UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO MATEMATICA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

Tesi di Laurea Magistrale in Matematica

An algorithmic approach to the random spanning forests

Relatrice:
Prof.ssa Alessandra Bianchi

Laureando:
Riccardo Michielan

16 ottobre 2020

Anno Accademico 2019-2020

Contents

Introduction

Given a graph it is possible to extract a subgraph without cycles, that is a tree, from it. Iterating this procedure we obtain a forest spanning the graph. How can we sample uniform or, more generally, weighted spanning forests? The answer follows from the works of D. Wilson [14] who provided in the 1990's an algorithm based on loop-erased random walks. This method turned out to be amazingly easy to perform, but at the same time extremely powerful, since it brings out the links between the study of random spanning forests and the theory of Markov chains. In this thesis, we collect the basic results of this approach to the random spanning forests and we focus on some interesting consequences in terms of network analysis.

Graphs are one of the most used mathematical structures, both for theoretical studies and applications. They are the basis of models of systems and processes studied in computer science, engineering, chemistry, molecular biology, social sciences, and many other applied fields. Indeed graphs are useful to highlight the inter connections between data and to get a very precise description of interacting systems. In our world, the amount of data available is growing enormously day by day, therefore in this context it is increasingly necessary to be able to process a huge amount of information with new, more efficient and more effective techniques. Hence the strong research interest in probabilistic approaches to Network theory, in particular in Markov process theory, that is a fundamental support to this application.
From a probabilistic point of view, the most immediate way to analyze a graph is to explore it by means of random walks with transition probabilities proportional to the weights assigned. The primary object of our study is Wilson's algorithm, which has precisely this idea behind it. Considering a finite connected weighted graph $\overline{\mathcal{G}}$, the algorithm can be summarized briefly as follows:

- Fix a vertex inside the graph, denoted by Δ. This vertex will be interpreted as an absorption point.

- Choose arbitrarily a starting vertex x_1 and perform a loop-erased random walk until hitting Δ. This is obtained by erasing the loops from the random walk as soon as they appear. Denote by γ_1 this self-avoiding path.

- For any $n \geq 2$, choose arbitrarily a starting vertex x_n and perform a new loop-erased random walk until we hit the set $\{\Delta, \gamma_1, ..., \gamma_{n-1}\}$. Denote by γ_n this self-avoiding path.

- Iterate this procedure until the entire graph is covered by the paths.

We call $\mathcal{W}$ the random variable associated with the random realization of this algorithm. Thus, $\mathcal{W}$ takes values in the set of all possible spanning trees rooted at Δ. The crucial point is that this procedure remains unbiased, no matter which method we use to arbitrarily choose the starting vertices of the random walks. Among other things, the latter property makes the Wilson's algorithm a prototype for bias-free maze generation.
Clearly, such a tool is of great interest in computational mathematics and computer science. However, it also plays a fundamental role in highlighting the features of spanning trees and framing them into the Markov theory. Quoting R. Lyons and Y. Peres [15], if "electrical networks and random walks are two faces of the same underlying object", studying spanning trees "we discover an appealing third face which appears at first to be completely unrelated". Therefore, in the thesis we aim to show the connections between these apparently different objects, approaching the study of spanning trees and of spanning forests by means of the Wilson's algorithm.
As the title suggests, our work mainly concerns with random spanning forests. A random spanning forest is a random variable Φ, taking values on the set of all possible rooted spanning forests, with law ν called "forest probability measure". Note that there exists a natural and intuitive way to switch from spanning trees to spanning forests: indeed it is sufficient to consider $\mathcal{G}$, which is the graph obtained removing from $\overline{\mathcal{G}}$ the point Δ; then any rooted spanning tree on $\overline{\mathcal{G}}$ is associated to a rooted spanning forest on $\mathcal{G}$. This simple remark gives space to new interesting characteristics of the problem, which are analyzable by applying the algorithmic approach.
On one side, it is possible to identify the set of roots of any given forest. Indeed, as we will see, one can prove that the set of roots of a random spanning forest Φ, denoted by $\rho(\Phi)$, is a determinantal process. That is, for any subset A of vertices of the graph,

$$\mathbb{P}(A \subset \rho(\Phi)) = \det_A(K)$$

where K is called kernel matrix. Moreover, introducing a random-walk-based notion of "well distributed points", L. Avena and A. Gaudillière showed in [10] that the set $\rho(\Phi)$ is well distributed inside the graph.
On the other side, one can look at the connected components of a spanning forest (i.e. the trees) forming a partition of the graph. As observed in [11] by L. Avena, F. Castell, A. Gaudillière, C. Mélot, if we denote by Π the partition given by the random spanning forest, then Π has the tendency to cluster together points that with high probability belong to the same trajectory of the random walk. This open the possibility to analyze

the correlation structure of the graph, through this spanning forest exploration. In particular one can look to the following two-points correlation: if x, y is any pair of vertices in the graph the we set

$$U(x, y) = \mathbb{P}(x \text{ and } y \text{ are in different blocks of } \Pi)$$

In the case of simple geometries, we are able to compute the value U, using an operative characterization and exploiting the symmetries.
There exists a flourishing literature concerning random spanning forests. As a matter of fact, interesting results have recently been achieved, for example on coarse graining schemes for networks, multi-resolution analysis on graphs, smoothing of graph signals [16, 17, 18]. This, once again, demonstrates the great importance of this topic within the research on network analysis.

The thesis is divided in four chapters. The first part is devoted to collect and outline the basic results regarding Markov processes. We make use, contextually, of both discrete-time and continuous-time Markov chains, therefore it is necessary to point out the relations between them. In particular it has to be clear how to switch from the discrete case to the continuous one, and vice versa, depending on the needs.
In the second chapter we focus on probabilistic techniques derived from potential theory. As we are interested in finite state spaces, first we need to provide a discretization of the Laplace equation. Then, introducing the notion of electrical networks, we understand how the study of harmonic functions is intimately linked to the Markov theory. We also show that electrical networks and random walks are indeed two sides of the same coin. Therefore we define the Green's function, an object of fundamental importance in the continuation of the rest of thesis.
Then, we go deeply inside the core of our work, and we start defining the forest probability measure ν on a weighted graph. Then, we introduce the notion of loop-erased trajectory and describe its law, using a famous and important result of P. Marchal [19]. In particular, we show that this result links our problem to the study of the Laplacian graph, via linear algebra techniques. At this point we are able to give a rigorous definition of Wilson's algorithm and to show that it is a sampling algorithm for the measure ν. Hence we review some results about the number of roots $|\rho(\Phi)|$ of a random spanning forest (that corresponds to the number of trees of the forest), and in particular we characterize its distribution and prove that the root process is determinantal.
Lastly we focus on a direct analysis of the loop-erased partitioning deriving from the Wilson's algorithm, by the study of the pairwise interaction potential U. In particular we compute the two-points correlation, provided by U, under two specific geometries. First we analyze the mean-field model, which corresponds to a complete graph with constant weights on edges, that represents a single-community model. In this geometry

the interaction potential is uniform, as it is invariant for any choice of pairs of points x, y in the graph. Also we prove that, after rescaling the parameters of the problem, the interaction potential U tends to the Gaussian distribution.
We then move to a more involved structure, which describes a multi-communities model. In this case, we consider the partition of the vertices of a complete graph into two disjoint sets, which correspond to two distinct communities. Then we set weight w_1 to edges connecting points inside the same community, and instead weight w_2 to edges connecting points in different communities. In this geometry, the interaction potential just depends on whether the points x, y are in the same community or not, thus there exists only two possible scenarios for U: if x and y are in the same community, we denote the interaction potential by $U(in)$; on the contrary, if x and y belong to different communities, we denote it by $U(out)$.
We conclude our work by showing the underlying similarities between the two models taken into consideration.

