UNIVERSIDAD DE LA REPUBLICA URUGUAY

PEDECIBA

Large Deviations for Exploration Processes on Random Graphs

Valeria Goicoechea Jackson

Programa de Posgrado en Matemática
PEDECIBA, Facultad de Ciencias
Universidad de la República

Montevideo – Uruguay
Mayo de 2022

UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

PEDECIBA

Large Deviations for Exploration Processes on Random Graphs

Valeria Goicoechea Jackson

Tesis de Doctorado presentada al Programa de Posgrado en Matemática, PEDECIBA de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Doctor en Matemática.

Director:
Dra. Prof. Paola Bermolen Romeo

Codirector:
Dr. Prof. Matthieu Jonckheere

Montevideo – Uruguay
Mayo de 2022

Goicoechea Jackson, Valeria

Large Deviations for Exploration Processes on Random Graphs / Valeria Goicoechea Jackson. - Montevideo: Universidad de la República, PEDECIBA, Facultad de Ciencias, 2022.

VIII, 106 p. 29, 7cm.

Director:

Paola Bermolen Romeo

Codirector:

Matthieu Jonckheere

Tesis de Doctorado – Universidad de la República, Programa en Matemática, 2022.

Referencias bibliográficas: p. 101 – 106.

1. Grandes Desvíos, 2. Algoritmos Greedy. I. Bermolen Romeo, Paola, *et al.* II. Universidad de la República, Programa de Posgrado en Matemática. III. Título.

INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

Dra. Prof. Paola Bermolen

Dra. Prof. Karine Bertin

Dr. Prof. Pablo Groisman

Dr. Prof. José R. León

Dr. Prof. Alain Roualt

Montevideo – Uruguay
Mayo de 2022

RESUMEN

En esta tesis nos enfocamos en el estudio de los grandes desvíos (GD) para sucesiones de procesos de Markov que describen el comportamiento de ciertos algoritmos de exploración *greedy* sobre grafos aleatorios con el fin de construir conjuntos independientes en esos grafos. Nos centramos en cuatro aspectos de los GD para estos procesos:

- Probar los GD para las trayectorias de dichos procesos de Markov,
- Deducir el límite fluido a partir de la función de tasa del GD,
- Encontrar la trayectoria que minimiza la función de tasa sobre un conjunto de trayectorias,
- Concluir resultados de GD para el tamaño del conjunto independiente construido mediante el algoritmo *greedy*.

Para demostrar el PGD (Principio de Grandes Desvíos) para las sucesiones de procesos de interés, utilizamos la estrategia propuesta por [Feng and Kurtz, 2006] para el estudio de GD de procesos estocásticos, la que se basa en la convergencia de semigrupos no lineales asociados a dichos procesos.

La tesis se desarrolla de la siguiente manera. Comenzamos presentando brevemente el trabajo de [Feng and Kurtz, 2006] en el contexto de procesos de Markov sobre espacios de estados compactos. En el Capítulo 3, analizamos los cuatro aspectos de los GD mencionados antes para una sucesión de procesos de Markov relacionados a un algoritmo *greedy* definido sobre un grafo de Erdös-Rényi dado con el objetivo de construir un conjunto independiente, cuando el tamaño del grafo tiene a infinito. En el Capítulo 4, repetimos este análisis para un algoritmo en el que simultáneamente se construye un grafo d-regular y un conjunto independiente en ese grafo. Finalmente, en el Capítulo 5 extendemos los resultados del capítulo anterior para grafos aleatorios uniformes más generales.

Además de presentar resultados originales sobre los GD para los procesos de interés, creemos que el aporte de este trabajo consiste en mostrar de forma entendible la herramienta poderosa propuesta en el trabajo de [Feng and Kurtz, 2006] para el estudio de GD de procesos, con posibles aplicaciones a diversas áreas.

Palabras claves:
Grandes Desvíos, Algoritmos Greedy.

ABSTRACT

In this thesis, we focus on the study of large deviations (LD) for sequences of Markov processes that describe the behaviour of greedy exploration algorithms on random graphs in order to build independent sets. We focus on four aspects of the LD for these processes:

- Proving the path-state LD for these Markov processes,
- Deducing the fluid limit from the LD rate function,
- Finding the trajectory that minimises the LD rate function over a set of trajectories,
- Concluding LD results for the size of the independent set constructed by the greedy algorithm.

To prove the LDP (*Large Deviation Principle*) for the sequences of processes of interest, we use the general strategy to study LDP of processes proposed by [Feng and Kurtz, 2006], which is based on the convergence of non-linear semigroups.

The thesis is developed as follows. We begin by briefly introducing the work of [Feng and Kurtz, 2006] in the context of Markov processes on compact state spaces. Then, in Chapter 3, we analyse the four aspects of LD mentioned above for a sequence of Markov processes related to a greedy algorithm defined on a given Erdös-Rényi graph to construct an independent set when the size of the graph goes to infinity. In Chapter 4, we repeat this analysis for an algorithm that simultaneously constructs a d-regular graph and an independent set on that graph. Finally, in Chapter 5, we extend the results of the previous chapter to more general uniform random graphs.

In addition to presenting original results on large deviations for the processes of interest, we believe that the contribution of this thesis is to present in a comprehensive way the powerful tool proposed in the work of [Feng and Kurtz, 2006] for the study of large deviations of processes, with possible applications to many areas.

Keywords:
Large Deviations, Greedy Algorithms.

Contents

1 Introduction **1**

1.1 What are the large deviations? . . . 1

1.2 Random graphs . . . 2

1.3 Random graphs exploration algorithms analysed in this thesis . . . 4

1.4 Why study large deviations for these random graphs exploration algorithms? . . 7

1.5 Objective and main results obtained . . . 8

1.6 Thesis structure . . . 12

2 Large deviations for Markov processes **14**

2.1 Generalities about large deviations . . . 15

2.2 Large deviations for Markov processes - the work of Feng and Kurtz . . . 20

2.3 Theory of viscosity solutions for Hamilton-Jacobi equations . . . 31

2.4 Rate function optimization . . . 36

2.5 Fluid limit . . . 37

2.6 The theory of Freidlin and Wentzell in the context of the work of Feng and Kurtz 39

3 Large deviations for the greedy exploration algorithm over Erdös-Rényi graphs **46**

3.1 Introduction . . . 47

3.2 Greedy exploration algorithm . . . 49

3.3 Main Results . . . 51

3.3.1 Large Deviation Principle . . . 51

3.3.2 Large deviations for the independent set size . . . 56

3.3.3 On the size of the maximum independent set . . . 58

3.4 Proof of Theorem 3.3.1 . . . 59

4 Large deviations for the greedy exploration process on d-regular graphs **66**

4.1 Introduction . . . 67

4.2 Description of the dynamics 67
4.3 Main Results 70
4.3.1 LDP for $\left\{X^N\right\}_N$ 70
4.3.2 LDP for $\left\{E^N\right\}_N$ 72
4.3.3 Optimization of the rate function $\hat{I}$ 74
4.3.4 Large deviations for the independent set size 77
4.4 Proof of Theorem 4.3.1 78

5 Large deviations for the greedy exploration process on configuration models 84
5.1 Introduction 85
5.2 Description of the model 85
5.3 Main Results 88
5.3.1 LDP for $\left\{X^N\right\}_N$ 88
5.3.2 Fluid limit of the process $\left\{X_t^N\right\}_t$ 91
5.3.3 Optimization of the rate function. 92
5.3.4 Large deviations for the independent set size 93
5.4 Proof of Theorem 5.3.1 94

Bibliography 101

Chapter 1

Introduction

In this thesis, we study large deviations for Markov processes modelling certain random graph exploration algorithms in order to build independent sets. This introduction briefly describes what large deviations are, what a random graph is, what an independent set is, which random graph exploration algorithms are considered in this thesis and why the study of large deviations may be of interest for these algorithms.

1.1 What are the large deviations?

The study of large deviations (LD) is concerned with studying the probabilities of very rare events. To understand why certain rare events might matter, one need only think of the enormous impact that winning the lottery (if we played) would have on our lives. Of course, this is in the case of a rare event with positive repercussions. But, on the other hand, we could also think of the enormous impact that rare events with catastrophic consequences can have on our lives, whether in terms of environment, economy, means of transport, communication, and so on.

Already the history of the study of large deviations begins with a very practical problem in Esscher's work [Esscher, 1932], in which rare events are analysed for a financial situation where the total claim amount exceeds the reserved amount (in insurance terms). In that case, the number of claim amounts was modelled by independent random variables with identical Poisson distribution. The rare event was that the average of these variables was much larger than the expected value of this variable. This was to be a preamble to Cramer's future work in [Cramér, 1938], in which the same question is analysed for the average of random variables with any distribution.

Since then, it has been necessary to define mathematically what it means for an event to be rare. A unified formalisation of the large deviation theory was developed in 1966 in a paper by Varadhan [Varadhan, 1966]. The large deviation principle (LDP) characterises the limiting behaviour, as $N \to +\infty$, of a sequence of probability measures $\{\mathbb{P}_N\}_N$ defined on $(\mathcal{X}, \mathcal{B})$ in terms of a rate function I. This characterisation is via asymptotic upper and lower exponential bounds on the values that $\mathbb{P}_N$ assigns to measurable sets of $\mathcal{X}$.

Starting with Donsker and Varadhan, a general foundation was laid that allowed one to point out several "general tricks" that seem to work in diverse situations. Moreover, large deviations estimates have proved to be the crucial tool required to handle many questions in statistics, engineering, statistical mechanics, and applied probability. So then, because of the diversity of applications, the theory has evolved one step at a time to cover more general scenarios.

Now, there are at least two approaches in the literature to prove an LDP. The traditional approach to LDP is via the so-called *change of measure method*. Indeed, beginning with the work of [Cramér, 1938] and including the fundamental work on large deviations for stochastic processes by [Freidlin and Wentzell, 1984] and [Donsker and Varadhan, 1975], much of the results has been obtained from a change of measure techniques. In this approach, a tilted or reference measure is identified under which the event of interest has a high probability. Then, the probability of the event under the original measure is bounded in terms of the Radon-Nikodym density relating to both measures.

Another approach is analogous to the Prohorov compactness approach to weak convergence of probability measures (by studying the tightness of these measures). It is sometimes referred to as the *exponential tightness method*. This has been established by [Puhalskii, 1994], [O'Brien and Vervaat, 1995], [de Acosta, 1997], [Dupuis and Ellis, 1997], [Fleming, 1985], [Evans and Ishii, 1985], and others. In this approach, exponential tightness plays the same role in large deviations theory as tightness does in weak convergence theory, exploiting the idea that large deviations can be thought of as a type of weak convergence at an exponential level.

In this thesis, we focus on the study of LD for sequences of Markov processes that describe the behaviour of greedy exploration algorithms on random graphs in order to build independent sets. To prove an LDP for those sequences, we use the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], based on the convergence of non-linear semigroups. This strategy belongs to the exponential tightness method mentioned above.

1.2 Random graphs

A graph is a structure composed by vertices and edges. The usual notation is $G = (V, E)$, where $V = \{v_i\}_i$ is the set of vertices v_i (finite or not), and $E = \{(v_i, v_j)\}_{i,j}$ denotes the edges connecting those vertices. All the graphs considered in this thesis are undirected (which means that all the graph edges are bidirectional), being the size of V a parameter N that goes to infinity.

Given two vertices $v_1, v_2 \in V$, we say that v_1 is a neighbour of v_2 if $(v_1, v_2) \in E$ (and $(v_2, v_1) \in E$ since the graph is undirected). We say that v has *degree* k if it has k neighbours.

We are interested in constructing *independent sets* of a graph G, i.e., subsets of vertices $S \subset V$ such that if the vertices v_1, v_2 belongs to S, then $(v_1, v_2) \notin E$. An independent set is said to be *maximal* if there is no larger independent set containing it as a subset, and it is *maximum* if there is no other independent set of larger size.

Both the construction of maximum independent sets and the calculation of its size are known to be NP-hard problems (see [Karp, 1972]), that is, problems for which all the existing algorithms that solve them have running times that grows faster than polynomially with N. A natural way to try to efficiently produce a large independent set in an input graph G is to output a maximal independent set. While in principle a badly chosen maximal independent set can be very small (like, for example, the star center in a star), one might hope that quite few of the maximal independent sets will have size comparable in some sense to the *independence number* of G (the size of the maximum independent set).

Random graphs were introduced by [Erdős, 1959] to give a probabilistic construction of a graph with large girth [1] and large chromatic number. [2] While there is earlier work using random graphs, for example, in [Moreno and Jennings, 1938] where a *"chance sociogram"* (a directed Erdös-Rényi model) was considered in studies comparing the fraction of reciprocated links in their network data with the random model, it was not until the work of [Erdős, 1959] and [Erdős and Rényi, 1959] that a systematic work began on random graphs as objects of interest in their own right, see [Frieze and Karoński, 2016].

Precisely, the first random graph considered in this thesis in Chapter 3 is the random graph known as the Erdös-Rényi graph. This is a finite graph obtained by setting an edge between each pair of vertices independently and with the same probability. An Erdös-Rényi graph is then $G(N, p)$, such that N is the number of vertices $V = \{v_1, \ldots, v_N\}$ of the graph and p is the probability that two vertices are connected (i.e. $(v_i, v_j) \in E$ with probability p). If X is the number of neighbours of any vertex (i.e., the number of vertices with which it shares an edge), then X has a Binomial distribution $\mathbf{Bin}(N-1, p)$.

In this case, it is possible to obtain bounds for the size of the maximum independent set by means of combinatorial theory tools. In 1962, Erdös proved that if α_N is the size of the

[1] The *girth* of a graph is the size of the smallest cycle contained in the graph.

[2] The *chromatic number* of a graph is the smallest number of colours needed to color the vertices of so that no two adjacent vertices share the same colour.

maximum independet set of $G(N, p_N)$ and $Np_N \geq 3$, then a.s. $\alpha_N \leq -2\frac{\log(Np_N)}{\log(1-p_N)}$. Then, for Erdös-Rényi graphs, we know at least a bound for the independence number (even if we do not know how to find it in polynomial time). On the other hand, for the case of Erdös-Rényi graphs $G\left(N, p_N = \frac{c}{N}\right)$ with $c < e$, the a.s. limiting proportion of the maximum independent set $\frac{\alpha_N}{N}$ is known, and it is given by $\sigma^*(c) = w(c) + \frac{c}{2}w(c)^2$ with $w(c) = e^{-W(c)}$ and $W(x)$ the Lambert function, see [Jonckheere and Saenz, 2019].

The second type of random graphs considered in this thesis are d-regular graphs. Those are graphs where each vertex v shares a fixed number d of edges with other graph vertices. The d neighbours of a vertex are uniformly randomly chosen. In this case, we allow a vertex to share more than one edge with the same neighbour (multi-edges) and share edges with itself (loops).

One only has to draw a few pictures (or watch Episode 407 of the TV serie **NUMB3RS**) to be convinced that it is not so easy to construct d-regular graphs. Therefore, we will consider a method known as *configurational model* for constructing d-regular random graphs.

Moreover, we can construct and work with more general uniform random graphs using this method. Therefore, we briefly define below how this method works, see [Bollobás, 1980] for example. Consider a sequence $d_1, \ldots, d_N$ of degrees ($d_i \geq 0$) such that $d_1 + \cdots + d_N$ is even. If $V = \{v_1, \ldots, v_N\}$ is the set of vertices, we assume that the vertex v_i has degree d_i, i.e. a number d_i of half-edges available to be paired with the half-edges of other vertices. Next we describe how these half-edges are paired as the random graph is sequentially constructed:

- select (somehow) a vertex v,
- match uniformly each unpaired half-edge from v with another one which is unpaired too,
- repeat until all vertices of the graph have paired their half-edges.

As mentioned before, this sequential construction results in a random graph that may contain multiple edges and loops. However, the probability of this happening converges to 0 when the size of the graph N goes to infinity (see [Brightwell et al., 2017]).

1.3 Random graphs exploration algorithms analysed in this thesis

First, it should be clarified that these algorithms are not the most efficient since they do not find independent sets of maximum size (see [Gamarnik and Sudan, 2017]). Still, they are quite intuitive and fit many practical models, where maximum independent sets are not always sought. Moreover, their importance lies in the fact that those algorithms are easy to program, analyse mathematically, and usually have polynomial complexity.

The random exploration algorithms considered in this thesis are usually referred to as *"greedy algorithms"*, see for instance the definition of an unweighted greedy algorithm in [Jungnickel, 2005]. Those algorithms build independents sets by selecting (somehow) sequentially a vertex $v \in V$, and adding v to the independent set if the resulting set does not span any edge.

Due to its simplicity, the greedy algorithms have been studied extensively by various authors in different fields (see [Krivelevich et al., 2020]), ranging from combinatorics [Wormald, 1995], probability [Rahman and Virág, 2017] and computer science [Fischer and Noever, 2020] to chemistry [Flory, 1939]. In communication sciences and wireless networks in particular, it allows representing the number of connections for CSMA-like algorithms in a given time-slot, for a given spatial configuration of terminals (see [Kleinrock and Takagi, 1985] for a classical reference on the protocol definition).

As early as 1931, this model was studied by chemists under the name *random sequential adsorption* (RSA), focusing primarily on d-dimensional grids. The 1-dimensional case was solved by [Flory, 1939], who showed that the expected value of the proportion of the independent set size obtained (known as the *greedy independence ratio*) converges to $T^* = \frac{1}{2}(1 - e^{-2})$ as the path length goes to infinity. A continuous analogue, in which "cars" of unit length "park" at random free locations on the interval $[0, T]$, was introduced (and solved) by [Rényi, 1958], under the name of car-parking process. The limiting density, as T goes to infinity, is therefore called Rényi's parking constant, and T^* may be considered as its discrete counterpart. Following this terminology, the final state of the car-parking process is often called the *jamming limit* of the graph, and the density of this state is called the *jamming constant*. In combinatorics, the greedy algorithm for finding a maximal independent set was analysed in order to give a lower bound on the (usually asymptotic) typical independence number of (random) graphs. The asymptotic greedy independence ratio of binomial random graphs with linear edge density was studied by [McDiarmid, 1990]. The asymptotic greedy independence ratio of random regular graphs was studied by [Wormald, 1995], who used the so-called differential equation method. His result was further extended in [Lauer and Wormald, 2007] for any sequence of regular graphs with growing girth. More recently, the case of uniform random graphs with given degree sequences was studied in [Bermolen et al., 2017b] and [Brightwell et al., 2017].

The first sequential exploration algorithm considered in this thesis works as follows. We start with a given graph $G = (V, E)$ for which $V = \{v_i\}_{i=1,\dots,N}$ is the set of N vertices and E is the set of edges. At each step $k = 0, 1, \dots,$ we consider that each vertex is either *active*, *blocked*, or *unexplored*. Accordingly, the set of vertices will be split into three components: the set of active vertices, the set of blocked vertices, and the set of unexplored vertices. Initially, all the vertices are declared as *unexplored*. At each step, this algorithm selects randomly and uniformly within the set of unexplored nodes a vertex and changes its state into *active*. After this, it takes all of its unexplored neighbours and changes their states into *blocked*. The *active* and *blocked* vertices are considered as *explored* and removed from the set of *unexplored* vertices. The algorithm keeps repeating this procedure until the step T_N^* at which all the vertices are either active or blocked (or equivalently, the set of the *unexplored* vertex is empty). Observe that at any step k, the active vertices conform to an independent set and that T_N^* is the size of the independent set constructed by the algorithm. Chapter 3 considered this greedy exploration algorithm on a given Erdös-Rényi graph.

The other sequential exploration algorithm considered in this thesis consists of building the independent set simultaneously with the random graph in a *configurational model*. It works as follows. Given a set of vertices $V = \{v_1, \ldots, v_N\}$ and a sequence of given degrees $d_1, \ldots, d_N$ (such that $d_1 + \cdots + d_N$ is even) where we assume that the vertex v_i has degree d_i, the random graph and the independent set are sequentially constructed in the following way. Initially, each vertex v_i has a number d_i of half-edges available to be paired with the half-edges of other vertices. At each step $k = 0, 1, \ldots,$ this algorithm:

- selects a vertex v uniformly from the set of vertices that still have all their half-edges unpaired, and puts it into the independent set (or changes its state into *active*),
- matches each unpaired half-edge from v uniformly with another one which is unpaired too.

The algorithm keeps repeating this procedure until the step T_N^* at which there are no more vertices with all their half-edges unpaired. At this point, there may still be some unpaired half-edges pointing out from vertices that have previously paired some of their half-edges. These may be paired off uniformly at random to complete the construction of the graph.

Chapter 4 considered this greedy exploration algorithm for constructing a d-regular graph simultaneously with an independet set in this graph. Then, in this case the degrees d_i are constants ($d_i = d$ for all i).

Chapter 5 considered this greedy exploration algorithm for constructing more general uniform random graphs simultaneously with independent sets in those graphs. In this case, we assume that each degree d_i is bounded.

1.4 Why study large deviations for these random graphs exploration algorithms?

To begin with, the study of large deviations is of interest in itself from a mathematical point of view. The richness of the LD partly derives from the fact that it is non-linear (since it is a type of weak convergence at the exponential level), which contributes to much of the difficulty of its analysis. Moreover, LD theory widely generalises the notion of convergence of probability measures.

Moreover, this theory has applications to many areas, including statistics, communication networks and queueing systems, information theory, statistical mechanics, risk-sensitive control, and finances. This is because, as mentioned above, it allows working on scenarios in which very rare events with catastrophic consequences occur. Moreover, it allows predicting the most likely scenario once a rare event has occurred.

In communication networks and queueing systems, see [Shwartz and Weiss, 1995], [Weiss, 1995] (and references therein) for an introduction to some LD techniques used to analyse models of communication networks. In most cases analysed there, communication networks are modelled by Poisson point processes.

In the context of equilibrium statistical mechanics, the theory of large deviations provides exponential-order estimates of probabilities that refine and generalise Einstein's theory of fluctuations, see [Touchette, 2009]. [Touchette, 2009] reviews this and other connections between large deviation theory and statistical mechanics to show that the mathematical language of statistical mechanics is the language of large deviation theory. In addition, large deviations make it possible to describe phase transitions considered in statistical mechanics.

[Dembo, 1994] explores some connections between LD and information theory. Moreover, connections between control problems and LD results were first made by [Fleming, 1985] and [Fleming, 1978]. [Dupuis and Ellis, 1997] systematically develop these connection showing that, in many cases, one can represent a large class of functionals of the processes as the minimal cost functions of stochastic control problems and then verify convergence of the functionals to the minimal cost functions of limiting deterministic control problem.

In finance, large deviations arise in various contexts. For example, they occur in risk management to compute the probability of significant losses of a portfolio subject to market risk or the default probabilities of a portfolio under credit risk. LD methods are largely used in rare events simulation and appear naturally in the approximation of option pricing, particularly for barrier options and far from the money options. See [Pham, 2007] for some applications of LD in finance and insurance. As mentioned above, the history of large deviations already begins with the analysis of a financial situation in Esscher's work [Esscher, 1932].

On the other hand, random graphs can model networks that are presented everywhere. From social networks such as Facebook, Twitter or LinkedIn, the World Wide Web and the Internet to the complex interactions between proteins in the cells of our bodies, we face the challenge of understanding their structure and development. In general, natural networks grow unpredictably, and this is often modelled by a random construction. We can then work on these random constructions to predict the occurrence of rare events. Moreover, as a consequence of obtaining results of large deviations on these constructions, the average behaviour can also be deduced.

1.5 Objective and main results obtained

In addition to presenting original results on large deviations for the processes of interest, we believe that the contribution of this thesis is to present in a comprehensive way the powerful tool proposed in the work of [Feng and Kurtz, 2006] for the study of large deviations of processes, with possible applications to many areas.

In this thesis, we focus on the study of LD for sequences of Markov processes $\left\{X^N\right\}_N$ with $X^N = \left\{X_t^N\right\}_{t\in[0,1]}$ that describe the behaviour of greedy exploration algorithms on random graphs in order to build independent sets. Informally, this means that if A is a set containing possible trajectories $\mathbf{x}$ of the process X^N, then there exists a rate function I such that the (possibly very small) probability that the process X^N belongs to A can be approximated by

$$\mathbb{P}(X^N \in A) \approx e^{-N \inf\limits_{\mathbf{x}\in A} I(\mathbf{x})}.$$

We focus on four aspects of the LD for these processes:

- Proving the path-state LD for these Markov processes,
- Deducing the *fluid limit* (a result of the type of the law of large numbers for stochastic processes which consists in a convergence of a sequence of stochastic processes to a deterministic function, usually subject to some type of rescaling, when a parameter N goes to infinity) from the LD rate function,
- Finding the trajectory that minimises the LD rate function over a set of trajectories,
- Deducing LD results for the size of the independent set constructed by the greedy algorithm.

In each chapter, we prove a large deviation principle (LDP) for the sequence of processes of interest using the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], based on the convergence of non-linear semigroups.

The remarkable work of [Feng and Kurtz, 2006] consists of combining the tools of probability, analysis, and control theory used in the works of [de Acosta, 1997], [Dupuis and Ellis, 1997], [Evans and Ishii, 1985], [Fleming, 1978], [Puhalskii, 1994], and others to propose a general strategy for the study of large deviations of processes. In the case of Markov processes, this program is carried out in four steps:

1. The first step consists of proving the convergence of non-linear generators H^N and derive the limit operator $\mathbf{H}$.
2. The second step consists of verifying the *exponential compact containment condition*.
3. The third step consists of proving that $\mathbf{H}$ generates a semigroup $\mathbf{V} = \{V_t\}_t$. This issue is nontrivial and follows, for example, by showing that the Hamilton-Jacobi equation $f(x) - \beta H(x, \nabla f(x)) - h(x) = 0$ has a unique solution f for sufficiently many $h \in C(E)$ and $\beta > 0$ in a viscosity sense, when $\mathbf{H}(f)(x) = H(x, \nabla f(x))$. The rate function is constructed in terms of that limit $\mathbf{V}$.
4. This limiting semigroup usually admits a variational form known as *Nisio semigroup* in control theory. Then, the fourth step consists of constructing a variational representation for the rate function from this variational representation of $\mathbf{V}$.

In a nutshell, as a consequence of the first two steps, the process verifies the *exponential tightness* condition, the third step assures the existence of an LDP, and the fourth step provides an applicable version of the rate.

In the cases analysed in this thesis, after working on the four steps mentioned before, we deduce a variational form of the rate function and prove that it can be expressed as an action integral of a cost function L, that is, if $\mathbf{x}$ is a possible trajectory for the process, then the rate function can be written as

$$I(\mathbf{x}) = \int_0^1 L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\, dt.$$

Additionally, in each case, the cost function L has a simple interpretation in terms of local deviations for the average of random variables, whose distribution approximates the jumps distribution of the process.

Moreover, we find the trajectory that minimises the LD rate function over a set of trajectories (i.e., the most probable trajectory) by studying the Hamiltonian dynamics associated with the rate function obtained.

The following is a brief description of the content of chapters 3, 4 and 5.

Large Deviations for the greedy exploration algorithm over Erdös-Rényi graphs

In Chapter 3, we consider the greedy exploration algorithm previously defined over Erdös-Rényi graphs. Let $G\left(N, \frac{c}{N}\right)$ be a given Erdös-Rényi graph, and Z_k^N be the number of explored vertices at time k. Thanks to the great amount of independence and symmetry of the edges' collection in this sparse Erdös-Rényi graph, the greedy exploration algorithm is characterized by the simple one-dimensional Markov process $\left\{Z_k^N\right\}_k$. Consequently, a functional law of large numbers described by a differential equation can be employed to get the macroscopic size of the constructed independent set when the number of vertices N goes to infinity (see [Bermolen et al., 2017a] and references in [McDiarmid, 1990]). Diffusion approximations for the process and central limit theorem derived from it for the size T_N^* of the associated independent set are also known, see [Bermolen et al., 2017a]. Moreover, in [Pittel, 1982], exponential bounds are proved for the probability that the stopping times $t_f(G(N, p/N))$ of the f-driven algorithms (in particular, T_N^*) belong to certain intervals. However, to the best of our knowledge, there is no characterization of a large deviation principle for both the discrete-time Markov process $\left\{Z_k^N\right\}_k$ and the random variable T_N^*, which can give various types of useful information both on the greedy exploration and on the independent set landscape. For example, this allows estimating the (low) probability of getting independent sets with sizes comparable to the independence number using a greedy algorithm. Moreover, it allows determining the most probable trajectory for which the independent set's size is bigger/smaller than selected bounds.

The topic of Chapter 3 is a refined analysis of this simple algorithm by studying the large deviations for a rescaling of the sequence of processes $\left\{Z_k^N\right\}_k$, when the number N of vertices goes to infinity.

To prove our main result, we use the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], which is presented in Section 2.2. The rate function can be expressed in a closed-form formula, and associated optimization problems can be solved explicitly, providing the large deviation trajectory. Also, we derive large deviations results for the size of the maximum independent set discovered by such an algorithm and analyse the probability that it exceeds known bounds for the maximal independent set. Moreover, we explore the link between these results and the landscape complexity of the independent set and the exploration dynamic.

The results of this chapter were accepted for publication in **ALEA** (*Latin American Journal of Probability and Mathematical Statistics*). It can be consulted at [Bermolen et al., 2021b].

Large Deviations for the greedy exploration process on d-regular graphs

In Chapter 4, we prove large deviations for the greedy exploration of configuration models by jointly constructing a d-regular graph and discovering an independent set in this graph. We consider a time-discretized version of the method proposed by [Bermolen et al., 2017b] and [Brightwell et al., 2017] for creating more general uniform random graphs from a given degree sequence. We consider a discrete-time Markov process describing the evolution of this algorithm and prove a large deviation principle for a rescaling of this process. In this case, the algorithm is described by the Markov process $\left\{X_k^N\right\}_k$ such that $X_k^N = \left(S_k^N, U_k^N, E_k^N\right)$, being

- S_k^N, the number of vertices that have already been placed into the independent set at step k,
- U_k^N, the total number of unpaired half-edges at step k,
- E_k^N, the number of vertices such that none of their d half-edges were paired at step k.

As a corollary, we derive large deviations results for the independent set size constructed by this algorithm. Finally, we retrieve known results about the independent set size obtained and the change that occurs in the dynamics when $d = 2$ or $d > 2$.

Large deviations for the greedy exploration process on configuration models

In Chapter 5, we extend the results presented in Chapter 4 to the greedy exploration of configuration models, building on a time-discretized version of the method proposed by [Bermolen et al., 2017b] and [Brightwell et al., 2017] by jointly constructing general uniform random graphs from a given degree sequence and its exploration.

We start with a set of vertices $\mathcal{V}^N = \{1, 2, \dots, N\}$ such that each vertex i has a bounded degree $\deg(i) \leq D < \infty$, and such that the initial distribution of degrees $\frac{1}{N}\#\left\{i : \deg(i) = j\right\}$ converges to $p_j \geq 0$, when the number of vertices N goes to infinity, for all $j = 0, \dots, D$ (with $\sum_{j=0}^{D} p_j = 1$). Each vertex i of the graph has a number $\deg(i)$ of half-edges available to be paired with the half-edges of other vertices. We prove an LDP for a rescaling of the sequence of Markov processes $\left\{X^N\right\}_N$, where $X^N = \left\{X_k^N\right\}_k$ and $X_k^N = \left(S_k^N, U_k^N, E_k^N(0), E_k^N(1), \dots, E_k^N(D)\right)$ is such that

- S_n^N is the number of vertices that have already been placed into the independent set at step k,

- U_k^N is the total number of unpaired half-edges at step k,
- $E_k^N(j)$ is the number of vertices with degree j such that none of their j half-edges were paired at step k.

Again, the proof of this result follows the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006]. As a corollary, we deduce the corresponding fluid limit and LD results for the independent set size discovered by this exploration algorithm.

The results of this chapter were submitted to *Electronic Communications in Probability* including the results of Chapter 4 as a particular case. It can be consulted at [Bermolen et al., 2021a].

1.6 Thesis structure

Finally, we briefly describe the structure of the thesis.

In Chapter 2, we present the main results used in this thesis to study large deviations for sequences of Markov processes defined on a compact state space.

In Section 2.1, we present some generalities about large deviations, including the definition of a *Large Deviation Principle*, the connection with Laplace's Principle, and the analogy with the weak convergence of probability measures. In Section 2.2, we present the main results of [Feng and Kurtz, 2006] for the study of large deviations of stochastic processes in the context of Markov processes defined on a compact state space. In Section 2.3, we present the tools we use to study the uniqueness of viscosity solutions of Hamilton-Jacobi equations. This uniqueness is an essential requirement for the study of large deviations, according to the work of [Feng and Kurtz, 2006]. Since in the cases analysed in this thesis the large deviations rate function can be written as an action functional, in Section 2.4 we present tools from the calculus of variations to find the trajectory that minimises the rate function over a set of trajectories (i.e. the optimal trajectory). In Section 2.5, we show that the *fluid limit* of a sequence of Markov processes can be deduced from the study of the large deviations. Finally, since it is impossible to talk about large deviations for Markov processes without mentioning the pioneering work of [Freidlin and Wentzell, 1984], we present in Section 2.6 a synthesis of their results in the context of the work of [Feng and Kurtz, 2006].

In Chapter 3, we study large deviations for a greedy exploration algorithm over Erdös-Rényi graphs when the number of vertices goes to infinity.

In Section 3.1, we introduce the greedy algorithm over Erdös-Rényi graphs and known results about the independent set size obtained by such an algorithm. In Section 3.2, we formally define the sequence of processes related to the greedy algorithm over a given Erdös-Rényi graph. In Section 3.3, we present the main result: a path-state LDP for the greedy exploration process. The proof of this result is deferred to Section 3.4. As a corollary, we obtain LD results for the size of the independent set discovered by such an algorithm and analyse its implications.

In Chapter 4, we study large deviations for the greedy exploration on a configuration model for constructing d-regular graphs when the number of vertices goes to infinity.

In Section 4.1, we give a brief introduction to the chapter. In Section 4.2, we define the dynamic analysed in this chapter. Moreover, we define a sequence of Markov processes related to this construction. In Section 4.3, we present the main result: a path-state LDP for the sequence of Markov processes defined in Section 4.2. The detailed proof is deferred to Section 4.4. As a corollary, we obtain the corresponding fluid limit and large deviations results for the size of the independent set constructed.

In Chapter 5, we extend the results from Chapter 4 to the greedy exploration of configuration models, building on a time-discretized version of the method proposed by [Bermolen et al., 2017b] and [Brightwell et al., 2017] for constructing a random graph from a given degree sequence and its exploration.

In Section 5.1, we give a brief introduction to the chapter. In Section 5.2, we define the dynamic analysed in this chapter. Moreover, we define a sequence of Markov processes related to this algorithm. In Section 5.3, we present the main result of this chapter: a path-state LDP for the sequence of Markov processes defined in Section 5.2 along with the heuristic that motivates the result. The detailed proof is deferred to Section 5.4. As a corollary, we deduce the fluid limit of the process and LD results for the size of the independent set constructed by the exploration algorithm.

Chapter 2

Large deviations for Markov processes

Abstract

In this chapter, we present the main results used in this thesis to study large deviations for sequences of Markov processes defined on a compact state space.

This chapter is organised as follows. In Section 2.1, we present some generalities about large deviations, including the definition of a *Large Deviation Principle*, the connection with Laplace's Principle, and the analogy with the weak convergence of probability measures. In Section 2.2, we present the main results of [Feng and Kurtz, 2006] for the study of large deviations of stochastic processes in the context of Markov processes defined on a compact state space. In Section 2.3, we present the tools we use to study the uniqueness of viscosity solutions of Hamilton-Jacobi equations. This uniqueness is an essential requirement for the study of large deviations, according to the work of [Feng and Kurtz, 2006]. Since in the cases analysed in this thesis the large deviations rate function can be written as an action functional, in Section 2.4 we present tools from the calculus of variations to find the trajectory that minimises the rate function over a set of trajectories (i.e. the optimal trajectory). In Section 2.5, we show that the *fluid limit* of a sequence of Markov processes can be deduced from the study of the large deviations. Finally, since it is impossible to talk about large deviations for Markov processes without mentioning the pioneering work of [Freidlin and Wentzell, 1984], we present in Section 2.6 a synthesis of their results in the context of the work of [Feng and Kurtz, 2006].

2.1 Generalities about large deviations

The theory of large deviations is concerned with the asymptotic estimation of probabilities of rare events. Consider a complete and separable metric space $(\mathcal{X}, d)$ and a sequence (it can be a net) of probability measures $\{\mathbb{P}_N\}_N$ defined on $\mathcal{B}(\mathcal{X})$, the σ-algebra of all Borel subsets of $\mathcal{X}$.

Definition 2.1.1 (Large Deviation Principle, Varadhan) *The sequence of probability measures $\{\mathbb{P}_N\}_N$ verifies a* large deviation principle *(LDP) if there exists a lower semicontinuous function $I : \mathcal{X} \to [0, +\infty]$ such that for each open set A,*

$$\liminf_{N\to+\infty} \frac{1}{N} \log \mathbb{P}_N(A) \geq -\inf_{x\in A} I(x),$$

and for each closed set B,

$$\limsup_{N\to+\infty} \frac{1}{N} \log \mathbb{P}_N(B) \leq -\inf_{x\in B} I(x).$$

I is called the rate function *for the large deviation principle. A rate function I is* good *if for each $a \in [0, +\infty)$ the set $\{x : I(x) \leq a\}$ is (in addition to being closed) a compact subset of $\mathcal{X}$.*

This definition can be thought of as a type of weak convergence at an exponential level. The analogy becomes much more clearly if we recall the following equivalent formulation of weak convergence:

Proposition 2.1.1 (Portmanteau) *The following are equivalents:*

1. *$\{\mathbb{P}_N\}_N$ converge weakly to $\mathbb{P}$*
2. *$\int f(x)d\mathbb{P}_N(x) \to \int f(x)d\mathbb{P}(x)$ for all $f \in C_b(\mathcal{X})$, being $C_b(\mathcal{X})$ the space of bounded and continuous functions $f : \mathcal{X} \to \mathbb{R}$.*
3. *$\liminf_{N\to+\infty} \mathbb{P}_N(A) \geq \mathbb{P}(A)$ for all open $A \in \mathcal{B}(\mathcal{X})$.*
4. *$\limsup_{N\to+\infty} \mathbb{P}_N(B) \leq \mathbb{P}(B)$ for all closed $B \in \mathcal{B}(\mathcal{X})$.*

That is, large deviations are an exponential version of properties *3* and *4* from the Portmanteau theorem, so it is not surprising that the study of large deviations involves many concepts that are analogous to concepts in the study of weak convergence. Note, however, that we assume both an upper and lower bound for a large deviation principle instead of the equivalence that holds in the case of weak convergence.

