by the Gradient Projection method.

b) An analogous consideration can be drawn for the average density of odd cycles: simulations show a decreasing trend of such quantity as long as α decreases.

As a final strategy, we performed an analysis of the variational problem over a specific class of generalized bipartite graphons with generic parameters p_i, p_b, a. This choice was motivated by

the guess that the natural way for the constant function for changing its structure was to split in a bipartite one. Optimizing over the parameters we observed that the only way for finding stationary points different from the constant solution (with $p_i = p_b$), was to settle $a = \frac{1}{2}$. This specific choice again revealed that, crossing a critical value of α, the optimizer of the variational problem, which is the homogeneous function up to such threshold, breaks its structure in a generalized bipartite one with parameters p_i and p_b respectively close to 0 and p.

Two conclusive remarks are in order:

1) all the critical values of α revealed, for a fixed p, by the Gradient Projection method and by the third analysis of the functional coincide (in both numerical approaches α varies with unitary step);

2) below the critical threshold, the values of $p_i(\alpha)$ and $p_b(\alpha)$ returned by the Gradient Projection method and by the analysis of the functional, are close, not equal to 0 and p. Moreover $p_i(\alpha)$ and $p_b(\alpha)$ progressively approach these two benchmarks as long as α decreases. In particular, for finite α, the fact that p_i does not vanish is coherent with what we know from the theory.

Our conjecture

As a final conclusion emerging from the investigation of the replica breaking regime, we conjecture the following behavior of the variational problem returning the scaled cumulant generating function of the triangle observable: for a fixed p, below a negative critical threshold $\alpha_c(h_p)$, the optimizer switches from the constant function to the generalized bipartite graphon with parameter $p_i(\alpha) \approx 0$, $p_b(\alpha) \approx p$ and $a = \frac{1}{2}$. Moreover, $p_i(\alpha)$ and $p_b(\alpha)$ respectively converge to 0 and p in the negative limit of α, thus recovering the expected result. This conjecture is portrayed in Figure 5.18.

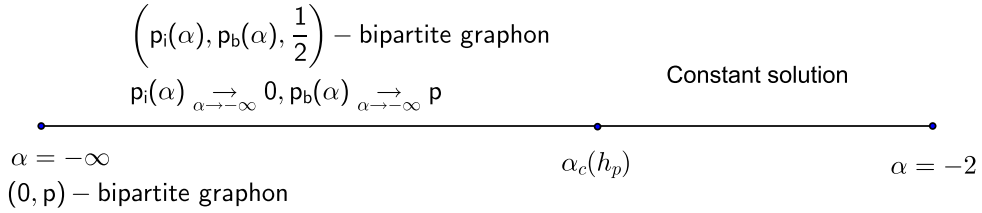


Figure 5.18: Representation of the conjectured solution of Problem 5.1.

Future work

Future work will concern two main aspects. First of all, the research of a more efficient implementation strategy aimed at reducing the computational cost of the extended Cloning algorithm. This approach can be twofold: on one hand we can look for a smart way of addressing the combinatorics underlying the computation of the cloning rate, thus performing the related sum in a more efficient way. On the other hand, we can go on the direction of making a change on the dynamics, for example allowing only one connection per iteration and consequently updating the variation of the increment. This approach requires more steps, more precisely $\binom{n}{2}$ instead of n, if n is the final size of the graph we want to reproduce. The hope is that this strategy could bring an improvement, introducing simplifications on the computation of the reproduction rates.
Finally, as a second goal, we aim at proving the thesis conjectured above, concerning the expression of the scaled cumulant generating function in the unresolved region.