Chapter 1

Markov chains

Let E be a countable set. Consider a *discrete-time (or continuous-time) stochastic process* on the state space E, namely a sequence $\{X_k\}$, with $k \in \mathbb{N}$ (or $\mathbb{R}^+$) denoting the time, of random variables with values in E. The set E is assumed to be countable, but most of the time we will work in finite state spaces.
If $X_k = i$ the process is said to be in state i at time k. The generic time for the process is denoted by n in the discrete case or by t in the continuous case.
We denote with $\mathbb{P}$ the law of the process on the set of trajectories and with $\mathbb{E}$ the related expected value. If we assume that the *initial state* X_0 of the stochastic process has law μ, called *initial distribution*, then we denote by $\mathbb{P}_\mu$ the law of its trajectories so that $\mathbb{P}_\mu(X_0 = x) = \mu(x)$, and with $\mathbb{E}_\mu$ the expectation. In particular if $\mu = 1_x$, the indicator function, then with a slight abuse of notation we write $\mathbb{P}_x$ and $\mathbb{E}_x$ instead of $\mathbb{P}_{1_x}$ and $\mathbb{E}_{1_x}$.

In this section we closely follow the discussion on Markov chains as given in the book of P. Brémaud [1], in particular sections 2, 3, 4 and 9 where a more detailed explanation of many of the results can be found. However regarding continuous time Markov chain we refer to the notes of S. F. Nielsen [2].

1.1 Discrete-time Markov chains

Definition 1.1. Let $\{X_n\}$ be a discrete-time stochastic process on the state space E. If $\forall n \geq 1$ and $\forall x_0, x_1 ..., x_{n+1} \in E$

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, ... X_1 = x_1, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) \tag{1.1}$$

the process $\{X_n\}$ is called *Markov chain* (MC). The condition (1.1) is called *Markov property* and has the following interpretation: given the present, the future evolution of the process is indipendent from the past.
If the right-hand side of (1.1) is indipendent of the temporal index n, the Markov chain is called *homogeneous* (HMC). In this case the stochastic matrix $P = \{p_{yx}\}_{y,x \in E}$ where

$$p_{yx} = \mathbb{P}(X_{n+1} = x | X_n = y) \tag{1.2}$$

is called *transition matrix*.
For any $n \geq 0$ we call *distribution at time n* of the MC the row vector ν_n with coordi-

nates

$$\nu_n(x) = \mathbb{P}(X_n = x). \tag{1.3}$$

Assume $\{X_n\}$ is a HMC. Note that from Bayes' rule $\mathbb{P}(X_n = x_n, ..., X_0 = x_0) = \mathbb{P}(X_0 = x)\mathbb{P}(X_1 = x_1|X_0 = x_0)\cdots\mathbb{P}(X_n = x_n|X_{n-1} = x_{n-1}, ..., X_0 = x_0)$; if ν_0 is the law of X_0, in view of the Markov property (1.1) we obtain

$$\mathbb{P}_{\nu_0}(X_{n+1} = x_{n+1}, ..., X_0 = x_0) = \nu_0(x_0)p_{x_0x_1}\cdots p_{x_{n-1}x_n}. \tag{1.4}$$

By observing that $\nu_{n+1}(x) = \sum_{y\in E}\nu_n(y)p_{yx}$ we also get the recursive formula $\nu_{n+1} = \nu_n P$ from which

$$\nu_n = \nu_0 P^n. \tag{1.5}$$

Thus, the initial distribution and the transition matrix of a Markov chain are sufficient to determine the law of the process (1.4) and the distribution at time n (1.3).
In fact, by writing $P^n = \{p_{yx}(n)\}_{y,x\in E}$ we have

$$p_{yx}(n) = \mathbb{P}(X_n = x|X_0 = y) = \mathbb{P}_y(X_n = x). \tag{1.6}$$

Stationary measure

We give a few definitions in order to classify the states of a Markov chain and introduce the concept of stationarity.
Let $x, y \in E$. The state y is said to be *accessible* from state x if there exists $M \geq 0$ such that $p_{xy}(M) > 0$, i.e. there exists at least one finite path between x and y of positive measure. The two states x and y are said to *communicate* if x is accessible from y and y is accessible from x.
It is trivial to verify that the communication relation between states is an equivalence relation, so it produces a partition of E into disjoint *communication classes*.

Definition 1.2. A HMC is said *irreducible* if there exists only one communication class, that is all states communicate with each other.

Definition 1.3. Let P be the transition matrix associated to a HMC. A probability distribution π is called *stationary distribution* if it satisfies

$$\pi = \pi P. \tag{1.7}$$

Iterating (1.7) gives $\pi = \pi P^n$ for all $n \geq 0$ if π is a stationary distribution. Then from (1.5) we deduce that if the initial distribution of the Markov chain is π then $v_n = \pi$ for all $n \geq 0$, i.e. if the chain starts with some stationary distribution, it keeps it forever. Therefore the term *stationary* comes by the observation that

$$\mathbb{P}_\pi(X_n = x_0, ..., X_{n+k} = x_k) = \pi(x_0)p_{x_0x_1}\cdots p_{x_{k-1}x_k} \tag{1.8}$$

is indipendent from n.
It is important to observe that the stationary distribution may not be unique. In order to achieve uniqueness some more additional assumptions need to be satisfied, e.g. irreducibility and *recurrence*. The reader can find in [1] (Chapter 3) more details.

Definition 1.4. A HMC with an initial stationary distribution π (nonnull) is called *reversible* if for all $x, y \in E$

$$\pi(x)p_{xy} = \pi(y)p_{yx}. \tag{1.9}$$

The interpretation of the latter definition is the following. Consider a HMC with transition matrix P. Take π as stationary initial distribution so that $\mathbb{P}(X_n = x) = \pi(x)$ for all n. Then by Bayes formula

$$\mathbb{P}(X_n = x|X_{n+1} = y) = \frac{\mathbb{P}(X_{n+1} = y|X_n = x)\mathbb{P}(X_n = x)}{\mathbb{P}(X_{n+1} = y)} = \frac{\pi(x)p_{xy}}{\pi(y)} = p_{yx}. \tag{1.10}$$

That is P is also the transition matrix of the Markov chain when the time is reversed. The equation system in (1.9) is called *detailed balance equations*.

By definition a Markov chain is a stochastic process that satisfies the Markov property. The following stronger result, involving stopping times, holds for a HMC.

Theorem 1.1. *Let $(X_n)_{n\geq 0}$ be an HMC on the state space E with transition matrix P that defines the law $\mathbb{P}$ of the process. Let τ be a stopping time with respect to the chain. Then for any state $x \in E$, given that $X_\tau = x$, the following hold:*

(a) the process after τ is indipendent from the process before τ;

(b) the process after τ is a HMC with law $\mathbb{P}_x$, i.e. it has transition matrix P.

For the proof see [1] Chapter 2, Theorem 7.1.

This theorem is called *Strong Markov property*. Indeed it generalizes the Markov property to the case when we use some stopping time τ as temporal index of the process instead of a deterministic time n.

1.2 Continuous-time Markov chains

In this section we will provide all the necessary instruments to define in all generality a continuous-time Markov chain. In the next chapters we will use only a very special class of MC (homogeneous MC with regular jumps) but nevertheless very large and important for applications.
In order to get a precise definition first we need to introduce a class of continuous-time

processes called Poisson processes.