If $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space and $\{X_N\}_N$ is a sequence of random variables $X_N : \Omega \to \mathcal{X}$, we say that $\{X_N\}_N$ verifies an LDP if their push-forward measures defined by $\mathbb{P}_N(A) = \mathbb{P}(X_N \in A)$ for all $A \in \mathcal{B}(\mathcal{X})$ verify an LDP.

Remark 2.1.1 Let us introduce some comments about the LDP definition.

1. ***Weak LDP:*** Definition of LDP with "closed set B" replaced by "compact set B" is called the *weak large deviation principle* (see [Dembo and Zeitouni, 1998], for example).
2. The LDP is equivalent to the assertion that for each $A \in \mathcal{B}(\mathcal{X})$ it is verified that

$$-\inf_{x\in \mathring{A}} I(x) \leq \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N(A) \leq \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N(A) \leq -\inf_{x\in \bar{A}} I(x),$$

where $\mathring{A}$ and $\bar{A}$ are respectively the interior and closure of A. Moreover, the set A is said to be a *good set for I* if in the previous equation the inequalities are all an equality.

3. ***Uniqueness of the rate function:*** Note that, if I is a lower semicontinuous function, then

$$I(x) = \lim_{\varepsilon\to 0^+} \inf_{y\in B_\varepsilon(x)} I(y),$$

being $B_\varepsilon(x) = \{y \in \mathcal{X} : d(x, y) < \varepsilon\}$. If $\{\mathbb{P}_N\}_N$ verifies an LDP with rate function I, then

$$-\inf_{y\in B_\varepsilon(x)} I(y) \leq \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N\left(B_\varepsilon(x)\right) \leq \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N\left(\overline{B_\varepsilon(x)}\right) \leq -\inf_{y\in \overline{B_\varepsilon(x)}} I(y).$$

This implies that

$$-I(x) = \lim_{\varepsilon\to 0} \liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N\left(B_\varepsilon(x)\right) = \lim_{\varepsilon\to 0} \limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N\left(\overline{B_\varepsilon(x)}\right),$$

and it follows that the semicontinuity requirement of I assures the uniqueness of the rate function.

4. ***An equivalent formulation of the LDP:*** The LDP is equivalent to the *Laplace principle* (LP). As we show below, the formulation of the LDP⇒LP implication is known as Varadhan's lemma. The formulation of LP⇒LDP appears as Theorem **1.2.3** in [Dupuis and Ellis, 1997]. Consider a sequence of random variables $\{X_N\}_N$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 2.1.2 (Laplace's Principle) *Let I be a rate function defined on $\mathcal{X}$. The sequence $\{X_N\}_N$ is said to satisfy the Laplace's Principle (LP) on $\mathcal{X}$ with rate function I if*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{-Nf(X_N)}\right] = -\inf_{x\in\mathcal{X}} \left\{f(x) + I(x)\right\}, \quad \forall f \in C_b(\mathcal{X}).$$

Varadhan's lemma generalises the well known method of Laplace for studying the asymptotics of certain integrals on $\mathbb{R}$. That is, given $f \in C_b[0, 1]$, Laplace's method states that

$$\lim_{N\to\infty}\frac{1}{N}\log\int_0^1 e^{-Nf(x)}dx = -\inf_{x\in[0,1]} f(x).$$

Theorem 2.1.2 (Varadhan's lemma) *Assume that the sequence $\{X_N\}_N$ satisfies an LDP with rate function I, then*

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-Nf(X_N)}\right] = -\inf_{x\in\mathcal{X}}\{f(x)+I(x)\}, \quad \forall f\in C_b(\mathcal{X}).$$

That is, LDP$\Rightarrow$ LP. The following non-rigorous calculation shows why this result is verified. If we summarize the LDP by the notation $\mathbb{P}(X_N\in dx)\asymp e^{-NI(x)}dx$, then

$$\mathbb{E}\left[e^{-Nf(X_N)}\right] = \int_{\mathcal{X}} e^{-Nf(x)}\mathbb{P}(X_N\in dx)\asymp\int_{\mathcal{X}} e^{-N(f(x)+I(x))}dx,$$

and

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-Nf(X_N)}\right] = \lim_{N\to\infty}\frac{1}{N}\log\int_{\mathcal{X}} e^{-N(f(x)+I(x))}dx = -\inf_{x\in\mathcal{X}}\{f(x)+I(x)\}.$$

That is, as in Laplace's method, Varadhan's lemma states that to exponential order the main contribution to the integral is due to the largest value of the exponent. The next theorem, proves the converse, highlighting a basic feature of the weak convergence approach:

Theorem 2.1.3 (Theorem 1.2.3. from [Dupuis and Ellis, 1997]) *If I is a rate function on $\mathcal{X}$ and the limit*

$$\Lambda(f) = \lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-Nf(X_N)}\right] = -\inf_{x\in\mathcal{X}}\{f(x)+I(x)\},$$

is valid for all $f\in C_b(\mathcal{X})$, then $\{X_N\}_N$ satisfies an LDP on $\mathcal{X}$ with rate function I.

This theorem is related to another converse to Varadhan's lemma due to Bryc. As we show below, Bryc's theorem states that if the sequence $\{X_N\}_N$ is *exponentially tight* and the limit

$$\Lambda(f) = \lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{-Nf(X_N)}\right]$$

exists for all $f\in C_b(\mathcal{X})$, then the rate function can be written as

$$I(x) = -\inf_{f\in C_b(\mathcal{X})}\{f(x)+\Lambda(f)\}.$$

In general, there are at least two approaches in the literature to prove an LDP. The traditional approach to LDP is via the so-called *change of measure method.* Indeed, beginning with the work of [Cramér, 1938] and including the fundamental work on large deviations for stochastic processes by [Freidlin and Wentzell, 1984] and [Donsker and Varadhan, 1975], much of the analysis has been based on a change of measure techniques. In this approach, a tilted or reference measure is identified under which the event of interest has a high probability. The probability of the event under the original measure is bounded in terms of the Radon-Nikodym density relating to both measures.

Another approach is analogous to the Prohorov compactness approach to weak convergence of probability measures (by studying the tightness of these measures). It is sometimes referred to as the *exponential tightness method.* This has been established by [Puhalskii, 1994], [O'Brien and Vervaat, 1995], [de Acosta, 1997], [Dupuis and Ellis, 1997], [Fleming, 1985], [Evans and Ishii, 1985], [Feng and Kurtz, 2006], and others.

According to this approach, we recall now the well-known Prohorov theorem to show the analogy between the study of weak convergence and large deviations.

Proposition 2.1.4 (Prohorov) *If the sequence of probabilities $\{\mathbb{P}_N\}_N$ defined on $\mathcal{X}$ is tight, then there exists a subsequence $\{N_k\}_k$ and a probability $\mathbb{P}$ such that $\{\mathbb{P}_{N_k}\}_k$ converges weakely to $\mathbb{P}$.*

Definition 2.1.3 (Tightness) *A sequence $\{\mathbb{P}_N\}_N$ of probability measures defined on $\mathcal{X}$ is tight if for each $\alpha > 0$ there exits a compact set $K_\alpha \subset \mathcal{X}$ such that $\sup_N \mathbb{P}_N(K_\alpha^c) \leq \alpha$.*

Then, the proof of weak convergence results typically involves the verification of relative compactness or *tightness* for the sequence and the unique characterization of the possible limit distribution. The analogous approach to large deviations involves verifying *exponential tightness* (that we define next) and unique characterization of the possible rate function.

Definition 2.1.4 (Exponential tightness) *A sequence $\{\mathbb{P}_N\}_N$ of probability measures defined on $\mathcal{X}$ is exponentially tight if for each $\alpha > 0$ there exits a compact set $K_\alpha \subset \mathcal{X}$ such that $\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N(K_\alpha^c) \leq -\alpha$.*

Exponential tightness plays the same role in large deviations theory as tightness does in weak convergence theory. The following theorem, proved separately by [Puhalskii, 1991] and [O'Brien and Vervaat, 1991], is the analogue of the Prohorov compactness theorem:

Proposition 2.1.5 (Puhalskii, O'Brien-Vervaat) *Let $(\mathcal{X}, d)$ be a metric space and $\{\mathbb{P}_N\}_N$ a sequence of tight probability measures on the Borel σ-algebra of $\mathcal{X}$ [1]. Suppose that $\{\mathbb{P}_N\}_N$ is*

[1] Recall that if $\mathcal{X}$ is complete and separable, then every probability measure is tight.

exponentially tight. Then, there exists a subsequence $\{N_k\}_k$ along which the LDP holds with a good rate function.

Moreover, one consequence of the exponential tightness is that if a sequence is exponentially tight, then the large deviation principle holds if and only if the weak large deviation principle holds (see [Dembo and Zeitouni, 1998], for example). Perhaps, the best-known result according to this approach is Bryc's theorem, or inverse of Varadhan's Lemma, which we recall below:

Proposition 2.1.6 (Bryc) *If the sequence of probability measures $\{\mathbb{P}_N\}_N$ defined on $\mathcal{X}$ is exponentially tight and the following limit exists:*

$$\Lambda(f) = \lim_{N\to\infty} \frac{1}{N} \log \left[\int e^{Nf(x)} d\mathbb{P}_N(x) \right]$$

for all $f \in C_b(\mathcal{X})$, then $\{\mathbb{P}_N\}_N$ verifies an LDP with rate function $I : \mathcal{X} \to [0, +\infty]$ given by

$$I(x) = \sup_{f\in C_b(\mathcal{X})} \{f(x) - \Lambda(f)\},$$

where $C_b(\mathcal{X})$ is the space of bounded and continuous functions $f : \mathcal{X} \to \mathbb{R}$.

That is, if $\mathcal{X}$ is a topological vector space, then $I(x) = \Lambda^*(x)$ with the product of duality $\langle f, x\rangle = f(x)$.

In this thesis, we are in the case in which the sequence of probability measures $\{\mathbb{P}_N\}_N$ comes from a sequence of Markov processes $\left\{X^N\right\}_N$ (in discrete or continuous-time) with state spaces E^N, being E^N a subset of a compact set E of $\mathbb{R}^n$ ($n \in \mathbb{N}$ fixed). Then, we consider $\mathcal{X} = D_E[0, \infty)$, equipped with the Skorohod topology. There are in the literature conditions that ensure the exponential tightness in this case, but the calculation of $\Lambda(f) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\left[e^{Nf(X^N)}\right]$ is difficult or impossible.

The work of [Feng and Kurtz, 2006], which we present in the following section, solves both this problem and that of exponential tightness. Since the transitions characterize the Markov process, instead of calculating $\Lambda(f)$, the time $t \geq 0$ is fixed, and the convergence of the following non-linear semigroups is considered: $V_t^N : Dom\left(V_t^N\right) \subset B(E^N) \to B(E^N)$ such that $V_t^N(x) = \frac{1}{N} \log \mathbb{E}\left[e^{Nf(X_t^N)} \middle| X_0^N = x\right]$, being $B(E^N)$ the space of bounded, Borel measurable functions (i.e. t is fixed and the domain of the functions f is E^N instead of the much more complex space $D_E[0, \infty)$). It is proved that, under certain assumptions, the convergence of these non-linear semigroups ensures an LDP.

Actually, instead of studying the convergence of the non-linear semigroups $V^N = \left\{V_t^N\right\}_t$, the convergence of their (non-linear) generators H^N (defined in the next section) is studied. The reason why we study the convergence of the sequence of operators H^N instead of V^N is because the variational representations for the sequence of functionals V^N can be difficult to obtain. Then, we first verify convergence of these semigroups by methods that only require convergence of the corresponding generators H^N and conditions on the limiting generator $\mathbf{H}$. Working only with the sequence of generators frequently provides conditions that are easier to verify than conditions that give a convergent sequence of variational representations. Variational representations for the limit $\{V_t\}_t$ are still important, however, as they provide methods for obtaining simple representations of the large deviation rate function.

2.2 Large deviations for Markov processes - the work of Feng and Kurtz

In this section, we present the main results of [Feng and Kurtz, 2006] for the study of large deviations of stochastic processes in the context of Markov processes defined on a compact state space.

The remarkable work of [Feng and Kurtz, 2006] consists of combining the tools of probability, analysis, and control theory used in the works of [de Acosta, 1997], [Dupuis and Ellis, 1997], [Evans and Ishii, 1985], [Fleming, 1978], [Fleming, 1999], [Puhalskii, 1994], and others to propose a general strategy for the study of large deviations of processes. In this subsection, we briefly describe their work in the context of Markov processes with state spaces included on a compact subset $E \subset \mathbb{R}^n$. However, many of the results presented in [Feng and Kurtz, 2006] apply to more general stochastic processes on metric spaces (E, r).

Throughout, E will be a compact subset of $\mathbb{R}^n$. $M(E)$ will denote the space of real-valued Borel measurable functions on E, and $B(E) \subset M(E)$, the space of bounded, Borel measurable functions. An operator $A = (A, Dom(A))$ is given by a domain $Dom(A) \subset M(E)$ and a map $A : Dom(A) \to M(E)$. Also we write A for the graph of the map $A = \{(f, A(f)) : f \in Dom(A)\}$ and write, for example, $A \subset C_b(E) \times C_b(E)$ if the domain $Dom(A)$ and the range $\mathcal{R}(A)$ are contained in $C_b(E)$. In some cases, A can be multi-valued and non-linear. The space of E-valued, càdlàg functions on $[0, \infty)$ with the Skorohod topology will be denoted by $D_E[0, \infty)$.

- **Continuous-time Markov processes:** An E-valued Markov process $X = \{X_t\}_t$ is usually characterized in terms of its *generator*, a linear operator $A \subset B(E) \times B(E)$. One approach to the characterization of X is to requiere that all the processes of the form

$$f\left(X_t\right) - f\left(X_0\right) - \int_0^t A(f)\left(X_s\right) ds, \tag{2.1}$$

be martingales w.r.t. some filtration $\{\mathcal{F}_t\}_t$ independent of f (see [Kurtz, 1971], for example). If X satisfies this condition is said to be a solution of the *martingale problem* for A. If f is bounded away from zero, then the process defined by Equation (2.1) is a martingale if and only if

$$\frac{f(X_t)}{f(X_0)} e^{-\int_0^t \frac{A(f)(X_s)}{f(X_s)} ds}$$

is a martingale. Consequently, if we define $Dom(\mathcal{H}) = \left\{ f \in B(E) : e^f \in Dom(A) \right\}$ and set

$$\mathcal{H}(f) = e^{-f} A\left(e^f\right),$$

then we can define the *exponential martingale problem* by requiring that

$$\exp\left\{ f(X_t) - f(X_0) - \int_0^t \mathcal{H}(f)(X_s)\, ds \right\} \tag{2.2}$$

to be a martingale w.r.t. a filtration independent of f. Moreover, the process defined by Equation (2.2) is a martingale if and only if (2.1) is a martingale. It follows that X is a solution of the linear martingale problem for A if and only if it is a solution of the exponential martingale problem for $\mathcal{H}$. As we will see, large deviations for sequences of Markov processes $\left\{X^N\right\}_N$ are strongly connected to the convergence of these operators $\mathcal{H}$.

Weak convergence results for a sequence of Markov processes $\left\{X^N\right\}_N$ can be based on convergence of the corresponding *linear-semigroups* $T^N = \left\{T^N_t\right\}_t$ defined by

$$T^N_t : Dom(T^N_t) \subset B(E^N) \to M(E^N), \quad T^N_t(f)(x) = \mathbb{E}\left[f\left(X^N_t\right) | X^N_0 = x\right].$$

This linear-semigroup T^N satisfies

$$\frac{d}{dt} T^N_t(f) = A^N\left(T^N_t\right)(f), \quad T^N_0(f) = f,$$

where A^N is the linear-generator for X^N. An analogous approach to large deviations results is suggested by [Fleming, 1985] using the nonlinear contraction (in the supremum norm) semigroup $V^N = \left\{V^N_t\right\}_t$ given by

$$V^N_t : Dom(V^N_t) \subset B(E^N) \to M(E^N), \quad V^N_t(f)(x) = \frac{1}{N} \log \mathbb{E}\left[e^{Nf\left(X^N_t\right)} | X^N_0 = x\right]. \tag{2.3}$$

Again, at least formally, V^N should satisfy

$$\frac{d}{dt}V_t^N(f) = \frac{1}{N}\mathcal{H}^N\left(NV_t^N(f)\right) = \frac{1}{N}e^{-NV_t^N(f)}A^N\left(e^{NV_t^N(f)}\right),$$

which leads us to define the *non-linear generator*

$$H^N(f) = \frac{1}{N}e^{-Nf}A^N\left(e^{Nf}\right). \tag{2.4}$$

Note that the process defined by

$$\exp\left\{Nf\left(X_t^N\right) - Nf\left(X_0^N\right) - \int_0^t NH^N(f)\left(X_s^N\right)ds\right\}$$

is a martingale (since the process defined by Equation (2.2) is a martingale).

- **Discrete-time Markov processes:** Consideration of discrete-time Markov processes leads to slightly different formulations of the two martingale problems defined before. As a consequence, the non-linear generators considered for the discrete-time case are slightly different. Let $\{Y_k\}_k$ be a time-homogeneous Markov chain with state space E and transition operator

$$T : Dom(T) \subset B(E) \to M(E), \quad T(f)(x) = \mathbb{E}\left[f\left(Y_{k+1}\right)|Y_k = x\right] = \mathbb{E}\left[f\left(Y_1\right)|Y_0 = x\right].$$

For $\varepsilon > 0$, define $X_t^\varepsilon = Y_{[t/\varepsilon]}$. Then, setting $\mathcal{F}_t^{X^\varepsilon}$ the σ-algebra generated by the sets $\{X_s^\varepsilon : s \le t\}$, for $f \in B(E)$ we have that the process given by

$$\begin{aligned} f\left(X_t^\varepsilon\right) - f\left(X_0^\varepsilon\right) - \sum_{k=0}^{[t/\varepsilon]-1}\left(T(f)\left(Y_k\right) - f\left(Y_k\right)\right) \\ = f\left(X_t^\varepsilon\right) - f\left(X_0^\varepsilon\right) - \int_0^{[t/\varepsilon]\varepsilon}\frac{1}{\varepsilon}\left(T - Id\right)(f)\left(X_s^\varepsilon\right)ds \end{aligned}$$

is an $\left\{\mathcal{F}_t^{X^\varepsilon}\right\}_t$-martingale, and the process

$$\exp\left\{f\left(X_t^\varepsilon\right) - f\left(X_0^\varepsilon\right) - \int_0^{[t/\varepsilon]\varepsilon}\frac{1}{\varepsilon}\log\left(e^{-f}T\left(e^f\right)\left(X_s^\varepsilon\right)\right)ds\right\}$$

is an $\left\{\mathcal{F}_t^{X^\varepsilon}\right\}_t$-martingale too.

Consider now the sequences of Markov chains $\left\{Y^N\right\}_N$ defined on E^N, and $\{\varepsilon_N\}_N$ with

$\varepsilon_N \to 0$. Then, for the sequence $X_t^N = Y_{[t/\varepsilon_N]}^N$, we define the operators

$$A^N(f) = \frac{1}{\varepsilon_N}\left(T^N - Id\right)(f) \text{ and } H^N(f) = \frac{1}{N\varepsilon_N}\log\left(e^{-Nf}T^N\left(e^{Nf}\right)\right), \tag{2.5}$$

(T^N is the transition operator of Y^N), so that

$$f\left(X_t^N\right) - f\left(X_0^N\right) - \int\limits_0^{[t/\varepsilon_N]\varepsilon_N} A^N(f)\left(X_s^N\right)ds$$

and

$$\exp\left\{Nf\left(X_t^N\right) - Nf\left(X_0^N\right) - \int\limits_0^{[t/\varepsilon_N]\varepsilon_N} NH^N(f)\left(X_s^N\right)ds\right\}$$

are martingales.

As mentioned before, one of our main assumptions should be the convergence of the sequence of non-linear generators $\left\{H^N\right\}_N$ defined on Equations (2.4) and (2.5) for the continuous and discrete-time case, where the type of convergence may depend on the particular problem.

The general idea is that if there is a functional $\mathbf{H}$ such that $H_N \to \mathbf{H}$, $\mathbf{H}$ generates a semigroup $\mathbf{V} = \{V_t\}_t$ and the *exponential compact containment condition* is verified, then the sequence of Markov processes $\left\{X^N\right\}_N$ (defined in the continuous or discrete-time case) verifies an LDP with rate function I that depends on $\mathbf{V}$. Moreover, if $\mathbf{H}$ is such that $\mathbf{H}(f)(x) = H\left(x, \nabla f(x)\right)$ for all $f \in C^1(E)$ and Conditions **8.9**, **8.10** and **8.11** from [Feng and Kurtz, 2006] are also verified, we obtain a variational representation of I. Moreover, in the cases analised in this thesis, the rate will be written as an action integral of the Legendre-Fenchel transform of H given by $L(x,\beta) = \sup\limits_{\alpha\in\mathbb{R}^n}\left\{\langle\alpha,\beta\rangle - H(x,\alpha)\right\}$.

A typical application of the results presented in [Feng and Kurtz, 2006] (here we only present the main results that we will use later) requires the following steps:

Step 1: Verify the convergence of the sequence of operators H^N and derive the limit operator H.

In general, the convergence of the sequence H^N may be in an *extended limit* (see Definition **A.12** from [Feng and Kurtz, 2006]) or graph sense, that is, if the process X^N is defined on E^N and we have maps $\eta_N : E^N \to E$, we can define new maps $\hat{\eta}_N : B(E) \to B(E^N)$ such that $\hat{\eta}_N(f) = f \circ \eta_N$. If $H^N \subset B(E^N) \times B(E^N)$, the extended limit of $\left\{H^N\right\}_N$ is the collection of

$(f,g) \in B(E) \times B(E)$ such that there exists $(f_N, g_N) \in H^N$ satisfying

$$\lim_{N\to\infty} \|f_N - \hat{\eta}_N(f)\| + \|g_N - \hat{\eta}_N(g)\| = 0.$$

In some examples, the limit is described in terms of a pair of operators, $(\mathbf{H}_\ddagger, \mathbf{H}_\dagger)$, where $\mathbf{H}_\dagger$ is the lim sup of $\left\{H^N\right\}_N$, and $\mathbf{H}_\ddagger$ is the lim inf.

In the cases analysed in this thesis, the operator $\mathbf{H}$ follows naturally from the definition of H^N and the fact that the domain considered for H^N is $C^1(E)$. Moreover, the operator $\mathbf{H}$ will be of the form $\mathbf{H}(f)(x) = H\left(x, \nabla f(x)\right)$ for all $f \in C^1\left(E\right)$, where $H : E \times R^n \to \mathbb{R}$ is sufficiently regular on $\mathring{E} \times \mathbb{R}^n$ and convex w.r.t. the second variable.

Step 2: Verify the exponential compact containment condition.

The convergence of operators H^N typically gives exponential tightness, provided we can verify the *exponential compact containment condition* that we define next.

Definition 2.2.1 (Exponential compact containment condition) *The sequence* $\left\{X^N\right\}_N$ *verifies the* exponential compact containment condition *if for all* $\alpha > 0$, *there exists* $K_\alpha \subset E$ *compact such that* $\limsup_N \frac{1}{N} \log \mathbb{P}\left(\left\{\exists\, t : X_t^N \notin K_\alpha\right\}\right) \leq -\alpha$.

Note that, unlike Definition 2.1.4, a compact subset of E is needed and not from the more complex space $D_E[0,\infty)$.

Step 3: Verify the comparison principle for the limiting operator H.

This step consists of proving that the limiting operator $\mathbf{H}$ generates a semigroup $\mathbf{V} = \{V_t\}_t$. This is the most technical step. By definition, V_t^N verifies

$$\frac{\mathrm{d}}{\mathrm{d}t} V_t^N(f) = H^N\left(V_t^N(f)\right), \qquad V_0^N(f) = f.$$

Then, $\mathbf{H}$ generates a semigroup if there exists a semigroup $\mathbf{V} = \{V_t\}_t$ such that for all $f \in Dom(\mathbf{V})$,

$$\frac{\mathrm{d}}{\mathrm{d}t} V_t(f) = \mathbf{H}\left(V_t(f)\right), \qquad V_0(f) = f.$$

The following theorem from [Crandall and Liggett, 1971] states that the map $\mu_t^N = \left(Id - \frac{t}{N}\mathbf{H}\right)^{-N}$ converges to the solution of the previous equation if $\mathbf{H}$ is *m-dissipative*.

Definition 2.2.2 (m-dissipativity) *We say that an operator* $(\mathbf{H}, Dom(\mathbf{H}))$ *defined on* $\mathcal{X} \times \mathcal{X}$

is dissipative *if for all* $\beta > 0$*, we have*

$$\|(f - \beta\mathbf{H}(f)) - (g - \beta\mathbf{H}(g))\| \geq \|f - g\|, \quad \forall f, g \in Dom(\mathbf{H}).$$

A dissipative operator $\mathbf{H}$ *is called* m-dissipative *if for all* $\beta > 0$ *the map* $(Id - \beta\mathbf{H})$ *is surjective on* $\mathcal{X}$ *(i.e., the range of* $Id - \beta\mathbf{H}$ *verifies* $\mathcal{R}(Id - \beta\mathbf{H}) = \mathcal{X}$*). We say that an operator* $(\mathbf{H}, Dom(\mathbf{H}))$ *satisfies the* range condition *if for all* $\beta > 0$ *the range of the map* $(Id - \beta\mathbf{H})$ *is dense in* $\mathcal{X}$*.*

It can be shown that the closure $\left(\overline{\mathbf{H}}, Dom\left(\overline{\mathbf{H}}\right)\right)$ (in the product topology of $\mathcal{X} \times \mathcal{X}$) of a dissipative operator $(\mathbf{H}, Dom(\mathbf{H}))$ is itself dissipative and satisfies $\mathcal{R}\left(Id - \beta\overline{\mathbf{H}}\right) = \overline{\mathcal{R}(Id - \beta\mathbf{H})}$. Hence, if an operator $(\mathbf{H}, Dom(\mathbf{H}))$ is dissipative and satisfies the range condition, its closure has the property that $\mathcal{R}\left(Id - \beta\overline{\mathbf{H}}\right) = \mathcal{X}$ for all $\beta > 0$. On the other hand, the map $Id - \beta\overline{\mathbf{H}}$ is injective by the dissipativity of $\overline{\mathbf{H}}$. Hence, we can invert the map and define (as in the well known Hille-Yosida theorem for the linear semigroups case):

$$R\left(\beta, \overline{\mathbf{H}}\right) : \mathcal{X} \to Dom(\overline{\mathbf{H}}) \qquad R\left(\beta, \overline{\mathbf{H}}\right)(f) = \left(Id - \beta\overline{\mathbf{H}}\right)^{-1}(f). \tag{2.6}$$

We present bellow the Crandall-Liggett theorem.

Theorem 2.2.1 *For a densely defined, dissipative operator* $(\mathbf{H}, Dom(\mathbf{H}))$ *on a Banach space* $\mathcal{X}$*, the following are equivalents:*

(i) The closure $\overline{\mathbf{H}}$ *of* $\mathbf{H}$ *generates a contraction semigroup in the sense that there exists*

$$V_t(f) = \lim_{N\to\infty} R\left(\frac{t}{N}, \overline{\mathbf{H}}\right)^N (f),$$

uniformly for t *in compact intervals.* $R\left(\beta, \overline{\mathbf{H}}\right)$ *is the map defined on Equation* (2.6).

(ii) The range condition holds, that is, the range of $(Id - \beta\mathbf{H})$ *is dense in* $\mathcal{X}$ *for some (hence all, as consequence of Lemma* ***5.2*** *from [Feng and Kurtz, 2006])* $\beta > 0$.

We say that $(\mathbf{H}, Dom(\mathbf{H}))$ is densely defined if $Dom(\mathbf{H})$ is dense in $\mathcal{X}$.

In this thesis, the limiting operator $\mathbf{H}$ is defined on the space $\mathcal{X} = C(E)$ (with the supremum norm). For operators defined on function spaces, the verification of the *dissipativity* can often be checked via the positive maximum principle:

Definition 2.2.3 (The positive maximum principle) *Let* E *be a compact subset of* $\mathbb{R}^n$*, and* $\mathbf{H} : Dom(\mathbf{H}) \subset C(E) \to C(E)$ *be an operator. We say that* $\mathbf{H}$ *satisfies the* positive maximum principle *if for any two functions* $f, g \in Dom(\mathbf{H})$*, we have:*

(i) If $x_0 \in E$ is such that $f(x_0) - g(x_0) = \sup_{x\in E} \{f(x) - g(x)\}$, then $\mathbf{H}(f)(x_0) - \mathbf{H}(g)(x_0) \leq 0$.

(ii) If $x_0 \in E$ is such that $f(x_0) - g(x_0) = \inf_{x\in E} \{f(x) - g(x)\}$, then $\mathbf{H}(f)(x_0) - \mathbf{H}(g)(x_0) \geq 0$.

Lemma 2.2.2 *If an operator $(\mathbf{H}, Dom(\mathbf{H}))$ satisfies the positive maximum principle, then it is dissipative.*

Moreover, the dissipativity of $\mathbf{H}$ follows from the fact that $\mathbf{H}$ is the limit of dissipative operators H^N.

On the other hand, checking the *range condition* for non-linear operators might to ve very hard or even impossible. In the cases analysed in this thesis, the non-linear operator $\mathbf{H}$ is such that $\mathbf{H}(f) = H(x, \nabla f(x))$ if $x \in E$ and $f \in C^1(E)$. Then, $Dom(\mathbf{H}) = C^1(E)$, $\mathbf{H} : C^1(E) \to C(E)$, and the range condition consists on proving that for sufficiently many $h \in C(E)$ and $\beta > 0$, there exists $f \in C^1(E)$ such that

$$f - \beta\mathbf{H}(f) = h. \tag{2.7}$$

However the verification of this property can be a formidable obstacle. Consider for example the case where $h \in C(E) \setminus C^1(E)$. If there exists $f \in C^1(E)$ such that $f(x) - \beta H(x, \nabla f(x)) = h(x)$ for all $x \in E$, then $f(x) = h(x) + \beta H(x, \nabla f(x))$. But unless h is canceled by some term of $\beta H(x, \nabla f(x))$ (which in principle is $C^0(E)$ too), this cannot happen because otherwise $f \notin C^1(E)$. One way out is to work with *viscosity solutions* (see Definition 2.3.1). Moreover, due to Theorem **6.14** of [Feng and Kurtz, 2006], it is enough to prove that the *comparison principle* (see Definition 2.3.2) is verified for the Hamilton-Jacobi equation defined by Equation (2.7) since this ensures the uniqueness of the viscosity solution $f \in C(E)$ constructed in the proof of Lemma **6.9**.

If the comparison principle is verified, then the operator $\mathbf{H}$ can be extended to $\hat{\mathbf{H}}$ such that $\hat{\mathbf{H}}$ is m-dissipative and generates a semigroup $\mathbf{V} = \{V_t\}_t$ (see Theorem **8.27** of [Feng and Kurtz, 2006]). As mentioned by [Feng and Kurtz, 2006], the verification of the comparison principle is an analytic issue and often gives the impression of being rather involved and disconnected from the probabilistic large deviations problems. An in-depth study of the comparison principle for Hamilton-Jacobi equations in this context is presented in [Kraaij, 2016], using results from [Crandall et al., 1992] and Chapter 9 of [Feng and Kurtz, 2006]. In Section 2.3, we present the main results from [Crandall et al., 1992], [Feng and Kurtz, 2006] and [Kraaij, 2016] that we use to prove the comparison principle in the cases analysed in this thesis.

Once we have verified the three steps mentioned before, Theorem **6.14** from [Feng and Kurtz, 2006] assures that the sequence of processes $\{X^N\}_N$ (actually, for the process

$\hat{X}_t^N = \eta^N \left(X_t^N\right))$, defined for the continuous or discrete-time Markov processes, is exponentially tight and satisfies an LDP with rate function I defined implicitly in terms of V_t.

Theorem 2.2.3 (6.14 from [Feng and Kurtz, 2006]) *Let $E^N, E \subset \mathbb{R}^n$. Let $\eta^N : E^N \to E$ be Borel measurable, and define $\hat{\eta}^N : B(E) \to B(E^N)$ by $\hat{\eta}^N(f) = f \circ \eta^N$. Assume that $E = \lim_N \eta^N(E^N)$. Assume that one of the following holds:*

1. ***Continuous-time case:*** *For each N, $A^N \subset B(E^N) \times B(E^N)$ and existence and uniqueness holds for the $D_{E^N}[0,\infty)-$ martingale problem for A^N. The process X^N is solution of the martingale problem for A^N, and*

$$H^N(f) = \frac{1}{N} e^{-Nf} A^N \left(e^{Nf}\right), \text{ if } e^{Nf} \in Dom(A^N).$$

2. ***Discrete-time case:*** *For each N, T^N is a transition operator on $B(E^N)$ for a Markov chain, $\varepsilon_N > 0$ and $\varepsilon_N \to 0$, $\left\{Y_k^N\right\}_k$ is a discrete-time Markov chain with time-step ε_N and transition operator T^N, and*

$$H^N(f) = \frac{1}{N\varepsilon_N} \log\left[e^{-Nf} T^N \left(e^{Nf}\right)\right], \qquad f \in B(E^N).$$

Let $\left\{V_t^N\right\}_t$ be the non-linear semigroup generated by H^N. Let $\mathbf{H} \subset C(E) \times B(E)$ with $Dom(\mathbf{H})$ dense in $C(E)$. Suppose that for each $f \in Dom(\mathbf{H})$, there exists $f_N \in Dom(H^N)$ such that $\left\|\hat{\eta}^N(f) - f_N\right\| \to 0$, $\sup_N \left\|H^N(f_N)\right\| < \infty$, and for each $x \in E$ and sequence $x^N \in E^N$ satisfying $\eta^N(x^N) \to x$, it is verified that

$$\lim_N H^N(f_N)(x^N) = \mathbf{H}(f)(x).$$

Fix $\beta_0 > 0$. Suppose that for each $0 < \beta < \beta_0$, there exists a dense subset $D_\beta \subset C(E)$ such that for each $h \in D_\beta$, the comparison principle holds for

$$\left(Id - \beta\mathbf{H}\right)(f) = h.$$

Define $\hat{X}^N = \eta^N(X^N)$. Suppose that the sequence of initial conditions $\left\{\hat{X}_0^N\right\}_N$ satisfies an LDP with good rate function I_0. Then,

1. *The operator $\hat{\mathbf{H}} = \bigcup_\beta \left\{\left(R(\beta,h), \frac{R(\beta,h)-h}{\beta}\right) : h \in C(E)\right\}$ generates a semigroup on $C(E)$ given by $V_t(h) = \lim_N R\left(\frac{t}{N}, h\right)^N$. With abuse of notation, $R(\beta,h)$ is the unique viscosity solution of Equation (2.7) (constructed in the proof of Lemma **6.9**) for $\beta > 0$ and $h \in C(E)$.*

2. *In the continuous-time case,* $\lim_{N\to\infty} \left\| \hat{\eta}^N \left(V_t(f)\right) - V_t^N(f_N) \right\| = 0$, *whenever* $f \in C(E)$, $f_N \in B(E^N)$, *and* $\left\|\hat{\eta}^N(f) - f_N\right\| \to 0$.
3. *In the discrete-time case,*

$$\lim_{N\to\infty} \left\| \hat{\eta}^N \left(V_t(f)\right) - V_t^N(f_N) \right\| + \left\| \hat{\eta}^N \left(V_t(f)\right) - V_{t+\varepsilon_N}^N(f_N) \right\| = 0,$$

whenever $f \in C(E)$, $f_N \in B(E^N)$, *and* $\left\|\hat{\eta}^N(f) - f_N\right\| \to 0$.
4. $\left\{\hat{X}^N\right\}_N$ *is exponentially tight and satisfies an LDP with rate funtion* $I : D_E[0,\infty) \to [0,+\infty]$ *given by*

$$I(\mathbf{x}) = \sup_{\{t_i\}_i \in \Delta_{\mathbf{x}}^c} I_0(\mathbf{x}(0)) + \sum_i I_{t_i - t_{i-1}} \left(\mathbf{x}(t_i) | \mathbf{x}(t_{i-1})\right), \tag{2.8}$$

where $0 \leq t_1 \leq t_2 \leq \cdots \leq t_i \leq \ldots$, $\Delta_{\mathbf{x}} \subset [0,\infty)$ *is the set of continuity points of* $\mathbf{x}$, *and*

$$I_t(y|x) = \sup_{f\in C(E)} \left\{f(y) - V_t(f)(x)\right\}.$$

This is a theoretical result but does not provide an applicable characterization of the rate. The next step provides a simplified version of the rate that can be used in practice.

Step 4: Construct a variational representation for the limiting semigroup $V = \{V_t\}_t$

Typically, we can identify the limiting semigroup $V = \{V_t\}_t$ as the Nisio semigroup for a control problem that is constructed in terms of the limiting operator $\mathbf{H}(f)(x) = H\left(x, \nabla f(x)\right)$. The control problem then gives an alternative and more explicit representation of the rate function.

Assume that $Dom(\mathbf{H}) = C^1(E)$ and $\mathbf{H}$ is such that $\mathbf{H}(f)(x) = H\left(x, \nabla f(x)\right)$, where $H : E \times \mathbb{R}^n$ is continuously differentiable on $\mathring{E} \times \mathbb{R}^n$ and convex w.r.t. the second variable. This step consists of proving that $V_t = \mathcal{V}_t$, where the semigroup $\mathcal{V}_t$ is given as a variational problem where one optimises a pay-off $f\left(\mathbf{x}(t)\right)$, but where a cost is paid that depends on the whole trajectory $\{\mathbf{x}(s) : 0 \leq s \leq t\}$. This cost is accumulated over time and is given in a Lagrangian form.