Appendix A

Matlab code for the extended Cloning algorithm

```
1 tic;
2 clear all;
3 close all;
4 rand('seed',123);
5 currentFolder=pwd;
6
7 %Main
8 %Variables declaration
9 %Main variables:
10 %dim_in  (integer) - size of the graphs in the initial population.
11 %M_2     (integer) - initial size of the population of graphs.
12 %N       (integer) - number of iterations.
13 %alfa    (vector)  - vector of parameters of the s.c. generating fct
       .
14 %p_graph (integer) - probability of the dense ER graph.
15 %mu      (vector)  - ratio.
16 %muAlfa  (vector)  - s.c. generating fct for each alfa.
17 %rates   (vector)  - reproduction rates.
18 %seq     (array)   - random starting population of M_2 01-matrices.
19 %num     (vector)  - integers numbers from 1 to the sum of rates.
20 %population_extr (vector) - extracted population.
21 %sum_r (integer)          - sum of all rates.
22 %rates_cum (integer)      - cumulative sum of the vector rates.
23 %disposition (array)      - possible evolutions of a graph.
24 %prob_cumsum (array)      - cumulative sum of the transition prob.
```

```
25 %rate_int (int)            - reproduction integer rate of the graph.
26
27 %Supporting variables:
28 %S,min,p,element,diff      -(integer variables)
29 %Graph,Graph_base          -(matrices)
30
31 dim_in= 2;
32 M_2=1000;
33 N=10;
34 p_graph=0.4;
35 alfa=-3:0.5:7;
36
37 %Initialization of vectors and arrays
38 muAlfa= zeros(1,length(alfa));
39 num=zeros(1,M_2);
40 rates=zeros(1,M_2);
41 seq= cell(1,M_2);
42 popolazione_extr= cell(1,M_2);
43
44 %Intialization of the population of graphs
45 Graph_base=zeros(dim_in,dim_in);
46 for l=1:M_2
47     element=randsample([0 1],1,true,[(1-p_graph) p_graph]);
48     Graph=Graph_base;
49     Graph(1,2)=element;
50     seq{1,l}= Graph;
51 end
52
53 for a=1:length(alfa)
54     disp(a);
55     %initialize population to seq.
56     population=seq;
57     mu= zeros(1,N);
58     for n=1:N
59         S=M_2;
60         %Evolution & Computation of reproduction rates
61         for l=1:length(population)
62             [disposition, prob_cumsum, rate_int]= Dispo(population
                   {1,l}, p_graph, alfa(a), N+dim_in);
63             rates(l)=rate_int;
```

```
64              %extract a transition proabability
65              p=rand();
66              diff= p-prob_cumsum;
67              min= find(diff<=0,1);
68              evolution_pop=disposition{1,min};
69              %add the column associated to the evolution
70              population{1,l}=[population{1,l} evolution_pop'; zeros
                    (1,length(evolution_pop)+1)];
71          end
72
73          sum_r=sum(rates); %new size of the population
74
75          if(sum_r==0)
76              fprintf('population size null: exit at time t=%d\n', n
                    );
77              disp(alfa(a));
78              exit;
79          end
80
81          %Extract M_2 numbers between 1 and sum_r
82          for kk=1:M_2
83              num(kk)=floor(1+rand()*(sum_r-1));
84          end
85          rates_cum = cumsum(rates);
86
87          %rates_cum allows to know in which partition the extracted
                numbers fall
88          for ii = 1:M_2
89            ind_min= find(num(ii)-rates_cum <= 0,1);
90            population_extr{1,ii}= population{1,ind_min};
91          end
92
93          %if all clones have been killed, exit
94          if(isempty(population_extr))
95              fprintf('population size null: exit at time t=%d\n', n
                    );
96              disp(alfa(a));
97              exit;
98          end
99
```

```
100         %update the population
101         population=population_extr;
102
103         %partial values of the scaled cumulant generating function
104         mu(n)= S/sum(rates);
105         S=sum(rates);
106     end
107
108     %scaled cumulant generating function
109     muAlfa(a)= -(1/nchoosek((N+dim_in),2))*sum(log(mu));
110     %clear variables before the cycle starts again
111     clear mu;
112     clear population_extr;
113     clear population;
114 end
115 %plot
116 plot(alfa,muAlfa,'*g');
117 xlabel('\alpha','FontSize',20,'FontWeight','bold','Color','r');
118 ylabel('\mu(\alpha)','FontSize',20,'FontWeight','bold','Color','r');
119 title(strcat('p=', num2str(p_graph)),'FontSize',20);
120 legend({strcat('Cl N=',num2str(N+dim_in))},'FontSize',20, 'Location'
    , 'northwest');
121 %
    ==========================================================================

122 % End of Main.m file
123 %
    ==========================================================================
```

```
1 function [disposition, prob_cumsum, rate_int]= Dispo(graph,p_graph,
    alfa, N)
2 % Dispo
3 %   This function computes:
4 %   1) the array of possible dispositions associated to a graph of
    the
5 %       graph
6 %   2) the cumulative sum vector of the transition probabilities
7 %   3) the integer rate associated to the graph
8 %
```

```
%
% SYNOPSIS
%    [disposition , prob_cumsum, rate_int] = Dispo(graph,p,alfa , N)
%
% INPUT
% graph      (upper triangular matrix) - graph of the population
% p_graph    (double)                  - probability of the ER graph
% alfa       (double)                  - parameter of the scaled c.g.
    fct
% N          (integer)                 - final size of the graph
%
% OUTPUT [disposition , prob_cumsum, rate_int]
% disposition (array) - array of dispositions(possible transitions
    of graph)
% prob_cumsum (vector)- cumulative vector of the transition
    probabilities
% rate_int (integer)  - integer rate associated to the graph

n=size(graph,1);
vector=ones(1,n);
disposition=cell(1,2^n);
vector_p_tran= zeros(1,2^n);

vector_p_tran(1)=(p_graph^n)*exp(alfa*sum(sum(graph))/N);
cont=1;
vector_supp=vector;
disposition{1,cont}=vector;

for i=1:n
    C=nchoosek(1:n,i);
    num_righe=size(C,1);
for j=1:num_righe
    %create the disposition
    cont=cont+1;
    indexes=C(j,:);
    vector(indexes)=0;
    disposition{1,cont}=vector;
    num_ones=sum(vector);
    num_zeros=n-num_ones;
    %increment of triangles
```

```
46      increment= vector*(graph*vector');
47      vector_p_tran(cont)=(p_graph^(num_ones))*exp(alfa*increment/N)
            *(1-p_graph)^(num_zeros);
48      vector=vector_supp;
49 end
50 clear C;
51 end
52 rate_non_int=sum(vector_p_tran);
53 prob_cumsum= cumsum(vector_p_tran/rate_non_int);
54 prand= rand();
55 rate_int= floor(rate_non_int+ prand);
56 end
57 %
```

```
58 % End of Dispo.m file
59 %
```