Recall that a continuous-time stochastic process on the state space E is a family $\{X_t\}$ with $t \in \mathbb{R}^+$ of random variables with values in E. With E countable the distribution of any stochastic process is determined by the probabilities

$$P(X_{t_0} = i_0, X_{t_1} = i_1, ..., X_{t_n} = i_n) \tag{1.11}$$

for any $n \in \mathbb{N}$, for $0 \leq t_0 < t_1 < ... < t_n$ and $i_0, i_1, ..., i_n \in E$.

Poisson processes

Definition 1.5. Let $\lambda > 0$ and $S_1, S_2, ...$ be a sequence of i.i.d. random variables with exponential distribution of rate $\lambda > 0$, i.e. $S_n \sim Exp(\lambda)$. Consider the random variables $T_n := \sum_{k=1}^{n} S_k$, $n \in \mathbb{N}$. Then the process $\{N(t)\}$ given by

$$N(t) = \sum_{n=1}^{\infty} 1_{\{T_n \leq t\}} \tag{1.12}$$

is called *homogeneous Poisson process* with intensity λ (HPP(λ)).

In the definition $\{T_n\}$ has to be interpreted as a sequence of increasing times at which the process *rings*. The random variables $\{S_n\}$ are also called *inter-event* times.

Proposition 1.2. *Let $\{N(t)\}$ be HPP(λ). Then for any $t > 0$, $N(t)$ is a random variable with Poisson distribution of parameter λt, i.e. $N(t) \sim Poisson(\lambda t)$.*

Proof. Note that T_n is the sum of n indipendent exponential random variables, thus it is a Gamma distributed random variable of parameters n, λ, i.e. $T_n \sim \Gamma(n, \lambda)$. Its probability distribution function is

$$p_{T_n}(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} 1_{\{x \geq 0\}}.$$

We aim to compute directly the cumulative distribution function for $N(t)$.
Now clearly $n = 0$ trivially implies by definition $P(N(t) < 0) = 0$ which gives us no information, hence suppose $n \geq 1$. Note that

$$F(n) = P(N(t) < n) = P(T_n > t) = \int_t^{\infty} \frac{\lambda^n}{\Gamma(n)} s^{n-1} e^{-\lambda s} \, ds \tag{1.13}$$

In the case $n = 1$ (1.9) gives:

$$F(1) = \int_t^{\infty} \lambda e^{-\lambda s} \, ds = \Big[-e^{-\lambda s} \Big]_t^{\infty} = e^{-\lambda t}. \tag{1.14}$$

Whereas in the case $n > 1$ integrating by parts we get the recursive formula

$$F(n) = \Big[-\frac{\lambda^{n-1}}{(n-1)!} s^{n-1} e^{-\lambda s} \Big]_t^{\infty} + \int_t^{\infty} \frac{\lambda^{n-1}}{\Gamma(n-1)} s^{n-2} e^{-\lambda s}\, ds$$
$$= \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} + F(n-1). \quad (1.15)$$

Finally combining (1.10) and (1.11) we have

$$P(N(t) < n) = F(n) = \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad (1.16)$$

that corresponds to the cumulative distribution function of a $Poisson(\lambda t)$. □

Consider now two (or more) indipendent HPPs. We could imagine that the different HPPs are competing with each other and we would like to know which one will prevail over the others. We can manage to get an answer to this problem by *merging* the processes together and retrieving some properties about it.

Let $\{N_1(t)\}$ and $\{N_2(t)\}$ be the two HPP with intensities λ_1 and λ_2. Consider the sequences $\{T_n^1\}$, $\{T_n^2\}$ obtained from definition 1.3 as the sum of i.i.d. exponential random variables with rate λ_1, λ_2.
We are then able to construct the *superposition* $\{T_n\}$ of the two sequences of times joining and sorting them in increasing order:

1. starting from $t = T_0(= 0)$ we choose $T_1 := \min\{T_1^1, T_1^2\}$;
2. consider for example $\min\{T_1^1, T_1^2\} = T_1^1$ (the other case is analogous). Then define a new sequence by $\hat{T}_n^1 := T_{n+1}^1$, i.e. the new sequence is obtained by *removing* from the original sequence the first event time reached;
3. starting from the time $t = T_1$ we choose $T_2 := \min\{\hat{T}_1^1, T_1^2\}$;
4. we iterate this process.

Simply by defining $N_1(t) + N_2(t) = N(t) := \sum_{n^1}^{\infty} 1_{\{T_n \le t\}}$ we obtain a new Poisson process with intensity $\lambda_1 + \lambda_2$, due to the fact that the minimum between two exponential random variables with rate λ_1, λ_2 is an exponential with rate $\lambda_1 + \lambda_2$.

This result can be generalized to the countable case, i.e.: if $\{N_i(t)\}_{i \in I}$ with I at most countable are Poisson process with intensities λ_i and if $\sum_{i \in I} \lambda = \lambda < \infty$ then $N(t) := \sum_{i \in I} N_i(t)$ is HPP(λ).

Theorem 1.3. ***Competition Theorem.*** *Let $\{N_i(t)\}_{i\in I}$ be indipendent Poisson processes, with I finite or countable and $\sum_{i\in I}\lambda_i = \lambda < \infty$. Denote by Z the first event time of $N = \sum_{i\in I} N_i$ and by K the index of the HPP responsible for such event (in particular T_1 is also the first event time for N_K). Then*

$$P(K = i, T_1 \geq a) = P(K = i)P(Z \geq a) = \frac{\lambda_i}{\lambda}e^{-\lambda a}. \tag{1.17}$$

Proof. a) First we prove the theorem for $|I| = n$ finite. By definition of N we have that $Z = \inf\{T_1^1, ..., T_1^n\}$. Thus in particular

$$P(Z \geq a) = P(T_1^1 \geq a, ..., T_1^n \geq a) = \prod_{j=1}^{n} P(T_1^j \geq a) = \prod_{j=1}^{n} e^{-\lambda_j a} = e^{-\lambda a} \tag{1.18}$$

because the first event times have exponential distribution with rate λ_j. Now fix $i \in \{1, ..., n\}$ and define $U = \inf\{T_1^1, ..., \hat{T_1^i}, ..., T_1^n\}$ (where $\hat{T_1^i}$ means that we are removing the element with index i from that sequence). Then we have

$$P(K = i, Z \geq a) = P(a \leq T_1^i < U) \tag{1.19}$$

that is the probability that the first event time is the one on position i and it occurs after a. By some computation we have

$$\begin{aligned} P(a \leq T_1^i < U) = \int_a^\infty P(U > x)\lambda_i e^{-\lambda_i x}\, dx &= \int_a^\infty e^{-(\lambda_1+...+\hat{\lambda_i}+...+\lambda_n)x}\lambda_i e^{-\lambda_i x}\, dx \\ &= \int_a^\infty \lambda_i e^{-\lambda x}\, dx = \frac{\lambda_i}{\lambda}e^{-\lambda a}. \end{aligned} \tag{1.20}$$

Also by letting $a = 0$ we have $P(K = i) = P(K = i, Z \geq a) = \frac{\lambda_i}{\lambda}$. From which we get the indipendence between K and Z.

b) Now consider the case $|I| = \infty$. Denote by $\{K_n = i, Z_n \geq a\}$ the events with $Z_n = \inf\{T_1^1, ..., T_1^n\}$ and K_n the index between 1 and n of the first event occured. This events are decreasing as $n \to \infty$ because the number of competing processes is increasing. Thus we have

$$P(K = i, Z \geq a) = \lim_{n\to\infty} P(K_n = i, Z_n \geq a) = \lim_{n\to\infty} \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} e^{-\sum_{j=1}^n \lambda_j a} = \frac{\lambda_i}{\lambda}e^{-\lambda a}.$$

□

Continuous-time homogeneous Markov Chains

Definition 1.6. Let $\{X_t\}$ be a continuous-time stochastic process on the state space E. If for all $i, j, i_1, ..., i_k \in E$, all $t, s \geq 0$ and all $0 \leq s_1 < s_2 < ... < s_k < s$ it holds

$$P(X_{t+s} = j | X_s = i, X_{s_k} = i_k, ..., X_{s_1} = i_1) = P(X_{t+s} = j | X_s = i) \tag{1.21}$$

the process $\{X_t\}$ is a *Markov chain*. If in addition the right hand side is indipendent of s the Markov chain is called *homogeneous*. In this case, for fixed $t \geq 0$, we define

$$P(t) := \{p_{ij}(t)\}_{i,j \in E} \tag{1.22}$$

where

$$p_{ij}(t) := P(X_{t+s} = j | X_s = i) = P_i(X_t = j). \tag{1.23}$$

$\{P(t)\}_{t \geq 0}$ is called *transition semigroup* of the continuous-time HMC.