Consider the Legendre-Fenchel transform of H given by

$$L\left(x,\beta\right) = \sup_{\alpha\in\mathbb{R}^n} \left\{\langle\alpha,\beta\rangle - H\left(x,\alpha\right)\right\} = \sup_{f\in C^1(E)} \left\{\langle\nabla f(x),\beta\rangle - \mathbf{H}(f)(x)\right\}.$$

Since H is continuous and convex w.r.t. α, the Fenchel-Moreau theorem states that L is continuous, convex w.r.t. β, and

$$H(x,\alpha) = \sup_{\alpha\in\mathbb{R}^n}\left\{\langle\alpha,\beta\rangle - L(x,\beta)\right\}$$

(we write $L \leftrightarrow H$ for short). This gives us the following variational representation of $\mathbf{H}$:

$$\mathbf{H}(f)(x) = H(x,\nabla f(x)) = \sup_{\beta\in\mathbb{R}^n}\left\{\langle\nabla f(x),\beta\rangle - L(x,\beta)\right\}. \tag{2.9}$$

Although we formally define the Nisio semigroup in terms of relaxed controls measures, using Jensen's inequality, it can be proved that in this case, the Nisio semigroup $\mathcal{V}_t$ turns out to be

$$\mathcal{V}_t(f)(x_0) = \sup_{\mathbf{x}\in\mathcal{AC}:\,\mathbf{x}(0)=x_0}\left\{f(\mathbf{x}(t)) - \int_0^t L(\mathbf{x}(s),\dot{\mathbf{x}}(s))\,ds\right\}, \tag{2.10}$$

where $\mathcal{AC}$ represents the space of all absolutely continuous functions $\mathbf{x} : [0,\infty) \to E$. The following informal calculation shows why $\mathcal{V}_t$ should be the semigroup generated by $\mathbf{H}$ (i.e., $\mathcal{V}_t = V_t$):

$$\begin{aligned}
\left[\frac{d}{dt}\mathcal{V}_t(f)(x_0)\right]_{t=0} &= \sup_{\mathbf{x}\in\mathcal{AC}:\,\mathbf{x}(0)=x_0}\frac{d}{dt}\left[f(\mathbf{x}(t)) - \int_0^t L(\mathbf{x}(s),\dot{\mathbf{x}}(s))\,ds\right]_{t=0}\\
&= \sup_{\mathbf{x}\in\mathcal{AC}:\,\mathbf{x}(0)=x_0}\left\{\langle\nabla f(x_0),\dot{\mathbf{x}}(0)\rangle - L(x_0,\dot{\mathbf{x}}(0))\right\}\\
&= \sup_{\beta\in\mathbb{R}^n}\left\{\langle\nabla f(x_0),\beta\rangle - L(x_0,\beta)\right\}\\
&= \mathbf{H}(f)(x_0) = \mathbf{H}(\mathcal{V}_0(f))(x_0).
\end{aligned}$$

Next, we formally define the Nisio semigroup $\mathcal{V}_t$. We present conditions **8.9**, **8.10**, and **8.11** from [Feng and Kurtz, 2006], and Theorem 2.2.7, which is a direct consequence of theorems **8.14**, **8.23**, **8.27**, and **8.29** from [Feng and Kurtz, 2006].

Definition 2.2.4 (Control set of a linear operator and Nisio semigroup) *Let U and E be complete and separable metric spaces. Let $A : Dom(A) \subset B(E) \to M(E\times U)$ be a single valued linear operator. Let $\mathcal{M}_m(U)$ be the space of Borel measures λ on $U\times[0,\infty)$ satisfying $\lambda(U\times[0,t]) = t$ for all $t \geq 0$. The measure λ is known as a* relaxed control. *We say that the pair $(\mathbf{x},\lambda) \in D_E[0,\infty)\times\mathcal{M}_m(U)$ satisfies the relaxed control equation for A if and only if:*

1. $\iint_{U\times[0,t]} |A(f)(\mathbf{x}(s),u)|\,\lambda(du\times ds) < \infty\ \forall f \in Dom(A),\ \forall t \geq 0$;

2. $f(\mathbf{x}(t)) - f(\mathbf{x}(0)) = \iint_{U\times[0,t]} A(f)(\mathbf{x}(s),u)\lambda(du\times ds)$ $\forall f \in Dom(A)$, $\forall t \geq 0$.

We denote the collection of pairs satisfying the above properties by $\mathcal{Y}$. If $\Gamma \subset E\times U$, define

$$\mathcal{Y}^{\Gamma} = \left\{(\mathbf{x},\lambda)\in\mathcal{Y} : \iint_{U\times[0,t]} \mathbf{1}_{\Gamma}(\mathbf{x}(s),u)\,\lambda(du\times ds) = t,\ t\geq 0\right\}.$$

The Nisio semigroup corresponding to the control problem determined by the linear operator A and the cost function $-L$ is:

$$\mathcal{V}_t(f)(x_0) = \sup_{\{(\mathbf{x},\lambda)\in\mathcal{Y}^{\Gamma}:\,\mathbf{x}(0)=x_0\}} \left\{f(\mathbf{x}(t)) - \iint_{U\times[0,t]} L(\mathbf{x}(s),u)\lambda(du\times ds)\right\} \tag{2.11}$$

for each $x_0 \in E$ (the supremum of an empty set is defined to be $-\infty$).

Note that the operator A appears in the definition of the control set. Next, we present conditions **8.9**, **8.10**, and **8.11** for the particular case at which $\mathbf{H}_{\ddagger} = \mathbf{H} = \mathbf{H}_{\dagger}$.

Condition 2.2.4 (Condition 8.9 from [Feng and Kurtz, 2006]) *The functions A, L defined before verify:*

1. *$A \subset C(E)\times C(E\times U)$ is single-valued and $Dom(A)$ separates points (that is, for any $x\neq y\in E$, there exists $f\in Dom(A)$ such that $f(x)\neq f(y)$).*
2. *$\Gamma\subset E\times U$ is closed, and for each $x_0\in E$, there exists $(\mathbf{x},\lambda)\in\mathcal{Y}^{\Gamma}$ such that $\mathbf{x}(0)=x_0$.*
3. *$L : E\times U \to [0,\infty]$ is a lower-semicontinuous function, and for each $c>0$ and compact $K\subset E$, the set $\{(x,u)\in\Gamma : L(x,u)\leq c\}\cap(K\times U)$ is relatively compact.*
4. *For each compact $K\subset E$, $T>0$, and $0\leq M<\infty$, there exists a compact $\hat{K} = \hat{K}(K,T,M)\subset E$ such that if $(\mathbf{x},\lambda)\in\mathcal{Y}^{\Gamma}$, $\mathbf{x}(0)\in K$, and*

$$\iint_{U\times[0,T]} L(\mathbf{x}(s),u)\,\lambda(du\times ds)\leq M,$$

   then $\mathbf{x}(t)\in\hat{K}$ for all $0\leq t\leq T$.
5. *For each $f\in Dom(A)$ and compact set $K\subset E$, there exits a right continuous, nondecreasing function $\psi_{f,K} : [0,\infty)\to[0,\infty)$ such that*

$$|A(f)(x,u)| \leq \psi_{f,K}(L(x,u)),\quad \forall (x,u)\in\Gamma\cap(K\times U),$$

   and $\lim_{r\to\infty}\frac{1}{r}\psi_{f,K}(r) = 0$.

Condition 2.2.5 (Condition 8.10 from [Feng and Kurtz, 2006]) *For each $x_0 \in E$, there exists $(\mathbf{x}, \lambda) \in \mathcal{Y}^\Gamma$ such that $\mathbf{x}(0) = x_0$ and*

$$\iint_{U\times[0,\infty)} L\left(\mathbf{x}(s), u\right) \lambda\left(du \times ds\right) = 0.$$

Condition 2.2.6 (Condition 8.11 from [Feng and Kurtz, 2006]) *For each $x_0 \in E$ and $f \in Dom(\mathbf{H})$, there exists $(\mathbf{x}, \lambda) \in \mathcal{Y}^\Gamma$ such that $\mathbf{x}(0) = x_0$, and*

$$\int_0^t \mathbf{H}(f)\left(\mathbf{x}(s)\right) ds = \iint_{U\times[0,t]} \left(A(f)\left(\mathbf{x}(s), u\right) - L\left(\mathbf{x}(s), u\right)\right) \lambda\left(ds \times du\right), \quad t \geq 0.$$

Finally, the much more useful version of the rate function that we present for the cases analysed in this thesis derives from the theorem that we present below. It is a presentation of theorems **8.14**, **8.23**, **8.27** and **8.29** from [Feng and Kurtz, 2006].

Theorem 2.2.7 *Assume that conditions from Theorem 2.2.3 are verified for the continuous or dicrete-time case. If Conditions **8.9**, **8.10** and **8.11** from [Feng and Kurtz, 2006] are also verified, then:*

1. *$V_t(f) = \mathcal{V}_t(f)$ for all $f \in Dom(\mathbf{V})$, where $\mathcal{V}_t = \{\mathcal{V}_t\}_t$ is the Nisio semigroup associated to the cost function $-L$ defined on Equation (2.11).*
2. *$I(\mathbf{x}) = I_0(x(0)) + \inf_{\{\lambda:\, (\mathbf{x},\lambda)\in\mathcal{Y}\}} \left\{\iint_{U\times[0,\infty)} L(\mathbf{x}(s), u)\lambda(du \times ds)\right\}$.*

2.3 Theory of viscosity solutions for Hamilton-Jacobi equations

In this section, we present the tools we use in this thesis to study the uniqueness of viscosity solutions of Hamilton-Jacobi equations. This uniqueness is an essential requirement for the study of large deviations, according to the results presented in the previous section.

Theories of viscosity solutions of (deterministic) differential equations and probability are doubly connected. On the one hand, some viscosity solutions are built from the limit of processes related to "probabilistic games". On the other hand, the study of viscosity solutions of differential equations allows solving probabilistic problems. The work of [Feng and Kurtz, 2006] focuses on this second aspect. That is, they use the theory of viscosity solutions to prove the existence of the semigroup $\{V_t\}_t$ associated with the non-linear operator $\mathbf{H}$.

The theory of viscosity solutions applies to certain differential equations of the form $F\left(x, u(x), Du(x), \ldots, D^k u(x)\right) = 0$. The primary virtues of this theory are that it allows merely continuous (or semi-continuous) functions to be solutions of fully nonlinear differential equations. In the expression $F\left(x, u(x), Du(x), \ldots, D^k u(x)\right)$, u is a real-valued function defined on some subset $E \subset \mathbb{R}^n$ (or even on a metric space), and $Du(x), \ldots, D^k u(x)$ corresponds to derivatives of u. However, these derivatives do not have a classical meaning, as we present below.

Let $E \subset \mathbb{R}^n$ be some compact set. Consider a function $F : E \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$. It is known that for many equations it is not possible to solve the first-order differential equation:

$$F\left(x, u(x), \nabla u(x)\right) = 0 \tag{2.12}$$

in a classical sense. A well known example is the Eikonal equation on $E = [-1, 1]$, given by

$$\begin{cases} |u'(x)| - 1 = 0, \\ u(-1) = u(1) = 0. \end{cases}$$

Classical solutions to this problem do not exist due to Rolle's theorem, then solutions are sought in a relaxed sense. For this example, there exists infinitely many solutions that solve this equation almost everywhere. For example, $u_1(x) = 1 - |x|$ or $u_2(x) = |x| - 1$, but only one of them is a viscosity solution in the sense that we define below.

Definition 2.3.1 (Viscosity solutions) *We say that u is a* viscosity subsolution *of Equation (2.12) if it is bounded, upper-semicontinuous and if for every $\phi \in C^1(E)$ and $x_0 \in E$ such that $u(x_0) - \phi(x_0) = \sup\limits_{x \in E} \{u(x) - \phi(x)\}$, it is verified that:*

$$F\left(x_0, u(x_0), \nabla\phi(x_0)\right) \leq 0.$$

We say that v is a viscosity supersolution *of Equation (2.12) if it is bounded, lower-semicontinuous and if for every $\psi \in C^1(E)$ and $x_0 \in E$ such that $v(x_0) - \psi(x_0) = \inf\limits_{x \in E} \{v(x) - \psi(x)\}$, it is verified that:*

$$F\left(x_0, v(x_0), \nabla\psi(x_0)\right) \geq 0.$$

We say that u is a viscosity solution *of Equation (2.12) if it is both a sub and supersolution.*

Note that an upper semi-continuous function always attains its supremum in a compact set (analogously, a lower semi-continuous function always attains its infimum). Consider the case at which $u(x_0) - \phi(x_0) = 0$, then Definition 2.3.1 says that u is a subsolution of Equation (2.12) if every time we touch it from above in $(x_0, u(x_0))$ by a function $\phi \in C^1(E)$, it is verified that

$F(x_0, u(x_0), \nabla\phi(x_0)) \leq 0$. Notice that we replace the derivatives of u (which might not exist) with the derivatives of ϕ. Analogously, v is a supersolution if every time we touch it from below in $(x_0, v(x_0))$ by a function $\psi \in C^1(E)$, it is verified that $F(x_0, v(x_0), \nabla\psi(x_0)) \geq 0$.

Remark 2.3.1 Let us introduce some comments about the definition of viscosity solutions.

1. There are other definitions of viscosity solutions in the literature for higher-order differential equations and the case in which the set E is not compact (so there would not have to exist x_0 such that $u(x_0) - \phi(x_0) = \sup_{x\in E}\{u(x) - \phi(x)\}$).
2. Since a viscosity solution u is an upper and lower-semicontinuous function, it must be a bounded and continuous function, which contrasts with the weak solution method based on Sobolev spaces. Sobolev spaces require that the function and its derivatives belong to L^p (to integrate them).
3. Of course, if the differential equation has a classical solution, this solution will also be a solution in the viscosity sense that we have just defined.

In the case of the Eikonal equation, the only viscosity solution turns out to be $u_1(x) = 1-|x|$. Since it is differentiable everywhere except in $x = 0$, the point of interest turns out to be $x_0 = 0$. Since every function $\phi \in C^1(E)$ that touches u_1 at $(0, 1)$ from above has to verify $\phi'(0) \in [-1, 1]$, u_1 is a viscosity subsolution of the Eikonal equation. Moreover, there exists no $\psi \in C^1(E)$ that touches u_1 from below at the point $(0, 1)$, which implies that u_1 is a viscosity supersolution too. With a similar argument, it can be proved that u_2 is a subsolution, but it can't be a supersolution since for any $\psi \in C^1(E)$ that touches u_2 below at the point $(0, -1)$, it is verified that $\psi'(0) \in (-1, 1)$. The uniqueness of u_1 as viscosity solution of the Eikonal equation is established via the *comparison principle* that we defined below.

Definition 2.3.2 (Comparison Principle) *We say that Equation (2.12) satisfies the* comparison principle *if for any subsolution u and supersolution v it is verified that $u \leq v$.*

Remark 2.3.2 Note that if the comparison principle is satisfied, then a viscosity solution is unique: if u_1 and u_2 are viscosity solutions, then $u_1 \leq u_2$ since u_1 is subsolution and u_2 is supersolution. Moreover, $u_2 \leq u_1$ since u_2 is subsolution and u_1 is supersolution. Then, $u_1 = u_2$.

Next, we state the results from [Crandall et al., 1992], [Kraaij, 2016], and Chapter 9 from [Feng and Kurtz, 2006] to prove the comparison principle for the cases analysed in this thesis. As mentioned before, in these cases, the Hamilton-Jacobi equations are of the form

$$f(x) - \beta H(x, \nabla f(x)) - h(x) = 0, \quad x \in E, \tag{2.13}$$

where $h \in C(E)$ and $\beta > 0$ are fixed, $E \subset \mathbb{R}^n$ is compact, and $H : E \times \mathbb{R}^n \to \mathbb{R}$ is continuous and convex w.r.t. the second variable. Then, in this case, the function F is $F_{\beta,h}(x, u, \alpha) = u - \beta H(x, \alpha) - h(x)$, and the Hamilton-Jacobi equation is

$$F_{\beta,h}(x, f(x), \nabla f(x)) = 0, \quad x \in E.$$

Let u and v be respectively a viscosity subsolution and supersolution of Equation (2.13). We want to prove that $u - v \leq 0$. The first step to prove this inequality, first proposed by [Crandall et al., 1992] and then in Chapter 9 of [Feng and Kurtz, 2006] for more general spaces, consists in constructing sequences x_α and y_α ($\alpha \to \infty$) that converge to a maximising point $z \in E$ such that $u(z) - v(z) = \sup_{x \in E} \{u(x) - v(x)\}$.

Proposition 2.3.1 ([Crandall et al., 1992], Lemma 3.1 and Proposition 3.7) *Let E be a compact subset of $\mathbb{R}^n$, let $u : E \to \mathbb{R}$ be an upper-semicontinous function, $v : E \to \mathbb{R}$ be a lower-semicontinous function, and let $\psi : E \times E \to \mathbb{R}$ be a lower-semicontinuous function such that $\psi \geq 0$ and $\psi(x, y) = 0$ if and only if $x = y$. For $\alpha > 0$ ($\alpha \to \infty$), let x_α, $y_\alpha \in E$ such that* [1]*:*

$$u(x_\alpha) - v(y_\alpha) - \alpha\psi(x_\alpha, y_\alpha) = \sup_{x,y \in E} \{u(x) - v(y) - \alpha\psi(x, y)\}.$$

Then, the following hold:

1. $\lim_{\alpha \to \infty} \alpha\psi(x_\alpha, y_\alpha) = 0$,
2. *all the limiting points of the sequence (x_α, y_α) are of the form (z, z), and for these points it is verified that $u(z) - v(z) = \sup_{x \in E} \{u(x) - v(x)\}$.*

The entire Chapter 9 of [Feng and Kurtz, 2006] is devoted to extend the conditions imposed in [Crandall et al., 1992] to ensure the comparison principle for more general spaces, even for the case in which there is not a limit operator $\mathbf{H}$, but there is an upper bound $\mathbf{H}_\dagger$ and a lower bound $\mathbf{H}_\ddagger$, which are presented as graphs (instead of operators). Since, in our case, there exists a limit operator $\mathbf{H}$, and it verifies $\mathbf{H}(f)(x) = H(x, \nabla f(x))$, with $H(x, \alpha)$ continuous and convex w.r.t. α, we use the simplest conditions presented as Proposition 2 and Lemma 5 in [Kraaij, 2016].

Definition 2.3.3 (Good penalization function) *The function $\psi : E \times E \to \mathbb{R}$ is a* good penalization function *if $\psi \geq 0$, $\psi(x, y) = 0$ if and only if $x = y$, it is continuously differentiable in both components, and if $\psi_x(x, y) = -\psi_y(x, y)$ for all $x, y \in E$. ψ_x and ψ_y denote the derivatives of ψ w.r.t. x and y.*

[1] There exist since $E \times E$ is compact and the function $g_\alpha(x, y) = u(x) - v(y) - \alpha\psi(x, y)$ is upper-semicontinuous.

Proposition 2.3.2 (Proposition 2 and Lemma 5 from [Kraaij, 2016])
Let $(\mathbf{H}, Dom(\mathbf{H}))$ be an operator such that $Dom(\mathbf{H}) = C^1(E)$ and $\mathbf{H}(f)(x) = H(x, \nabla f(x))$. Let u be a subsolution and v a supersolution of Equation (2.13) for some $\beta > 0$ and $h \in C(E)$. Let ψ be a good penalization function and let x_α, y_α $(\alpha \to \infty)$ satisfying

$$u(x_\alpha) - v(y_\alpha) - \alpha\psi(x_\alpha, y_\alpha) = \sup_{x,y\in E} \{u(x) - v(y) - \alpha\psi(x,y)\}.$$

1. *If*
$$\liminf_{\alpha\to\infty} H(x_\alpha, \alpha\psi_x(x_\alpha, y_\alpha)) - H(y_\alpha, \alpha\psi_x(x_\alpha, y_\alpha)) \le 0, \tag{2.14}$$
then $u \le v$. i.e. Equation (2.13) satisfies the comparison principle.
2. *Moreover, $\sup_\alpha H(y_\alpha, \alpha\psi_x(x_\alpha, y_\alpha)) < \infty$.*

Since the proof of this proposition is typical for the comparison principle problem, we include the proof of the first part of this proposition for completeness.

Proof. Let $\beta > 0$ and $h \in C(E)$ be fixed. Let u be a subsolution and v a supersolution to Equation (2.13). We want to prove that $\sup_{x\in E}\{u(x) - v(x)\} \le 0$ (i.e., $u \le v$). Suppose by contradiction that $\delta = \sup_{x\in E}\{u(x) - v(x)\} > 0$. For $\alpha > 0$ $(\alpha \to \infty)$, let x_α, y_α be such that

$$u(x_\alpha) - v(y_\alpha) - \alpha\psi(x_\alpha, y_\alpha) = \sup_{x,y\in E} \{u(x) - v(y) - \alpha\psi(x,y)\}.$$

Since $\alpha\psi(x_\alpha, y_\alpha) \to 0$ and for any limiting point z we have $u(z) - v(z) = \sup_{x\in E}\{u(x) - v(x)\} = \delta > 0$, then for α large enough, $u(x_\alpha) - v(y_\alpha) \ge \frac{\delta}{2}$.

For every α, the map $\phi_{1,\alpha}(x) := v(y_\alpha) + \alpha\psi(x, y_\alpha)$ is in $C^1(E)$ and $u(x) - \phi_{1,\alpha}(x)$ has a maximum at x_α.

On the other hand, $\phi_{2,\alpha}(y) := u(x_\alpha) - \alpha\psi(x_\alpha, y)$ is also in $C^1(E)$ and $v(y) - \phi_{2,\alpha}(y)$ has a minimum at y_α.

As u is subsolution and v is supersolution to Equation (2.13), we have:

$$\frac{u(x_\alpha) - h(x_\alpha)}{\beta} \le H(x_\alpha, \nabla\phi_{1,\alpha}(x_\alpha)) = H(x_\alpha, \alpha\psi_x(x_\alpha, y_\alpha)), \tag{2.15}$$

and

$$\frac{v(y_\alpha) - h(y_\alpha)}{\beta} \ge H(y_\alpha, \nabla\phi_{2,\alpha}(y_\alpha)) = H(y_\alpha, -\alpha\psi_y(x_\alpha, y_\alpha)) = H(y_\alpha, \alpha\psi_x(x_\alpha, y_\alpha)). \tag{2.16}$$

It follows that for α large enough we have

$$0 < \frac{1}{\beta}\frac{\delta}{2} \leq \frac{u\left(x_\alpha\right) - v\left(y_\alpha\right)}{\beta} = \frac{u\left(x_\alpha\right) - h\left(x_\alpha\right)}{\beta} - \frac{v\left(y_\alpha\right) - h\left(y_\alpha\right)}{\beta} + \frac{1}{\beta}\left(h\left(x_\alpha\right) - h\left(y_\alpha\right)\right) \tag{2.17}$$

$$\leq H\left(x_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) - H\left(y_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) + \frac{1}{\beta}\left(h\left(x_\alpha\right) - h\left(y_\alpha\right)\right). \tag{2.18}$$

As $h \in C(E)$, $\lim_\alpha \left(h\left(x_\alpha\right) - h\left(y_\alpha\right)\right) = 0$ (since $x_\alpha, y_\alpha \to z$). Together with the assumption of the proposition, we find that the lim inf as $\alpha \to \infty$ is bounded above by 0, which contradicts the assumption that $\delta > 0$. $\square$

Then, for the cases analysed in this thesis, it is enough to prove that the function $H\left(x, \alpha\right)$ verifies the inequality presented in Equation (2.14).

2.4 Rate function optimization

Since in the cases analysed in this thesis, the large deviation rate function can be written as an action functional, in this section, we present tools from the calculus of variations to find the trajectory that minimises the rate function over a set of trajectories (i.e. the optimal trajectory).

We focus in to solve the following optimization problem: given a set of possible trajectories $A \subset D_E[0,1]$ and a functional $I : D_E[0,1] \to [0,\infty]$ given by

$$I\left(\mathbf{x}\right) = \begin{cases} \int_0^1 L\left(\mathbf{x}(t), \dot{\mathbf{x}}(t)\right) dt, & \text{if } \mathbf{x} \in \mathcal{AC}, \\ +\infty, & \text{in other cases,} \end{cases}$$

we want to found the trajectory (or trajectories) that solves

$$(P_1) \quad \inf_{\mathbf{x} \in A} I(\mathbf{x}). \tag{2.19}$$

We use the notation $\mathcal{AC}$ to refer to the set of all absolutely continuous functions $\mathbf{x} : [0,1] \to E$. L is a *cost function* given in a *Lagrangian* form $L : E \times \mathbb{R}^n \to \mathbb{R}$ such that

$$L\left(x, \beta\right) = \sup_{\alpha \in \mathbb{R}^n} \left\{\langle \alpha, \beta \rangle - H\left(x, \alpha\right)\right\}.$$

The function $H : E \times \mathbb{R}^n \to$ is differentiable on $\mathring{E} \times \mathbb{R}^n$ and convex w.r.t. α. As $\alpha \mapsto H\left(x, \alpha\right)$

is convex and continuous, it follows by the Fenchel-Moreau theorem that also

$$H(x,\alpha) = \sup_{\beta\in\mathbb{R}^n} \{\langle\alpha,\beta\rangle - L(x,\beta)\}.$$

We use the notation $H \leftrightarrow L$ for short and call H the *Hamiltonian* function.

Techniques from classical mechanics can be used to obtain information about the rate function I which would be very difficult to get from the law of the related stochastic process itself.

Euler-Lagrange equations, presented in Equation (2.20), give us conditions for a curve $\mathbf{x}$ to be a *stationary curve* for the functional I (see [Arnold, 1987], for example):

$$L_x(\mathbf{x},\dot{\mathbf{x}}) - \frac{\mathrm{d}}{\mathrm{d}t} L_\beta(\mathbf{x},\dot{\mathbf{x}}) = 0 \qquad \text{(Euler-Lagrange)}. \tag{2.20}$$

Here L_x and L_β denote the derivatives of L w.r.t. the first and second coordinate. Note that Equation (2.20) turns out to be a second-order differential equation.

In many cases (as in Chapter 5), we cannot find the function L explicitly, but we do explicitly have the Hamiltonian function H. Then, we can switch Equation (2.20) to the easier first-order *Hamilton equations* by doubling the dimension of the problem. By adding the variable $\alpha(t) = L_\beta(\mathbf{x}(t),\dot{\mathbf{x}}(t))$, and rewriting the Euler-Lagrange equations, we find that $(\mathbf{x}(t),\alpha(t))$ must satisfy the Hamilton equations (see [Arnold, 1987], for example):

$$\begin{cases} \dot{\mathbf{x}} = H_\alpha(x,\alpha), & \text{(Hamilton)} \\ \dot{\alpha} = -H_x(x,\alpha). \end{cases} \tag{2.21}$$

Similar to the notation for L, H_x and H_α denote the derivatives of H w.r.t. the first and second coordinate.

Then, problem (P_1) becomes problem (P_2), where, in our case, the closure of the set $A \subset D_E[0,1]$ is considered w.r.t. the Skorohod topology in $D_E[0,1]$:

$$(P_2) \quad \inf\left\{I(x_\alpha) : x_\alpha \text{ is solution of Equation (2.21) and } x_\alpha \in \bar{A}\right\}. \tag{2.22}$$

2.5 Fluid limit

In this section, we show that the fluid limit of a sequence of Markov processes can be deduced from the study of the large deviations.

In each chapter of this thesis, we deduce fluid limit results as a corollary from the LDP for the sequences of Markov processes of interest. These results are nothing more than a particular case of the theorem that we recall below, which ensures that large deviations imply almost-sure convergence to the set of minimizers of the rate function.

Proposition 2.5.1 (LD imply almost-sure convergence) *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and* $\{X_N\}_N$, $X_N : \Omega \to \mathcal{X}$, *be a sequence of random variables in a complete and separable metric space* $(\mathcal{X}, d)$. *Assume that* $\{X_N\}_N$ *satisfies an LDP with good rate function* $I : \mathcal{X} \to [0, \infty]$. *Then,*

$$d\left(X_N, \{I = 0\}\right) \to 0 \text{ almost-sure as } N \to \infty,$$

where $\{I = 0\} = \{x \in \mathcal{X} : I(x) = 0\}$ *is the set of minimizers of* I.

Since only the definition of large deviations and Borel-Cantelli lemma are required to prove this theorem, we include a simple proof of this result for completeness.

Proof. It is enought to prove that the sum $\sum_{N=1}^{\infty} \mathbb{P}\left(d\left(X_N, \{I = 0\}\right) \geq \varepsilon\right)$ is finite, since Borel-Cantelli Lemma assures that $\mathbb{P}\left(\{d\left(X_N, \{I = 0\}\right) \geq \varepsilon\} \text{ infinitely often }\right) = 0$, and almost-sure convergence follows as

$$\begin{aligned}
\mathbb{P}\left(d\left(X_N, \{I = 0\}\right) \to 0 \text{ is not true}\right) &= \mathbb{P}\left[\bigcup_{k=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{N \geq m} \left\{d\left(X_N, \{I = 0\}\right) \geq \frac{1}{k}\right\}\right] \\
&\leq \sum_k \mathbb{P}\left[\bigcap_{m=1}^{\infty} \bigcup_{N \geq m} \left\{d\left(X_N, \{I = 0\}\right) \geq \frac{1}{k}\right\}\right] \\
&= \sum_k \mathbb{P}\left[\left\{d\left(X_N, \{I = 0\}\right) \geq \frac{1}{k}\right\} \text{ infinitely often }\right] \\
&= 0.
\end{aligned}$$

Then, we show that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for large N, we have:

$$\mathbb{P}\left[d\left(X_N, \{I = 0\}\right) \geq \varepsilon\right] \leq e^{-N\frac{\delta}{2}}.$$

Let $\varepsilon > 0$. The set $U_\varepsilon = \{x \in \mathcal{X} : d\left(x, \{I = 0\}\right) < \varepsilon\}$ is an open neighborhood of the set of minimizers $\{I = 0\}$. Since I is a good rate function, for any open nighborhood U of the set of minimizers, there exists a $\delta > 0$ such that $\{I \leq \delta\} \subset U$. Thus, there exists a $\delta > 0$ such that $\{I \leq \delta\} \subset U_\varepsilon$. By the LD upper bound, for any closed set B and any $\eta > 0$, there exists an integer $N_0 = N_0(\eta)$ such that for all $N \geq N_0$, we have:

$$\mathbb{P}\left[X_N \in B\right] \leq e^{-N(I(B) - \eta)},$$

where $I(B) = \inf_{x \in B} I(x)$. As $\mathcal{X} \setminus U_\varepsilon$ is closed and is contained in $\{I > \delta\}$, we obtain for $N_0 = N_0\left(\frac{\delta}{2}\right)$ large enough that for all $N \geq N_0$,

$$\mathbb{P}\left[X_N \in \mathcal{X} \setminus U_\varepsilon\right] \leq e^{-N\left(I(\mathcal{X}\setminus U_\varepsilon)-\frac{\delta}{2}\right)} \leq e^{-N\left(I(\{I>\delta\})-\frac{\delta}{2}\right)} \leq e^{-N\frac{\delta}{2}},$$

since $\inf_{x\in\{I>\delta\}} I(x) \geq \delta$. $\square$

In our case, the random variables of interest are stochastic processes, more precisely, Markov processes. Then, the previous theorem assures that the fluid limit for these processes is contained in the set of minimizers $\{\mathbf{x} \in D_E[0,1] : I(x) = 0\}$ of the good rate function $I : D_E[0,1] \to [0, +\infty]$. We use the term *fluid limit* (first introduced in [Kurtz, 1971]) to refer to results of the type of the law of large numbers for sequences of Markov processes.

2.6 The theory of Freidlin and Wentzell in the context of the work of Feng and Kurtz

Finnaly, since it is impossible to talk about large deviations for Markov processes without mentioning the pioneering work of [Freidlin and Wentzell, 1984], we present in this section a synthesis of their results in the context of the work of [Feng and Kurtz, 2006].

Discontinuous Markov processes which can be considered as a result of random perturbations of dynamical systems arise in various problems. Freidlin and Wentzell consider a sequence of diffusion processes $\left\{X^N\right\}_N$, with $X_t^N \in \mathbb{R}^n$ (n fixed), satisfying the Itô equation:

$$X_t^N = x + \frac{1}{\sqrt{N}} \int_0^t \sigma\left(X_s^N\right) dW_s + \int_0^t b\left(X_s^N\right) ds,$$

where $\{W_s\}_s$ is a standar Brownian motion, $b : \mathbb{R}^n \to \mathbb{R}^n$ is such that $b(x) = (b_1(x), \ldots, b_n(x))$, and $\sigma : \mathbb{R}^n \to \mathcal{M}_{n\times n}(\mathbb{R})$ is such that $\sigma(x) = \left(\sigma_{i,j}(x)\right)_{i,j}$ (here $\mathcal{M}_{n\times n}(\mathbb{R})$ is the set of real matrices $n \times n$). For large values of N, the corresponding Markov process $\left\{X_t^N\right\}_t$ is essentially a solution of the ordinary differential equation

$$\dot{\mathbf{x}} = b(\mathbf{x}),$$

and the corresponding large deviation theory is concerned with the probabilities of sample paths significantly different from the solution of this equation. [Freidlin and Wentzell, 1984] impose conditions to a Lagrangian function L related to the process $X_\cdot^N$ in order to prove an LDP. However, we don't have the function L explicitly in many cases, so it can't be easy to verify these conditions. From [Feng and Kurtz, 2006], we can analyse those conditions from the point of view of the function H.

Next, we present results for this type of processes, from the simplest case of perturbation of a Brownian motion to the more general case considered in [Freidlin and Wentzell, 1984].

Schilder theorem

Let $\{W_t\}_{t\in[0,1]}$ denote a standard Brownian motion in $\mathbb{R}^n$. Consider the process $W_t^\varepsilon = \sqrt{\varepsilon} W_t$ with $\varepsilon \to 0$, and let ν^ε be the probability measure induced by W^ε on $C_0[0,1]$, the space of all continuous functions $\mathbf{x} : [0,1] \to \mathbb{R}^n$ such that $\mathbf{x}(0) = 0$, equipped with the supremum norm topology.

Starting from approximating the process W_t^ε by an average of random variables with normal distribution and Cramér's theorem (see Mogulskii's theorem in [Dembo and Zeitouni, 1998], for example), Schilder's theorem states that the sequence of measures $\{\nu^\varepsilon\}_\varepsilon$ satisfies an LDP in $\mathcal{X} = C_0[0,1]$ with good rate function $I : \mathcal{X} \to [0,+\infty]$ given by

$$I(\mathbf{x}) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{\mathbf{x}}(t)|^2 \, dt, & \text{if } \mathbf{x} \in H_1 := \left\{ \int_0^t f(s) ds \, : \, f \in L^2[0,1] \right\}, \\ +\infty, & \text{otherwise.} \end{cases}$$

$L^2[0,1]$ is the space of square integrable functions $f : [0,1] \to \mathbb{R}^n$, and $|.|$ denotes the usual Euclidean norm in $\mathbb{R}^n$.

Schilder theorem plus Contraction principle

Let $\{X_t^\varepsilon\}_{t\in[0,1]}$ be the diffusion process that is the unique solution of the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t^\varepsilon = b\left(X_t^\varepsilon\right) dt + \sqrt{\varepsilon}\mathrm{d}W_t, \\ X_0^\varepsilon = 0, \end{cases} \tag{2.23}$$

where $b : \mathbb{R}^n \to \mathbb{R}^n$ is a uniformly continuous function. The existence and uniqueness of the strong solution of Equation (2.23) is standard.

Let $\tilde{\mu}^\varepsilon$ denote the probability measure induced by X_t^ε on $C_0[0,1]$. Then, $\tilde{\mu}^\varepsilon = \mu^\varepsilon \circ F^{-1}$, where μ^ε is the measure induced by $\{\sqrt{\varepsilon} W_t\}_t$, and $F : C_0[0,1] \to C_0[0,1]$ is the deterministic function given by $F(g) = f$ if f is the unique continuous solution of

$$f(t) = \int_0^t b\left(f(s)\right) ds + g(t) \quad \forall t \in [0,1].$$

Since F is a continuous function, the LDP associated with $\{X^\varepsilon\}_\varepsilon$ is a direct application of the *contraction principle* (see [Dembo and Zeitouni, 1998], for example) w.r.t. the map F. $\{X^\varepsilon\}_\varepsilon$ satisfies an LDP in $\mathcal{X} = C_0[0,1]$ with good rate function $I : \mathcal{X} \to [0,+\infty]$ given by

$$I(\mathbf{x}) = \begin{cases} \frac{1}{2}\int_0^1 \left|\dot{\mathbf{x}}(t) - b(\mathbf{x}(t))\right|^2 dt, & \text{if } \mathbf{x} \in H_1, \\ +\infty, & \text{if } \mathbf{x} \notin H_1. \end{cases}$$

LDP for diffusion processes with non-constant diffusion coefficients (Chapters 3 and 4 from [Freidlin and Wentzell, 1984])

Now, let $\left\{X_t^N\right\}_{t\in[0,1]}$ be the diffusion process that is the unique solution of the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t^N = b\left(X_t^N\right)dt + \frac{1}{\sqrt{N}}\sigma\left(X_t^N\right)\mathrm{d}W_t, \\ X_0^N = x, \end{cases} \tag{2.24}$$

where $x \in \mathbb{R}^n$ is fixed, $b : \mathbb{R}^n \to \mathbb{R}^n$ is a uniformly Lipschitz continuous function, all elements of the diffusion matrix σ are bounded, uniformly Lipschitz continuous functions, and W_t is a standar Brownian motion in $\mathbb{R}^n$ (the existence and uniqueness of the strong solution $\left\{X_t^N\right\}_t$ of Equation (2.24) is standard). Note that the map defined by the process $X_\cdot^N$ on $C[0,1]$ is measurable but need not be continuous, therefore the contraction principle can not be used directly. Indeed, this non-continuity is strikingly demonstrated by the fact that the solution of Equation (2.24), when W_t is replaced by its polygonal approximation, differs in the limit from $X_\cdot^N$ by a non-zero correction term. On the other hand, this correction term is of the orden of $\frac{1}{N}$, so it is non expected to influence the LD results. In [Freidlin and Wentzell, 1984] it is proved that the solution of Equation (2.24) satisfies an LDP on $\mathcal{X} = C[0,1]$ with good rate function $I : \mathcal{X} \to [0,\infty]$ given by

$$I_x(\mathbf{x}) = \inf\left\{\frac{1}{2}\int_0^1 |\dot{g}(t)|^2\, dt : \ g \in H_1, \text{ and } \mathbf{x}(t) = x + \int_0^t b(\mathbf{x}(s))\,ds + \int_0^t \sigma(\mathbf{x}(s))\,\dot{g}(s)ds\right\}, \tag{2.25}$$

where the infimum over an empty set is taken as $+\infty$.