Bibliography

[1] S. Asmussen and C. Klüppelberg. Large deviations results for subexponential tails, with applications to insurance risk. *Stochastic processes and their applications*, 64(1):103–125, 1996.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[3] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.

[4] D. P. Bertsekas. *Nonlinear programming*. Athena scientific Belmont, 1999.

[5] S. Bhamidi, G. Bresler, and A. Sly. Mixing time of exponential random graphs. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 803–812. IEEE, 2008.

[6] K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, and S. McKay. Monte carlo simulation in statistical physics. *Computers in Physics*, 7(2):156–157, 1993.

[7] B. Bollobás. Random graphs. In *Modern graph theory*, pages 215–252. Springer, 1998.

[8] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

[9] B. e. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[10] S. Bonettini, I. Loris, F. Porta, M. Prato, and S. Rebegoldi. On the convergence of a linesearch based proximal-gradient method for nonconvex optimization. *Inverse Problems*, 33(5):055005, 2017.

[11] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Counting graph homomorphisms. In *Topics in discrete mathematics*, pages 315–371. Springer, 2006.

[12] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

[13] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs ii. multiway cuts and statistical physics. *Annals of Mathematics*, pages 151–219, 2012.

[14] S. Chatterjee. An introduction to large deviations for random graphs. *Bulletin of the American Mathematical Society*, 53(4):617–642, 2016.

[15] S. Chatterjee. *Large Deviations for Random Graphs: École D'Été de Probabilités de Saint-Flour XLV-2015*, volume 2197. Springer, 2017.

[16] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299:396–450, 2016.

[17] S. Chatterjee, P. S. Dey, et al. Applications of stein's method for concentration inequalities. *The Annals of Probability*, 38(6):2443–2485, 2010.

[18] S. Chatterjee, P. Diaconis, et al. Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461, 2013.