Proposition 1.4. *$\{P(t)\}_{t \geq 0}$ is a semigroup.*

Proof. For $t = 0$ it holds $p_{ij}(0) = P_i(X_0 = j) = 1_{\{j=i\}}$. Thus $P(0) = I$.
For $t, s \geq 0$ it holds

$$\begin{aligned} p_{ij}(t+s) = P_i(X(t+s) = j) &= \sum_{k \in E} P_i(X(t+s) = j, X(t) = k) \\ &= \sum_{k \in E} P_i(X(t+s) = j | X(t) = k) P_i(X(t) = k) = \sum_{k \in E} p_{kj}(t) p_{ik}(s) \end{aligned}$$

from which we deduce $P(t+s) = P(t)P(s) = P(s)P(t)$. □

The distribution at time t is the row vector μ^t with $\mu^t(i) = P(X(t) = i)$. Set μ the distribution of X_0, then using the Markov property of the process at time 0 it holds

$$\mu^t = \mu P(t). \tag{1.24}$$

Also the probability distribution of a continuous-time HMC is completely determined by its initial distribution and its transition semigroup because for all $0 \leq t_1 \leq ... \leq t_k$ and for all $i_1, ..., i_k \in E$ it holds

$$P(X(t_1) = i_1, ..., X(t_k) = i_k) = \sum_{i_0 \in E} \mu(i_0) \prod_{j=1}^{k} p_{i_{j-1}, i_j}(t_j - t_{j-1}). \tag{1.25}$$

Definition 1.7. A stochastic process $X = \{X(t)\}_{t \geq 0}$ is called *regular jump* process if

- for all $t \geq 0$ there exists $\varepsilon > 0$ such that

$$X(t+s) = X(t) \qquad \text{a.s. for all } s \in [t, t+\varepsilon); \tag{1.26}$$

- denoting by A the set of discontinuities of the function $t \mapsto X(t)$, then for all $c \geq 0$

$$|A \cap [0, c]| < \infty. \tag{1.27}$$

In particular condition (1.26) means that, at any time t the process remains stable for at least a short time (so we can notice the *jumps*). So it is possible to define the sequence of times $\{\tau_n\}_{n\geq 0}$, with

$$0 = \tau_0 < \tau_1 < ... < \tau_n < ... \quad ,$$

and a sequence $\{\hat{X}_n\}_{n\geq 0}$ in E such that

$$X(t) = \hat{X}_n \qquad \text{if } \tau_n \leq t < \tau_{n+1}.$$

The random variables $\{\tau_n\}_{n\geq 0}$ are called *transition times*. Besides, condition (1.27) means that the set of discontinuity points for X is σ-discrete, and it implies that

$$\tau_\infty = \lim_{n\to\infty} \tau_n = \infty.$$

The request for a HMC to be a regular jump process seems reasonable, but there exist examples of HMC that are not: in such case the states are unstable! Anyway we will not work with them.

Finally the *Strong Markov property* holds also for continuous regular jump HMC.

Theorem 1.5. *Let $X = \{X(t)\}_{t\geq 0}$ be a regular jump HMC on E with transition semigroup $\{P(t)\}_{t\geq 0}$. Let τ be a stopping time with respect to X and k an arbitrary state of E. Then, given $X(\tau) = k$,*

(a) the chain after τ and the chain before τ are indipendent;

(b) the chain after τ is a regular jump HMC with transition semigroup $\{P(t)\}_{t\geq 0}$.

Proof. See page 346 of [1]. □

Uniform Markov chains

Starting from a discrete-time HMC it is possible to construct a continuous-time HMC that moves exactly in the same way as the discrete one does.

Definition 1.8. Let $\hat{X} = \{\hat{X}_n\}_{n\geq 0}$ be a discrete-time HMC on the state space E and denote with $\hat{P}$ its transition matrix. Let N be an HPP on $\mathbb{R}^+$ with intensity $\lambda > 0$. Assume N and $\hat{X}$ are indipendent. The process $X = \{X_t\}$ defined by

$$X_t := \hat{X}_{N(t)} \tag{1.28}$$

is called *uniform Markov chain*. The discrete-time HMC $\hat{X}$ is called *subordinated chain*.

In other words, each time an event T_i of the Poisson process occurs the discrete Markov chain moves one step; at time T_n we have that $X_{T_n} = \hat{X}_n$, i.e. the Markov chain has moved n steps forward.
Observe that for some step the Markov chain can remain in the same position, since it may be $\hat{p}_{ii} > 0$ for some $i \in E$, that is $\mathbb{P}(\hat{X}_{n+1} = \hat{X}_n) > 0$.

A uniform Markov chain $X = \{X_t\}$ is in fact a continuous-time HMC due to the *memoryless* property of the exponential distribution that characterizes the inter-event times of the Poisson process. The transition semigroup of X is given by

$$P(t) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \hat{P}^n \tag{1.29}$$

because

$$\begin{aligned} p_{ij}(t) = P_i(X_t = j) = P_i(\hat{X}_{N(t)} = j) &= \sum_{n=0}^{\infty} P_i(N(t) = n, \hat{X}_n = j) \\ &= \sum_{n=0}^{\infty} P(N(t) = n) P_i(\hat{X}_n = j) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} (\hat{P}^n)_{ij} \end{aligned}$$

Given a discrete HMC $\hat{X}$ with transition matrix $\hat{P} = (\hat{p}_{ij})_{i,j \in E}$ there is another way to define the uniform Markov chain: we associate to each jump $i \to j$ in E an $\mathrm{HPP}(\lambda \hat{p}_{ij})$ that rules the time of the jump, provided the chain is in i.
Observe that starting from i there could be many different competing Poisson processes. Nevertheless, for any fixed $i \in E$, we have

$$\sum_{j \in E} \lambda \hat{p}_{ij} = \lambda,$$

and from the Competition Theorem 1.3 it holds that:

1. at any point $i \in E$ there is an $\mathrm{HPP}(\lambda)$ that rules the time at which the next jump is performed;
2. the probability that the next jump will be $i \to j$ is $\hat{p}_{ij}$.

In conclusion the two different constructions are equivalent.

Infinitesimal generator

Let $\{P(t)\}_{t \geq 0}$ be a transition semigroup, i.e. for any t, s the matrix $P(t)$ is stochastic, $P(0) = I$ and $P(t+s) = P(t)P(s)$.
Suppose moreover the continuity of the transition semigroup, that is

$$\lim_{h \to 0} p_{ij}(t+h) = p_{ij}(t) \tag{1.30}$$

for any $t \geq 0$, $i, j \in E$. In fact the continuity at the origin ($\lim_{h\to 0} P(h) = I$) suffices to guarantee the continuity for the whole semigroup.