Let $a(x) = \sigma(x)\cdot\sigma^t(x)$, where $\sigma^t(x)$ denotes the transpose of the matrix (or vector) $\sigma(x)$, and $\cdot$ denotes the product of matrices. Note that for this case, the linear generator of the process $X_\cdot^N$ is

$$A^N(f)(x) = \frac{1}{2N}\sum_{i,j} a_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x),$$

where we can take the domain of A^N to be the space of twice continuously differentiable functions with compact support $C_c^2(\mathbb{R}^n)$. Then, the non-linear generator suggested by

[Feng and Kurtz, 2006] is

$$H^N(f)(x) = \frac{1}{2N}\sum_{i,j} a_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \frac{1}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial f}{\partial x_i}(x)\frac{\partial f}{\partial x_j}(x) + \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x).$$

Consequently, the limiting non-linear generator is $\mathbf{H}(f) = \lim_{N\to\infty} H^N(f)$ such that

$$\mathbf{H}(f)(x) = \frac{1}{2}\left(\nabla f(x)\right)^t \cdot a(x)\cdot \nabla f(x) + (b(x))^t \cdot \nabla f(x) = H\left(x, \nabla f(x)\right),$$

being $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that

$$H\left(x,\alpha\right) = \frac{1}{2}\alpha^t \cdot a(x)\cdot\alpha + (b(x))^t \cdot \alpha = \frac{1}{2}\left|\sigma^t(x)\cdot\alpha\right|^2 + (b(x))^t\cdot\alpha.$$

Since $H\left(x,\alpha\right)$ is convex w.r.t. α, $H \leftrightarrow L$, being $L\left(x,\beta\right) = \sup_{\alpha\in\mathbb{R}^n}\left\{\alpha\cdot\beta - H\left(x,\alpha\right)\right\}$. Therefore, the limiting non-linear generator can be written as

$$\mathbf{H}(f)(x) = \sup_{u\in\mathbb{R}^n}\left\{A(f)(x,u) - L(x,u)\right\},$$

where $A(f)(x,u) = \left(\nabla f(x)\right)^t \cdot u$ for $f \in C_c^2\left(\mathbb{R}^n\right)$. Applying Corollary **8.28** from [Feng and Kurtz, 2006], the rate function for the sequence $\left\{X^N_\cdot\right\}$, where $X^N_\cdot$ is solution of Equation (2.24), is

$$I(\mathbf{x}) = \inf_{\{u:(x,u)\in\mathcal{Y}\}} \int_0^1 L\left(\mathbf{x}(s), u(s)\right)ds,$$

where $\mathcal{Y}$ is the collection of solutions of the functional equation:

$$f\left(\mathbf{x}(t)\right) - f(x) = \int_0^t A(f)\left(\mathbf{x}(s), u(s)\right)ds, \quad \forall f\in C_c^2\left(\mathbb{R}^n\right), 0\le t\le 1.$$

In the present setting, this equation reduces to $\dot{x}(t) = u(t)$, then the rate function can be written as

$$I(\mathbf{x}) = \begin{cases} \int_0^1 L\left(\mathbf{x}(s), \dot{\mathbf{x}}(s)\right)ds, & \text{if } \mathbf{x}\in\mathcal{AC}, \\ +\infty, & \text{otherwise.}\end{cases}$$

One can use other variational representations of the operator $\mathbf{H}$ and arrive at different expressions for the rate function I. For example, if we choose

$$A(f)(x,u) = u^t\cdot\left(\sigma^t(x)\cdot\nabla f(x)\right) + (b(x))^t\cdot\nabla f(x), \quad \forall f\in C_c^2\left(\mathbb{R}^n\right),$$

$L(x,u) = \frac{1}{2}|u|^2$, and define $\mathcal{H}(f)(x) = \sup_{u\in\mathbb{R}^n} \{A(f)(x,u) - L(x,u)\}$, then $\mathbf{H} = \mathcal{H}$ and the rate function can be expressed as

$$I_x(\mathbf{x}) = \inf \left\{ \frac{1}{2}\int_0^1 |u(t)|^2 \, dt : \; u \in L^2[0,1], \text{ and } \mathbf{x}(t) = x + \int_0^t b\left(\mathbf{x}(s)\right) ds + \int_0^t \sigma\left(\mathbf{x}(s)\right) u(s) ds \right\},$$

which coincides with the form of the rate presented in Equation (2.25).

More general case of Freidlin-Wentzell theory

More generally, we can consider $\left\{X_t^N\right\}_{t\geq 0}$ to be a Markov process in $\mathbb{R}^n$ with right continuous trajectories and infinitesimal generator A^N defined for twice continuously differentiable functions with compact support by the formula

$$A^N(f)(x) = N \int_{\mathbb{R}^n} \left(f\left(x + \frac{1}{N}z\right) - f(x) - \frac{1}{N} z^t \cdot \nabla f(x) \right) \eta\left(x, dz\right) \tag{2.26}$$

$$+ \frac{1}{2N} \sum_{i,j} a_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \left(b(x)\right)^t \cdot \nabla f(x), \tag{2.27}$$

where for each $x \in \mathbb{R}^n$, $\eta\left(x,\cdot\right)$ is a measure defined on $\mathcal{B}\left(\mathbb{R}^n\right)$ such that

$$\int_{\mathbb{R}^n\setminus\{0\}} |z|^2 \, \eta\left(x, dz\right) < +\infty,$$

and for each $B \in \mathcal{B}\left(\mathbb{R}^n\right)$, $\eta\left(\cdot, B\right)$ is a Borel-measurable function. Note that, if X^N corresponds to A^N, then $X_t^N = \frac{1}{N} X_{Nt}$, where the process $\left\{X_t\right\}_t$ has the generator A given by

$$A(f)(x) = \int_{\mathbb{R}^n} \left(f\left(x+z\right) - f(x) - z^t \cdot \nabla f(x) \right) \eta\left(x, dz\right)$$

$$+ \frac{1}{2} \sum_{i,j} a_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \left(b(x)\right)^t \cdot \nabla f(x).$$

In Chapter 5 of [Freidlin and Wentzell, 1984], the following recipe is presented to study the large deviations in this case: assume that for all $\alpha = (\alpha_1, \dots, \alpha_n)$, the expression

$$H(x,\alpha) = \sum_i b_i(x)\alpha_i + \frac{1}{2}\sum_{i,j} a_{i,j}(x)\alpha_i\alpha_j \tag{2.28}$$

$$+ \int\limits_{\mathbb{R}^n\setminus\{0\}} \left(e^{\sum_i \alpha_i z_i} - 1 - \sum_i \alpha_i z_i \right) \eta(x,dz) \tag{2.29}$$

is finite. The function H is convex and analytic w.r.t. α, and it vanishes at zero. [Freidlin and Wentzell, 1984] describe the connection of H with the Markov process X^N in the following way: if we apply the operator A^N defined in Equation (2.26) to the function $e^{\sum_i \alpha_i x_i}$, then we obtain $NH\left(x, \frac{1}{N}\alpha\right) e^{\sum_i \alpha_i x_i}$. If $L(x,\beta)$ denote the Legendre-Fenchel transform of $H(x,\alpha)$ w.r.t. α, then [Freidlin and Wentzell, 1984] prove in Theorem **2.1** of Chapter 5 that the sequence $\left\{X^N\right\}_N$ verifies an LDP if the following conditions on L are verified:

I. There exists an everywhere finite nonnegative convex function $\bar{H}(\alpha)$ such that $\bar{H}(0) = 0$ and $H(x,\alpha) \leq \bar{H}(\alpha)$ for all x, α.

II. The function $L(x,\beta)$ is finite for all values of the arguments, for any $R > 0$ there exists positive constants M and m such that $L(x,\beta) \leq M$, $|L_\beta(x,\beta)| \leq M$, $\sum_{i,j} \frac{\partial^2 L}{\partial\beta_i\partial\beta_j}(x,\beta)\, c_i c_j \geq m \sum_i c_i^2$ for all $x, c \in \mathbb{R}^n$ and all β with $|\beta| < R$.

III. $\Delta L(\delta) := \sup\limits_{|y-y'|<\delta} \sup\limits_{\beta} \frac{L(y',\beta)-L(y,\beta)}{1+L(y,\beta)} \to 0$ if $\delta \to 0$.

It is proved that the functional I given by

$$I(\mathbf{x}) = I_{T_1,T_2}(\mathbf{x}) = \begin{cases} \int_{T_1}^{T_2} L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\, dt, & \text{if } \mathbf{x} \in \mathcal{AC} \text{ and the integral is convergent,} \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.30}$$

is the LD rate function (or $NI(\mathbf{x})$ is an *action functional* in terms of [Freidlin and Wentzell, 1984]) for the sequence $\left\{X^N\right\}_N$ with $X^N = \left\{X_t^N\right\}_{T_1 \leq t \leq T_2}$. They prove the upper and lower bounds from a suitable change of measure.

However, we do not have the function L explicitly in many cases, so it can be a formidable taks to verify conditions I, II, and III in practice. From [Feng and Kurtz, 2006], we can analyse those conditions from the point of view of the function H. In this case, the non-linear generator of the process is

$$H^N(f)(x) = \int\limits_{\mathbb{R}^n} \left(e^{N\left(f\left(x+\frac{1}{N}z\right)-f(x)\right)} - 1 - (\nabla f(x))^t \cdot z\right) \eta\,(x,dz) + (b(x))^t \cdot \nabla f(x)$$
$$+ \frac{1}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial f}{\partial x_i}(x)\frac{\partial f}{\partial x_j}(x) + \frac{1}{N}\sum_{i,j} a_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Now, assuming that $\int_{\mathbb{R}^n}\left(e^{\alpha^t\cdot z} - 1 - \alpha^t \cdot z\right)\eta\,(x,dz) < \infty$ for each $\alpha \in \mathbb{R}^n$, the limiting semigroup $\mathbf{H}(f) = \lim\limits_N H^N(f)$ is given by

$$\mathbf{H}(f)(x) = \int\limits_{\mathbb{R}^n} \left(e^{(\nabla f(x))^t\cdot z} - 1 - (\nabla f(x))^t \cdot z\right) \eta\,(x,dz) + \frac{1}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial f}{\partial x_i}(x)\frac{\partial f}{\partial x_j}(x) + (b(x))^t \cdot \nabla f(x).$$

A variational representation of $\mathbf{H}$ can be constructed as before. Define

$$H\,(x,\alpha) = \int\limits_{\mathbb{R}^n} \left(e^{\alpha^t\cdot z} - 1 - \alpha^t \cdot z\right) \eta\,(x,dz) + \frac{1}{2}\left|\sigma^t(x)\alpha\right|^2 + (b(x))^t \cdot \alpha,$$

and $L\,(x,\beta) = \sup\limits_{\alpha\in\mathbb{R}^n}\{\alpha\cdot\beta - H\,(x,\alpha)\}$. Then

$$\mathbf{H}(f)(x) = H\,(x,\nabla f(x)) = \sup_{u\in\mathbb{R}^n}\left\{A(f)(x,u) - L(x,u)\right\},$$

where $A(f)(x,u) = (\nabla f(x))^t \cdot u$ for all $f \in C_c^2\,(\mathbb{R}^n)$, and the rate function is given by

$$I\,(\mathbf{x}) = \begin{cases} \int_0^\infty L\,(\mathbf{x}(t),\dot{\mathbf{x}}(t))\,dt, & \text{if } \mathbf{x} \in \mathcal{AC} \text{ and the integral is convergent,} \\ +\infty, & \text{otherwise,} \end{cases}$$

which coincides with the rate function defined in Equation (2.30).

Chapter 3

Large deviations for the greedy exploration algorithm over Erdös-Rényi graphs

Abstract

In this chapter, we prove large deviations for a greedy exploration process over Erdös-Rényi (ER) graphs, when the number of nodes goes to infinity. To prove our main result, we use the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], which is presented in Section 2.2. The rate function can be expressed in a closed-form formula, and associated optimization problems can be solved explicitly, providing the large deviation trajectory. Also, we derive large deviations results for the size of the maximum independent set discovered by such an algorithm and analyse the probability that it exceeds known bounds for the maximal independent set. Moreover, we explore the link between these results and the landscape complexity of the independent set and the exploration dynamic.

The results of this chapter were accepted for publication in **ALEA** (*Latin American Journal of Probability and Mathematical Statistics*). They are currently visible in [Bermolen et al., 2021b].

This chapter is organized as follows. In Section 3.1, we introduce the greedy algorithm over Erdös-Rényi graphs and known results about the independent set size obtained by such an algorithm. In Section 3.2, we formally define the sequence of processes related to the greedy algorithm over a given Erdös-Rényi graph. In Section 3.3, we present the main result: a path-state LDP for the greedy exploration process. The proof of this result is deferred to Section 3.4. As a corollary, we obtain an LDP for the size of the independent set discovered by such an algorithm and analyse its implications.

3.1 Introduction

In this section, we introduce the greedy algorithm over Erdös-Rényi graphs and known results about the independent set size obtained by such an algorithm.

Consider a finite, possibly random, graph G for which V is the set of N nodes or vertices. A typical sequential exploration algorithm, usually referred to as *"greedy algorithm"* [1] works as follows. Initially, all the vertices are declared as *unexplored*. At each step, it selects a vertex and changes its state into *active*. After this, it takes all of its unexplored neighbours and changes their states into *blocked*. The *active* and *blocked* vertices are considered as *explored* and removed from the set of *unexplored* vertices. The algorithm keeps repeating this procedure until the step T_N^* at which all vertices are either active or blocked (or equivalently, the set of the *unexplored* vertex is empty). Observe that at any step k, the active vertices conform to an independent set (i.e. there are no edges between the nodes of this set) and that T_N^* is the size of the independent set constructed by the algorithm. Let Z_k^N be the number of explored nodes at time k, then $Z_{T_N^*}^N = N$.

Our motivation to study such an exploration process on random graphs is twofold. On the one hand, exploration processes have received a great amount of attention in spatial structures. It has been considered on discrete structures like $\mathbb{Z}^d$ (see [Ritchie, 2006, Ferrari et al., 2002]) and point processes (see [Penrose, 2001, Baccelli and Tien Viet, 2012]). In physics and biological sciences, where it is usually referred to as random sequential absorption, it models phenomena of deposition of colloidal particles or proteins on surfaces (see [Evans, 1993]). In communication sciences and wireless networks in particular, it allows representing the number of connections for CSMA-like algorithms in a given time-slot, for a given spatial configuration of terminals (see [Kleinrock and Takagi, 1985] for a classical reference on the protocol definition).

On the other hand, these dynamics are the simplest procedure to construct (maximal) independent sets and have been extensively studied for specific graphs. Explicit results for the size of these sets have been obtained for regular graphs in [Wormald, 1995], exploiting their particular structure; see also [Gamarnik and Sudan, 2017] for graphs with large girths, and [Bermolen et al., 2017b] for more general configuration models. In this context, the greedy algorithm is the simplest instance of a local algorithm, i.e., using only local information available at each vertex and using some randomness. Recently, it was proven in [Gamarnik and Sudan, 2017] that contrary to previously stated conjectures (for instance, in

[1] It is called *greedy* although there is no policy to choose the optimal vertex in each step, see for instance the definition of an unweighted greedy algorithm in [Jungnickel, 2005].

[Hatami et al., 2014]), local algorithms can not discover asymptotically maximum independent sets (independent set of maximum size) and stay sub-optimal, up to multiplicative constant, for regular graphs with large girth. Hence, it is natural to look at related questions for Erdös-Rényi (ER) graphs: we focus on giving estimates of reaching a given size of maximum independent sets by studying the large deviations of the exploration process.

Thanks to the great amount of independence and symmetry of the edges' collection in a sparse ER graph $G(N, c/N)$, the greedy exploration algorithm is characterized by the simple one-dimensional Markov process $\{Z_k^N\}_k$. Consequently, a functional law of large numbers described by a differential equation can be employed to get the macroscopic size of the constructed independent set when the number of nodes goes to infinity (see [Bermolen et al., 2017a] and references in [McDiarmid, 1990]). Diffusion approximations for the process and central limit theorem derived from it for the size T_N^* of the associated independent set are also known, see [Bermolen et al., 2017a]. Moreover, in [Pittel, 1982], exponential bounds are proved for the probability that the stopping times $t_f(G(N, p/N))$ of the f-driven algorithms (in particular, T_N^*) belong to certain intervals. However, to the best of our knowledge, there is no characterization of a large deviation principle for both the discrete-time Markov process $\{Z_k^N\}_k$ and the random variable T_N^*, which can give various types of useful information both on the greedy exploration and on the independent set landscape. For example, it allows determining the most probable trajectory for which the independent set's size is bigger/smaller than selected bounds.

The topic of this chapter is a refined analysis of this simple algorithm by studying the large deviations for the sequence of processes $\{Z_k^N\}_k$. As a corollary, we obtain LD results for the size of the independent set constructed by the algorithm.

Although $\{Z_k^N\}_k$ is a simple Markov process, as far as we know, computing its LDP does not directly follow from classical results. Indeed, the well-known work of [Freidlin and Wentzell, 1984] is not directly applicable to this process since both the drift and the jump measure involved in the underlying stochastic differential equation depend on the scaling parameter. An LD upper bound for a general family of processes, including processes whose (discontinuous) drift and jump measure depends on the scaling, is presented in [Dupuis et al., 1991]. However, the authors do not provide sufficient conditions to ensure that the general upper bound obtained for simpler processes is still valid for this case.

We prove that its LD upper bound not only works for a continuous-time version of $\{Z_k^N\}_k$, but is also effectively the LD rate function. To prove this LDP, we use the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], presented in Section 2.2. After working on the four steps presented in Section 2.2, we deduce not only a variational form of the rate function but also prove that it can be expressed as an action integral of a cost function L. Moreover, by solving the associated Hamilton's equations, the rate optimization over a set of trajectories can be transformed into a real parametric function optimization.

Additionally, the cost function L has a simple interpretation in terms of local deviations for the average of Poisson random variables. As such, this is a first step to understand how such local algorithms behave in complicated landscapes.

This result also allows us to derive quantitative results about the independent set's size constructed by this algorithm. For instance, we can compute the probability that this size lets be larger than the asymptotic Erdös bound for the maximum independent set when $c \geq 3$ and for the maximum independent set's exact value when $c < e$. In particular, it sheds light on the relation between the complexity of the landscape and the exploration algorithm. It is known (and coined as the e-phenomena in [Spitzer, 1975, Jonckheere and Saenz, 2019]) that for $G(N, c/N)$ with $c < e$, an improved local algorithm [1] is asymptotically optimal. The computation of LD estimates for the greedy exploration (using the asymptotic Erdös bound) allows us to give evidence of a phase transition for the independent set landscape around e (we lose some precision here because of using a bound instead of the true asymptotic value of the independent set), but it hints at an interesting connection between complexity phase transitions and explicit large deviations results.

3.2 Greedy exploration algorithm

In this section, we formally define the sequence of processes related to the greedy algorithm over a given Erdös-Rényi graph.

Let $G\left(N, \frac{c}{N}\right)$ be a sparse Erdös-Rényi graph for which V is the set of N vertices. At any step $k = 0, 1, 2, \ldots$, we consider that each vertex is either *active*, *blocked*, or *unexplored*. Accordingly, the set of vertices will be split into three components: the set of active vertices $\mathcal{A}_k$, the set of blocked vertices $\mathcal{B}_k$, and the set of unexplored vertices $\mathcal{U}_k$.

The greedy exploration algorithm in discrete-time on a graph G can be described as follows. Initially, it sets $\mathcal{U}_0 = V$, $\mathcal{A}_0 = \emptyset$ and $\mathcal{B}_0 = \emptyset$. To explore the graph, at the $(k+1)$-th step it selects uniformly a vertex $i_{k+1} \in \mathcal{U}_k$ and changes its state into active. After this, it takes all of its unexplored neighbors, i.e. the set $\mathcal{N}_{i_{k+1}} = \{w \in \mathcal{U}_k | i_{k+1} \text{ shares and edge with } w\}$, and changes their states into blocked. This means that the resulting set of vertices will be given by $\mathcal{U}_{k+1} = \mathcal{U}_k \backslash \{i_{k+1} \cup \mathcal{N}_{i_{k+1}}\}$, $\mathcal{A}_{k+1} = \mathcal{A}_k \cup \{i_{k+1}\}$ and $\mathcal{B}_{k+1} = \mathcal{B}_k \cup \mathcal{N}_{i_{k+1}}$. The algorithm iterates this procedure until the step T_N^* at which all vertices are either active or blocked (or equivalently $\mathcal{U}_{T_N^*} = \emptyset$). Observe that at any step k, the active vertices conform an independent set and that $\mathcal{A}_{T_N^*}$ is a maximal independent set (because each of the vertices in $V \setminus A_{T_N^*}$ is a

[1] The degree-greedy algorithm, which is an improvement of the modification of the greedy algorithm presented in the earlier paper of [Karp and Sipser, 1981].

neighbour of at least one vertice of $A_{T_N^*}$).

Let $Z_k^N = \left|\mathcal{A}_k^N \cup \mathcal{B}_k^N\right|$ be the number of explored vertices at step k. By construction, $Z_{k+1}^N = Z_k^N + 1 + \zeta_{k+1}^N$, where ζ_{k+1}^N is the number of unexplored neighbors of the selected active vertex at step $k+1$. The distribution of ζ_{k+1}^N depends only on the number of already explored vertices Z_k^N, that is the distribution is Binomial with updated parameter $N - Z_k^N - 1$ and the same edge probability c/N. The process $\left\{Z_k^N\right\}_k$ is then a discrete-time Markov chain with state space $\{0, 1, 2, ..., N\}$, increasing, time-homogeneous and with an absorbing state N. We are interested in $T_N^* \in \{0, 1, 2, ..., N\}$, the time at which $\left\{Z_k^N\right\}_k$ reaches N, since T_N^* coincides with the size of the maximal independent set constructed by this algorithm.

We use the notation presented in Section 2.2 for the discrete-time Markov processes case. Let $Y^N = \left\{Y_k^N\right\}_{k\geq 0}$ be a scaled version of the described process: $Y_k^N = \frac{Z_k^N}{N}$. The transition operator of the process Y^N for $x \in E^N = \left\{\frac{k}{N} : k = 0, 1, ..., N\right\}$ is:

$$T_N(f)(x) := T_{Y^N}(f)(x) = \mathbb{E}\left[f\left(x + \frac{1}{N} + \frac{1}{N}\zeta_{N,x}\right)\right], \tag{3.1}$$

where $\zeta_{N,x}$ is the number of unexplored neighbors of the selected active vertex given that there are already Nx explored vertices. Then $\zeta_{N,x}$ has a Binomial distribution with parameters $n = N - Nx - 1$ and $p = \frac{c}{N}$. We consider the embedding maps $\eta_N : E^N \to E$, where $E = [0, 1]$. Define the following continuous process:

$$X_t^N = Y_{[Nt]}^N = \frac{Z_{[Nt]}}{N} \text{ if } t \in [0, 1]. \tag{3.2}$$

This process is a semimartingale. Moreover, it can be decomposed as

$$X_t^N = \int_0^t \left[1 + c\left(1 - X_s^N - \frac{1}{N}\right)\right] \mathrm{d}s + \frac{M_{tN}^N}{N},$$

where $\left\{M_t^N\right\}_t$ is a $\mathbb{F}^N = \left\{\mathcal{F}_t^N\right\}_t$ martingale with $\mathcal{F}_t^N = \sigma\left(Z_{[Ns]}^N : 0 \leq s \leq t\right)$.

In [Bermolen et al., 2017a] it is proved that the sequence of processes $\left\{X^N\right\}_N$, contained in the space of càdlàg functions $D_E[0, 1]$, converges in the Skorohod topology to $\{\hat{z}(t)\}_t$, where

$$\hat{z}(t) = \begin{cases} z(t), & \text{if } t \leq T^*, \\ 1, & \text{if } t > T^*, \end{cases} \text{ being } z(t) \text{ the solution of } \begin{cases} \dot{z} = 1 + c(1 - z), \\ z(0) = 0, \end{cases} \tag{3.3}$$

and $T^* = \inf\{t \in [0, 1] : z(t) \geq 1\}$. This equation has an explicit solution given by $z(t) = \frac{1+c}{c}\left(1 - e^{-ct}\right)$. Moreover, a law of large numbers can be deduced for the proportion of vertices that conform to the independent set constructed by the algorithm. In particular, it is proved that $\frac{T_N^*}{N}$ converges in probability to $T^* = \frac{1}{c}\log(1+c)$.

In the same paper [Bermolen et al., 2017a] and for a different scaling of the process, a diffusion result is also proved from which a central limit theorem for $\frac{T_N^*}{N}$ is deduced: $\sqrt{N}\left(\frac{T_N^*}{N} - T^*\right)$ converges in distribution to a centered normal random variable with variance $\sigma^2 = \frac{c}{2(c+1)^2}$. Now, we study an LDP for both the sequence of processes $\left\{X^N\right\}_N$ and for $\left\{\frac{T_N^*}{N}\right\}_N$. It is known that results of central limit theorems and large deviations types are independent of each other, and neither is stronger than the other. However, as we mentioned in Section 2.5, the LDP also automatically provides results of the law of large numbers type.

3.3 Main Results

In this section, we present the main results of this chapter. In Subsection 3.3.1, we present an LDP for the sequence of processes $\left\{X^N\right\}_N$ given by $X^N = \left\{X_t^N\right\}_{0\leq t\leq 1}$ and deduce its fluid limit. Moreover, we provide a way to find the trajectory that minimices the LD rate function over a set of trajectories. In Subsection 3.3.2, we prove large deviations for the sequence of random variables $\left\{\frac{T_N^*}{N}\right\}_N$. This theorem provides quantitative results for the probability of the independent set's size being bigger/smaller than selected bounds, which are presented in Subsection 3.3.3.

3.3.1 Large Deviation Principle

Now, we present a more refined analysis of the simple exploration algorithm presented in the previous section. As a corollary, in the next subsection, we deduce an LDP for the sequence of random variables $\left\{\frac{T_N^*}{N}\right\}_N$.

Theorem 3.3.1 (LDP for $\left\{X^N\right\}_N$) *The sequence $\left\{X^N\right\}_N$ with $X^N = \left\{X_t^N\right\}_{0\leq t\leq 1}$, where $X_t^N = \frac{Z_{[Nt]}^N}{N}$, verifies an LDP on $D_E[0,1]$ with good rate function $I : D_E[0,1] \to [0,+\infty]$ such that*

$$I(\mathbf{x}) = \begin{cases} \int_0^1 L\left(\mathbf{x}(t), \dot{\mathbf{x}}(t)\right) dt & \text{if } \mathbf{x} \in \mathcal{H}_L, \\ +\infty & \text{in other case,} \end{cases} \tag{3.4}$$

where $E = [0,1]$, $L : E \times \mathbb{R} \to \mathbb{R}$ is the cost function given by

$$L(x,\beta) = \begin{cases} (\beta-1)\left[\log\left(\frac{\beta-1}{c(1-x)}\right)-1\right]+c(1-x), & \text{if } x<1 \text{ and } \beta>1,\\ c(1-x), & \text{if } x<1 \text{ and } \beta=1,\\ 0, & \text{if } x=1 \text{ and } \beta=0,\\ +\infty & \text{in other cases }, \end{cases} \tag{3.5}$$

and $\mathcal{H}_L$ is the set of all absolutely continuous function $\mathbf{x} : [0,1] \to [0,1]$ with value 0 at 0 and such that the integral $\int_0^1 L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\, dt$ exists and it is finite.

The proof is deferred to Section 3.4.

Remark 3.3.1 Let us introduce some comments about the cost function L. The function defined in Equation (3.5) is the Legendre transform w.r.t the second variable of the function $H : E \times \mathbb{R} \to \mathbb{R}$ given by

$$H(x,\alpha) = \begin{cases} \alpha + c(1-x)(e^{\alpha}-1), & \text{if } 0 \le x < 1,\\ 0, & \text{if } x=1, \end{cases} \tag{3.6}$$

that is $L(x,\beta) = \sup\limits_{\alpha\in\mathbb{R}}\{\alpha\beta - H(x,\alpha)\}$. Since $H(x,\alpha)$ is convex w.r.t. α, the function L is also convex w.r.t. β and verifies $H(x,\alpha) = \sup\limits_{\beta\in\mathbb{R}}\{\alpha\beta - L(x,\beta)\}$.

Then, as a consequence of Theorem 3.3.1 and Proposition 2.5.1, we can deduce the fluid limit for the sequence $\{X^N\}_N$.

Corollary 3.3.2 (Fluid limit of $\{X^N\}_N$) *The sequence of processes $\{X^N\}_N$ converges almost-sure as $N \to \infty$ to the function $\hat{z}(t)$ defined in Equation (3.3).*

Proof. As $L(x,\beta) = 0$ if and only if $\beta = H_\alpha(x,0)$, where $H_\alpha(x,\alpha)$ is the partial derivative of $H(x,\alpha)$ w.r.t. α, the trajectories with zero cost are the ones that verify $\dot{\mathbf{x}} = H_\alpha(\mathbf{x},0) = 1 + c(1-\mathbf{x})$. For the initial condition $\mathbf{x}(0) = 0$, as expected, the unique trajectory that has zero cost is the fluid limit $\hat{z}$ defined in Equation (3.3), i.e. $I(\hat{z}) = 0$ and $I(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \hat{z}$. □

The following proposition gives an intuitive interpretation of the cost function $L(x,\beta)$ in terms of the rate function for the average of independent Poisson random variables.

Proposition 3.3.3 *For $x < 1$ and $\beta > 1$, it is verified that $L(x,\beta) = \Lambda^*_{c(1-x)}(\beta-1)$, where $\Lambda^*_\lambda(u)$ is the LD rate function for the average of independent Poisson random variables with parameter λ.*

Proof. The rate function given by Crámer's theorem for the average of independent random variables Poisson with parameter λ is $\Lambda_\lambda^*(u) = u\left(\log\left(\frac{u}{\lambda}\right) - 1\right) + \lambda$ (see [Dembo and Zeitouni, 1998] for example). To complete the proof it is enough to observe that $L(x,\beta)$ coincides with $\Lambda_\lambda^*(u)$ when $\lambda = c(1-x)$ and $u = \beta - 1$. $\square$

The previous result can be explained using the following heuristics (which, of course, are far from a proof but give some intuition):

- The graph sparsity implies that the graph is locally tree-like and that the exploration does not see neighbours of a given vertex being neighbours between them.
- The asymptotic distribution of the number of unexplored neighbours of the selected active vertex is Poisson with a time-varying mean. In other words, the exploration does not change the Poisson nature of the degree distribution, which can be explained by the fact that the biased size distribution of Poisson distribution is again Poisson.

More precisely, the cost of a given curve $\mathbf{x}(t)$ such that $\mathbf{x} \in \mathcal{H}_L$ with $\dot{\mathbf{x}}(t) > 1$ for all $t \in [0,1]$ is given by $L\left(\mathbf{x}(t), \dot{\mathbf{x}}(t)\right) = \Lambda_{\lambda(t)}^*\left(\dot{\mathbf{x}}(t) - 1\right)$, with $\lambda(t) = c\left(1 - \mathbf{x}(t)\right)$. For a fixed $t \in (0,1)$, the curve $\mathbf{x}(t)$ represents the macroscopic proportion of explored vertices at time t. Then, the infinitesimal increment $\dot{\mathbf{x}}(t) \approx \frac{\mathbf{x}(t+h)-\mathbf{x}(t)}{h}$ corresponds to the mean number of new explored nodes in one step (the new active node and its unexplored blocked neighbours), that is:

$$\frac{X_{t+h}^N - X_t^N}{h} = \frac{1}{Nh} \sum_{k=[Nt]+1}^{[Nt+Nh]} \left(1 + \zeta_k^N\right) \approx 1 + \frac{1}{Nh} \sum_{k=[Nt]+1}^{[Nt+Nh]} \zeta_k^N,$$

where ζ_k^N has a Binomial distribution with parameters $N - Z_k - 1$ and $\frac{c}{N}$. For large values of N and $k \in [[Nt]+1, [Nt+Nh]]$, if $\frac{Z_k}{N}$ is close to $\mathbf{x}(t)$, then ζ_k^N can be approximated by a Poisson random variable with parameter $(N - Z_k - 1)\frac{c}{N} \approx c(1 - \mathbf{x}(t))$. Observe that, in particular, the mean macroscopic behavior $z(t)$ should verify $\dot{z}(t) = 1 + c(1 - z(t))$, which is the fluid limit we have already seen. Moreover, the global cost of a deviation to a trajectory $\mathbf{x}(t)$ can be interpreted as a consequence of the accumulated cost of microscopic deviations of the average of Poisson random variables with parameter $c(1 - \mathbf{x}(t))$.

Rare event probability estimation.

We now use the previous theorem to estimate probabilities of rare events related to $\left\{X^N\right\}_N$. In the next section, we apply these results to derive an LDP for the size of the independent set constructed by the algorithm.

As a consequence of Theorem 3.3.1, if $A \subset D_E[0,1]$ is a *good set* for I (or an *I-continuous* *set*, see [Dembo and Zeitouni, 1998]), then $\lim_N \frac{1}{N} \log \mathbb{P}\left(X^N \in A\right) = -\inf_{\mathbf{x}\in A} I(\mathbf{x})$. The next proposition will facilitate the computation of this infimum for the sets A of interest.

Proposition 3.3.4 (Rate function optimization) *Let $I : D_E[0,1] \to [0,+\infty]$ be the rate function defined in Equation (3.4). Then,*

1. *The optimization problem for the rate over a set of trajectories $A \subset D_E[0,1]$ can be reduced to a one-dimensional optimization problem:*

$$\inf_{\mathbf{x}\in A} I(\mathbf{x}) = \inf_{\{\alpha_0\in\mathbb{R}:\, \hat{x}_{\alpha_0}\in\bar{A}\}} F(\alpha_0),$$

where the closure of A is considered w.r.t. the Skorohod topology. The real function F is defined by

$$F(\alpha_0) = \int_0^{T_{\alpha_0}} L\left(x_{\alpha_0}(t), \dot{x}_{\alpha_0}(t)\right) dt, \tag{3.7}$$

being x_{α_0} the solution of the following differential equation:

$$\begin{cases} \dot{x} = 1 + c(1-x)e^{\alpha}, & x(0) = 0, \\ \dot{\alpha} = c(e^{\alpha}-1), & \alpha(0) = \alpha_0, \end{cases} \tag{3.8}$$

$T_{\alpha_0} = \inf\{t \in [0,1] : x_{\alpha_0}(t) \geq 1\}$, *and* $\hat{x}_{\alpha_0}(t) = \begin{cases} x_{\alpha_0}(t), & \text{if } t \leq T_{\alpha_0}, \\ 1, & \text{if } t > T_{\alpha_0}. \end{cases}$

2. *The explicit solution of Equation (3.8) is the fluid limit (3.3) when $\alpha_0 = 0$. For $\alpha \neq 0$ it is given by:*

$$x_{\alpha_0}(t) = \left[\frac{1}{ck_0}\log\left(\frac{1-k_0}{1-k_0e^{ct}}\right) + \frac{1}{e^{-ct}-k_0} - \frac{1}{1-k_0}\right]\left(e^{-ct}-k_0\right), \tag{3.9}$$

where $k_0 = 1 - e^{-\alpha_0}$. In this case, $F(\alpha_0)$ can be written as

$$F(\alpha_0) = \int_0^{T_{\alpha_0}} c\left(1 - x_{\alpha_0}(t)\right)\left[e^{\alpha(t)}\left(\alpha(t)-1\right)+1\right] dt, \tag{3.10}$$

where $\alpha(t) = -\log\left(1-k_0e^{ct}\right)$.

Then, in other words, Theorem 3.3.1 and the previous proposition ensure that, given that the process X^N belongs to A, one might expect that $\sup_{t\in[0,1]} \left|X_t^N - \hat{x}_{\alpha_0^*}(t)\right| \approx 0$ for some α_0^* such that $\hat{x}_{\alpha_0^*} \in \bar{A}$.

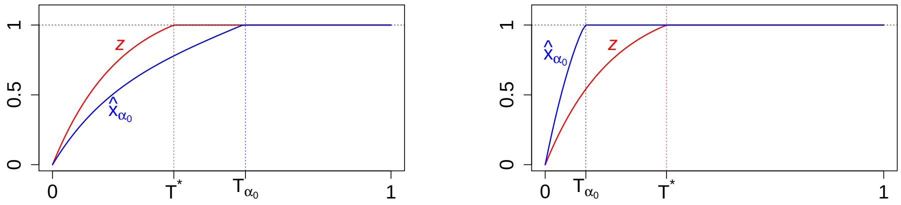

Figure 3.1: Graph of $\hat{x}_{\alpha_0}$ for same value of $\alpha_0 < 0$ (left graph) and $\alpha_0 > 0$ (right graph) compared with the fluid limit $\hat{z}$.