[19] S. Chatterjee and S. S. Varadhan. The large deviation principle for the erdős-rényi random graph. *European Journal of Combinatorics*, 32(7):1000–1017, 2011.

[20] G. Dematteis, T. Grafke, and E. Vanden-Eijnden. Rogue waves and large deviations in deep sea. *Proceedings of the National Academy of Sciences*, page 201710670, 2018.

[21] A. Dembo. Zeitouni, 0.(1998). large deviations techniques and applications. *Applications of Mathematics (New York)*, 38, 1994.

[22] F. Den Hollander. *Large deviations*, volume 14. American Mathematical Soc., 2008.

[23] P. Ehrenfest and T. Ehrenfest-Afanassjewa. *Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem*. Hirzel, 1907.

[24] R. S. Ellis. *Entropy, large deviations, and statistical mechanics*. Springer, 2007.

[25] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[26] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. Probability of second law violations in shearing steady states. *Physical review letters*, 71(15):2401, 1993.

[27] K. S. T. H. A. First. Course in stochastic processes. 2nd, 1975.

[28] C. Giardinà, J. Kurchan, V. Lecomte, and J. Tailleur. Simulating rare events in dynamical processes. *Journal of statistical physics*, 145(4):787–811, 2011.

[29] C. Giardinà, J. Kurchan, and L. Peliti. Direct evaluation of large-deviation functions. *Physical review letters*, 96(12):120603, 2006.

[30] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[31] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media, 2013.

[32] J. Hammersley. *Monte carlo methods*. Springer Science & Business Media, 2013.

[33] A. Iusem. On the convergence properties of the projected gradient method for convex optimization. *Computational & Applied Mathematics*, 22(1):37–52, 2003.

[34] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen. On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8(1-2):131–159, 1999.

[35] K. Kurdyka. On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–784. Chartres: L'Institut, 1950-, 1998.

[36] J. L. Lebowitz and H. Spohn. A gallavotti–cohen-type symmetry in the large deviation functional for stochastic dynamics. *Journal of Statistical Physics*, 95(1-2):333–365, 1999.

[37] S. Łojasiewicz. Sur la géométrie semi-et sous-analytique. *Ann. Inst. Fourier*, 43(5):1575–1595, 1993.

[38] L. Lovász. Large networks and graph limits, volume 60 of american mathematical society colloquium publications. *American Mathematical Society, Providence, RI*, 22, 2012.

[39] L. Lovász and B. Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.

[40] L. Lovász and B. Szegedy. Szemerédi's lemma for the analyst. *GAFA Geometric And Functional Analysis*, 17(1):252–270, 2007.

[41] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. *Random Structures & Algorithms*, 47(1):109–146, 2015.

[42] C. Matthews, J. Weare, A. Kravtsov, and E. Jennings. Umbrella sampling: a powerful method to sample tails of distributions. *arXiv preprint arXiv:1712.05024*, 2017.

[43] M. Newman. *Networks*. Oxford university press, 2018.

[44] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[45] J. Park and M. Newman. Solution for the properties of a clustered network. *Physical Review E*, 72(2):026136, 2005.

[46] M. Prato, S. Bonettini, I. Loris, F. Porta, and S. Rebegoldi. On the constrained minimization of smooth kurdyka—łojasiewicz functions with the scaled gradient projection method. In *Journal of Physics: Conference Series*, volume 756, page 012001. IOP Publishing, 2016.

[47] C. Radin and M. Yin. Phase transitions in exponential random graphs. *The Annals of Applied Probability*, pages 2458–2471, 2013.

[48] R. Rockafellar. Convex analysis, princeton university press, princeton, 1970. *MATH Google Scholar*, 1997.

[49] E. Seneta. *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.

[50] A. Shiryaev. Probability. 1996. *Graduate Texts in Mathematics*, 1996.

[51] J. v. Tiel. *Convex analysis*. John Wiley, 1984.

[52] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[53] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[54] S. S. Varadhan. Asymptotic probabilities and differential equations. *Communications on Pure and Applied Mathematics*, 19(3):261–286, 1966.

[55] Y. Zhao. On the lower tail variational problem for random graphs. *Combinatorics, Probability and Computing*, 26(2):301–320, 2017.