Theorem 1.6. *Given a continuous transition semigroup $\{P(t)\}_{t\geq 0}$ on a countable state space E, for any $i, j \in E$, there exist*

$$\alpha(i) := \lim_{h\to 0} \frac{1 - p_{ii}(h)}{h} \in [0, \infty], \tag{1.31}$$

$$\alpha(i, j) := \lim_{h\to 0} \frac{p_{ij}(h)}{h} \in [0, \infty). \tag{1.32}$$

Definition 1.9. For each state $i \in E$ define $\alpha(i, i) = -\alpha(i)$. The quantities $\alpha(i, j)$ for $i, j \in E$ are the *local characteristics* of the semigroup. Moreover the matrix

$$L = \{\alpha(i, j)\}_{i,j\in E} \tag{1.33}$$

is the *infinitesimal generator* of the semigroup and from the previous theorem it holds

$$L = \lim_{h\to 0} \frac{P(h) - P(0)}{h} \tag{1.34}$$

Consider for example the Uniform HMC $X_t = \hat{X}_{N(t)}$ where N is an HPP(λ) and $\hat{X}$ is a discrete-time Markov chain with transition matrix $\hat{P}$. Fix $i \in E$. Then the transition semigroup is given in (1.29) and the infinitesimal generator is

$$L = \lim_{h\to 0} \frac{P(h) - I}{h} = \lim_{h\to 0} \frac{\sum_{n=0}^{\infty} e^{-\lambda h} \frac{(\lambda h)^n}{n!} \hat{P}^n - I}{h}. \tag{1.35}$$

We can interchange the limit and the sum and what we get is

$$\alpha(i, j) = \lim_{h\to 0} \frac{(1 - \lambda h)(\lambda h)\hat{p_{ij}} + o(h)}{h} = \lambda \hat{p_{ij}} \quad \text{if } i \neq j \tag{1.36}$$

$$\alpha(i) = \lim_{h\to 0} \frac{1 - (1 - \lambda h)(1 + \lambda h \hat{p_{ij}}) + o(h)}{h} = \lambda(1 - \hat{p_{ii}}) \tag{1.37}$$

that in compact notation is

$$L = \lambda(\hat{P} - I). \tag{1.38}$$

Definition 1.10. Let $\hat{X}$ be a discrete Markov chain with transition matrix P and X the uniform Markov chain associated to it, with rate $\lambda = 1$. $L = P - I$ is the infinitesimal generator of X. Then we define with L also the infinitesimal generator of the discrete time process $\hat{X}$.

From (1.35), the infinitesimal generator can be interpreted as the derivative in 0 of the function $t \mapsto P(t)$. Also, in view of the semigroup properties, for any $t \geq 0$, $h \geq 0$

$$\frac{P(t+h) - P(t)}{h} = P(t)\frac{P(h) - I}{h} = \frac{P(h) - I}{h}P(t). \tag{1.39}$$

If the passage to the limit is allowed (e.g. when the state space E is finite) we obtain the differential equations

$$\frac{d}{dt}P(t) = P(t)L, \qquad \frac{d}{dt}P(t) = LP(t). \tag{1.40}$$

Thus the dynamic of $P(t)$ is obtained by solving the so-called *Kolmogorov's backward* or *forward* differential systems in (1.40) with the initial condition $P(0) = I$.
If E is countable infinite, the passage to the limit in (1.34) is allowed under suitable conditions (see [1]).

Moreover, if E is finite, the unique solution to the Kolmogorov's backward/forward equation with initial condition $P(0) = I$ is

$$P(t) = e^{tL}. \tag{1.41}$$

Definition 1.11. A *stationary distribution* of the HMC X with transition semigroup $\{P(t)\}_{t\geq 0}$ is a probability measure π on E such that for any $t \geq 0$

$$\pi = \pi P(t) \tag{1.42}$$

As in the discrete case, from (1.24) we have that if $X(0)$ has law π, then also $X(t)$ is distributed according to π. Hence the chain is said to be in a *stationary regime* (or *equilibrium*), because for any $0 \leq t_1 \leq ... \leq t_k$ and any states $i_1, ..., i_k$

$$P_\pi(X(t_1 + s) = i_1, ..., X(t_k + s) = i_k) = P_\pi(X(t_1) = i_1, ..., X(t_k) = i_k) \tag{1.43}$$

for any $s \geq 0$. In fact the quantity in (1.43) is equal to $\pi(i_1)p_{i_1 i_2}(t_2 - t_1) \cdots p_{i_{k-1} i_k}(t_k - t_{k-1})$ and is indipendent from s.
The stationary distribution may not exists and need not to be unique. Anyway for continuous time HMC on a finite state space E, where (1.34) is satisfied, there is an easy characterization of the stationary distribution, that is

$$\pi L = 0 \tag{1.44}$$

where L is the infinitesimal generator of the HMC. In expanded form

$$\pi(i)\alpha(i) = \sum_{j \neq i} \pi(j)\alpha(j, i). \tag{1.45}$$

Embedded process

We have seen how to define the (continuous) uniform Markov chain from a discrete time HMC. For regular jump HMCs there exists a canonical way to reverse the procedure, in order to get a discretization of the dynamics at continuous time.

Definition 1.12. Let $0 = \tau_0 < \tau_1 < ... < \tau_n < ...$ be the increasing sequence of transition times (which are stopping times) for a regular jump process $X = \{X(t)\}_{t\geq 0}$. By convention we say that $X(\infty) = \Delta$ where Δ is an arbitrary element that is not contained in E. Consider the extended state space $E_\Delta = E \cup \{\Delta\}$.
The process $\tilde{X} = \{\tilde{X}_n\}_{n\geq 0}$ with values in E_Δ defined by

$$\tilde{X}_n := X(\tau_n) \tag{1.46}$$

is called *embedded process* of the HMC.

Notice that if X is a regular jump HMC then from the Strong Markov property it follows that, given $X(\tau_n) = k \in E$ then $\{X(\tau_n + t)\}_{t\geq 0}$ and $\{X(\tau_n \wedge t)\}_{t\geq 0}$ are indipendent. Therefore, given $\tilde{X}_n = k$, we have that $\{\tilde{X}_{n+1}, \tilde{X}_{n+2}, ...\}$ are independent from $\{\tilde{X}_1, ..., \tilde{X}_{n-1}\}$, that is $\tilde{X}$ is a discrete Markov chain and, again by the Strong Markov property, it is homogeneous.

Moreover the following result, that is called *Regenerative Structure* of the embedded process, holds.

Theorem 1.7. *Let $X = \{X(t)\}_{t\geq 0}$ be a regular jump HMC with infinitesimal generator $L = \{\alpha(i,j)\}_{i,j\in E}$, transition times $\{\tau_n\}_{n\geq 0}$ and embedded process $\tilde{X} = \{\tilde{X}_n\}_{n\geq 0}$.*

(a) $\tilde{X}$ is a discrete-time HMC with state space E_Δ and transition matrix $\tilde{P} = \{\tilde{p}_{ij}\}_{i,j\in E_\Delta}$ given by

$$\tilde{P} = \left(\begin{array}{c|c} \tilde{P}|_E & * \\ \hline 0 \ \cdots \ 0 & 1 \end{array}\right) \tag{1.47}$$

where $\tilde{p}_{ij} = \frac{\alpha(i,j)}{\alpha(i)}$ if $i, j \in E$ and $\alpha(i) > 0$, $\tilde{p}_{ij} = 0$ if $i, j \in E$ and $\alpha(i) = 0$, whereas in $*$ there are the elements $\tilde{p}_{i\Delta} = 0$ if $\alpha(i) > 0$, $\tilde{p}_{i\Delta} = 1$ if $\alpha(i) = 0$.