Proof. To prove the first statement, note that if $\mathbf{x} \in \mathcal{H}_L$ is such that $\mathbf{x}(t) = 1$ for all $t \geq t_0$, then $I(\mathbf{x}) = \int_0^1 L(\mathbf{x}, \dot{\mathbf{x}})\mathrm{d}t = \int_0^{t_0} L(\mathbf{x}, \dot{\mathbf{x}})\mathrm{d}t$, so just consider the Euler-Lagrange (EL) equation presented in Equation (2.20) for $x < 1$ and $\beta > 1$. In this case, the path $\{\mathbf{x}(t)\}_t$ is a stationary curve of I if it satisfies the following ODE:

$$\begin{cases} (x-1)\ddot{x} + (cx - (1+c))\,\dot{x} - cx + (1+c) = 0, \\ x(0) = 0, \\ \dot{x}(0) = v_0. \end{cases} \tag{3.11}$$

To solve Equation (3.11), we consider Hamilton's equations, which are equivalent to (EL):

$$\begin{cases} \dot{x} = H_\alpha(x, \alpha), \\ \dot{\alpha} = -H_x(x, \alpha), \end{cases} \qquad \text{(Hamilton)}, \tag{3.12}$$

where α is an auxiliary function. H_x and H_α are the partial derivatives of H w.r.t. x and α. In our case these equations give Equation (3.8). We are interested in solutions x_{α_0} of Equation (3.8) until the time at which they reach the value 1, then we take $\hat{x}_{\alpha_0}$ as in the proposition and get $\inf_{\mathbf{x} \in A} I(\mathbf{x}) = \inf_{\{\alpha_0 : \hat{x}_{\alpha_0} \in \bar{A}\}} I(\hat{x}_{\alpha_0})$. The uniqueness of solution for Equation (3.11) ensures that a monotony property with respect to the initial condition α_0 holds. This implies that $x_{\alpha_0}(t) > t$ for all t if $\alpha_0 > -\infty$, then $T_{\alpha_0} = \inf\{t \in [0,1] : x_{\alpha_0}(t) \geq 1\} \leq 1$ and $I(\hat{x}_{\alpha_0}) = F(\alpha_0)$ with $F(\alpha_0)$ defined in Equation (3.10). Figure 3.1 contains the graph of $\hat{x}_{\alpha_0}$ for same value of $\alpha_0 < 0$ and $\alpha_0 > 0$ compared with the fluid limit $\hat{z}$. To prove the second part of the proposition, observe that the fluid limit (3.3) (until it reaches $x = 1$) is a solution of $\dot{z} = 1 + c(1-z)$, so it is a solution of Equation (3.8) with $\alpha = 0$. If $\alpha_0 \neq 0$, the solution x_{α_0} can be found explicitly and it is given in Equation (3.9). We use that x_{α_0} is solution of Equation (3.8) for the simplification

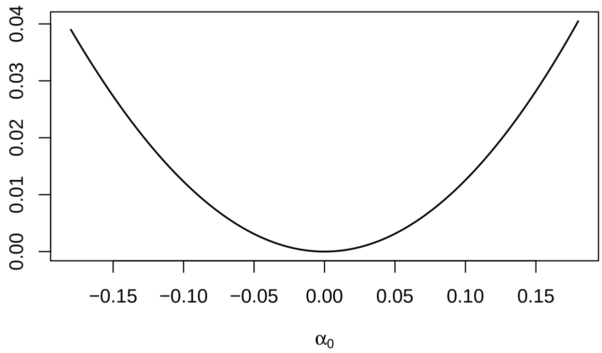

Figure 3.2: Graph of the function $F(\alpha_0) = I(\hat{x}_{\alpha_0})$, that is, a parametric version of the rate function.

of the cost function $L(x_{\alpha_0}, \dot{x}_{\alpha_0})$. Figure 3.2 contains the graph of $F(\alpha_0) = I(\hat{x}_{\alpha_0})$ as a function of α_0. $\square$

Remark 3.3.2 Let us introduce some comments about the previous result:

1. The ODE continuity theorem is verified with respect to the initial condition for Equation (3.8). Then, the solution $\hat{x}_{\alpha_0}$ with initial conditions $x(0) = 0$ and $\alpha(0) = \alpha_0 \approx 0$, is close to the fluid limit $\hat{z}$.
2. The system defined by Equation (3.12) is conservative: if $u(t) = (x(t), \alpha(t))$ is the solution of (3.12) with initial conditions $u_0 = (0, \alpha_0)$, it verifies $\dot{u} = J\nabla H(u)$ with $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Since J is an antisymmetric matrix, it results that $\frac{\mathrm{d}}{\mathrm{d}t}H(u) = (\nabla H(u))^t J\nabla H(u) = 0$ for all t. Then, the solutions of the general equation of (3.12) are contained in the level sets of the Hamiltonian H.

3.3.2 Large deviations for the independent set size

In the previous subsection, we present a path-space LDP for the exploration process defined in Section 3.1. Now, we derive from this theorem and the previous proposition about the rate optimization over a specific set, LD results for the sequence of random variables $\left\{\frac{T_N^*}{N}\right\}_N$. This theorem provides quantitative results for the probability of the independent set's size being bigger/smaller than selected bounds, which are presented in the next subsection.

Theorem 3.3.5 *Consider T_N^* defined before as the stopping time of the greedy exploration process over $G(N, \frac{c}{N})$.*

1. *If $\varepsilon > 0$ is such that $T^* + \varepsilon < 1$, then*

$$\lim_N \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \geq T^* + \varepsilon\right) = -F\left(\alpha_0(T^* + \varepsilon)\right),$$

   where $\alpha_0(T^* + \varepsilon)$ is the unique real number $\alpha_0 < 0$ such that $T_{\alpha_0} = T^* + \varepsilon$.

2. *If $\varepsilon > 0$ is such that $T^* - \varepsilon > 0$, then*

$$\lim_N \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \leq T^* - \varepsilon\right) = -F\left(\alpha_0(T^* - \varepsilon)\right),$$

   where $\alpha_0(T^* - \varepsilon)$ is the unique real number $\alpha_0 > 0$ such that $T_{\alpha_0} = T^* - \varepsilon$.

In both cases $F(\alpha_0)$ and T_{α_0} are as in Proposition 3.3.4.

Proof. We only prove the first statement because the proof of the second one is analogous. Define the set A_ε such that $A_\varepsilon = \{\mathbf{x} \in D_E[0,1] : \mathbf{x}(0) = 0, \mathbf{x}$ is increasing, $0 \leq \mathbf{x}(t) \leq 1$ for all t and $\inf\{t : \mathbf{x}(t) = 1\} \geq T^* + \varepsilon\}$. By construction, A_ε is a good set for I, then

$$\lim_N \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \geq T^* + \varepsilon\right) = \lim_N \frac{1}{N} \log \mathbb{P}\left(X^N \in A_\varepsilon\right) = -\inf_{\left\{\alpha_0 : \hat{x}_{\alpha_0} \in \bar{A}_\varepsilon\right\}} F\left(\alpha_0\right).$$

Let x_{α_0} be the solution of the homogenous ODE defined in Equation (3.11) with initial velocity $v_0 = \dot{x}_{\alpha_0}(0) = 1 + ce^{\alpha_0}$. The uniqueness of the solution ensures that the following monotony property with respect to the initial condition is verified:

$$\text{if } \alpha_0 < \alpha_1 \Rightarrow x_{\alpha_0}(t) < x_{\alpha_1}(t) \text{ for all } t \Rightarrow T_{\alpha_0} > T_{\alpha_1}.$$

In addition, it can be seen that for all $T \in (T^*, 1)$, there exists a unique value $\alpha_0 = \alpha_0(T) < 0$ such that $x_{\alpha_0}(T) = 1$ (i.e. $T = T_{\alpha_0}$). Then, there is only one $\alpha_0^* < 0$ such that $x_{\alpha_0^*}(T^* + \varepsilon) = 1$ and

- if $\alpha_0 \leq \alpha_0^* \Rightarrow T_{\alpha_0} \geq T^* + \varepsilon \Rightarrow \hat{x}_{\alpha_0} \in A_\varepsilon$,
- if $\alpha_0 > \alpha_0^* \Rightarrow T_{\alpha_0} < T^* + \varepsilon \Rightarrow \hat{x}_{\alpha_0} \notin A_\varepsilon$,

which implies that $\inf_{\{\alpha_0 : \hat{x}_{\alpha_0} \in \bar{A}_\varepsilon\}} F\left(\alpha_0\right) = \inf_{\{\alpha_0 \leq \alpha_0^*\}} F\left(\alpha_0\right)$. To complete the proof, it suffices to prove that $\inf_{\{\alpha_0 \leq \alpha_0^*\}} F\left(\alpha_0\right) = F(\alpha_0^*)$. Let $h(\alpha_0, t) = L(x_{\alpha_0}, \dot{x}_{\alpha_0})$ and $\alpha_1 < \alpha_2 < 0$. Using the monotony that we mentioned before, it can be seen that $\frac{\partial}{\partial \alpha_0} h(\alpha_0, t) < 0$ for all $\alpha_0 < 0$ and $t \in [0,1]$, that is $h(\alpha_1, t) > h(\alpha_2, t)$ for all t. Finally, since $T_{\alpha_1} > T_{\alpha_2}$ we obtain:

$$F(\alpha_2) = \int_0^{T_{\alpha_2}} L(x_{\alpha_2}, \dot{x}_{\alpha_2}) \mathrm{d}t \leq \int_0^{T_{\alpha_2}} L(x_{\alpha_1}, \dot{x}_{\alpha_1}) \mathrm{d}t < \int_0^{T_{\alpha_1}} L(x_{\alpha_1}, \dot{x}_{\alpha_1}) \mathrm{d}t = F(\alpha_1),$$

which completes the proof. □

3.3.3 On the size of the maximum independent set

The problem of finding the *maximum* independent sets in deterministic graphs is known to be NP-hard. Thus, an interesting research question is to find classes on random graphs where finding *maximum* independent sets can be (at least at the first order in N) obtained with polynomial complexity. This question is, of course, an instance of a more general viewpoint which aims at identifying phase transitions in the analysis of combinatorial optimization problems, allowing to describe drastically different scenarios depending on a few macroscopic parameters, sometimes called order-parameters.

This type of results has been proven to hold for Erdös-Rényi graphs and configuration models in [Spitzer, 1975] and [Jonckheere and Saenz, 2019]. The order-parameter being c, the mean number of neighbours of a given node. Interestingly the phase transition does not correspond for the graph to be subcritical ($c < 1$) but to a much finer property of the landscape of maximal independent sets. The phase transition corresponds to $c < e$ and differentiates between regimes where a simple degree-greedy algorithm reaches (asymptotically) the maximum independent set or not. This same phase transition is reflected in the properties of the spectrum of the graph, see [Coste and Salez, 2018].

We conjecture that the large deviations characteristics of the greedy algorithm for discovering the maximum independent set also have an interesting transition for values of c around e. Since the exact optimal order-one asymptotic value of the maximal independent set's size is known only for values of $c < e$, we cannot yet display a complete characterization of this phenomenon. We can, however, obtain interesting numerical results by using the Erdös bound, instead of the true value. Let σ_N the *maximum* size of the independent set of an ER graph $G(N, c/N)$, then a.s. $\sigma_N \leq \frac{2\log(c)}{c} N(1 + o(N))$ if $c \geq 3$. In Figure 3.3 we compute the large deviation rate corresponding to the event $\{\frac{T_N^*}{N} \geq \sigma_i^*(c)\}$ for $i = 1, 2$. Here σ_1^* is the exact proportion of the maximum independent set of an ER graph $G(N, c/N)$ when $c < e$ and it is given by $\sigma_1^*(c) = w(c) + \frac{c}{2}(w(c))^2$ with $w(c) = e^{-W(c)}$ and $W(x)$ the Lambert function, see [Jonckheere and Saenz, 2019]. The value $\sigma_2^*(c) = \frac{2}{c}\log(c)$ is the Erdös upper bound for the proportion of the maximum independent set for $c \geq 3$.

Though the numerical computations for $c > e$ could give largely overestimated values; we believe it nevertheless illustrates the clear change of regime around the value e. It shows that the independent sets geometry changes, leading to significantly greater large deviations constants for the greedy exploration when c gets larger than e. This characterization of the “energy" landscape is a usual situation in statistical physics where interesting phase transitions can be

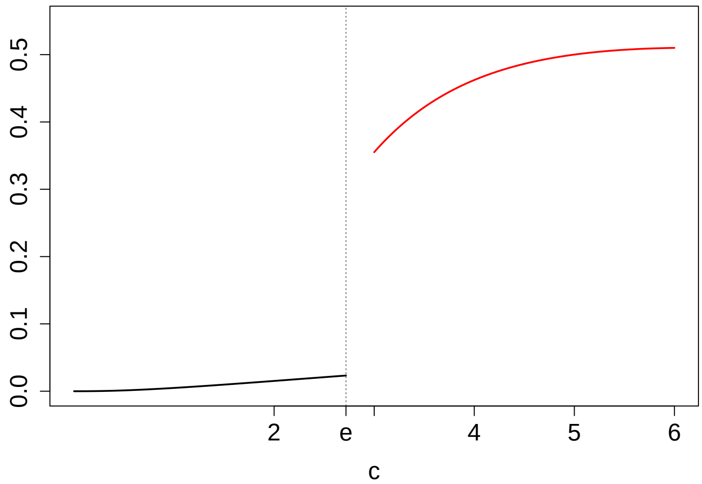

Figure 3.3: Evolution of $F(\alpha_0(\sigma_1^*(c)))$ for $0 < c < e$ and $F(\alpha_0(\sigma_2^*(c)))$ for $c \geq 3$.

well described through large deviations, see [Touchette, 2009].

3.4 Proof of Theorem 3.3.1

In this section, we prove that the previously defined sequence of processes $\{X_N\}_N$ verifies the assumptions from [Feng and Kurtz, 2006] presented in Section 2.2. We organize the proof of Theorem 3.3.1 using the steps mentioned in Section 2.2, which are presented as propositions.

In a nutshell, as a consequence of the first two steps, the process $\{X^N\}_N$ verifies the exponential tightness condition. Then, Step 3 assures an LDP via the comparison principle, and finally, Step 4 provides an applicable version of the rate.

Step 1: Convergence of nonlinear operators

Let T^N be the transition operator defined in Equation (3.1), and $H_N : Dom(H_N) \subset B(E) \to B(E)$ given by $H_N(f) = \log\left[e^{-Nf(x)}T_N\left(e^{Nf}\right)(x)\right]$.

Proposition 3.4.1 *There exists a functional* $\mathbf{H}$ *such that* H_N *converges to* $\mathbf{H}$ *when* $N \to \infty$ *in the following sense:* $\lim\limits_{N\to\infty} \sup\limits_{x\in E^N} |H_N(f)(x) - \mathbf{H}(f)(x)| = 0$ *for all* $f \in C^1(E)$. *The functional*

$\mathbf{H} : C^1(E) \to B(E)$ *is such that* $\mathbf{H}(f)(x) = H(x, f'(x))$, *where* $H : E \times \mathbb{R} \to \mathbb{R}$ *is defined by*

$$H(x,\alpha) = \begin{cases} \alpha + c(1-x)\left(e^{\alpha} - 1\right), & \textit{if } x < 1, \\ 0, & \textit{if } x = 1. \end{cases} \tag{3.13}$$

Proof. Let us first consider the case where $f \in C^2(E)$. Let $x \in E^N$ with $x \neq 1$, and $\zeta_{N,x}$ be the number of unexplored neighbours of the selected vertex in one step of the algorithm, given that there are already Nx explored vertices. Then, $\zeta_{N,x}$ has a Binomial distribution with parameters $N(1-x) - 1$ and $p = \frac{c}{N}$, and operator H^N can be written as

$$H^N(f)(x) = \begin{cases} \log \mathbb{E}\left[e^{N\left(f\left(x + \frac{1}{N} + \frac{\zeta_{N,x}}{N}\right) - f(x)\right)}\right], & \text{for } 0 \le x < 1 \\ 0, & \text{for } x = 1. \end{cases}$$

It is enough to prove that

$$\lim_{N\to\infty} \sup_{x \in E^N \setminus \{1\}} \mathbb{E}\left[e^{N\left(f\left(x + \frac{1}{N} + \frac{\zeta_{N,x}}{N}\right) - f(x)\right)}\right] - \mathbb{E}\left[e^{f'(x)\left(1+\zeta_{N,x}\right)}\right] = 0,$$

which is presented as Proposition 3.4.2 below. Then,

$$\lim_N H^N(f)(x) = \lim_N \log \mathbb{E}\left[e^{f'(x)\left(1+\zeta_{N,x}\right)}\right] = f'(x) + c(1-x)\left(e^{f'(x)} - 1\right) = H\left(x, f'(x)\right),$$

with $H : E \times \mathbb{R} \to \mathbb{R}$ defined in Equation (3.13). If $x = 1$, then $H^N(f)(1) = \mathbf{H}(f)(1) = 0$ for all N.

The result can be extended for $f \in C^1(E)$ by taking a sequence $\{f_m\}_m \subset C^2(E)$ such that $\lim_{m\to\infty} \sup_{x\in E} |f_m(x) - f(x)| = 0$ and the triangular inequality. □

Proposition 3.4.2 *If* $f \in C^2(E)$, *then*

$$\lim_{N\to\infty} \sup_{x \in E^N \setminus \{1\}} \mathbb{E}\left[e^{N\left(f\left(x + \frac{1}{N} + \frac{\zeta_{N,x}}{N}\right) - f(x)\right)}\right] - \mathbb{E}\left[e^{f'(x)\left(1+\zeta_{N,x}\right)}\right] = 0.$$

Proof. Let $f \in C^2(E)$ be fixed, then $N\left(f\left(x + \frac{k+1}{N}\right) - f(x)\right) = f'(x)(k+1) + f''\left(\theta_{k,x}\right)\frac{(k+1)^2}{2N}$, with $\theta_{k,x} \in \left(x, x + \frac{k+1}{N}\right)$. Then, it is enough to prove that

$$\mathbb{E}\left[e^{f'(x)(\zeta_{N,x}+1)}\left(e^{f''\left(\theta_{N,x}\right)\frac{\left(\zeta_{N,x}+1\right)^2}{2N}} - 1\right)\right]$$

converges to zero, being $\theta_{N,x}$ (with abuse of notation) the r.v. $\theta_{\zeta_{N,x},x} \in \left(x, x + \frac{\zeta_{N,x}+1}{N}\right)$ defined from Taylor's theorem. To prove this convergence, we bound $f''(\theta_{N,x})$ by $M_f = \sup_{\theta\in[0,1]} |f''(\theta)| < \infty$, and get

$$\mathbb{E}\left[e^{f'(x)(\zeta_{N,x}+1)}\left(e^{-\frac{M_f}{2}\frac{(\zeta_{N,x}+1)^2}{N}} - 1\right)\right] \leq \mathbb{E}\left[e^{f'(x)(\zeta_{N,x}+1)}\left(e^{\frac{f''(\theta_{N,x})}{2}\frac{(\zeta_{N,x}+1)^2}{N}} - 1\right)\right]$$
$$\leq \mathbb{E}\left[e^{f'(x)(\zeta_{N,x}+1)}\left(e^{\frac{M_f}{2}\frac{(\zeta_{N,x}+1)^2}{N}} - 1\right)\right].$$

Then, it is enough to prove that both $\mathbb{E}\left[e^{f'(x)(\zeta_{N,x}+1)}\left(e^{\pm\frac{M_f}{2}\frac{(\zeta_{N,x}+1)^2}{N}} - 1\right)\right]$ converge to zero. Thus, we prove the first convergence, and the second one can be proved analogously.

Let $U_{N,f} = e^{f'(x)(\zeta_{N,x}+1)}$ and $V_{N,f} = \left(e^{\frac{M_f}{2}\frac{(\zeta_{N,x}+1)^2}{N}} - 1\right)$. As $(\mathbb{E}(U_{N,f}V_{N,f}))^2 \leq \mathbb{E}\left(U_{N,f}^2\right)\mathbb{E}\left(V_{N,f}^2\right)$, just prove that $\mathbb{E}\left(U_{N,f}^2\right)\mathbb{E}\left(V_{N,f}^2\right) \to 0$. Observe that

$$\mathbb{E}\left(U_{N,f}^2\right) = \mathbb{E}\left[e^{2f'(x)(\zeta_{N,x}+1)}\right] = e^{2f'(x)}\left[(e^{2f'(x)} - 1)\frac{c}{N} + 1\right]^{N(1-x)-1} \to e^{2f'(x)+c(1-x)\left(e^{2f'(x)}-1\right)},$$

which is bounded, then it is enough to prove that $\mathbb{E}(V_{N,f}^2)$ converges to zero.

Let $\varphi_{N,f}(t) = \left(e^{\frac{M_f}{2}\frac{(t+1)^2}{N}} - 1\right)^2$ and $B_N > 0$, then $\mathbb{E}(V_{N,f}^2) = \mathbb{E}\left(\varphi_{N,f}\left(\zeta_{N,x}\right)\right)$ can be decomposed as

$$\mathbb{E}(V_{N,f}^2) = \mathbb{E}\left(\varphi_{N,f}\left(\zeta_{N,x}\right)\mathbf{1}_{\{\zeta_{N,x}\leq B_N\}}\right) + \mathbb{E}\left(\varphi_{N,f}\left(\zeta_{N,x}\right)\mathbf{1}_{\{\zeta_{N,x}>B_N\}}\right). \tag{3.14}$$

For the first term, a generalization of Jensen's inequality allows to get the following upper bound:

$$\mathbb{E}\left(\varphi_{N,f}\left(\zeta_{N,x}\right)\mathbf{1}_{\{\zeta_{N,x}\leq B_N\}}\right) \leq \varphi_{N,f}(\mathbb{E}(\zeta_{N,x})) + \frac{1}{2}\left(\sup_{t\in[0,B_N]}\varphi''_{N,f}(t)\right)\mathbf{var}\left(\zeta_{N,x}\right),$$

which converges to zero if we take B_N such that $\frac{B_N^2}{N} \to 0$. For the second term, we have:

$$\mathbb{E}\left(\varphi_{N,f}\left(\zeta_{N,x}\right)\mathbf{1}_{\{\zeta_{N,x}>B_N\}}\right) \leq \sum_{k=B_N+2}^{N(1-x)}\left(e^{\frac{M_f}{2}\frac{k^2}{N}} - 1\right)^2\mathbb{P}(\zeta_{N,x} = k),$$

if $B_N > \mathbb{E}(\zeta_{N,x}) = c(1-x) - \frac{c}{N}$. Let $a_{N,k} = \left(e^{\frac{M_f}{2}\frac{k^2}{N}} - 1\right)^2\mathbb{P}(\zeta_{N,x} = k)$. Then, $0 \leq a_{N,k} \leq$

$(e^{\frac{M_f}{2}k} - 1)^2 \left(c(1-x)\right)^k \frac{1}{k!} := c_k$. As $\lim_k \frac{c_{k+1}}{c_k} = 0$, we have:

$$0 \leq \mathbb{E}\left(\varphi_{N,f}(\zeta_{N,x}) \mathbf{1}_{\{\zeta_{N,x} > B_N\}}\right) \leq \sum_{k=B_N+2}^{N(1-x)} a_{N,k} \leq \sum_{k=B_N+2}^{+\infty} c_k,$$

which converges to zero since it is the tail of a convergent serie if $B_N \to +\infty$. Finally, taking $B_N = c(1-x) + N^{\frac{1}{3}}$ we obtain that both terms in Equation (3.14) converge to zero. □

Step 2: Verify the exponential compact containment condition

In this case, $E = [0,1]$ is compact, so the *exponential compact containment condition* from Definition 2.2.1 is trivially verified by taking $K_\alpha = E$.

Step 3: Comparison principle

As mentioned in Section 2.2, the verification that for all $\beta > 0$ and sufficiently many $h \in C(E)$ there exists a solution $f \in C^1(E)$ for the following ODE

$$f(x) - \beta H\left(x, f'(x)\right) - h(x) = 0, \quad x \in E, \tag{3.15}$$

is difficult or impossible. An alternative is to prove the existence (and uniqueness) of viscosity solutions. Moreover, due to Theorem **6.14** of [Feng and Kurtz, 2006], it is enough to prove that the *comparison principle* is verified for this Hamilton-Jacobi equation. In Section 2.3, we presented the tools taken from [Kraaij, 2016], Chapter 9 of [Feng and Kurtz, 2006], and [Crandall et al., 1992] to prove the comparison principle in this case.

Proposition 3.4.3 *For each $\beta > 0$ and $h \in C(E)$ the comparison principle is satisfied for Equation (3.15).*

Proof. Let μ be a subsolution and v a supersolution of Equation (3.15). Let $\psi : [0,1]^2 \to \mathbb{R}^+$ be the good penalization function given by $\psi(x,y) = \frac{1}{2}(x-y)^2$, and let $x_\alpha, y_\alpha \in E$ be such that

$$\mu(x_\alpha) - v(y_\alpha) - \alpha\psi(x_\alpha, y_\alpha) = \sup_{x,y \in E} \left\{\mu(x) - v(y) - \alpha\psi(x,y)\right\}.$$

As a consequence of Proposition 2.3.2, it is enough to prove that the following inequality holds:

$$\liminf_{\alpha\to\infty} H\left(x_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) - H\left(y_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) \leq 0,$$

where ψ_x is the derivative of ψ w.r.t. x. If $z \in [0,1)$, then

$$H\left(x_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) - H\left(y_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) = -c\left(e^{\alpha(x_\alpha - y_\alpha)} - 1\right)\left(x_\alpha - y_\alpha\right).$$

By Proposition 2.3.1, we know that $x_\alpha - y_\alpha \to 0$ and due to the second part of Proposition 2.3.2 we have:

$$\sup_\alpha H\left(y_\alpha, \alpha\left(x_\alpha - y_\alpha\right)\right) = \sup_\alpha \alpha\left(x_\alpha - y_\alpha\right) + c\left(1 - y_\alpha\right)\left(e^{\alpha(x_\alpha - y_\alpha)} - 1\right) < \infty,$$

which implies that $\sup_\alpha \alpha\left(x_\alpha - y_\alpha\right) < \infty$. Then $\left\{\alpha\left(x_\alpha - y_\alpha\right)\right\}_\alpha$ has a convergent subsequence. Let A be its limit. Then,

$$\liminf_{\alpha\to\infty} H\left(x_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) - H\left(y_\alpha, \alpha\psi_x\left(x_\alpha, y_\alpha\right)\right) \le H(z, A) - H(z, A) = 0.$$

For $z = 1$, we repeat the previous analysis, being careful with cases in which $x_\alpha = 1$ or $y_\alpha = 1$ after a certain α_0. □

Step 4: Variational representation of the rate function

Finally, we prove that the rate function can be written as an action functional. As a consequence of the results presented in Subsection 2.2, it is enough to prove that Conditions **8.9**, **8.10**, and **8.11** from [Feng and Kurtz, 2006] (which are presented as Conditions 2.2.4, 2.2.5, and 2.2.6) are verified in this case. We present them as propositions. Also, the role of absolutely continuous functions in the definition of the rate function I is explained.

As $\mathbf{H}(f)(x) = H\left(x, f'(x)\right)$ for each $x \in E = [0,1]$ and $H \leftrightarrow L$, we have that $\mathbf{H}$ can be written as

$$\mathbf{H}(f)(x) = \sup_{u\in U}\left\{A(f)(x,u) - L(x,u)\right\},$$

where $U = \mathbb{R}$ and $A : C^1(E) \to M\left(E \times U\right)$ is the linear operator $A(f)(x,u) = f'(x)u$. As L is convex w.r.t. β, it follows that a deterministic control $\lambda\left(\mathrm{d}u \times \mathrm{d}s\right) = \delta_{u(s)}(\mathrm{d}u)\mathrm{d}s$ is allways the control with smallest cost by Jensen's inequality. Moreover, if $\mathbf{x} : E \to \mathbb{R}$ is an absolutely continuous function, then

$$f\left(\mathbf{x}(t)\right) - f\left(\mathbf{x}(0)\right) = \int_0^t f'\left(\mathbf{x}(s)\right)\dot{\mathbf{x}}(s)\mathrm{d}s = \iint_{\mathbb{R}\times[0,t]} f'\left(\mathbf{x}(s)\right) u\lambda(\mathrm{d}u \times \mathrm{d}s),$$

if we define $\lambda = \lambda(\mathbf{x})$ such that $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_{\dot{\mathbf{x}}(s)}(\mathrm{d}u)\mathrm{d}s$. Let $\Gamma = E \times U$. Then, the supremum in Equation (2.11) for the Nisio semigroup definition is reached on $\{(\mathbf{x}, \lambda) : \mathbf{x} \in \mathcal{AC},\ \mathbf{x}(0) = x_0\} \subset \mathcal{Y}^\Gamma$.

Proposition 3.4.4 *Conditions* **8.9** *of [Feng and Kurtz, 2006] (Condition 2.2.4) are verified.*

Proof.

1. It is trivially verified since A is a function and $Id \in Dom(A) = C^1(E)$.
2. Γ is closed and for all $x_0 \in E$ there exists $\mathbf{x} \in \mathcal{AC}$ such that $\mathbf{x}(0) = x_0$.
3. $L(x, \beta)$ is a lower-semicontinuous function since $L(x, \beta) \leftrightarrow H(x, \alpha)$ (if $\{f_\lambda\}_\lambda$ is a family of continuous functions, then $f(x) = \sup_\lambda f_\lambda(x)$ is l.s.c.), then $L^{-1}(\{c\}) = \{(x, u) : L(x, u) \leq c\}$ is a closed subset of $[0, 1] \times \mathbb{R}$. Then, it is enough to prove that it is bounded too. Suppose that it is not bounded, then there exist $\{u_n\}_n \subset \mathbb{R}$ such that $u_n \to \infty$ and $L(x, u_n) \leq c$ for all n. But $L(x, u_n) = \sup_\alpha \{u_n\alpha - H(x, \alpha)\} \geq u_n - H(x, 1) \to +\infty$, which is a contradiction.
4. It is trivially verified since we can always take the compact set $\hat{K} = E$.
5. We construct Ψ_f as in Lemma **10.21** from [Feng and Kurtz, 2006]: if $f \in Dom(A) = C^1(E)$, then there exists C_f such that $|f'(x)| \leq C_f$ for all $x \in E$. For each $s \geq 0$, define:

$$\varphi(s) := s \inf_{x \in [0,1]} \inf_{|u| \geq s} \frac{L(x, u)}{|u|},$$

and for each $r \geq 0$ define the function $\Psi_f(r) = C_f \varphi^{-1}(r)$. Finally, as it works for all $(x, u) \in \Gamma$, we can take the function $\psi_{f,K} = \psi_f$.

□

Proposition 3.4.5 *Condition **8.10** from [Feng and Kurtz, 2006] (Condition 2.2.5) is verified.*

Proof. Since $L(x, u) = 0 \Leftrightarrow u = H_\alpha(x, 0)$, the function $q(x) = H_\alpha(x, 0) = 1 + c(1 - x)$ solves the equation $L(x, q(x)) = 0$ for all $x \in E$. Note that the fluid limit verifies $\dot{z} = q(z)$ with the initial condition $z(0) = 0$. Given $x_0 \in E$, there exists $t_0 \in [0, 1]$ such that $z(t_0) = x_0$, then $z(t + t_0)$ is the solution of $\dot{x} = q(x)$ with $x(0) = x_0$. Define $\mathbf{x}(t) = z(t + t_0) \wedge 1$ for all $t \in [0, 1]$ and λ such that $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_{\{q(\mathbf{x}(s))\}}(\mathrm{d}u) \times \mathrm{d}s$, then $(\mathbf{x}, \lambda) \in \mathcal{Y}^\Gamma$ and verifies the required condition. □

Proposition 3.4.6 *Condition* **8.11** *from [Feng and Kurtz, 2006] (Condition 2.2.6) is verified.*

Proof. Let $x_0 \in E$ and $f \in C^1(E)$ be fixed. Then, we need to find $(\mathbf{x}, \lambda) \in \mathcal{Y}^\Gamma$ such that $\mathbf{x}(0) = x_0$ and

$$\int_0^t H\left(\mathbf{x}(s), f'\left(\mathbf{x}(s)\right)\right) \mathrm{d}s = \iint_{U\times[0,t]} \left(f'\left(\mathbf{x}(s)\right) u - L\left(\mathbf{x}(s), u\right)\right) \lambda\left(\mathrm{d}s \times \mathrm{d}u\right), \tag{3.16}$$

for all $t \in [0,1]$. If we define $q_f(x) = H_\alpha(x, f'(x))$, then $H(x, f'(x)) = f'(x)q_f(x) - L\left(x, q_f(x)\right)$ and Equation (3.16) is verified if we take $\lambda\left(\mathrm{d}u \times \mathrm{d}s\right) = \delta_{\left\{q_f(\mathbf{x}(s))\right\}}(\mathrm{d}u)\ \mathrm{d}s$. Now we have to add conditions so that in addition $(\mathbf{x}, \lambda)$ belongs to $\mathcal{Y}^\Gamma$. In particular, $(\mathbf{x}, \lambda)$ must verify:

$$\int_0^t g'\left(\mathbf{x}(s)\right) q_f\left(\mathbf{x}(s)\right) \mathrm{d}s = g\left(\mathbf{x}(t)\right) - g\left(\mathbf{x}(0)\right)\ \ \forall t \in [0,1],\ \forall g \in C^1(E).$$

Then we look for a path that solves the following problem:

$$\begin{cases} \mathbf{x} \text{ is differentiable almost everywhere and } \dot{\mathbf{x}}(t) = q_f\left(\mathbf{x}(t)\right), \\ \mathbf{x}(0) = x_0, \\ \mathbf{x}(t) \in [0,1] \text{ for all } t \geq 0. \end{cases} \tag{3.17}$$

Let $x_0 \in [0,1)$. Note that $q_f(x) = 1 + c(1-x)e^{f'(x)} > 1$ is continuous, then from Peano's theorem (see [Crandall, 1972]), we know that the ODE $\begin{cases} \dot{\mathbf{x}}(t) = q_f\left(\mathbf{x}(t)\right) \\ \mathbf{x}(t_0) = x_0 \end{cases}$ has a local solution $\mathbf{x} : J \to \mathbb{R}$, being J an open neighbourhood of t_0, it is also increasing and verifies $\mathbf{x}(t) \geq t$ for all $t \in [0,1]$. Since we are looking for a càdlàg function, we can paste these local solutions until the time T_{x_0} at which it reaches 1 and define $\mathbf{x}(t) = 1$ for $T_{x_0} \leq t \leq 1$. If $x_0 = 1$, we take $\mathbf{x} \equiv 1$. $\square$

Chapter 4

Large deviations for the greedy exploration process on d-regular graphs

Abstract

In this chapter, we prove large deviations for the greedy exploration on a configuration model by jointly constructing a d-regular graph and discovering an independent set in this graph. We consider a time-discretized version of the method proposed by [Bermolen et al., 2017b] and [Brightwell et al., 2017] for creating more general uniform random graphs from a given degree sequence. We consider a discrete-time Markov process describing the evolution of this algorithm and prove a large deviation principle for a rescaling of this process. The proof of this result follows the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], which is presented in Section 2.2. As a corollary, we derive large deviations results for the independent set size constructed by this algorithm. Finally, we retrieve known results about the independent set size obtained and the change that occurs in the dynamics when $d = 2$ or $d > 2$.

This chapter is organised as follows. In Section 4.1, we give a brief introduction to the chapter. In Section 4.2, we define the dynamic analysed in this chapter, which consists of simultaneously constructing a d-regular graph and an independent set in this graph. Moreover, we define a sequence of Markov processes related to this construction. In Section 4.3, we present the main result: a path-state LDP for the sequence of Markov processes defined in Section 4.2. The detailed proof is deferred to Section 4.4. As a corollary, we obtain the corresponding fluid limit and large deviations results for the size of the independent set constructed.

4.1 Introduction

In this chapter, we analyse a simple greedy algorithm to construct an independent set in a d-regular graph. We use a simultaneous construction of the random graph in a *configuration model* and an exploration discovering an independent set. This idea was first used by [Wormald, 1995] for d-regular graphs and then for [Bermolen et al., 2017b] and [Brightwell et al., 2017] for more general uniform random graphs. We consider a time-discretized version of the algorithm proposed by [Brightwell et al., 2017]. Moreover, in Chapter 5, we extend the results presented in this chapter for random graphs constructed in the same way from an initial degree sequence.

We prove a large deviation principle, when the number N of vertices of the graph goes to infinity, for a rescaled version $X_t^N = \frac{X_{[Nt]}^N}{N}$ ($t \in [0,1]$) of the Markov chain $\left\{X_n^N\right\}_n$ that counts the number of vertices that have already been placed into the independent set, and the number of *empty* (or non-explored) vertices at each step n of the algorithm.

The proof of this result follows the general strategy to study large deviations of processes proposed by Feng and Kurtz in [Feng and Kurtz, 2006], which is presented in Section 2.2.

In our case, after working on the four steps mentioned in Section 2.2, we prove that the rate function can be expressed as an action integral. Moreover, its cost function has an intuitive interpretation in terms of Crámer's theorem for the average of random variables approximating the distribution of new explored vertices, conditioned to the number of explored vertices, in each step of the algorithm.

We provide a way to find the trajectory that minimises the LD rate function over a set of trajectories (i.e., the most probable trajectory) by studying the Hamiltonian dynamics associated with the rate function obtained. As a corollary, we deduce LD results for the independent set size obtained. Finally, we retrieve known results about the independent set size and the changes in the dynamics when $d = 2$ or $d > 2$.

4.2 Description of the dynamics

In this section, we define the dynamic analysed in this chapter, which consists of simultaneously constructing a d-regular graph and an independent set in this graph. Moreover, we define a sequence of Markov processes related to this construction.

We consider the following greedy algorithm for constructing an independent set $\mathcal{S}^N$ in an N-vertex d-regular (possibly multi) graph $G(N, d)$. We start with a set of vertices $\mathcal{V}^N = \{1, 2, \ldots, N\}$, each one with degree d. At each step $n = 0, 1, \ldots, N$, the set $\mathcal{V}^N$ is partitioned into three classes:

- a set $\mathcal{S}_n^N$ of vertices that have already been placed into the independent set, with all their half-edges paired with vertices out of $\mathcal{S}_n^N$;

- a set $\mathcal{B}_n^N$ of *blocked vertices*, where at least one of its half-edges has been paired with a half-edge from $\mathcal{S}_n^N$;
- a set $\mathcal{E}_n^N$ of *empty vertices*, from which no half-edge has yet been paired.

Initially, all the vertices are *empty*, i.e. $\mathcal{E}_0^N = \mathcal{V}^N$ and $\mathcal{S}_0^N = \emptyset$. At step n, a vertex v is selected uniformly from $\mathcal{E}_n^N$, it is placed into $\mathcal{S}_n^N$, and all its half-edges are paired, drawing uniformly within the available half-edges. This pairing procedure results in the following updates:

1. v is moved from $\mathcal{E}_n^N$ to $\mathcal{S}_n^N$,
2. each half-edge pointing out from v is paired in turn with some other uniformly randomly chosen vertices among the currently unpaired half-edges,
3. the vertices in $\mathcal{E}_n^N$ with some half-edges already paired with a half-edge from v are moved to $\mathcal{B}_n^N$.

Note that some half-edges pointing out from v may be paired with half-edges from $\mathcal{B}_n^N$, or indeed with other half-edges from v, and no change in the status of those vertices results from such pairings. At each step n, the only paired edges are those with at least one endpoint in $\mathcal{S}_n^N$.

This algorithm terminates at the first step $n = T_N^*$ at which $\mathcal{E}_n^N = \emptyset$. At this point, there may still be some unpaired half-edges attached to blocked vertices. These may be paired off uniformly at random to complete the construction of the graph $G(N, d)$. Note that T_N^* coincides with the size of the independent set constructed by this algorithm. The expected value of $\frac{T_N^*}{N}$ is usually called the *jamming constant* of the graph.

For each $n \in \{0, 1, \ldots, N\}$, let be $X_n^N = \left(S_n^N, U_n^N, E_n^N\right)$ with:

- $S_n^N = \left|\mathcal{S}_n^N\right|$, the number of vertices that have already been placed into the independent set at step n;
- U_n^N, the total number of unpaired half-edges (corresponding to empty or blocked nodes) at step n;
- $E_n^N = \left|\mathcal{E}_n^N\right|$, the number of empty vertices at step n.

For each N, $\left\{X_n^N\right\}_n$ is a discrete-time Markov process. By construction, this Markov chain is updated in step $n+1$ as follows:

- The vertex v is placed into $\mathcal{S}_n^N$. Then $S_{n+1}^N = S_n^N + 1$.

- Each one of the half-edges pointing out from v is paired in turn with some other uniformly randomly chosen among the currently unpaired half-edge. Let H^N be the number of half-edges from v that are paired with another vertex different from v (blocked or empty), i.e. that do not form loops. Then, $U_{n+1}^N = U_n^N - d - H^N$. Note that H^N has a Hypergeometric distribution
$$\textbf{Hyper}\left(U_n^N, U_n^N - d, d\right).$$
- We have to distribute those H^N half-edges between the $U_n^N - dE_n^N$ half-edges corresponding to blocked vertices and the dE_n^N half-edges corresponding to empties. Let B^N be the number of half-edges from v that are paired to blocked vertices, then B^N has a Hypergeometric distribution (conditioned to H^N)
$$\textbf{Hyper}\left(U_n^N - d, U_n^N - dE_n^N, H^N\right).$$
- Finally, let $W^N = H^N - B^N$ be the number of half-edges pointing out from v that are paired to some empty vertex, and let $\tilde{W}^N$ be the number of empty vertices that share at least one edge with v. Then, $E_{n+1}^N = E_n^N - 1 - \tilde{W}^N$.

According to the following lemma, the distribution of $\tilde{W}^N$ can be approximated by the distribution corresponding to W^N.

Lemma 4.2.1 *Let $\hat{x} = (\hat{s}, \hat{u}, \hat{e})$ be an element in the state space of $\left\{X_n^N\right\}_n$, then*
$$\lim_{N\to\infty} \mathbb{P}\left(\tilde{W}^N = \omega | X_n^N = \hat{x}, H^N = h, B^N = b, W^N = \omega\right) = 1$$
for all $\omega = h - b$ with $0 \leq b \leq h \leq d$.

Proof. See Equation **17** from [Bermolen et al., 2017b]. In the article notation: $W^N = Y\left(\mu_{t^-}\right)(d)$ and $\tilde{W}^N = \tilde{Y}\left(\mu_{t^-}\right)(d)$. ☐

Let $X_t^N := \frac{X_{[Nt]}^N}{N}$ be a rescaled version of X_n^N with $t \in [0, 1]$. The state-space of this process is E^N, being
$$E^N = \left\{\frac{1}{N}(\hat{s}, \hat{u}, \hat{e}) : \hat{s}, \hat{e} \in \{0, \ldots, N\};\ \hat{u} \in \{0, \ldots, dN\};\ \hat{u} \geq d\hat{e}\right\},$$
which is included in the compact set $E := \{(x_1, x_2, x_3) \in [0, 1] \times [0, d] \times [0, 1] : x_2 \geq dx_3\}$.

In next section, we study an LDP for both $\left\{X^N\right\}_N$ and $\left\{\frac{T_N^*}{N}\right\}_N$. As a corollary, we deduce results of the law of large numbers type for both sequences and retrieve known results about the independence set size obtained and the change that occurs in the dynamics of the graph construction when $d = 2$ or $d > 2$.

4.3 Main Results

In this section, we present the main results of this chapter. In Subsection 4.3.1, we present an LDP for the sequence of processes $\left\{X^N\right\}_N$ given by $X^N = \left\{X_t^N\right\}_{0\leq t\leq 1}$ and deduce its fluid limit. In Subsection 4.3.2, we deduce an LDP for the sequence that counts the vertices remaining in the empty set and interpret the cost function $\hat{L}$ in terms of local deviations for the average of Binomial random variables. In Subsection 4.3.3, we provide a way to find the trajectory that minimises the LD rate function over a set of trajectories. Finally, in Subsection 4.3.4, we deduce large deviations results for the size of the independent set constructed by such an algorithm.

4.3.1 LDP for $\left\{X^N\right\}_N$

The theorem that we state below in the most important result in this chapter.

Theorem 4.3.1 (LDP for $\left\{X^N\right\}_N$) *The sequence $\left\{X^N\right\}_N$ with $X^N = \left\{X_t^N\right\}_{0\leq t\leq 1}$ verifies an LDP on $D_E[0,1]$ with good rate function $I\colon D_E[0,1]\to[0,+\infty]$ such that*

$$I(\mathbf{x}) = \begin{cases} \int_0^1 L\left(\mathbf{x}(t),\dot{\mathbf{x}}(t)\right)dt, & \text{if } \mathbf{x}\in\mathcal{H}_L,\\ +\infty, & \text{in other case.}\end{cases} \tag{4.1}$$

$L\colon E\times\mathbb{R}^3\to\mathbb{R}$ is the cost function given by

$$L(x,\beta) = \begin{cases} (\beta_3+1)\alpha_3^* - d\log\left[1+\left(e^{-\alpha_3^*}-1\right)\frac{dx_3}{x_2}\right] & \\ \quad \text{with } \alpha_3^* = \log\left[\frac{dx_3}{dx_3-x_2}\left(\frac{d}{\beta_3+1}+1\right)\right], & \text{if } \beta_1=1, \beta_2=-2d, \beta_3\geq-(d+1);\\ 0, & \text{if } x_3=\beta_3=0;\\ +\infty, & \text{in other cases,}\end{cases} \tag{4.2}$$

and $\mathcal{H}_L$ is the set of all absolutely continuous function $\mathbf{x}:[0,1]\to E$, $\mathbf{x}(t)=(x_1(t),x_2(t),x_3(t))$ with initial value $\mathbf{x}(0)=(0,d,1)$ and such that the integral $\int_0^1 L\left(\mathbf{x}(t),\dot{\mathbf{x}}(t)\right)dt$ exists and it is finite.

The proof of this theorem, which is based in the results from [Feng and Kurtz, 2006] presented in Section 2.2, is deferred to Section 4.4.

Remark 4.3.1 Let us introduce some comments about the cost function L. The function defined in Equation (4.2) is the Legendre transform w.r.t α of the function $H\colon E\times\mathbb{R}^3\to\mathbb{R}$

given by

$$H(x,\alpha) = \begin{cases} \alpha_1 - 2d\alpha_2 - \alpha_3 + d\log\left[1 + \left(e^{-\alpha_3} - 1\right)\frac{dx_3}{x_2}\right], & \text{if } x_3 > 0, \\ 0, & \text{if } x_3 = 0, \end{cases} \tag{4.3}$$

where $x = (x_1, x_2, x_3)$ and $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, i.e. $L(x,\beta) = \sup_{\alpha\in\mathbb{R}^3}\{\langle\alpha,\beta\rangle - H(x,\alpha)\}$.

Since $H(x,\alpha)$ is convex w.r.t. α (this result is easily proved in Subsection 4.4), the function L is also convex w.r.t. β and verifies $H(x,\alpha) = \sup_{\beta\in\mathbb{R}^3}\{\langle\alpha,\beta\rangle - L(x,\beta)\}$.

Then, as a consequence of Theorem 4.3.1 and Proposition 2.5.1, we deduce the fluid limit for the sequence $\left\{X^N\right\}_N$.

Corollary 4.3.2 (Fluid limit of $\left\{X^N\right\}_N$) *Let $\left\{X^N\right\}_N$ be the sequence of processes defined before. Then,*

1. *The sequence $\left\{X^N\right\}_N$ converges almost-sure, as $N \to \infty$, to the deterministic function $\hat{\mathbf{x}} : [0,1] \to E$, given by $\hat{\mathbf{x}}(t) = \begin{cases} \mathbf{x}(t), & \text{if } t \le T^*, \\ (T^*, 0, 0), & \text{if } t > T^*. \end{cases}$ The function $\mathbf{x}(t) = (s(t), u(t), e(t))$ is the solution of the following ODE:*

$$\begin{cases} \dot{s} = 1, \\ \dot{u} = -2d, \\ \dot{e} = -1 - \frac{d^2 e}{u}, \\ \mathbf{x}(0) = (0, d, 1), \end{cases} \tag{4.4}$$

and T^* is defined by $T^* = \inf\{t \in [0,1] : e(t) = 0\}$.

2. *Moreover, if $d \ge 3$, the unique solution of Equation (4.4) is $\mathbf{x}(t) = (s(t), u(t), e(t))$ with:*

$$\begin{cases} s(t) = t, \\ u(t) = d(-2t+1), \\ e(t) = \frac{1}{d-2}\left[2t - 1 + (d-1)(1-2t)^{\frac{d}{2}}\right], \end{cases} \tag{4.5}$$

and the jamming constant is $T^* = \frac{1}{2}\left[1 - \left(\frac{1}{d-1}\right)^{\frac{2}{d-2}}\right]$.

3. If $d = 2$, then $\mathbf{x}(t) = (s(t), u(t), e(t))$ with:

$$\begin{cases} s(t) = t, \\ u(t) = 2(-2t+1), \\ e(t) = (1-2t)\left[\frac{1}{2}\log(1-2t)+1\right], \end{cases} \tag{4.6}$$

and the jamming constant is $T^* = \frac{1-e^{-2}}{2}$.

Remark 4.3.2 The value of the jamming constant T^* for the case $d \geq 3$ coincides with the known result from [Wormald, 1995], and the value for $d = 2$ coincides with the known result from the earlier work of [Flory, 1939].

Proof. As $L(x, \beta) = 0$ if and only if $\beta = H_\alpha(x, 0)$, where $H_\alpha(x, \alpha)$ are the partial derivatives of $H(x, \alpha)$ w.r.t. $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, the trajectories with zero cost are the ones that verify $\dot{\mathbf{x}} = H_\alpha(\mathbf{x}, 0)$. If $\mathbf{x}(t) = (s(t), u(t), e(t))$, then $\mathbf{x}$ verifies $\dot{\mathbf{x}} = H_\alpha(\mathbf{x}, 0)$ if it verifies the ODE presented in Equation (4.4). For the initial condition $\mathbf{x}(0) = (0, d, 1)$, the unique solution of Equation (4.4) is $\mathbf{x}(t)$, which is presented in Equation (4.5) and (4.6) for the case $d \geq 3$ and $d = 2$. □

4.3.2 LDP for $\{E^N\}_N$

Since the trajectories with positive probability for X_t^N, when $N \to \infty$, are those $\mathbf{x}(t) = (s(t), u(t), e(t))$ such that $s(t) = t$ and $u(t) = d(-2t+1)$, we can directly deduce an LDP for a rescale of the process that counts the number of unexplored vertices in each step of the algorithm.

Theorem 4.3.3 (LDP for $\{E^N\}_N$) *Let $\{E_t^N\}_t$ be a rescale of the process that counts the number of unexplored vertices in each step given by $E_t^N := \frac{E_{[Nt]}^N}{N}$. Then, the sequence of processes $\{E^N\}_N$ such that $E^N = \{E_t^N\}_{t\in[0,1]}$ verifies an LDP in $D_{[0,1]}[0,1]$ with good rate function $\hat{I} : D_{[0,1]}[0,1] \to [0, +\infty]$ such that $\hat{I}(x) = \int_0^1 \hat{L}(t, x(t), \dot{x}(t))\, dt$ if x is an absolutely continuous function with initial condition $x(0) = 1$, and it is $+\infty$ in other case. The cost function $\hat{L}$ is given by*

$$\begin{aligned} \hat{L}(t, x, y) &= L((t, u(t), x), (1, -2d, y)) \\ &= (y+1)\log\left[\frac{dx}{dx - u(t)} \frac{y+1+d}{y+1}\right] - d\log\left[\frac{d(u(t)-dx)}{u(t)(y+1+d)}\right], \end{aligned}$$

where $u(t) = d(-2t+1)$.

Proof. It is deduced directly from Theorem 4.3.1. □

In this case, the cost function has a simple interpretation in terms of the LD rate function for the average of random variables with Binomial distribution, which approximates the distribution corresponding to the number of the new explored vertices in each step of the algorithm.

Proposition 4.3.4 *The cost function $\hat{L}(t,x,y)$ verifies $\hat{L}(t,x,y) = \Lambda^*_{W_{t,x}}(y)$, being $\Lambda^*_{W_{t,x}}(y)$ the LD rate function for the average of i.i.d. random variables $\{W^i_{t,x}\}_{i\in\mathbb{N}}$, where $W_{t,x} = B_{t,x} - d - 1$ and $B_{t,x}$ has a Binomial distribution with parameters $n = d$ and $p = 1 - \frac{dx}{u(t)}$.*

Proof. The rate function given by Crámer's theorem for the average of the random variables $\{W^i_{t,x}\}_{i\in\mathbb{N}}$ is:

$$\Lambda^*_{W_{t,x}}(y) = \sup_{\alpha\in\mathbb{R}} \left\{\alpha y - \Lambda_{W_{t,x}}(\alpha)\right\},$$

where $\Lambda_{W_{t,x}}(\alpha) = \log \mathbb{E}\left(e^{\alpha W_{t,x}}\right)$(see [Dembo and Zeitouni, 1998] for example). To complete the proof it is enough to observe that $\Lambda^*_{W_{t,x}}(y) = \Lambda^*_{B_{t,x}}(y + d + 1)$ coincides with the cost function $\hat{L}(t,x,y)$. □

Remark 4.3.3 The previous result can be explained using the following heuristics. Consider a curve $x(t)$ such that $0 < dx(t) \le u(t)$ for all $t \in [0,1]$, and $E^N_t \approx x(t)$. Then, the infinitesimal increment $\dot{x}(t)$ corresponds to the mean number of new explored vertices in one step, that is:

$$\dot{x}(t) \approx \frac{x(t+h) - x(t)}{h} \approx \frac{E^N_{[N(t+h)]} - E^N_{[Nt]}}{Nh} = \frac{1}{Nh} \sum_{k=[Nt]+1}^{[Nt+Nh]} -\left(1 + \tilde{W}^N_k\right), \tag{4.7}$$

where $\tilde{W}^N_k$ is the number of new explored nodes that share at least one half-edge with v, the selected node in step k. As mentioned before, with high probability, $\tilde{W}^N_k = W^N_k$, being W^N_k the number of half-edges pointing out from v paired with non-explored vertices at step k. If none of the half-edges of v form loops (which happens with high probability, see [Brightwell et al., 2017]), then $W^N_k = d - B^N_k$, where B^N_k is the number of half-edges from v paired with a previously blocked node. The r.v. B^N_k has a Hypergeometric distribution with parameters $U^N_k - d$, $U^N_k - dE^N_k$ and d. For large values of N and $k \in [[Nt]+1, [Nt+Nh]]$, if $\frac{E^N_k}{N}$ is closed to $x(t)$, then B^N_k can be approximated by a Binomial random variable with parameters $n = d$ and $p = \lim_N \frac{U^N_k - dE^N_k}{U^N_k - d} = 1 - \frac{dx(t)}{u(t)}$. Then,

$$\dot{x}(t) \approx \frac{1}{Nh} \sum_{k=[Nt]+1}^{[Nt+Nh]} -\left(1 + \tilde{W}^N_k\right) \approx -1 - d + \frac{1}{Nh} \sum_{k=[Nt]+1}^{[Nt+Nh]} B^N_k. \tag{4.8}$$

Observe that, in particular, the mean macroscopic behaviour $e(t)$ should verify

$$\dot{e}(t) = -1 - d + \mathbb{E}\left(B_{t,e(t)}\right) = -1 - d + d\left(1 - \frac{de(t)}{u(t)}\right) = -1 - \frac{d^2 e(t)}{u(t)},$$

which is the fluid limit we have already seen in Equation (4.4).

Then, the global cost of a deviation of the process E_t^N to a trajectory $x(t)$ can be interpreted as a consequence of the accumulated cost of microscopic deviations of the average of Binomial random variables with parameters $n = d$ and $p = 1 - \frac{dx(t)}{u(t)}$.

4.3.3 Optimization of the rate function $\hat{I}$

Moreover, the optimization problem for the rate $\hat{I}$ over a set of trajectories $A \subset D_{[0,1]}[0,1]$ can be reduced to a one-dimensional optimization problem:

Proposition 4.3.5 (Rate function optimization) *Let $\hat{I}$ be the rate function defined in Theorem 4.3.3. Then,*

1. *If A is a subset of $D_{[0,1]}[0,1]$, then $\inf_{x\in A} \hat{I}(x) = \inf_{\left\{\alpha_0 \in \mathbb{R} : \hat{x}_{\alpha_0} \in \bar{A}\right\}} F(\alpha_0)$, where the closure of A is considered w.r.t. the Skorohod topology, $F(\alpha_0) = \int_0^{T_{\alpha_0}} \hat{L}\left(t, x_{\alpha_0}(t), \dot{x}_{\alpha_0}(t)\right) dt$, and x_{α_0} is such that $(x_{\alpha_0}, y_{\alpha_0})$ is the solution of the following ODE*

$$\begin{cases} \dot{x} = -1 + \frac{dx}{e^y(2t-1+x)-x}, \\ \dot{y} = \frac{d(1-e^y)}{e^y(2t-1+x)-x}, \\ x(0) = 1,\ y(0) = \alpha_0. \end{cases} \tag{4.9}$$

The time T_{α_0} is defined by $T_{\alpha_0} = \inf\left\{t \in [0,1] : x_{\alpha_0}(t) \leq 0\right\}$, and

$$\hat{x}_{\alpha_0}(t) = \begin{cases} x_{\alpha_0}(t), & \text{if } 0 \leq t \leq T_{\alpha_0}, \\ 0, & \text{if } t > T_{\alpha_0}. \end{cases}$$

2. *Moreover, the real function $F(\alpha_0)$ can be written as*

$$F(\alpha_0) = \int_0^{T_{\alpha_0}} \frac{d}{e^{y_{\alpha_0}}(2t-1+x_{\alpha_0}) - x_{\alpha_0}} \left[x_{\alpha_0} y_{\alpha_0} + \left(1 - e^{y_{\alpha_0}}\right)\left(\frac{d}{e^{y_{\alpha_0}}(2t-1+x_{\alpha_0}) - x_{\alpha_0}} + 1\right)\right] dt,$$

being $(x_{\alpha_0}(t),\ y_{\alpha_0}(t))$ the solution of Equation (4.9).

Remark 4.3.4 Let us introduce some comments about the previous proposition.

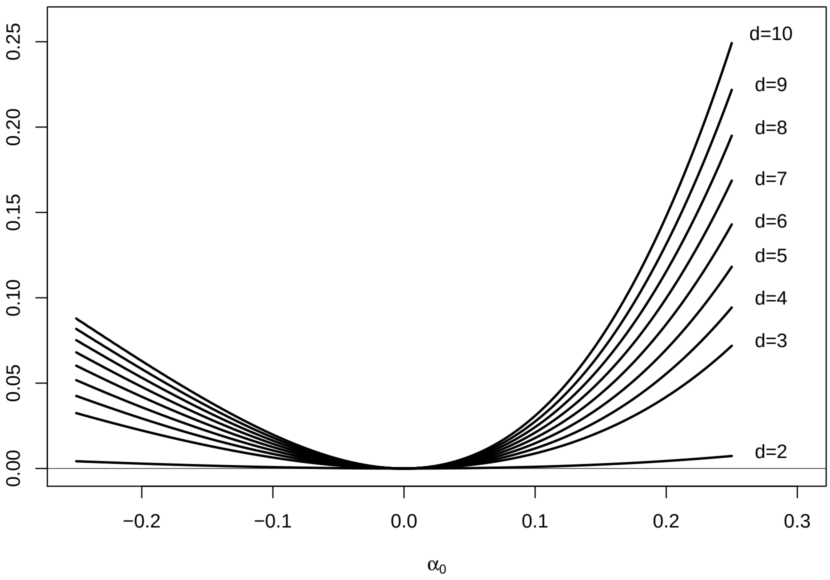

Figure 4.1: Evolution of $F(\alpha_0)$ as function of α_0 for $d = 2, \ldots, 10$.

1. In Figure 4.1, the evolution of F as a function of α_0 is represented for $d = 2, 3, \ldots, 10$. The change that occurs in the graph of $F(\alpha_0)$ for $d = 2$ and $d > 2$ appears to reflect the well-known abrupt change in the geometry for d-regular graphs when $d = 2$ or $d > 2$.

2. As expected, for $\alpha_0 = 0$, $\hat{x}_{\alpha_0}(t)$ coincides with the fluid limit $e(t)$ (i.e., $F(0) = 0$) and $y_{\alpha_0}(t) = 0$ for all t. In Figure 4.2, the graphs of $\hat{x}_{\alpha_0}(t)$ and the fluid limit $e(t)$ are compared for $\alpha_0 < 0$ and $\alpha_0 > 0$.

3. Moreover, note from Equation (4.9) that for any α_0, the initial velocity of $x_{\alpha_0}(t)$ is $\dot{x}_{\alpha_0}(0) = -1 - d$. This makes sense since there are no blocked nodes in the initial step of the process, then $d + 1$ empty vertices will always have to be blocked (since the probability of having multiple-edges converges to zero).

Proof. To prove the first part of Proposition 4.3.5, note that if $\mathbf{x} = (x_1, x_2, x_3) \in \mathcal{H}_L$ is such that $x_3(t) = 0$ for all $t \geq t_0$, then $I(\mathbf{x}) = \int_0^1 L(\mathbf{x}, \dot{\mathbf{x}})\mathrm{d}t = \int_0^{t_0} L(\mathbf{x}, \dot{\mathbf{x}})\mathrm{d}t$, then just consider Hamilton's equations:

$$\begin{cases} \dot{\mathbf{x}} = H_\alpha(\mathbf{x}, \alpha), \\ \dot{\alpha} = -H_x(\mathbf{x}, \alpha), \\ \mathbf{x}(0) = (0, d, 1),\ \alpha(0) = (\alpha_1(0), \alpha_2(0), \alpha_3(0)), \end{cases} \qquad \text{(Hamilton)}, \qquad (4.10)$$

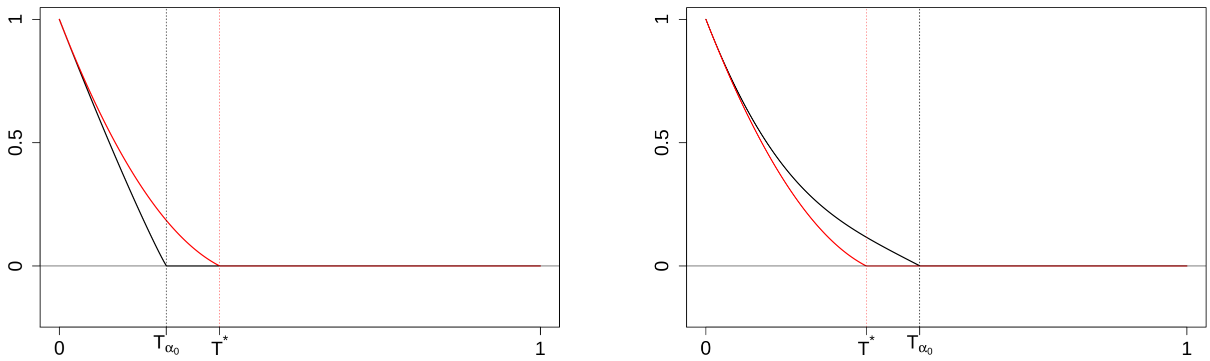

Figure 4.2: Graph of $\hat{x}_{\alpha_0}$ for same value of $\alpha_0 < 0$ (left) and $\alpha_0 > 0$ (right) compared with the fluid limit $e(t)$.

for the case $x_3 > 0$. In Equation (4.10), α is an auxiliary function, and $H_{\mathbf{x}}$, H_α are the vectors of partial derivatives of H w.r.t. x and α. The function H is defined in Equation (4.3) and Equation (4.10) becomes:

$$\begin{cases} \dot{x}_1 = 1 \Rightarrow x_1(t) = t, \\ \dot{x}_2 = -2d \Rightarrow x_2(t) = d(-2t+1), \\ \dot{x}_3 = -1 - \frac{d^2 x_3 e^{-\alpha_3}}{x_2 + \left(e^{-\alpha_3}-1\right) d x_3} = -1 - \frac{d x_3}{e^{\alpha_3}(2t-1+x_3) - x_3}, \\ \dot{\alpha}_1 = 0, \\ \dot{\alpha}_2 = \frac{d^2\left(e^{-\alpha_3}-1\right) d x_3}{x_2\left(x_2 + \left(e^{-\alpha_3}-1\right) d x_3\right)}, \\ \dot{\alpha}_2 = \frac{-d^2\left(e^{-\alpha_3}-1\right)}{x_2 + \left(e^{-\alpha_3}-1\right) d x_3} = \frac{d(1-e^{\alpha_3})}{e^{\alpha_3}(2t-1+x_3) - x_3}, \\ x_1(0) = 0,\ x_2(0) = d,\ x_3(0) = 1. \end{cases} \tag{4.11}$$

Since the equations in α are auxiliary, with the change of notation $x = x_3$, $y = \alpha_3$, the system that we need to solve is presented in Equation (4.9). We are interested in the solution x_{α_0} of (4.9) until the time at which it reaches the value 0, then we take $\hat{x}_{\alpha_0}$ as in the proposition, and get $\inf_{x \in A} I(x) = \inf_{\left\{\alpha_0 : \hat{x}_{\alpha_0} \in \bar{A}\right\}} I\left(\hat{x}_{\alpha_0}\right).$

To prove the second part of the proposition, we use that x_{α_0} is solution of Equation (4.9) for the simplification of the cost function $\hat{L}\left(t, x_{\alpha_0}, \dot{x}_{\alpha_0}\right)$. $\square$

4.3.4 Large deviations for the independent set size

Let $T_N^* \in \{0, \dots, N\}$ be the first discrete-time at which the set of empty vertices remains empty. Note that $\frac{T_N^*}{N}$ coincides with the proportion size of the independent set constructed by the greedy algorithm. In this subsection, we derive large deviation results from Theorem 4.3.3 and Proposition 4.3.5 for the sequence of random variables $\left\{\frac{T_N^*}{N}\right\}_N$.

Theorem 4.3.6 *Let T_N^* be the stopping time of the greedy exploration process over a d-regular graph $G(N,d)$.*

1. *If $\varepsilon > 0$ is such that $T^* + \varepsilon < 1$, then*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \geq T^* + \varepsilon\right) = -F\left(\alpha_0(T^* + \varepsilon)\right),$$

where $\alpha_0(T^* + \varepsilon)$ is the unique real number $\alpha_0 > 0$ such that $T_{\alpha_0} = T^* + \varepsilon$.

2. *If $\varepsilon > 0$ is such that $T^* - \varepsilon > 0$, then*

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \leq T^* - \varepsilon\right) = -F\left(\alpha_0(T^* - \varepsilon)\right),$$

where $\alpha_0(T^* - \varepsilon)$ is the unique real number $\alpha_0 < 0$ such that $T_{\alpha_0} = T^* - \varepsilon$.

In both cases $F(\alpha_0)$ and T_{α_0} are as in Proposition 4.3.5.

Figure 4.3 shows the evolution of $F\left(\alpha_0(T^* \pm \varepsilon)\right)$ as a function of $\varepsilon \in \left[0, \frac{T^*}{4}\right]$ for $d = 3$, 4, 5, 6, 7, 8 and 9, compared with $\varepsilon \in \left[0, \frac{T^*}{6}\right]$ for $d = 2$. Note that in each case the time T^* depends on d. Again, the abrupt change in the dynamics is observed for $d = 2$ and $d > 2$.

Proof. We only prove the first statement. Define the set A_ε such that $A_\varepsilon = \{x \in D_{[0,1]}[0,1] : x(0) = 1, x \text{ is decreasing}, 0 \leq x(t) \leq 1 \text{ for all } t \text{ and } \inf\{t : x(t) = 0\} \geq T^* + \varepsilon\}$. By construction,

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \geq T^* + \varepsilon\right) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(E^N \in A_\varepsilon\right) = -\inf_{\left\{\alpha_0 : \hat{x}_{\alpha_0} \in \bar{A}_\varepsilon\right\}} F\left(\alpha_0\right).$$

Let x_{α_0} be the solution of Equation (4.9) with $y(0) = \alpha_0$. The uniqueness of the solution ensures that the following monotony property with respect to the initial condition α_0 is verified:

$$\text{if } \alpha_0 < \alpha_1 \Rightarrow x_{\alpha_0}(t) < x_{\alpha_1}(t) \text{ for all } t \Rightarrow T_{\alpha_0} < T_{\alpha_1}.$$

In addition, it can be seen that for all $T \in (T^*, 1)$, there exists a unique value $\alpha_0 = \alpha_0(T) > 0$ such that $x_{\alpha_0}(T) = 1$ (i.e. $T = T_{\alpha_0}$). Then, there is only one $\alpha_0^* > 0$ such that $x_{\alpha_0^*}(T^* + \varepsilon) = 1$ and

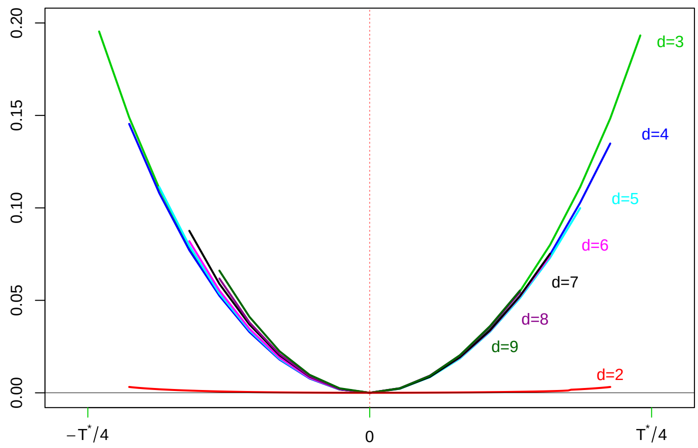

Figure 4.3: Evolution of $F\left(\alpha_0^*(T^* \pm \varepsilon)\right)$ as a function of $\varepsilon > 0$.

- if $\alpha_0 \geq \alpha_0^* \Rightarrow T_{\alpha_0} \geq T^* + \varepsilon \Rightarrow \hat{x}_{\alpha_0} \in A_\varepsilon$,
- if $\alpha_0 < \alpha_0^* \Rightarrow T_{\alpha_0} < T^* + \varepsilon \Rightarrow \hat{x}_{\alpha_0} \notin A_\varepsilon$,

which implies that $\inf_{\{\alpha_0 : \hat{x}_{\alpha_0} \in \bar{A}_\varepsilon\}} F(\alpha_0) = \inf_{\{\alpha_0 \geq \alpha_0^*\}} F(\alpha_0) = F(\alpha_0^*)$. □

4.4 Proof of Theorem 4.3.1

In this section, we prove that the sequence of processes $\left\{X^N\right\}_N$ defined in Section 4.2 verifies the assumptions from [Feng and Kurtz, 2006] presented in Section 2.2. We organize the proof of Theorem 4.3.1 in the steps mentioned in Section 2.2, which are presented as propositions.

As mentioned before, Step 1 and Step 2 assure that $\left\{X^N\right\}_N$ verifies the exponential tightness condition. Then, Step 3 ensures an LDP via the comparison principle, and finally, Step 4 provides the applicable version of the rate, given as an action integral function.

Step 1: Convergence of the nonlinear operators

Let T^N be the linear generator of $\left\{\frac{X_n^N}{N}\right\}_n$, being $\left\{X_n^N\right\}_n$ the discrete-time Markov process defined in Section 4.2 by $X_n^N = \left(S_n^N, U_n^N, E_n^N\right)$. Let $H_N : Dom(H_N) \subset B(E) \to B(E)$ be the non-linear generator given by $H_N(f)(x) = \log\left[e^{-Nf(x)} T^N\left(e^{Nf}\right)(x)\right]$.

Proposition 4.4.1 *There exists a functional* $\mathbf{H}$ *such that* H_N *converges to* $\mathbf{H}$ *when* $N \to \infty$ *in the following sense:* $\lim_{N\to\infty} \sup_{x \in E^N} \left|H_N(f)(x) - \mathbf{H}(f)(x)\right| = 0$ *for all* $f \in C^1(E)$. *The functional*

$\mathbf{H} : C^1(E) \to B(E)$ *is such that* $\mathbf{H}(f)(x) = H(x, \nabla f(x))$*, where* $H : E \times \mathbb{R}^3 \to \mathbb{R}$ *is given by*

$$H(x, \alpha) = \begin{cases} \alpha_1 - 2d\alpha_2 - \alpha_3 + d \log \left[1 + \left(e^{-\alpha_3} - 1\right) \frac{dx_3}{x_2}\right], & \text{if } x_3 > 0, \\ 0, & \text{if } x_3 = 0. \end{cases} \tag{4.12}$$

Proof. Let $x = \frac{1}{N}(\hat{s}_N, \hat{u}_N, \hat{e}_N)$ be an element of E^N with $\hat{s}_N, \hat{e}_N \in \{0, 1, \dots, N\}$, $\hat{u}_N \in \{0, 1, \dots, dN\}$, and $\hat{u}_N \geq d\hat{e}_N > 0$. Then,

$$\begin{aligned} T^N(f)(x) &= \mathbb{E}\left[f\left(\frac{X_{n+1}^N}{N}\right) \Big| \frac{X_n^N}{N} = x\right] \\ &= \mathbb{E}\left[f\left(\frac{1}{N}\left(S_n^N + 1, U_n^N - d - H^N, E_n^N - 1 - \tilde{W}^N\right)\right) \Big| \frac{X_n^N}{N} = x\right] \\ &= \mathbb{E}\left[f\left(x + \frac{1}{N}\left(1, -d - H^N, -1 - \tilde{W}^N\right)\right) \Big| \frac{X_n^N}{N} = x\right] \\ &= \sum_{h=0}^{d} \sum_{b=0}^{h} \sum_{\tilde{w}=0}^{h-b} f\left(x + \frac{1}{N}(1, -d - h, -1 - \tilde{w})\right) p_N(x, h, b, \tilde{w}), \end{aligned}$$

since $\tilde{W}^N \leq W^N = H^N - B^N$. By construction (see Section 4.2),

$$\begin{aligned} p_N(x, h, b, \tilde{w}) &= \mathbb{P}\left(\frac{X_{n+1}^N}{N} = x + \frac{1}{N}(1, -d - h, -1 - \tilde{w}) \Big| \frac{X_n^N}{N} = x\right) \\ &= \mathbb{P}\left(\tilde{W}^N = \tilde{w} \Big| B^N = b;\ H^N = h;\ \frac{X_n^N}{N} = x\right) \\ &\quad \times \mathbb{P}\left(B^N = b \Big| H^N = h;\ \frac{X_n^N}{N} = x\right) \times \mathbb{P}\left(H^N = h \Big| \frac{X_n^N}{N} = x\right). \end{aligned}$$

As is mentioned in Section 4.2, the random variable H^N has a Hypergeometric distribution

$$H^N \sim \mathbf{Hyper}\left(Nu, N(u - de), d\right),$$

B^N has a (conditioned) Hypergeometric distribution

$$B^N \sim \mathbf{Hyper}\left(Nu - d, N(u - de), H^N\right),$$

and

$$\lim_N \mathbb{P}\left(\tilde{W}^N = \tilde{w} \Big| B^N = b;\ H^N = h;\ \frac{X_n^N}{N} = x\right) = \lim_N \mathbb{P}\left(W^N = \tilde{w} \Big| B^N = b;\ H^N = h;\ \frac{X_n^N}{N} = x\right),$$

by Lemma 4.2.1. Then,

$$e^{-Nf(x)}T^N\left(e^{Nf}\right)(x)\approx\sum_{h=0}^{d}\sum_{b=0}^{h}e^{N\left(f\left(x+\frac{1}{N}(1,-d-h,-1-(h-b))\right)-f(x)\right)}p_N(x,h,b,h-b).$$

If $f\in C^2(E)$, then

$$\lim_{N\to\infty}e^{-Nf(x)}T^N\left(e^{Nf}\right)(x)=\sum_{h=0}^{d}\sum_{b=0}^{h}e^{\langle\nabla f(x),(1,-d-h,-1-h+b)\rangle}\lim_{N\to\infty}p_N(x,h,b,h-b).$$

Using Stirling's formula, we obtain that

$$\lim_{N\to\infty}p_N(x,h,b,h-b)=\begin{cases}\frac{d!}{(d-b)!b!}\left(1-\frac{de}{u}\right)^b\left(\frac{de}{u}\right)^{d-b}, & \text{if } 0<b\le h=d,\\ 0, & \text{if } h<d.\end{cases}$$

If $\nabla f(x)=\alpha=(\alpha_1,\alpha_2,\alpha_3)$, then

$$\begin{aligned}\lim_{N\to\infty}e^{-Nf(x)}T^N\left(e^{Nf}\right)(x)&=e^{\alpha_1-2d\alpha_2+(1-d)\alpha_3}\sum_{b=0}^{d}e^{\alpha_3 b}\frac{d!}{(d-b)!b!}\left(1-\frac{de}{u}\right)^b\left(\frac{de}{u}\right)^{d-b}\\&=e^{\alpha_1-2d\alpha_2+(1-d)\alpha_3}\left(\frac{de}{u}+\left(1-\frac{de}{u}\right)e^{\alpha_3}\right)^d.\end{aligned}$$

Then, $\lim_{N\to\infty}H^N(f)(x)=H\left(x,\nabla f(x)\right)$ with $H(x,\alpha)$ defined in Equation (4.12).