(b) Given $\tilde{X}$ the sequence random variables $\{\tau_{n+1} - \tau_n\}_{n\geq 0}$ are indipendent and for any $s > 0$

$$P(\tau_{n+1} - \tau_n > s \mid \tilde{X}) = e^{-s\cdot\alpha(\tilde{X}_n)}. \tag{1.48}$$

The proof for the regenerative structure for the embedded process of a HMC can be found in [1] (Chapter 8, Theorem 4.2).

In conclusion from a regular jump HMC it is possible to retrieve the skeleton structure of the Markov chain, that is the discrete time HMC given by the embedded process. The embedded process of a continuous HMC only partially reverse the building procedure of the uniform Markov chain from a discrete HMC. In fact: take any irreducible discrete HMC $\hat{X}$ with transition matrix $\hat{P}$ on the state space E and build from it the uniform Markov chain X (which has regular jumps) with parameter λ. Then the embedded process of X (restricted to E) may have different law respect to $\hat{X}$, and they have the same law if and only if $\hat{p}_{ii} = 0$ for all $i \in E$.

An immediate corollary of Theorem 1.7 is the following.

Theorem 1.8. *Two regular jump HMCs with the same infinitesimal generator and the same initial distribution are probabilistically equivalent, that is the infinitesimal generator characterizes the semigroup.*

1.3 Linear operators on Markov chains

Let $\mathcal{X}$ be a finite or countable space. Consider the set

$$D = \{f : \mathcal{X} \to \mathbb{R}\} \tag{1.49}$$

which is a vector space over $\mathbb{R}$.

Consider now a linear operator $K : D \to D$. We can think to K as an operator acting on the left or on the right, in the following way: if $f, \phi \in D$

$$(Kf)(x) := \sum_{y \in \mathcal{X}} K(x, y) f(y) \tag{1.50}$$

$$(\phi K)(x) := \sum_{y \in \mathcal{X}} \phi(y) K(y, x) \tag{1.51}$$

That is, K is a $|\mathcal{X}| \times |\mathcal{X}|$ real matrix with components $K_{xy} = K(x, y)$ and the equation (1.37)-(1.38) are the classical matricial product using f as column vector and ϕ as row vector. The matrix

$$K = (K(x, y))_{x,y \in \mathcal{X}} \tag{1.52}$$

is called *kernel* of the operator.
Given a linear operator we can always describe its kernel matrix; but also, viceversa, given a matrix we can always think it to be the kernel of a linear operator (on right or

on left).
Thus, if the context is clear, without any specification we'll call K both the linear operator and the matrix (1.39) of the linear operator.

Actually to associate K to a matrix we should have introduced before an appropriate basis for D: in the finite case this can be done in a very obvious way, but in the infinite (countable) case it's slightly more difficult and we would need some additional assumption on D. Thus for the moment we assume $\mathcal{X}$ is finite.

Let X be a discrete Markov chain on the state space $\mathcal{X}$ with transition matrix $P = (p_{xy})_{x,y\in\mathcal{X}}$. Then P can be seen as the linear operator $P : D \to D$

$$(Pf)(x) = \sum_{y\in\mathcal{X}} p_{xy} f(y) = E_x[f(X_1)] \tag{1.53}$$

thus Pf is the mean value of f after one step. Similarly $P^n = (p_{xy}(n))_{x,y\in\mathcal{X}}$ is also the linear operator $P^n : D \to D$

$$(P^n f)(x) = \sum_{y\in\mathcal{X}} p_{xy}(n) f(y) = E_x[f(X_n)] \tag{1.54}$$

that is $P^n f$ is the mean value of f after n steps.

Now let X be a continuous Markov chain on $\mathcal{X}$ with transition semigroup $(P(t))_{t\geq 0}$ where, for any $t \geq 0$, $P(t) = (P^t(x,y))_{x,y\in\mathcal{X}}$ and infinitesimal generator $L = (\alpha(x,y))_{x,y\in\mathcal{X}}$. Then $P^t : D \to D$

$$(P^t f)(x) = \sum_{y\in\mathcal{X}} P^t(x,y) f(y) = E_x[f(X_t)] \tag{1.55}$$

and $P^t f$ is the mean value of f after time t. At last, recalling that by definition of infinitesimal generator we have $\alpha(x,x) = -\alpha(x) = -\sum_{y\neq x} \alpha(x,y)$, then $L : D \to D$ with

$$(Lf)(x) = \sum_{y\in\mathcal{X}} \alpha(x,y) f(y) = \sum_{y\neq x} \alpha(x,y)[f(y) - f(x)]. \tag{1.56}$$

In the examples above all the matrices are associated to left operators. But we can also think to those matrices as right operators. In particular if we consider

$$\mathcal{M} : \{\mu : \mathcal{X} \to [0,1] \mid \mu \text{ is a probability measure}\} \tag{1.57}$$

we can think to any stochastic matrix as an operator $\mathcal{M} \to \mathcal{M}$. In the discrete-time case $P^n : \mathcal{M} \to \mathcal{M}$ with

$$(\nu P^n)(x) = \sum_{y\in\mathcal{X}} \nu(y) p_{yx}(n) = P_\nu(X_n = x) \tag{1.58}$$

and in the continuous-time case $P^t : \mathcal{M} \to \mathcal{M}$ with

$$(\mu P^t)(x) = \sum_{y \in \mathcal{X}} \nu(y) p_{yx}(t) = P_\mu(X_t = x) \tag{1.59}$$

are coherent with (1.5) and (1.20) thinking to ν and μ as initial distributions.

Chapter 2

Elements of potential theory

2.1 Discrete Laplace equation

The study of steady state phenomena is often related to the Laplace equation $\Delta f = 0$, a second order partial differential equation. The potential theory is the study of the solutions to this problem.

Let $\mathcal{U} \subset \mathbb{R}^d$ be an open connected set.

Definition 2.1. A function $f \in \mathcal{C}(\mathcal{U}, \mathbb{R})$ is called *harmonic* is $\Delta f(x) = 0$ for all $x \in \mathcal{U}$ where

$$\Delta f(x) := \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} f(x) \tag{2.1}$$

Definition 2.2. Let $\mathcal{U} \subset \mathbb{R}$ be an open subset and $g \in \mathcal{C}(\partial\mathcal{U}, \mathbb{R})$. Then

$$\begin{cases} \Delta f(x) = 0 & \text{if } x \in \mathcal{U}, \\ f(x) = g(x) & \text{if } x \in \partial\mathcal{U} \end{cases} \tag{2.2}$$

is called *Dirichlet problem.*

Proving well-posedness for the Dirichlet problem can be a difficult task, but still many important results are obtained by an analytical approach. These results can be found in various books, for example in [4] (Chapter 3).

Theorem 2.1. *Let $B_r(x) := \{y \in \mathbb{R}^d : |x - y| < r\}$, the ball of radius r and center x, such that $B_r(x) \subset\subset \mathcal{U}$. If f is harmonic in $\mathcal{U}$ then the following mean value formulas hold:*

$$f(x) = \frac{1}{|B_r(x)|} \int_{B_r(x)} f(y)\, dy \tag{2.3}$$

$$f(x) = \frac{1}{|\partial B_r(x)|} \int_{\partial B_r(x)} f(y)\, d\sigma(y) \tag{2.4}$$

It is easy to observe that (2.3) and (2.4) are equivalent, since for any integrable function u it holds:

$$\int_{B_r(x)} u(y)\, dy = \int_0^r ds \int_{\partial B_s(x)} u(y)\, d\sigma(y).$$

A continuous function for which (2.3) or (2.4) holds is said to satisfy the *mean value property* and the reason is deduced directly from the two expressions.