If $x=\frac{1}{N}\left(\hat{s}_N,\hat{u}_N,0\right)$, then $T^N(f)(x)=f(x)$ and $H^N(f)(x)=0$. Finally, this result is extended for $f\in C^1(E)$ by taking a sequence $\{f_m\}_m\subset C^2(E)$ such that $\lim_{m\to\infty}\sup_{x\in E}|f_m(x)-f(x)|=0$ and the triangular inequality. $\square$

Proposition 4.4.2 *The function H defined in Equation (4.12) is convex w.r.t. α.*

Proof. Let x be fixed, and $M(x,\alpha)$ be the Hessian matrix of the function $H(x,\alpha)$ w.r.t. $\alpha=(\alpha_1,\alpha_2,\alpha_3)$. Then,

$$M(x,\alpha)=\left(H_{\alpha_i\alpha_j}(x,\alpha)\right)_{i,j=1,2,3}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&m(x,\alpha)\end{pmatrix},$$

being $m(x,\alpha)=\frac{-d^2x_3(x_2-dx_3)e^{\alpha_3}}{(e^{\alpha_3}(x_2-dx_3)+dx_3)^2}\ge 0$. Since the eigenvalues of $M(x,\alpha)$ are 0 and $m(x,\alpha)\ge 0$, $M(x,\alpha)$ is a positive semi-definite matrix and this implies that H is convex w.r.t. α. $\square$

Step 2: Verify the exponential compact containment condition

In this case, $E = [0, 1]$ is compact, so the *exponential compact containment condition* from Definition 2.2.1 is trivially verified by taking $K_\alpha = E$.

Step 3: Comparison principle

In this subsection, we prove that for each $\beta > 0$ and $h \in C(E)$ the *comparison principle* (see Definition 2.3.2) is verified for the following equation:

$$f(x) - \beta H(x, \nabla f(x)) - h(x) = 0. \tag{4.13}$$

Proposition 4.4.3 *For each $\beta > 0$ and $h \in C(E)$ the comparison principle is satisfied for Equation (4.13).*

Proof. Let μ be a subsolution and v a supersolution of Equation (4.13). Let $\psi : E \times E \to \mathbb{R}^+$ be the good penalization function given by $\psi(x, y) = \frac{1}{2}\|x - y\|^2$, and consider the sequences x^α, y^α (with $\alpha \to +\infty$) defined by (see Section 2.3)

$$\mu(x^\alpha) - v(y^\alpha) - \alpha\psi(x^\alpha, y^\alpha) = \sup_{x,y\in E} \{\mu(x) - v(y) - \alpha\psi(x, y)\}.$$

By Proposition 2.3.1, the sequence (x^α, y^α) converges to (z, z) and $z = (z_1, z_2, z_3)$ verifies $\mu(z) - v(z) = \sup_{x\in E} \{\mu(x) - v(x)\}$. As a consequence of Proposition 2.3.2, it is enough to prove that the following inequality holds:

$$\liminf_{\alpha\to\infty} H(x^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) - H(y^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) \le 0,$$

where $\psi_x(x, y) = (\nabla\psi(., y))(x)$ is the vector of partial derivatives of ψ w.r.t. $x = (x_1, x_2, x_3)$. If $z_3 > 0$, then

$$H(x^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) - H(y^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) = d\log\left[\frac{1 + \left(e^{-\alpha(x_3^\alpha - y_3^\alpha)} - 1\right)\frac{dx_3^\alpha}{x_2^\alpha}}{1 + \left(e^{-\alpha(x_3^\alpha - y_3^\alpha)} - 1\right)\frac{dy_3^\alpha}{y_2^\alpha}}\right],$$

and $\liminf_{\alpha\to\infty} H(x^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) - H(y^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) = 0$. For $z_3 = 0$, we repeat the previous analysis, being careful with cases in which $x_3^\alpha = 0$ or $y_3^\alpha = 0$ after a certain α_0 (i.e. $H(x^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) = 0$ or $H(y^\alpha, \alpha\psi_x(x^\alpha, y^\alpha)) = 0$ for all $\alpha > \alpha_0$). □

Step 4: Variational representation of the rate function

Finally, we prove that the rate function can be written as an action functional. As a consequence of the results presented in Subsection 2.2, it is enough to prove that Conditions **8.9**, **8.10**, and **8.11** from [Feng and Kurtz, 2006] (which are presented as Conditions 2.2.4, 2.2.5, and 2.2.6) are verified in this case. We present them as propositions.

In this case, as $\mathbf{H}(f)(x) = H(x, \nabla f(x))$ for each $x \in E$ and $H \leftrightarrow L$, we have that $\mathbf{H}$ can be written as $\mathbf{H}(f)(x) = \sup_{u \in U} \{A(f)(x,u) - L(x,u)\}$, where $U = \mathbb{R}^3$ and $A : C^1(E) \to M(E \times U)$ is the linear operator given by $A(f)(x,u) = \langle \nabla f(x), u \rangle$. As L is convex w.r.t. β, it follows that a deterministic control $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_{u(s)}(\mathrm{d}u)\mathrm{d}s$ is allways the control with smallest cost by Jensen's inequality. Moreover, if $\mathbf{x} : E \to \mathbb{R}^3$ is an absolutely continuous function , then

$$f(\mathbf{x}(t)) - f(\mathbf{x}(0)) = \int_0^t \langle \nabla f(\mathbf{x}(s)), \dot{\mathbf{x}}(s) \rangle \, \mathrm{d}s = \iint_{\mathbb{R}^3 \times [0,t]} A(f)(\mathbf{x}(s),\, u)\, \lambda(\mathrm{d}u \times \mathrm{d}s),$$

if define $\lambda = \lambda(\mathbf{x})$ such that $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_{\dot{\mathbf{x}}(s)}(\mathrm{d}u)\mathrm{d}s$. Let $\Gamma = E \times U$. Then, the supremum in Equation (2.11) for the Nisio semigroup definition is reached on $\{(\mathbf{x}, \lambda) : \mathbf{x} \in \mathcal{AC},\ \mathbf{x}(0) = x_0\} \subset \mathcal{Y}^\Gamma$.

Proposition 4.4.4 *Conditions* **8.9** *from [Feng and Kurtz, 2006] (Condition 2.2.4) are verified.*

The proof of this proposition is identical to that of Proposition 3.4.4, so we omit it.

Proposition 4.4.5 *Condition* **8.10** *from [Feng and Kurtz, 2006] (Condition 2.2.5) is verified.*

Proof. Since $L(x, \beta) = 0 \Leftrightarrow \beta = H_\alpha(x, 0)$, the function $q(x) = H_\alpha(x, 0)$ solves the equation $L(x, q(x)) = 0$ for all $x \in E$. Note that the fluid limit $x(t) = (s(t), u(t), e(t))$ verifies $\dot{x} = q(x)$ with the initial condition $x(0) = (0, d, 1)$. If x is solution of $\dot{x} = q(x)$ with initial condition $x(0) = x_0$ and define λ by $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_{\{q(\mathbf{x}(s))\}}(\mathrm{d}u) \times \mathrm{d}s$, then $(\mathbf{x}, \lambda) \in \mathcal{Y}^\Gamma$ and verifies the required condition. ∎

Proposition 4.4.6 *Condition* **8.11** *from [Feng and Kurtz, 2006] (Condition 2.2.6) is verified.*

Proof. Let $x_0 = (s_0, u_0, e_0) \in E$ and $f \in C^1(E)$ be fixed with $e_0 > 0$. Since $\mathbf{H}(f)(x) = \sup_{\beta \in \mathbb{R}^3} \{\langle \nabla f(x), \beta \rangle - L(x, \beta)\}$, we need to find $(\mathbf{x}, \lambda) \in \mathcal{Y}^\Gamma$ such that

$$\int_0^t H(\mathbf{x}(s), \nabla f(\mathbf{x}(s)))\, \mathrm{d}s = \iint_{U \times [0,t]} \left(\langle \nabla f(\mathbf{x}(s)), u \rangle - L(\mathbf{x}(s), u)\right) \lambda(\mathrm{d}s \times \mathrm{d}u), \qquad (4.14)$$

for all $t \in [0,1]$ and $\mathbf{x}(0) = x_0$. If define $q_f(x) = H_\alpha(x, \nabla f(x))$, then $H(x, \nabla f(x)) = \langle \nabla f(x), q_f(x)\rangle - L(x, q_f(x))$ and Equation (4.14) is verified for any path $\mathbf{x}$ if define $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_{\{q_f(\mathbf{x}(s))\}}(\mathrm{d}u)\,\mathrm{d}s$. Now, we have to add conditions so that in addition $(\mathbf{x}, \lambda)$ belongs to $\mathcal{Y}^\Gamma$ with $\mathbf{x}(0) = x_0$. In particular, $(\mathbf{x}, \lambda)$ must verify:

$$\int_0^t \langle \nabla g(\mathbf{x}(s)), q_f(\mathbf{x}(s))\rangle \,\mathrm{d}s = g(\mathbf{x}(t)) - g(\mathbf{x}(0)) \;\; \forall t \in [0,1], \; \forall g \in C^1(E).$$

Then, we look for a path that solves the following problem:

$$\begin{cases} \mathbf{x} \text{ is differentiable almost everywhere and } \dot{\mathbf{x}}(t) = q_f(\mathbf{x}(t)), \\ \mathbf{x}(0) = x_0, \\ \mathbf{x}(t) \in E \text{ for all } t \geq 0. \end{cases} \tag{4.15}$$

If x verifies Equation (4.15), then the other conditions for (x, λ) to be in $\mathcal{Y}^\Gamma$ are easily verified. $\mathbf{x}(t) = (x_1(t), x_2(t), x_3(t))$ verifies $\dot{x} = q_f(x)$ if and only if:

$$\begin{cases} \dot{x}_1(t) = 1, \\ \dot{x}_2(t) = -2d, \\ \dot{x}_3(t) = -1 - \frac{d^2 x_3(t)}{e^{\frac{\partial f}{\partial x_3}(x_1(t), x_2(t), x_3(t))}(x_2(t) - d x_3(t)) + d x_3(t)}, \\ \mathbf{x}(0) = (s_0, u_0, e_0) \,(\text{ with } u_0 \geq d e_0). \end{cases}$$

Then, $x_1(t) = s_0 + t$, $x_2(t) = -2dt + u_0$ and $x_3(t)$ verifies $\dot{x}_3 = h_f(t, x_3)$ with $h_f(t, x_3) = -1 - \frac{d^2 x_3}{e^{\frac{\partial f}{\partial x_3}(x_1(t), x_2(t), x_3)}(x_2(t) - d x_3) + d x_3}$. Note that $h_f(t, x_3)$ is a continuous function (it does not have to be Lipschitz) and $h_f(t, x_3) \leq 0$. Then, $x_3(t)$ must be decreasing and we can paste local solutions from Peano's Theorem (see [Crandall, 1972]) until the time T_0 at which $x_3(T_0) = 0$.

If $e_0 = 0$, the only possible initial condition is $x_0 = (0, 0, 0)$ and the equality is verified with $\lambda(\mathrm{d}u \times \mathrm{d}s) = \delta_0(u)\mathrm{d}s$. □

Chapter 5

Large deviations for the greedy exploration process on configuration models

Abstract

In this chapter, we extend the results presented in Chapter 4 to the greedy exploration on configuration models, building on a time-discretized version of the method proposed by [Bermolen et al., 2017b] and [Brightwell et al., 2017] by jointly constructing a random graph from a given degree sequence and discovering an independent set in this graph. We prove an LDP for a sequence of Markov processes related to this exploration. The proof of this result follows the general strategy to study large deviations of processes proposed by [Feng and Kurtz, 2006], which is presented in Section 2.2. Moreover, we provide an intuitive interpretation of the LD cost function using Crámer's theorem for the average of random variables with appropriate distribution, depending on the explored vertices distribution. As a corollary, we deduce the corresponding fluid limit and LD results for the independent set size discovered by this exploration algorithm.

The results of this chapter were submitted to *Electronic Communications in Probability* including the results of Chapter 4 as a particular case.

This chapter is organised as follows. In Section 5.1, we give a brief introduction to the chapter. In Section 5.2, we define the dynamic analysed in this chapter, which consists of simultaneously constructing a random graph and an independent set from an initial distribution of degrees. Moreover, we define a sequence of Markov processes related to this algorithm. In Section 5.3, we present the main result of this chapter: a path-state LDP for the sequence of Markov processes defined in Section 5.2 along with the heuristic that motivates the result. The detailed proof is deferred to Section 5.4. As a corollary, we deduce the fluid limit of the process and LD results for the size of the independent set constructed by the exploration algorithm.

5.1 Introduction

In this chapter, we analyse a simple greedy algorithm to construct an independent set over a random graph chosen uniformly from those with a given degree. We use a simultaneous construction of the random graph from a given degree sequence (i.e., a *configuration model*) and an independent set exploration. This idea was first used by [Wormald, 1995] for d-regular graphs and then for [Bermolen et al., 2017b] and [Brightwell et al., 2017] for more general uniform random graphs. We consider a time-discretized version of the algorithm proposed by [Brightwell et al., 2017] for a bounded degree sequence.

We prove a large deviation principle, when the number N of vertices of the graph goes to infinity, for a rescaled version $X_t^N = \frac{X_{[Nt]}^N}{N}$ $(t \in [0,1])$ of the multidimensional Markov chain $\left\{X_n^N\right\}_n$ that counts the number of vertices that have already been placed into the independent set, and the number of *empty* (or non-explored) vertices from each degree at each step n of the algorithm.

The proof of this result follows the general strategy to study large deviations of processes proposed by Feng and Kurtz in [Feng and Kurtz, 2006], which is presented in Section 2.2.

In this case, after working on the four steps mentioned in Section 2.2, we prove that the rate function can be expressed as an action integral. Moreover, its cost function has an intuitive interpretation in terms of Crámer's theorem for the average of random variables with appropriate distribution, depending on the explored vertices distribution in each step of the algorithm.

As a corollary, we deduce the corresponding fluid limit and LD results for independent set size discovered by such an algorithm.

5.2 Description of the model

In this section, we present the dynamics considered in this chapter, which consist of a simultaneous construction of a random graph and an independent set from an initial distribution of degrees.

We start with a set of vertices $\mathcal{V}^N = \{1, 2, \ldots, N\}$, such that $\deg(i) \leq D < \infty$ for all i and such that the initial distribution of degrees $\frac{1}{N}\#\{i : \deg(i) = j\}$ converges to $p_j \geq 0$, when the number of vertices N goes to infinity, for all $j = 0, \ldots D$ with $\sum_{j=0}^{D} p_j = 1$. Let us denote $\lambda = \sum_j j p_j$. Each vertex i of the graph has a number $\deg(i)$ of half-edges available to be paired with the half-edges of other vertices. Next we describe how these half-edges are paired as the random graph is constructed.

At each step $n = 0, 1, \dots, N$, the set $\mathcal{V}^N$ is partitioned into three classes:

- a set $\mathcal{S}_n^N$ of vertices that have already been placed into the independent set, with all half-edges paired with vertices out of $\mathcal{S}_n^N$;
- a set $\mathcal{B}_n^N$ of *blocked vertices*, where at least one of its half-edges has been paired with a half-edge from $\mathcal{S}_n^N$;
- a set $\mathcal{E}_n^N$ of *empty vertices*, from which no half-edge has yet been paired. $\mathcal{E}_n^N$ can be decomposed as $\mathcal{E}_n^N = \bigcup_{j=0}^{D} \mathcal{E}_n^N(j)$, where $\mathcal{E}_n^N(j)$ is the set of empty vertices of degree j at step n.

Initially, all vertices are *empty*, i.e. $\mathcal{E}_0^N = \mathcal{V}^N$ and $\mathcal{S}_0^N = \emptyset$. At step n, a vertex v is selected uniformly from $\mathcal{E}_n^N$, it is placed into $\mathcal{S}_n^N$, and all its half-edges are paired, drawing uniformly within the available half-edges. This pairing procedure results in the following updates:

- v is moved from $\mathcal{E}_n^N$ to $\mathcal{S}_n^N$,
- each half-edge incident to v (if it has some) is paired with some other uniformly randomly chosen vertices among the currently unpaired half-edges,
- all vertices in $\mathcal{E}_n^N$ with some half-edges already paired with a half-edge from v are moved to $\mathcal{B}_n^N$.

Note that some half-edges from v may be paired with half-edges from $\mathcal{B}_n^N$, or indeed with other half-edges from v, and no change in the status of a vertex results from such pairings. At each step n, the only paired edges are those with at least one endpoint in $\mathcal{S}_n^N$. This is the main difference between the dynamics described in [Bermolen et al., 2017b] and [Brightwell et al., 2017] for a continuous-time version of this algorithm. In [Bermolen et al., 2017b], the neighbours of blocked vertices are revealed at each step, meaning that degrees of empty vertices can change over time. For simplicity, we do not do this.

The algorithm terminates at the first step $n = T_N^*$ at which $\mathcal{E}_n^N = \emptyset$. At this point, there may still be some unpaired half-edges pointing out from blocked vertices. These may be paired off uniformly at random to complete the construction of the graph $G\left(N, (\deg(1), \dots, \deg(N))\right)$. Note that T_N^* coincides with the size of the independent set constructed by the algorithm. As we mentioned before, the expected value of $\frac{T_N^*}{N}$ is the *jamming constant* of the graph.

For each $n \in \{0, 1, \dots, N\}$, let us denote $X_n^N = \left(S_n^N, U_n^N, E_n^N(0), E_n^N(1), \dots, E_n^N(D)\right)$ with:

- $S_n^N = \left|\mathcal{S}_n^N\right|$, the number of vertices that have already been placed into the independent set at step n;

- U_n^N, the total number of unpaired half-edges (corresponding to empty or blocked vertices) at step n;
- $E_n^N(j) = \left|\mathcal{E}_n^N(j)\right|$, the number of empty vertices with degree j at step n.

$\left\{X_n^N\right\}_n$ is a discrete-time Markov process in $\mathbb{R}^{D+3}$. By construction, it is updated at step $n+1$ as follows:

- The vertex v is placed into $\mathcal{S}_n^N$. Then, $S_{n+1}^N = S_n^N + 1$.
- If $v \in \mathcal{E}_n^N(k)$ with $k \neq 0$, then:
  1. Each one of the k half-edges pointing out from v is paired in turn with some other uniformly randomly chosen between the currently unpaired half-edges. Let H^N be the number of half-edges from v that are paired with another vertex different from v (blocked or empty), i.e., that do not form loops. Then, $U_{n+1}^N = U_n^N - k - H^N$. Note that H^N has a Hypergeometric distribution:
  $$H^N \sim \textbf{Hyper}\left(U_n^N, U_n^N - k, k\right).$$
  2. We have to distribute those H^N half-edges between the $U_n^N - \sum_j jE_n^N(j)$ half-edges corresponding to blocked vertices and the $\sum_j jE_n^N(j)$ half-edges corresponding to empties. Let B^N be the number of half-edges of v that are paired to blocked vertices, then B^N has a Hypergeometric (conditioned to H^N) distribution:
  $$B^N \sim \textbf{Hyper}\left(U_n^N - k,\, U_n^N - \sum_{j=1}^{D} jE_n^N(j),\, H^N\right).$$
  3. Now, if $H^N = h$ (with $h \leq k$) and $B^N = b$ (with $b \leq h$), there are $h - b$ half-edges to distribute among the empties. Let W_j^N be the number of half-edges from v that are connected to some $w \in \mathcal{E}_n^N(j)$. Note that $\left(W_1^N, \ldots, W_D^N\right)$ has a (multivariate) Hypergeometric distribution:
  $$\textbf{Hyper}\left(\sum_j jE_n^N(j) - k, E_n^N(1), \ldots, k\left(E_n^N(k) - 1\right), \ldots, DE_n^N(D), h - b\right).$$
  4. Finally, let $\tilde{W}_j^N$ be the number of empty vertices of degree j that share at least one edge with v. Then, $E_{n+1}^N(0) = E_n^N(0)$, $E_{n+1}^N(j) = E_n^N(j) - \tilde{W}_j^N$ if $j \neq k$ and $E_{n+1}^N(k) = E_n^N(k) - 1 - \tilde{W}_k^N$.
- If $deg(v) = 0$, then $S_{n+1}^N = S_n^N + 1$, $U_{n+1}^N = U_n^N$, $E_{n+1}^N(0) = E_n^N(0) - 1$ and $E_{n+1}^N(j) = E_n^N(j)$ for all $j \neq 0$.

As in the previous chapter, the following lemma assures that the distribution of $\left(\tilde{W}_1, \ldots, \tilde{W}_D\right)$ can be approximated by the Hypergeometric distribution corresponding to $(W_1, \ldots, W_D)$.

Lemma 5.2.1 *Let $\hat{x} = (\hat{s}, \hat{u}, \hat{e}_0, \ldots, \hat{e}_D)$ be an element in the state space of $\left\{X_n^N\right\}_n$, and $(\omega_j)_j$ with $0 \le \omega_j \le \hat{e}_j$ such that $\sum_j \omega_j \le h - b$. Then,*

$$\lim_{N\to\infty} \mathbb{P}\left(\left(\tilde{W}_j^N\right)_j = (\omega_j)_j \,\middle|\, X_n^N = \hat{x}; \mathbf{deg}(v) = k; H^N = h; B^N = b; \left(W_j^N\right)_j = (\omega_j)_j\right) = 1$$

Proof. See Equation **17** from [Bermolen et al., 2017b]. In the article notation: $W_j^N = Y(\mu_{t^-})(j)$ and $\tilde{W}_j^N = \tilde{Y}(\mu_{t^-})(j)$. $\square$

Let $X_t^N := \frac{X_{[Nt]}^N}{N}$ be a rescaled version of X_n^N with $t \in [0,1]$. The state space of X_t^N is $E^N = \left\{\frac{1}{N}(\hat{s}, \hat{u}, \hat{e}_0, \ldots, \hat{e}_D) : \hat{s}, \hat{e}_i \in \{0, \ldots, N\}; \sum_j j\hat{e}_j \le \hat{u}\right\}$ which is included in the compact set $E := \left\{(s, u, e_0, e_1, \ldots, e_D) \in [0,1] \times \mathbb{R} \times [0,1]^{D+1} : \sum_j je_j \le u \le \lambda\right\}$. Moreover, the size of the independent set constructed by such an algorithm is given by

$$T_N^* = \inf\left\{n : \sum_j E_n^N(j) = 0\right\} = N \inf\left\{t \in [0,1] : \sum_j E_{[Nt]}^N(j) = 0\right\}.$$

In this chapter, we extend the results presented in the previous chapter for more general uniform random graphs by analysing large deviations results for both $\left\{X_\cdot^N\right\}_N$ and $\left\{\frac{T_N^*}{N}\right\}_N$.

5.3 Main Results

In this section, we present the main results of this chapter. In Subsection 5.3.1, we present an LDP for $X^N = \left\{X_t^N\right\}_{0\le t\le 1}$ and a heuristic description to derive this result. The proof of this theorem is based on the work of [Feng and Kurtz, 2006] and is deferred to Section 5.4. In Subsection 5.3.2, we deduce the corresponding fluid limit. In Subsection 5.3.3, we provide a way of finding the trajectory that minimises the LD rate function over a set of trajectories (i.e., the most probable trajectory) by studying the Hamiltonian dynamics associated with the rate function obtained. Finally, in Subsection 5.3.4, we deduce large deviations results for the independent set size discovered by such an algorithm.

5.3.1 LDP for $\left\{X^N\right\}_N$

The theorem that we state below is the most important result of this chapter.

Theorem 5.3.1 (LDP for $\{X^N\}_N$) *The sequence $\{X^N\}_N$ with $X^N = \{X_t^N\}_{0\le t\le 1}$ verifies an LDP on $D_E[0,1]$ with good rate function $I\colon D_E[0,1] \to [0,+\infty]$ such that $I(\mathbf{x}) = \int_0^1 L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\,dt$ if $\mathbf{x} \in \mathcal{H}_L$ and it is $+\infty$ in other case. $L\colon E \times \mathbb{R}^{D+3} \to \mathbb{R}$ is the cost function*

$$L(x,\beta) = \sup_{\alpha\in\mathbb{R}^{D+3}} \{\langle \alpha, \beta\rangle - H(x,\alpha)\}, \tag{5.1}$$

where $H\colon E\times\mathbb{R}^{D+3} \to \mathbb{R}$ is the convex function w.r.t. the second variable given by

$$H(x,\alpha) = \begin{cases} \log\left[\sum_k e^{\alpha_s - 2k\alpha_u - \alpha_k}\left(1 + \sum_{j=1}^{D}\left(e^{-\alpha_j} - 1\right)\frac{je_j}{u}\right)^k \frac{e_k}{\sum_j e_j}\right], & \text{if } \sum_j e_j > 0 \\ 0, & \text{if } \sum_j e_j = 0, \end{cases} \tag{5.2}$$

with $x = (s, u, e_0, e_1, \ldots, e_D)$ and $\alpha = (\alpha_s, \alpha_u, \alpha_0, \alpha_1, \ldots, \alpha_D)$. $\mathcal{H}_L$ is the set of all absolutely continuous function $\mathbf{x} : [0,1] \to E$, $\mathbf{x}(t) = (s(t), u(t), e_0(t), e_1(t), \ldots, e_D(t))$ with initial value $\mathbf{x}(0) = (0, \lambda, p_0, p_1, \ldots, p_D)$ and such that $s(t)$ is increasing, $u(t)$ and $e_j(t)$ are decreasing, and the integral $\int_0^1 L(\mathbf{x}(t), \dot{\mathbf{x}}(t))\,dt$ exists and it is finite.

A detailed proof of this theorem is deferred to Section 5.4. The convexity of $H(x,\alpha)$ w.r.t. α is stated in Proposition 5.4.2.

Remark 5.3.1 Though we believe that the case of unbounded degree distribution support would not raise different conclusions (under appropriate conditions on tails), the proof would become much more technical. We leave it for future work.

In what follows we provide an intuitive way to construct the cost function L in terms of the rate function provided by Cramer's theorem for the average of the approximated distribution of the new explored vertices in one step conditioning to the number of explored vertices.

Consider a curve $\mathbf{x}(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$ contained in E and such that $X_t^N \approx \mathbf{x}(t)$. Then, the infinitesimal increments $\dot{\mathbf{x}}(t) = (\dot{s}(t), \dot{u}(t), \dot{e}_0(t), \ldots, \dot{e}_D(t))$ correspond to the mean number of new explored vertices from each degree in one step, as can be deduced from the following statement:

$$\dot{\mathbf{x}}(t) \approx \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h} \approx \frac{X^N_{[N(t+h)]} - X^N_{[Nt]}}{Nh} = \frac{1}{Nh}\sum_{n=[Nt]}^{[N(t+h)]-1}\left(X^N_{n+1} - X^N_n\right).$$

Proposition 5.3.2 *The distribution of the number of new explored vertices in one step $X_{n+1}^N - X_n^N$, conditioning to $X_t^N \approx x(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$, can be approximated by the* *random vector:*

$$Z^{\mathbf{x}(t)} = \begin{cases} (1,0,-1,0,\ldots,0)\,, & \textit{with probability } \frac{e_0(t)}{\sum_j e_j(t)}, \\ (1,-2k,0,-M_1,\ldots,-1-M_k,\ldots,-M_D)\,, & \textit{with probability } \frac{e_k(t)}{\sum_j e_j(t)}\ (1 \le k \le D). \end{cases} \tag{5.3}$$

In Equation (5.3), the coordinates M_i correspond to

$$(M_1,\ldots,M_D) \sim \mathbf{Mult}\,(K-B,q_1,\ldots,q_D)\,,$$

a Multinomial vector depending on $K \in \{0,\ldots,D\}$ such that $\mathbb{P}(K=k) = \frac{e_k(t)}{\sum_j e_j(t)}$, $B = B(K) \sim \mathbf{Bin}\left(K, 1-\frac{\sum_j je_j(t)}{u(t)}\right)$, and $q_i = \frac{ie_i(t)}{\sum_j je_j(t)}$.

Proof. At step n, the vertex $v \in \mathcal{E}_n^N(k)$ is drawn uniformly.
If $k \neq 0$, then $X_{n+1}^N - X_n^N = \left(1,-k-H^N,0,-\tilde{W}_1^N,\ldots,-1-\tilde{W}_k^N,\ldots,-\tilde{W}_D^N\right)$, where

- H^N is the number of half-edges from v that are joined to another vertex different from v. As the probability of loops converges to zero (see [Brightwell et al., 2017]), then $H^N \approx k$.
- Lemma 5.2.1 assures that $\tilde{W}_j^N \approx W_j^N$, where $\left(W_1^N,\ldots,W_D^N\right)$ has a (multivariate) Hypergeometric distribution with parameters $\sum_j jE_n^N(j) - k$, $E_n^N(1)$, $\ldots$, $jE_n^N(j)$, $\ldots$, $k\left(E_n^N(k)-1\right)$, $\ldots$, $DE_n^N(D)$, and $k - B^N$.
  Note that B^N can be approximated by a Binomial random variable B with parameters $n = k$ and $p = \lim_N \frac{U_n^N - \sum_j jE_n^N(j)}{U_n^N - k} = 1 - \frac{\sum_j je_j(t)}{u(t)}$. Moreover, $\left(W_1^N,\ldots,W_D^N\right)$ can be approximated by the Multinomial random vector $(M_1,\ldots,M_D) \sim \mathbf{Mult}\,(k-B,q_1,\ldots,q_D)$, with $q_i = \frac{ie_i(t)}{\sum_j je_j(t)}$.

If $v \in \mathcal{E}_n^N(0)$, then $X_{n+1}^N - X_n^N = (1,0,-1,0,\ldots,0)$. □

Proposition 5.3.3 *The cost function $L\left(\mathbf{x}(t),\dot{\mathbf{x}}(t)\right)$ defined in Equation (5.1) coincides with the LDP rate function for the average of i.i.d random variables $\left\{Z_i^{\mathbf{x}(t)}\right\}_{i\in\mathbb{N}}$ distributed as $Z^{\mathbf{x}(t)}$:*

$$L\left(\mathbf{x}(t),\dot{\mathbf{x}}(t)\right) = \sup_{\alpha\in\mathbb{R}^{D+3}} \left\{\langle\alpha,\dot{\mathbf{x}}(t)\rangle - H\left(\mathbf{x}(t),\alpha\right)\right\} = \Lambda^*_{Z^{\mathbf{x}(t)}}\left(\dot{\mathbf{x}}(t)\right).$$

Proof. Assuming that the random variables $\left\{Z_i^{\mathbf{x}(t)}\right\}_i$ are i.i.d., Cramér's theorem states that the LDP rate function for the average of these variables is

$$I(x) = \Lambda^*_{Z^{\mathbf{x}(t)}}(x) = \sup_{\alpha\in\mathbb{R}^{D+3}} \left\{\langle\alpha,x\rangle - \Lambda_{Z^{\mathbf{x}(t)}}(\alpha)\right\},\ \text{with } \Lambda_{Z^{\mathbf{x}(t)}}(\alpha) = \log\mathbb{E}\left[e^{\left\langle\alpha,Z^{\mathbf{x}(t)}\right\rangle}\right].$$

In this case, with $\alpha = (\alpha_s, \alpha_u, \alpha_0, \ldots, \alpha_D)$, we have that $\Lambda_{Z^{\mathbf{x}(t)}}(\alpha) = H(\mathbf{x}(t), \alpha)$, being $H(x, \alpha)$ ($H : E \times \mathbb{R}^{D+3} \to \mathbb{R}$) the log of the moment-generating function of the (conditioned) Multinomial vector Z^x (with $x \in E$) evaluated in α, which is presented in Equation (5.2).

Defining $L : E \times \mathbb{R}^{D+3} \to \mathbb{R}$ as in Equation (5.1), results that $L(\mathbf{x}(t), \dot{\mathbf{x}}(t)) = \sup_\alpha \{\langle \alpha, \dot{\mathbf{x}}(t)\rangle - H(\mathbf{x}(t), \alpha)\}$ coincides with $\Lambda^*_{Z^{\mathbf{x}(t)}}(\dot{\mathbf{x}}(t))$. $\square$

This means that the global cost of a deviation of $\{X_t^N\}_t$ to a trajectory $\mathbf{x}(t)$ can be interpreted as a consequence of the accumulated cost of microscopic deviations of the average of (conditioned) Multinomial random vectors, representing the degrees of the new explored nodes in one step.

Remark 5.3.2 (Fluid limit) Observe that, in particular, the mean macroscopic behaviour $\mathbf{x}(t)$ should verify:

$$\dot{\mathbf{x}}(t) \approx \mathbb{E}\left(Z^{\mathbf{x}(t)}\right) = (1, 0, -1, 0, \ldots, 0)\,\frac{e_0(t)}{\sum_j e_j(t)}$$
$$+ \sum_{k=1}^{D}\left(1, -2k, 0, -k\frac{e_1(t)}{u(t)}, \ldots, -k\frac{je_j(t)}{u(t)}, \ldots, -1-k\frac{ke_k(t)}{u(t)}, \ldots, -k\frac{De_D(t)}{u(t)}\right)\frac{e_k(t)}{\sum_j e_j(t)},$$

which coincides with the fluid limit that we formally prove in the following subsection.

5.3.2 Fluid limit of the process $\{X_t^N\}_t$

In this subsection, we formally deduce the fluid limit of $\{X_t^N\}_t$ as a consequence of Theorem 5.3.1 and Proposition 2.5.1.

Proposition 5.3.4 (Fluid limit) *The sequence of processes $\{X^N\}_N$ converges almost-sure, as $N \to \infty$, to the deterministic function $\hat{x} : [0, 1] \to E$ given by*

$$\hat{x}(t) = \begin{cases} (s(t), \hat{u}(t), \hat{e}_0(t), \ldots, \hat{e}_D(t)), & \text{if } t \le T^*, \\ (T^*, 0, \ldots, 0), & \text{if } t > T^*, \end{cases} \quad \text{where } \hat{e}_i(t) = \begin{cases} e_i(t), & \text{if } t \le t_i, \\ 0, & \text{if } t > t_i. \end{cases}$$

The times t_i are defined by $t_i = \inf\{t \in [0, 1] : e_i(t_i) \le 0\}$ and $x(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$ is (the) solution of the following ordinary differential equation:

$$\begin{cases} \dot{s} = 1, \\ \dot{u} = \frac{-2\sum_k ke_k}{\sum_k e_k}, \\ \dot{e}_i = \frac{-e_i - \frac{ie_i}{u}\sum_k ke_k}{\sum_k e_k}, \quad i = 0, \ldots, D, \\ s(0) = 0,\ u(0) = \lambda,\ e_i(0) = p_i. \end{cases} \tag{5.4}$$

$\hat{u}$ *is the solution of Equation* (5.4) *replacing* e_i *by* $\hat{e}_i$ *and* $T^* = \inf\{t \in [0,1] : \sum_k \hat{e}_k(t) = 0\} = \max\{t_0, \dots, t_D\}$.

Note that the dynamics described by Equation (5.4) and the fluid limit for the process considered in [Brightwell et al., 2017] differ only because of the term $\sum_k e_k$, which corresponds to the rate at which the clock of an unexplored vertex rings in the continuous-time case.

Proof. The cost function $L(x, \beta)$ defined in Equation (5.1) satisfies $L(x, \beta) \geq 0$ and $L(x, \beta) = 0$ if and only if $\beta = H_\alpha(x, 0)$, where $H_\alpha(x, \alpha)$ are the partial derivatives of $H(x, \alpha)$ w.r.t. $\alpha = (\alpha_s, \alpha_u, \alpha_0, \dots, \alpha_D)$. Then, the trajectories with zero cost are the ones that verify $\dot{x} = H_\alpha(x, 0)$. If in addition we impose the condition $x(0) = (0, \lambda, p_0, \dots, p_D)$, we obtain the autonomous Equation (5.4). Cauchy-Peano existence theorem ensures the existence of at least one solution of such equation. Let $\mathcal{D} = \{x \in E : e_i > 0\, \forall i\}$ and $f(x) = H_\alpha(x, 0)$. Then f is a C^1-function on $\mathcal{D}$, i.e. it is a locally Lipschitz continuous function on $\mathcal{D}$. This implies the uniqueness of solutions $e_i(t)$ for equation $\begin{cases} \dot{x} = f(x), \\ x(0) = x_0 \in \mathcal{D}, \end{cases}$ until the time t_i at which $e_i(t_i) = 0$, and then we take the solution $e_i(t) = 0$ for all $t \geq t_i$. $\square$

5.3.3 Optimization of the rate function.

The following proposition consists transforms the optimization problem of the rate function I over a set of trajectories into a real optimization problem.

Proposition 5.3.5 (Rate function optimization) *Let A be a subset of $D_E[0,1]$. Then,*

$$\inf_{x \in A} I(x) = \inf_{\left\{\alpha_0 \in \mathbb{R}^{D+3} : \hat{x}_{\alpha_0} \in \bar{A}\right\}} I\left(\hat{x}_{\alpha_0}\right),$$

where the closure of A is considered w.r.t. the Skorohod topology,

$$\hat{x}_{\alpha_0}(t) = \begin{cases} \left(s_{\alpha_0}(t), \hat{u}_{\alpha_0}(t), \hat{e}_{0,\alpha_0}(t), \dots, \hat{e}_{D,\alpha_0}(t)\right), & \text{if } t \leq T_{\alpha_0}, \\ \left(T_{\alpha_0}, 0, \dots, 0\right), & \text{if } t > T_{\alpha_0}, \end{cases}$$

$$\hat{e}_{i,\alpha_0}(t) = \begin{cases} e_{i,\alpha_0}(t), & \text{if } t \leq t_{i,\alpha_0}, \\ 0, & \text{if } t > t_{i,\alpha_0}, \end{cases} \qquad t_{i,\alpha_0} = \inf\{t \in [0,1] : e_{i,\alpha_0}(t) \leq 0\},$$

and $x_{\alpha_0}(t) = (s_{\alpha_0}(t), u_{\alpha_0}(t), e_{0,\alpha_0}(t), \dots, e_{D,\alpha_0}(t))$ is (the) solution of the following ordinary differential equation:

$$\begin{cases} \dot{x} = H_\alpha(x, \alpha), \\ \dot{\alpha} = -H_x(x, \alpha), \\ x(0) = (0, \lambda, p_0, \dots, p_D), \ \alpha(0) = \alpha_0. \end{cases} \tag{5.5}$$

H_x and H_α are the vectors of partial derivatives of H w.r.t. x and α, and $T_{\alpha_0} = \inf\{t \in [0,1] : \sum_k \hat{e}_{k,\alpha_0}(t) = 0\}$.