This property actually characterizes the harmonic functions since the reverse is also true.

Theorem 2.2. *Let $f \in \mathcal{C}(\mathcal{U})$. If f satisfies the mean value property, then $f \in \mathcal{C}^\infty$ and it is harmonic in $\mathcal{U}$.*

Theorem 2.3. *If f has the mean value property and attains its maximum or minimum at $p \in \mathcal{U}$, then f is constant. In particular, if $\mathcal{U}$ is bounded and $f \in \mathcal{C}(\overline{\mathcal{U}})$ is not constant, then, for every $x \in \mathcal{U}$*

$$f(x) < \max_{\partial \mathcal{U}} f \qquad \text{and} \qquad f(x) > \min_{\partial \mathcal{U}} f. \tag{2.5}$$

The last result is usually called *maximum principle*. Two immediate consequences of this principle are the *uniqueness* and the *stability* for the solution to the Dirichlet problem. However it is not trivial to prove *existence* for the solution working from an analytical point of view.

On the other hand, the strength of the probabilistic method lies in being able to write immediately a solution to the problem. Kakutani [6] in 1944 provided the following candidate solution to the problem (2.2):

$$h(x) := \mathbb{E}_x[g(W_{\tau_{\partial U}})] \tag{2.6}$$

where $\mathbb{P}_x$ is the law of a d-dimensional Brownian motion starting from x and τ_A is the hitting time defined by

$$\tau_A := \inf\{t \geq 0 : W_t \in A\}. \tag{2.7}$$

Indeed by definition of $\tau_{\partial \mathcal{U}}$ we have $h(x) = g(x)$ if $x \in \partial \mathcal{U}$. Moreover if we consider $B_r(x)$ such that $\overline{B_r(x)} \subset \mathcal{U}$ using the strong Markov property for W at time $\tau_{\partial B_r(x)}$ we obtain:

$$\begin{aligned} h(x) &= \mathbb{E}_x\left[g(W_{\tau_{\partial U}})\right] \\ &= \mathbb{E}_x\left[\mathbb{E}_x\left[g(W_{\tau_{\partial U}})|W_{\tau_{\partial B_r(x)}}\right]\right] \\ &= \int_{\partial B_r(x)} \mathbb{E}_y\left[g(W_{\tau_{\partial U}})\right]\, d\mathbb{P}_x(W_{\tau_{\partial B_r(x)}} = y) \\ &= \frac{1}{|\partial B_r(x)|} \int_{\partial B_r(x)} h(y)\, d\sigma(y) \end{aligned} \tag{2.8}$$

that is the mean value property. Thus, as a consequence of Theorem 2.2, the last thing to check in order to prove that h defined in (2.6) is harmonic is the continuity. We refer to [5], section 4.2, for a deeper discussion on regularity.

More in general one could consider the larger class of *semi-elliptic* partial differential equations, i.e. second order operators of the form:

$$\mathcal{L} = \sum_{i=1}^{n} b_i(x)\frac{\partial}{\partial x_i} + \sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} \tag{2.9}$$

where b_i and $a_{ij} = a_{ji}$ are continuous and all the eigenvalues of the symmetric matrix $a = (a_{ij})_{i,j}$ are non negative. The corresponding Dirichlet problem is

$$\begin{cases} \mathcal{L}f(x) = 0 & \text{if } x \in \mathcal{U} \\ f(x) = g(x) & \text{if } x \in \partial\mathcal{U} \end{cases} \tag{2.10}$$

and it can be solved similarly to the Laplacian case, by constructing an Itô diffusion process whose generator coincides with $\mathcal{L}$. In [7], chapter 9, B. Øksendal deals with the problem of finding the correct solution.

Our interests now will focus on the study of solutions for the Dirichlet problem in the *discrete* case. The first step is to define a discretized version of the Laplacian or, more generally, of a second order semi-elliptic operator.

Let $h > 0$ and $h\mathbb{Z}^d$ be the lattice of points $x = (x_1, ..., x_d)$ whose coordinates are integer multiples oh h. Consider the simple random walk in dimension d over the lattice, whose transition matrix is defined by

$$P_{xy} := \begin{cases} \frac{1}{2d} & \text{if } |x-y| = h \\ 0 & \text{if } |x-y| \neq h \end{cases} \tag{2.11}$$

As noticed in [4] if f is a twice continuously differentiable function then for any x it holds

$$\lim_{h\to 0} \left[\frac{1}{2d} \sum_{|x-y|=h} f(y) - f(x) \right] \frac{1}{h^2} = \frac{1}{2d}\Delta f(x). \tag{2.12}$$

The part inside square parenthesis in the latter limit can be rewritten as $Lf(x)$, coherently with the notation introduced in chapter 1, where $L = P - Id$ is the infinitesimal generator of the uniform Markov chain with rate $\lambda = 1$ associated to the simple random walk. This suggests to define the *discrete Laplace operator* as

$$\Delta_h^* := \frac{2d}{h^2} L \tag{2.13}$$

so that the *discrete Laplace equation* is $\Delta_h^* f = 0$ or, equivalently, $Lf = 0$.

In order to generalize the discretization procedure to any other semi-elliptic differential operator $\mathcal{L}$ the idea is to construct inside the lattice $h\mathbb{Z}^d$ a random walk whose infinitesimal generator is L and such that $\frac{2d}{h^2} L \to \mathcal{L}$ for $h \to 0$.

We may also extend this reasoning to more complicated discrete structures. Consider a generic irreducible discrete Markov chain X with transition matrix P on the state space E. The process induces a graph structure on E, in which the point of E are the nodes and edges (x, y) are identified whenever $p_{xy} > 0$.

Definition 2.3. Let X be a discrete irreducible Markov process on E with generator L and $\mathcal{U}$ any subset of E. Then, for the boundary condition $g : E \to \mathbb{R}$, we define the *discrete Dirichlet problem* associated to X as

$$\begin{cases} Lf(x) = 0 & \text{if } x \in \mathcal{U} \\ f(x) = g(x) & \text{if } x \in \mathcal{U}^c. \end{cases} \tag{2.14}$$

If $Lh(x) = 0$ for all $x \in \mathcal{U}$ then h is called *harmonic* for X in $\mathcal{U}$.

In the discrete case the harmonicity of a function is equivalent to a mean value property because:

$$\begin{aligned} Lf(x) = 0 \quad &\iff \quad \sum_{y \in E} p(x,y)[f(y) - f(x)] = 0 \\ &\iff \quad \sum_{y \in E} p(x,y) f(y) - f(x) = 0 \quad \iff \quad f(x) = \sum_{y \in E} p(x,y) f(y) \end{aligned} \tag{2.15}$$

that means f in x is equal to its average (through P) on the other vertices.

Observe that in compact notation (2.15) could be rewritten as $Pf = f$, that is f is harmonic if and only if it is eigenvector or P with eigenvalue 1.

The Kakutani's probabilistic form of the solution to the Dirichlet problem is very simple to state and prove, as done in [8], section 9.2, by D. Levin, Y. Peres and E. Wilmer.