Remark 5.3.3 As expected, for $\alpha_0 = (0, 0, \ldots, 0)$, $\hat{x}_{\alpha_0}(t)$ coincides with the fluid limit, which is solution of Equation (5.4), and $\alpha(t) = (0, \ldots, 0)$ for all t. Then $\inf_{x \in A} I(x) = 0$ if the fluid limit belongs to A.

Proof. Note that if $\mathbf{x} \in \mathcal{H}_L$ is such that $\mathbf{x}(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$ and $\sum_k e_k(t) = 0$ for all $t \geq t_0$, then $I(\mathbf{x}) = \int_0^1 L(\mathbf{x}, \dot{\mathbf{x}})\mathrm{d}t = \int_0^{t_0} L(\mathbf{x}, \dot{\mathbf{x}})\mathrm{d}t$, so just consider Hamilton's equations for the case $\sum_k e_k > 0$. Hamilton's equations, presented in Equation (5.5), give conditions for a function $\mathbf{x}$ to be a *stationary curve* of the functional I. Note that α is an auxiliary function. $\square$

5.3.4 Large deviations for the independent set size

From previous results, we can deduce an LDP for the sequence of stopping times $\frac{T_N^*}{N}$, which coincide with the proportion size of the independent set constructed by the algorithm.

Theorem 5.3.6 *Consider T_N^* defined before as the stopping time of the algorithm presented in Section 5.2.*

1. *If $\varepsilon > 0$ is such that $T^* + \varepsilon < 1$, then*

$$\lim_N \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \geq T^* + \varepsilon\right) = -F^+\left(T^* + \varepsilon\right),$$

   being $F^+\left(T^* + \varepsilon\right) = \inf\left\{I\left(\hat{x}_{\alpha_0}\right) : \ T_{\alpha_0} \geq T^* + \varepsilon, \alpha_0 \in \mathbb{R}^{D+3}\right\}$.

2. *If $\varepsilon > 0$ is such that $T^* - \varepsilon > 0$, then*

$$\lim_N \frac{1}{N} \log \mathbb{P}\left(\frac{T_N^*}{N} \leq T^* - \varepsilon\right) = -F^-\left(T^* - \varepsilon\right),$$

   being $F^-\left(T^* - \varepsilon\right) = \inf\left\{I\left(\hat{x}_{\alpha_0}\right) : \ T_{\alpha_0} \leq T^* - \varepsilon, \alpha_0 \in \mathbb{R}^{D+3}\right\}$.

In both cases $\hat{x}_{\alpha_0}$ and T_{α_0} are as in Proposition 5.3.5.

Proof. We only prove the first statement. Define the set A_ε, that contains the trajectories $\mathbf{x} \in D_E[0,1]$ such that $\mathbf{x}(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$, $\mathbf{x}(0) = (0, \lambda, p_0, \ldots, p_d)$, coordinates $e_i(t)$, $u(t)$ are decreasing, $s(t)$ is increasing, $0 \leq e_i(t), s(t) \leq 1$ for all t, and such that

$\inf\{t: \sum_k e_k(t)=0\} \geq T^*+\varepsilon$. Then, Proposition 5.3.5 implies that

$$\begin{aligned}\lim_N \frac{1}{N}\log\mathbb{P}\left(\frac{T_N^*}{N}\geq T^*+\varepsilon\right) &= \lim_N \frac{1}{N}\log\mathbb{P}\left(X_\cdot^N\in A_\varepsilon\right)\\ &= -\inf_{\left\{\alpha_0\in\mathbb{R}^{D+3}:\,\hat{x}_{\alpha_0}\in\bar{A}_\varepsilon\right\}} I\left(\hat{x}_{\alpha_0}\right) = F^+\left(T^*+\varepsilon\right).\end{aligned}$$

□

5.4 Proof of Theorem 5.3.1

Finally, in this section we prove that the sequence of processes $\{X^N\}_N$ defined in Section 5.2 verifies the assumptions from [Feng and Kurtz, 2006] presented in Section 2.2. We organize the proof of Theorem 5.3.1 in the steps mentioned in Section 2.2, which are presented as propositions.

Step 1: Convergence of the nonlinear operators

Let T^N be the linear generator of $\left\{\frac{X_n^N}{N}\right\}_n$, being $\{X_n^N\}_n$ the discrete-time Markov process defined in Section 5.2 by $X_n^N=\left(S_n^N,U_n^N,E_n^N(0),E_n^N(1),\ldots,E_n^N(D)\right)$. Let $H_N: Dom(H_N)\subset B(E)\to B(E)$ be the non-linear generator given by $H_N(f)(x)=\log\left[e^{-Nf(x)}T^N\left(e^{Nf}\right)(x)\right]$.

Proposition 5.4.1 *There exists a functional* $\mathbf{H}$ *such that* H_N *converges to* $\mathbf{H}$ *when* $N\to\infty$ *in the following sense:* $\lim\limits_{N\to\infty}\sup\limits_{x\in E^N}|H_N(f)(x)-\mathbf{H}(f)(x)|=0$ *for all* $f\in C^1(E)$. *The functional* $\mathbf{H}: C^1(E)\to B(E)$ *is such that* $\mathbf{H}(f)(x)=H(x,\nabla f(x))$, *where* $H: E\times\mathbb{R}\to\mathbb{R}$ *is given by*

$$H(x,\alpha)=\begin{cases}\log\left[\displaystyle\sum_{k=0}^{D}e^{\alpha_s-2k\alpha_u-\alpha_k}\left(1+\sum_{j=1}^{D}\left(e^{-\alpha_j}-1\right)\frac{je_j}{u}\right)^k\frac{e_k}{\displaystyle\sum_{j=0}^{D}e_j}\right], & \\ & \text{if } \sum_{j=0}^{D}e_j>0,\\ 0, & \text{if } \displaystyle\sum_{j=0}^{D}e_j=0.\end{cases}\tag{5.6}$$

Proof. Let be $x=(s,u,e_0,\ldots,e_D)\in E^N$ such that $x=\frac{1}{N}\left(\hat{s}_N,\hat{u}_N,\hat{e}_{0,N},\ldots,\hat{e}_{D,N}\right)$ with

$\hat{s}_N, \hat{e}_{i,N} \in \{0, 1, \ldots, N\}$, $\sum_j j\hat{e}_{j,N} \leq \hat{u}_N \leq \lambda N$, and $\sum_j \hat{e}_{j,N} > 0$. Then,

$$\begin{aligned}
T^N(f)(x) &= \mathbb{E}\left[f\left(\frac{X_{n+1}^N}{N} \right) \Big| \frac{X_n^N}{N} = x \right] \\
&= f\left(x + \frac{1}{N}(1, 0, -1, 0, \ldots, 0) \right) \frac{e_0}{\sum_j e_j} \\
&+ \sum_{k=1}^{D} \sum_{h=0}^{k} \sum_{b=0}^{h} \sum_{\tilde{w}_j : \sum \tilde{w}_j \leq h-b} f\left(x + \frac{1}{N}(1, -k-h, 0, -\tilde{w}_1, \ldots, -1-\tilde{w}_k, \ldots, -\tilde{w}_D) \right) \\
&\qquad \times p_N\left(x, k, h, b, (\tilde{w}_j)_j \right),
\end{aligned}$$

(see Section 5.2). The probability $p_N\left(x, k, h, b, (\tilde{w}_j)_j\right)$ is given by

$$\begin{aligned}
p_N\left(x, k, h, b, (\tilde{w}_j)_j \right) &= \mathbb{P}\left(\left(\tilde{W}_j^N \right)_j = (\tilde{w}_j)_j \Big| \frac{X_n^N}{N} = x; v \in \mathcal{E}_n^N(k); H^N = h; B^N = b; \left(W_j^N\right)_j = (\tilde{w}_j)_j \right) \\
&\times \mathbb{P}\left(\left(W_j^N\right)_j = (\tilde{w}_j)_j \Big| \frac{X_n^N}{N} = x; v \in \mathcal{E}_n^N(k); H^N = h; B^N = b \right) \\
&\times \mathbb{P}\left(B^N = b \Big| \frac{X_n^N}{N} = x; v \in \mathcal{E}_n^N(k); H^N = h \right) \\
&\times \mathbb{P}\left(H^N = h \Big| \frac{X_n^N}{N} = x; v \in \mathcal{E}_n^N(k) \right) \times \mathbb{P}\left(v \in \mathcal{E}_n^N(k) \Big| \frac{X_n^N}{N} = x \right),
\end{aligned}$$

where $H^N \sim \mathbf{Hyper}\,(Nu, Nu-k, k)$, $B^N \sim \mathbf{Hyper}\left(Nu-k, Nu - \sum_j jNe_j, H^N\right)$, and

$$\lim_{N\to\infty} \mathbb{P}\left(\left(\tilde{W}_j^N \right)_j = (\tilde{w}_j)_j \Big| \frac{X_n^N}{N} = x; v \in \mathcal{E}_n^N(k); H^N = h; B^N = b; \left(W_j^N\right)_j = (\tilde{w}_j)_j \right) = 1$$

by Lemma 5.2.1. Then,

$$\begin{aligned}
e^{-Nf(x)} T^N\left(e^{Nf}\right)(x) &= e^{N\left(f\left(x+\frac{1}{N}(1,0,-1,0,\ldots,0)\right)-f(x)\right)} \frac{e_0}{\sum_j e_j} \\
&+ \sum_{k=1}^{D} \sum_{h=0}^{k} \sum_{b=0}^{h} \sum_{\tilde{w}_j : \sum \tilde{w}_j \leq h-b} e^{N\left(f\left(x+\frac{1}{N}(1,-k-h,0,-\tilde{w}_1,\cdots -1-\tilde{w}_k,\ldots,-\tilde{w}_d)\right)-f(x)\right)} \\
&\qquad \times p_N\left(x, k, h, b, (\tilde{w}_j)_j \right).
\end{aligned}$$

If $f \in C^2(E)$, then

$$\lim_{N\to\infty} e^{-Nf(x)}T^N\left(e^{Nf}\right)(x) = e^{\langle \nabla f(x),(1,0,-1,0,\ldots,0)\rangle}\frac{e_0}{\sum_j e_j}$$
$$+\sum_{k=1}^{D}\sum_{h=0}^{k}\sum_{b=0}^{h}\sum_{\tilde{w}_j:\sum \tilde{w}_j\le h-b} e^{\langle \nabla f(x),(1,-k-h,0,-\tilde{w}_1,\ldots,-1-\tilde{w}_k,\ldots,-\tilde{w}_D)\rangle}$$
$$\times \lim_{N\to\infty} p_N\left(x,k,h,b,(\tilde{w}_j)_j\right).$$

Using Stirling's formula, we obtain that

$$\lim_{N\to\infty} p_N\left(x,k,h,b,(\tilde{w}_j)_j\right) = \begin{cases} \frac{e_k}{\sum_j e_j}\frac{C_b^k\left(u-\sum_j je_j\right)^b\left(\sum_j je_j\right)^{k-b}}{u^k} \times \frac{(k-b)!}{\tilde{w}_1!\ldots\tilde{w}_D!}\prod_{j:e_j>0}\frac{\left(je_j\right)^{\tilde{w}_j}}{\left(\sum_j je_j\right)^{k-b}}, & \\ & \text{if } h=k \text{ and } \sum_j \tilde{w}_j = h-b, \\ 0, & \text{in other cases.}\end{cases} \tag{5.7}$$

If $\nabla f(x) = \alpha = (\alpha_s, \alpha_u, \alpha_0, \ldots, \alpha_D)$, then

$$\lim_{N\to\infty} e^{-Nf(x)}T^N\left(e^{Nf}\right)(x) = \sum_{k=0}^{D} e^{\alpha_s-2k\alpha_u-\alpha_k}\left(1+\sum_{j=1}^{D}\left(e^{-\alpha_j}-1\right)\frac{je_j}{u}\right)^k\frac{e_k}{\sum_j e_j},$$

and $\lim_{N\to\infty} H^N(f)(x) = H\left(x,\nabla f(x)\right)$ with $H(x,\alpha)$ defined in Equation (5.6).

If $x = \frac{1}{N}\left(\hat{s}_N, \hat{u}_N, 0, \ldots, 0\right)$, then $T^N(f)(x) = f(x)$ and $H^N(f)(x) = 0$. This result is extended to $f \in C^1(E)$ by taking a sequence $\{f_m\}_m \subset C^2(E)$ such that $\limsup_m \sup_{x\in E} |f_m(x) - f(x)| = 0$ and the triangular inequality. □

As we mentioned before, for each $x \in E$ the function $H(x,\alpha)$ obtained is convex w.r.t. α. We state this result in the following proposition.

Proposition 5.4.2 *The function $H : E \times R^{D+3} \to \mathbb{R}$ defined in Equation (5.6) is convex w.r.t. α.*

Proof. Let $x \in E$ be fixed. We want to prove that for all $\alpha, \beta \in \mathbb{R}^{D+3}$, and $\lambda \in [0,1]$ it is verified that

$$H\left(x,\ \lambda\alpha + (1-\lambda)\beta\right) \le \lambda H(x,\alpha) + (1-\lambda)H(x,\beta).$$

For each $\alpha \in \mathbb{R}^{D+3}$, define the linear function $f_\alpha : E \to \mathbb{R}$ given by $f_\alpha(x) = \langle \alpha, x\rangle$, then

$f_\alpha \in C^1(E)$ and

$$H(x,\alpha) = H(x, \nabla f_\alpha(x)) = \mathbf{H}(f_\alpha)(x) = \lim_{N\to\infty} H^N(f_\alpha)(x).$$

Then, it is enough to prove that

$$H^N\left(f_{\lambda\alpha+(1-\lambda)\beta}\right)(x) \leq \lambda H^N(f_\alpha)(x) + (1-\lambda)H^N(f_\beta)(x), \quad x \in E^N.$$

Note that for each α, $H^N(f_\alpha)(x) = \log \mathbb{E}\left[e^{\left\langle \alpha, X_{k+1}^N - X_k^N \right\rangle} \middle| \frac{X_k^N}{N} = x\right]$. Then,

$$\begin{aligned}
&\lambda H^N(f_\alpha)(x) + (1-\lambda)H^N(f_\beta)(x) \\
&\qquad = \log\left[\left(\mathbb{E}\left(e^{\left\langle \alpha, X_{k+1}^N - X_k^N \right\rangle} \middle| \frac{X_k^N}{N} = x\right)\right)^\lambda\right] + \log\left[\left(\mathbb{E}\left(e^{\left\langle \beta, X_{k+1}^N - X_k^N \right\rangle} \middle| \frac{X_k^N}{N} = x\right)\right)^{1-\lambda}\right] \\
&\qquad = \log\left[\left(\mathbb{E}\left(e^{\left\langle \alpha, X_{k+1}^N - X_k^N \right\rangle} \middle| \frac{X_k^N}{N} = x\right)\right)^\lambda \left(\mathbb{E}\left(e^{\left\langle \beta, X_{k+1}^N - X_k^N \right\rangle} \middle| \frac{X_k^N}{N} = x\right)\right)^{1-\lambda}\right] \\
&\qquad \geq \log \mathbb{E}\left[\left(e^{\left\langle \alpha, X_{k+1}^N - X_k^N \right\rangle}\right)^\lambda \left(e^{\left\langle \beta, X_{k+1}^N - X_k^N \right\rangle}\right)^{1-\lambda} \middle| \frac{X_k^N}{N} = x\right],
\end{aligned}$$

since log is an increasing function and Hölder's inequality allows us to prove that for each random variable X and $\lambda \in [0,1]$ it is verified that

$$\left(\mathbb{E}(f(X))\right)^\lambda \left(\mathbb{E}(g(X))\right)^{1-\lambda} \geq \mathbb{E}\left[f(X)^\lambda g(X)^{1-\lambda}\right], \quad \forall f, g.$$

Finally,

$$\begin{aligned}
\lambda H^N(f_\alpha)(x) + (1-\lambda)H^N(f_\beta)(x) &\geq \log \mathbb{E}\left[e^{\left\langle \lambda\alpha+(1-\lambda)\beta, X_{k+1}^N - X_k^N \right\rangle} \middle| \frac{X_k^N}{N} = x\right] \\
&= H^N\left(f_{\lambda\alpha+(1-\lambda)\beta}\right)(x).
\end{aligned}$$

□

Step 2: Verify the exponential compact containment condition

Since E is a compact subset of $\mathbb{R}^{D+3}$, the *exponential compact containment condition* from Definition 2.2.1 is trivially verified by taking $K_\alpha = E$.

Step 3: Comparison principle

In this subsection, we prove that for each $\beta > 0$ and $h \in C(E)$ the *comparison principle* (see Definition 2.3.2) is verified for the following equation:

$$f(x) - \beta H(x, \nabla f(x)) - h(x) = 0. \tag{5.8}$$

Proposition 5.4.3 *For each $\beta > 0$ and $h \in C(E)$ the comparison principle is satisfied for Equation (5.8).*

Proof. Let μ be a subsolution and v a supersolution of Equation (5.8). Let $\psi : E \times E \to \mathbb{R}^+$ be the good penalization function given by $\psi(x, y) = \frac{1}{2}\|x - y\|^2$, and consider the sequences x^α, y^α (with $\alpha \to +\infty$) defined by

$$x^\alpha = \left(s^{x^\alpha}, u^{x^\alpha}, e_0^{x^\alpha}, \ldots, e_D^{x^\alpha}\right) \text{ and } y^\alpha = \left(s^{y^\alpha}, u^{y^\alpha}, e_0^{y^\alpha}, \ldots, e_D^{y^\alpha}\right)$$

such that (see Section 2.3)

$$\mu(x^\alpha) - v(y^\alpha) - \alpha\psi(x^\alpha, y^\alpha) = \sup_{x,y\in E} \left\{\mu(x) - v(y) - \alpha\psi(x, y)\right\}.$$

By Proposition 2.3.1, the sequence (x^α, y^α) converges to (z, z) and $z = (z_s, z_u, z_0, \ldots, z_D)$ verifies $\mu(z) - v(z) = \sup_{x\in E}\{\mu(x) - v(x)\}$. As a consequence of Proposition 2.3.2, it is enough to prove that the following inequality holds:

$$\liminf_{\alpha\to\infty} H\left(x^\alpha, \alpha\psi_x\left(x^\alpha, y^\alpha\right)\right) - H\left(y^\alpha, \alpha\psi_x\left(x^\alpha, y^\alpha\right)\right) \leq 0,$$

where $\psi_x(x, y) = \left(\nabla\psi\left(., y\right)\right)(x)$ is the vector of partial derivatives of ψ w.r.t. $x = (s, u, e_0, \ldots, e_D)$. If $\sum_j z_j > 0$, then

$$\begin{aligned}
&\liminf_{\alpha\to\infty} H\left(x^\alpha, \alpha\left(x^\alpha - y^\alpha\right)\right) - H\left(y^\alpha, \alpha\left(x^\alpha - y^\alpha\right)\right) \\
&\quad\leq \liminf_{\alpha\to\infty} \log\left[\frac{\sum_k e^{-2k\alpha\left(u^{x^\alpha} - u^{y^\alpha}\right) - \alpha\left(e_k^{x^\alpha} - e_k^{y^\alpha}\right)}\left(1 + \sum_j\left(e^{-\alpha\left(e_j^{x^\alpha} - e_j^{y^\alpha}\right)} - 1\right)\frac{jz_j}{z_u}\right)^k \frac{z_k}{\sum_j z_j}}{\sum_k e^{-2k\alpha\left(u^{x^\alpha} - u^{y^\alpha}\right) - \alpha\left(e_k^{x^\alpha} - e_k^{y^\alpha}\right)}\left(1 + \sum_j\left(e^{-\alpha\left(e_j^{x^\alpha} - e_j^{y^\alpha}\right)} - 1\right)\frac{jz_j}{z_u}\right)^k \frac{z_k}{\sum_j z_j}}\right] \\
&\quad= 0.
\end{aligned}$$

For $\sum_j z_j = 0$, we repeat the previous analysis, being careful with the cases in which $\sum_j e_j^{x^\alpha} = 0$ or $\sum_j e_j^{y^\alpha} = 0$ after a certain α_0 (i.e. $H\left(x^\alpha, \alpha\psi_x\left(x^\alpha, y^\alpha\right)\right) = 0$ or $H\left(y^\alpha, \alpha\psi_x\left(x^\alpha, y^\alpha\right)\right) = 0$ for all $\alpha > \alpha_0$). $\square$

Step 4: Variational representation of the rate function

Finally, we prove that the rate function can be written as an action functional. As a consequence of the results presented in Subsection 2.2, it is enough to prove that Conditions **8.9**, **8.10**, and **8.11** from [Feng and Kurtz, 2006] (which are presented as Conditions 2.2.4, 2.2.5, and 2.2.6) are verified in this case. We present them as propositions.

In this case, as $\mathbf{H}(f)(x) = H\left(x, \nabla f(x)\right)$ for each $x \in E$ and $H \leftrightarrow L$, the functional $\mathbf{H}$ can be written as $\mathbf{H}(f)(x) = \sup_{u\in U}\left\{A(f)(x,u) - L(x,u)\right\}$, where $U = \mathbb{R}^{D+3}$ and $A : C^1(E) \to M\left(E \times U\right)$ is the linear operator given by $A(f)(x,u) = \langle \nabla f(x), u\rangle$. Since $H(x,\alpha)$ is convex w.r.t. α, it follows that $L(x,\beta)$ is convex w.r.t. β and a deterministic control $\mu\left(\mathrm{d}u \times \mathrm{d}s\right) = \delta_{u(s)}(\mathrm{d}u)\mathrm{d}s$ is allways the control with smallest cost by Jensen's inequality. Moreover, if $\mathbf{x} : E \to \mathbb{R}^{D+3}$ is an absolutely continuous function, then

$$f\left(\mathbf{x}(t)\right) - f\left(\mathbf{x}(0)\right) = \int_0^t \left\langle \nabla f\left(\mathbf{x}(s)\right), \dot{\mathbf{x}}(s)\right\rangle \mathrm{d}s = \iint_{\mathbb{R}^{D+3}\times[0,t]} A(f)\left(\mathbf{x}(s),\, u\right)\mu(\mathrm{d}u \times \mathrm{d}s),$$

if define $\mu = \mu(\mathbf{x})$ such that $\mu(\mathrm{d}u \times \mathrm{d}s) = \delta_{\dot{\mathbf{x}}(s)}(\mathrm{d}u)\mathrm{d}s$. Let $\Gamma = E \times U$. Then, the supremum in Equation (2.11) for the Nisio semigroup definition is reached on $\{(\mathbf{x},\mu) :\, \mathbf{x} \in \mathcal{AC},\; \mathbf{x}(0) = x_0\} \subset \mathcal{Y}^\Gamma$.

Proposition 5.4.4 *Conditions* **8.9** *from [Feng and Kurtz, 2006] (Condition 2.2.4) are verified.*

The proof of this proposition is identical to that of Proposition 3.4.4, so we omit it.

Proposition 5.4.5 *Condition* **8.10** *from [Feng and Kurtz, 2006] (Condition 2.2.5) is verified.*

Proof. Since $L(x,\beta) = 0 \Leftrightarrow \beta = H_\alpha(x,0)$, the function $q(x) = H_\alpha(x,0)$ solves the equation $L(x,q(x)) = 0$ for all $x \in E$. Note that the fluid limit $x(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$ verifies $\dot{x} = q(x)$ with the initial condition $x(0) = (0, \lambda, p_0, \ldots, p_D)$. If x is solution of $\dot{x} = q(x)$ with initial condition $x(0) = x_0$ and define μ by $\mu\left(\mathrm{d}u \times \mathrm{d}s\right) = \delta_{\{q(\mathbf{x}(s))\}}(\mathrm{d}u) \times \mathrm{d}s$, then $(\mathbf{x},\mu) \in \mathcal{Y}^\Gamma$ and verifies the required condition. $\square$

Proposition 5.4.6 *Condition* **8.11** *from [Feng and Kurtz, 2006] (Condition 2.2.6) is verified.*

Proof. Let $x_0 = (s_0, u_0, e_{0,0}, \ldots, e_{D,0}) \in E$ and $f \in C^1(E)$ be fixed with $\sum_j e_{j,0} > 0$. Since $\mathbf{H}(f)(x) = \sup_{\beta \in \mathbb{R}^{D+3}} \{\langle \nabla f(x), \beta \rangle - L(x, \beta)\}$, we look for a pair $(\mathbf{x}, \mu) \in \mathcal{Y}^\Gamma$ that verifies $\mathbf{x}(0) = x_0$ and

$$\int_0^t H\left(\mathbf{x}(s), \nabla f\left(\mathbf{x}(s)\right)\right) \mathrm{d}s = \iint_{U \times [0,t]} \left[\langle \nabla f\left(\mathbf{x}(s)\right), u \rangle - L\left(\mathbf{x}(s), u\right)\right] \mu\left(\mathrm{d}s \times \mathrm{d}u\right), \quad \forall t. \tag{5.9}$$

If define $q_f(x) = H_\alpha\left(x, \nabla f(x)\right)$, then $H\left(x, \nabla f(x)\right) = \langle \nabla f(x), q_f(x) \rangle - L\left(x, q_f(x)\right)$ and Equation (5.9) is verified for any path $\mathbf{x}$ if take $\mu\left(\mathrm{d}u \times \mathrm{d}s\right) = \delta_{\left\{q_f(\mathbf{x}(s))\right\}}(\mathrm{d}u)\, \mathrm{d}s$. Now we have to add conditions such that in addition $(\mathbf{x}, \mu)$ belongs to $\mathcal{Y}^\Gamma$ with $\mathbf{x}(0) = x_0$. In particular, $(\mathbf{x}, \mu)$ must verify:

$$\int_0^t \langle \nabla g\left(\mathbf{x}(s)\right), q_f\left(\mathbf{x}(s)\right) \rangle \mathrm{d}s = g\left(\mathbf{x}(t)\right) - g\left(\mathbf{x}(0)\right) \;\; \forall t \in [0,1], \;\forall g \in C^1(E).$$

Then, we look for a path that solves the following problem:

$$\begin{cases} \mathbf{x} \text{ is differentiable almost everywhere and } \dot{\mathbf{x}}(t) = q_f\left(\mathbf{x}(t)\right), \\ \mathbf{x}(0) = x_0, \\ \mathbf{x}(t) \in E \text{ for all } t \geq 0. \end{cases} \tag{5.10}$$

$\mathbf{x}(t) = (s(t), u(t), e_0(t), \ldots, e_D(t))$ verifies $\dot{x} = q_f(x)$ if and only if:

$$\begin{cases} \dot{s} = 1, \\ \dot{u} = H_{\alpha_u}\left(x, \nabla f(x)\right), \\ \dot{e}_i = H_{\alpha_i}\left(x, \nabla f(x)\right), \\ \mathbf{x}(0) = (s_0, u_0, e_{0,0}, \ldots, e_{D,0}) \text{ with } u_0 \geq \sum_j j e_{j,0}. \end{cases}$$

Note that for each i: $h_{f,i}(x) = H_{\alpha_i}\left(x, \nabla f(x)\right)$ is a continuous function and $h_{f,i}(x) \leq 0$ if $x \in E$, then $e_i(t)$ is decreasing and we can paste local solutions from Peano's Theorem (see [Crandall, 1972]) until the time t_i at which $e_i(t_i) = 0$, and put $e_i(t) = 0$ for all $t \geq t_i$.

If $\sum_j e_{j,0} = 0$, the only possible initial condition is $x_0 = (0, 0, 0, \ldots, 0)$ and the equality is verified by taking $\mu\left(\mathrm{d}u \times \mathrm{d}s\right) = \delta_0(u)\mathrm{d}s$. $\square$

Bibliography

[Arnold, 1987] Arnold, B. I. (1987). *Métodos Matemáticos da Mecanica Clásica.* Editora Mir Moscovo.

[Baccelli and Tien Viet, 2012] Baccelli, F. and Tien Viet, N. (2012). Generating Functionals of Random Packing Point Processes: From Hard-Core to Carrier Sensing. *Computing Research Repository (CoRR).* abs/1202.0225.

[Bermolen et al., 2021a] Bermolen, P., Goicoechea, V., and Jonckheere, M. (2021a). Large deviations for the greedy exploration process on configuration models.

[Bermolen et al., 2021b] Bermolen, P., Goicoechea, V., Jonckheere, M., and Mordecki, E. (2021b). Large deviation principle for the greedy exploration algorithm over erdös-rényi graphs.

[Bermolen et al., 2017a] Bermolen, P., Jonckheere, M., and J., S. (2017a). Scaling Limits and Generic Bounds for Exploration Processes. *Journal of Statistical Physics.*, 5:989–1018.

[Bermolen et al., 2017b] Bermolen, P., Jonckheere, M., and Moyal, P. (2017b). The jamming constant of uniform random graphs. *Stochastic Processes and their Applications.*, 7:2138–2178.

[Bollobás, 1980] Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316.

[Brightwell et al., 2017] Brightwell, G., Janson, S., and Luczak, M. (2017). The greedy independent set in a random graph with given degrees. *Random Structures Algorithms*, 51(4):565–586.

[Coste and Salez, 2018] Coste, S. and Salez, J. (2018). Emergence of extended states at zero in the spectrum of sparse random graphs.

[Cramér, 1938] Cramér, H. (1938). Sur un nouveau théoreme-limite de la theorie des probabilitiés. *Acta. Sci. et Ind.*, 736:5–23.

[Crandall, 1972] Crandall, M. G. (1972). A generalization of Peano's existence theorem and flow invariance. *Proc. Amer. Math. Soc.*, 36:151–155.

[Crandall et al., 1992] Crandall, M. G., Ishii, H., and Lions, P.-L. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67.

[Crandall and Liggett, 1971] Crandall, M. G. and Liggett, T. M. (1971). Generation of semi-groups of nonlinear transformations on general Banach spaces. *Amer. J. Math.*, 93:265–298.

[de Acosta, 1997] de Acosta, A. (1997). Exponential tightness and projective systems in large deviation theory. In *Festschrift for Lucien Le Cam*, pages 143–156. Springer, New York.

[Dembo, 1994] Dembo, A. (1994). Large deviations in information theory and statistics. In *Proceedings of 1994 Workshop on Information Theory and Statistics*, pages 7–.

[Dembo and Zeitouni, 1998] Dembo, A. and Zeitouni, O. (1998). *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition.

[Donsker and Varadhan, 1975] Donsker, M. D. and Varadhan, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I, II (and III). *Comm. Pure Appl. Math.*, 28:1–47; ibid. 28 (1975), 279–301.

[Dupuis and Ellis, 1997] Dupuis, P. and Ellis, R. S. (1997). *A weak convergence approach to the theory of large deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication.

[Dupuis et al., 1991] Dupuis, P., Ellis, R. S., and Weiss, A. (1991). Large deviations for Markov processes with discontinuous statistics. I. General upper bounds. *Ann. Probab.*, 19(3):1280–1297.

[Erdős, 1959] Erdős, P. (1959). Graph theory and probability. *Canadian J. Math.*, 11:34–38.

[Erdős and Rényi, 1959] Erdős, P. and Rényi, A. (1959). On random graphs. I. *Publ. Math. Debrecen*, 6:290–297.

[Esscher, 1932] Esscher, F. (1932). On the probability function in the collective theory of risk. *Scandinavian Actuarial Journal*, 15:175–195.

[Evans, 1993] Evans, J. (1993). Random and cooperative sequential adsorption. *Reviews of Modern Physics.*, 65(4):1281–1329. .

[Evans and Ishii, 1985] Evans, L. C. and Ishii, H. (1985). A PDE approach to some asymptotic problems concerning random differential equations with small noise intensities. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 2(1):1–20.

[Feng and Kurtz, 2006] Feng, J. and Kurtz, T. G. (2006). *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

[Ferrari et al., 2002] Ferrari, P. A., Fernández, R., and Garcia, N. L. (2002). Perfect simulation for interacting point processes, loss networks and Ising models. *Stochastic Process. Appl.*, 102(1):63–88.

[Fischer and Noever, 2020] Fischer, M. and Noever, A. (2020). Tight analysis of parallel randomized greedy MIS. *ACM Trans. Algorithms*, 16(1):Art. 6, 13.

[Fleming, 1999] Fleming, W. (1999). Stochastic analysis, control optimization and applications. A volume in honor of W.H. Fleming.

[Fleming, 1978] Fleming, W. H. (1978). Exit probabilities and optimal stochastic control. *Appl. Math. Optim.*, 4(4):329–346.

[Fleming, 1985] Fleming, W. H. (1985). A stochastic control approach to some large deviations problems. In *Recent mathematical methods in dynamic programming (Rome, 1984)*, volume 1119 of *Lecture Notes in Math.*, pages 52–66. Springer, Berlin.

[Flory, 1939] Flory, P. J. (1939). Intramolecular reaction between neighboring substituents of vinyl polymers. *Journal of the American Chemical Society*, 61(6):1518–1521.

[Freidlin and Wentzell, 1984] Freidlin, M. I. and Wentzell, A. D. (1984). *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York. Translated from the Russian by Joseph Szücs.

[Frieze and Karoński, 2016] Frieze, A. and Karoński, M. (2016). *Introduction to random graphs*. Cambridge University Press, Cambridge.

[Gamarnik and Sudan, 2017] Gamarnik, D. and Sudan, M. (2017). Limits of local algorithms over sparse random graphs. *Ann. Probab.*, 45(4):2353–2376.

[Hatami et al., 2014] Hatami, H., Lovász, L., and Szegedy, B. (2014). Limits of locally-globally convergent graph sequences. *Geom. Funct. Anal.*, 24(1):269–296.

[Jonckheere and Saenz, 2019] Jonckheere, M. and Saenz, M. (2019). Asymptotic optimality of degree-greedy discovering of independent sets in configuration model graphs.

[Jungnickel, 2005] Jungnickel, D. (2005). *Chapter 5, The Greedy Algorithm, In:Graph, Networks and Algorithm*, volume 5 of *Algorithm and Computation in Mathematics*. Springer-Verlag Berlin Heidelberg, second edition.

[Karp, 1972] Karp, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972)*, pages 85–103.

[Karp and Sipser, 1981] Karp, R. M. and Sipser, M. (1981). Maximum matching in sparse random graphs. In *22nd Annual Symposium on Foundations of Computer Science (sfcs 1981)*, pages 364–375.

[Kleinrock and Takagi, 1985] Kleinrock, L. and Takagi, H. (1985). Throughput analysis for persistent csma systems. *IEEE Transactions on Communications*, 33(7):627–638.

[Kraaij, 2016] Kraaij, R. (2016). Large deviations for finite state Markov jump processes with mean-field interaction via the comparison principle for an associated Hamilton-Jacobi equation. *J. Stat. Phys.*, 164(2):321–345.

[Krivelevich et al., 2020] Krivelevich, M., Mészáros, T., Michaeli, P., and Shikhelman, C. (2020). Greedy maximal independent sets via local limits. In *31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 159 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 20, 19. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern.

[Kurtz, 1971] Kurtz, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probability*, 8:344–356.

[Lauer and Wormald, 2007] Lauer, J. and Wormald, N. (2007). Large independent sets in regular graphs of large girth. *J. Combin. Theory Ser. B*, 97(6):999–1009.

[McDiarmid, 1990] McDiarmid, C. (1990). Colourings of random graphs. In *Graph colourings (Milton Keynes, 1988)*, volume 218 of *Pitman Res. Notes Math. Ser.*, pages 79–86. Longman Sci. Tech., Harlow.

[Moreno and Jennings, 1938] Moreno, J. L. and Jennings, H. H. (1938). Statistics of social configurations. *Sociometry*, 1(3/4):342–74.

[O'Brien and Vervaat, 1991] O'Brien, G. L. and Vervaat, W. (1991). Capacities, large deviations and loglog laws. In *Stable processes and related topics (Ithaca, NY, 1990)*, volume 25 of *Progr. Probab.*, pages 43–83. Birkhäuser Boston, Boston, MA.

[O'Brien and Vervaat, 1995] O'Brien, G. L. and Vervaat, W. (1995). Compactness in the theory of large deviations. *Stochastic Process. Appl.*, 57(1):1–10.

[Penrose, 2001] Penrose, M. D. (2001). Random parking, sequential adsorption, and the jamming limit. *Comm. Math. Phys.*, 218(1):153–176.

[Pham, 2007] Pham, H. (2007). Some applications and methods of large deviations in finance and insurance. In *Paris-Princeton Lectures on Mathematical Finance 2004*, volume 1919 of *Lecture Notes in Math.*, pages 191–244. Springer, Berlin.

[Pittel, 1982] Pittel, B. (1982). On the probable behaviour of some algorithms for finding the stability number of a graph. *Mathematical Proceedings of the Cambridge Philosophical Society*, 92(3):511–526.

[Puhalskii, 1991] Puhalskii, A. (1991). On functional principle of large deviations. In *New trends in probability and statistics, Vol. 1 (Bakuriani, 1990)*, pages 198–218. VSP, Utrecht.

[Puhalskii, 1994] Puhalskii, A. (1994). The method of stochastic exponentials for large deviations. *Stochastic Process. Appl.*, 54(1):45–70.

[Rahman and Virág, 2017] Rahman, M. and Virág, B. (2017). Local algorithms for independent sets are half-optimal. *Ann. Probab.*, 45(3):1543–1577.

[Rényi, 1958] Rényi, A. (1958). On a one-dimensional problem concerning random space filling. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 3(1-2):109–127.

[Ritchie, 2006] Ritchie, T. (2006). Construction of the thermodynamic jamming limit for the parking process and other exclusion schemes on Z^d. *Journal of Statistical Physics*, 122:381–398.

[Shwartz and Weiss, 1995] Shwartz, A. and Weiss, A. (1995). *Large deviations for performance analysis*. Stochastic Modeling Series. Chapman & Hall, London. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[Spitzer, 1975] Spitzer, F. (1975). Markov random fields on an infinite tree. *Ann. Probability*, 3(3):387–398.

[Touchette, 2009] Touchette, H. (2009). The large deviation approach to statistical mechanics. *Phys. Rep.*, 478(1-3):1–69.

[Varadhan, 1966] Varadhan, S. R. S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286.

[Weiss, 1995] Weiss, A. (1995). An introduction to large deviations for communication networks. *IEEE Journal on Selected Areas in Communications*, 13(6):938–952.

[Wormald, 1995] Wormald, N. C. (1995). Differential equations for random processes and random graphs. *Ann. Appl. Probab.*, 5(4):1217–1235.