Theorem 2.4. *The unique solution to the Dirichlet problem (2.14) is*

$$h(x) := \mathbb{E}_x[g(X_{\tau_{\mathcal{U}^c}})] \tag{2.16}$$

Proof. Clearly $h(x) = g(x)$ for all $x \in \mathcal{U}^c$ by definition of $\tau_{\mathcal{U}^c}$. Thus, thanks to observation in (2.15), it is sufficient to prove that the function h defined above satisfies

the mean value property in $\mathcal{U}$.
If $x \in \mathcal{U}$ then $\mathbb{P}_x(\tau_{\mathcal{U}^c} \geq 1) = 1$. Hence using the Markov property for X it holds:

$$h(x) = \mathbb{E}_x[g(X_{\tau_{\mathcal{U}^c}})] = \sum_{y \in E} p(x,y)\mathbb{E}_x\left[g(X_{\tau_{\mathcal{U}^c}})|X_1 = y\right]$$
$$= \sum_{y \in E} p(x,y)\mathbb{E}_y[g(X_{\tau_{\mathcal{U}^c}})] = \sum_{y \in E} p(x,y)h(y) \quad (2.17)$$

that is h has the mean value property in x.

In order to get uniqueness we first prove the maximum principle, i.e. if h is harmonic for X in $\mathcal{U}$ then h reaches its maximum and minimum values on $\mathcal{U}^c$.

Define the set of maximum points for h in E

$$A := \{u \in E : h(u) = \max_{x \in E} h(x)\}. \quad (2.18)$$

Assume there exists $x \in \mathcal{U} \cap A$, otherwise there is nothing to prove, and choose $y \in E$ with $x \neq y$ such that $p(x,y) > 0$. Note that irreducibility allows to find such a point y. By harmonicity

$$h(x) = \sum_{z \in E} p(x,z)h(z) = p(x,y)h(y) + \sum_{z \neq y} p(x,z)h(z) \quad (2.19)$$

Now $h(z) \leq h(x)$ because $x \in A$, hence if $h(y) < h(x)$ we have from (2.19) $h(x) < \sum_{z \in E} p(x,z)h(x) = h(x)$ that is a contradiction. This means that $h(y) = h(x)$, i.e. $y \in A$. To conclude notice that by irreducibility for any x, y there exists a path of positive probability, hence choosing $x \in \mathcal{U} \cap A$ and $y \in \mathcal{U}^c$ and iterating the procedure above we discover that $y \in A$, that is the maximum value of h is reached in $\mathcal{U}^c$. Proof for the minimum is analogous.

Therefore, if h_1 and h_2 are two solutions of (2.14) then by linearity the difference $h_1 - h_2$ is solution to the problem $Lf = 0$ in $\mathcal{U}$ with boundary condition $f = 0$ in $\mathcal{U}^c$. Using maximum/minimum principle we observe that:

$$\max_{x \in E}(h_1 - h_2)(x) = \max_{x \in \mathcal{U}^c}(h_1 - h_2)(x) = 0 \quad \Longrightarrow \quad h_1 - h_2 \leq 0 \quad (2.20)$$
$$\min_{x \in E}(h_1 - h_2)(x) = \min_{x \in \mathcal{U}^c}(h_1 - h_2)(x) = 0 \quad \Longrightarrow \quad h_1 - h_2 \geq 0 \quad (2.21)$$

hence $h_1 = h_2$.

□

As noticed in [8], without using the maximum/minimum principle, an alternative way to prove uniqueness of the solution is to observe that condition $Lf(x) = 0$ in (2.14), rewritten as the mean value formula $f(x) = \sum_{y \in E} p(x,y) f(y)$, is a system of $|\mathcal{U}|$ linear equations in $|\mathcal{U}|$ unknowns. Thus existence of the solution automatically implies uniqueness.

2.2 Electrical networks

Definition 2.4. Let $\mathcal{X}$ be a countable set and $c : \mathcal{X} \times \mathcal{X} \to [0, +\infty)$ a real non-negative symmetric function. The pair $(\mathcal{X}, c)$ is called *electrical network* if:

(i) for any $x \in \mathcal{X}$
$$\mu(x) := \sum_{y \in \mathcal{X}} c(x,y) < +\infty \tag{2.22}$$

(ii) for all distinct $x, y \in \mathcal{X}$ there exist $x = z_1, z_2, ..., z_n = y$ such that $c(z_i, z_{i+1}) > 0$ for any $i = 1, ..., n-1$.

An electrical network can be interpreted as a weighted graph with positive weights. The elements of $\mathcal{X}$ are called *nodes*: Two nodes $x, y \in \mathcal{X}$ are connected if $c(x,y) > 0$. An *edge* is an element of $\mathcal{E}$, the set containing all the ordered pairs of connected nodes

$$\mathcal{E} = \{(x,y) \in \mathcal{X} \times \mathcal{X} : x \neq y, c(x,y) > 0\}. \tag{2.23}$$

Often can be useful to refer to an edge by calling it $e \in \mathcal{E}$. If $e = (x,y)$ then we use the notation $e^- = x, e^+ = y$, so that it is clear what is the direction of the edge; obviously by symmetry if $e = (x,y) \in \mathcal{E}$ then also $-e = (y,x) \in \mathcal{E}$.

The term *electrical* comes from the representation of such object in a physical point of view. For x, y different nodes $c(x,y)$ is called *conductance*, whereas

$$r(x,y) = \frac{1}{c(x,y)} \tag{2.24}$$

is called *resistance*.
In the case $(x,y) \notin \mathcal{E}$ then $c(x,y) = 0$, thus we set $r(x,y) = +\infty$.

Given any electrical network $(\mathcal{X}, c)$ it is possible to construct a discrete Markov chain X on the state space $\mathcal{X}$ with transition matrix $P = (p(x,y))_{x,y \in \mathcal{X}}$ such that

$$p(x,y) := \frac{c(x,y)}{\mu(x)} \tag{2.25}$$

X is ergodic and reversible with respect to the measure μ.

Conversely, any ergodic reversible discrete Markov chain is associated to some electrical network. Indeed, if μ is a reversible measure for X, using the transition matrix P we recover the weights $c(x,y) := \mu(x)p(x,y)$.
As the reversible measure is defined up to a multiplicative constant, there exist an infinite number of networks associated to X. Anyway, if the reversible measure is finite, the canonical choice is to take the network for which μ is a probability distribution.

We call *potential* any real function on the network $\mathcal{X}$.

Definition 2.5. Let $(\mathcal{X}, c)$ be an electrical network, $\mathcal{U} \subset \mathcal{X}$ and $g : \mathcal{X} \to \mathbb{R}$ a potential. The function $V : \mathcal{X} \to \mathbb{R}$ is an *equilibrium potential* with respect to the constraint $V(x) = g(x)$ for any $x \in \mathcal{U}^c$ if it satisfies:

- **Ohm's law:** the *current* $i : \mathcal{E} \to \mathbb{R}$ on the network associated to V is
$$i(x,y) = \frac{V(x) - V(y)}{r(x,y)} \tag{2.26}$$
- **Kirchoff's law:** for all $x \in \mathcal{U}$
$$\sum_{\{y:(x,y)\in\mathcal{E}\}} i(x,y) = 0. \tag{2.27}$$

Equivalently the equilibrium potential is solution of the problem:

$$\begin{cases} \sum_{y\in\mathcal{X}} c(x,y)[V(x) - V(y)] = 0 & \text{for } x \in \mathcal{U} \\ V(x) = g(x) & \text{for } x \in \mathcal{U}^c \end{cases} \tag{2.28}$$

Now let X be the discrete Markov chain associated with the network $(\mathcal{X}, c)$. If $\mathcal{L}$ is the infinitesimal generator of a uniform Markov chain over X with rate $\lambda = 1$ it holds

$$\mathcal{L}f(x) := \sum_{y\in\mathcal{X}} p(x,y)[f(y) - f(x)] = \sum_{y\in\mathcal{X}} \frac{c(x,y)}{\mu(x)}[f(y) - f(x)] \tag{2.29}$$

Observe that (2.7) is equivalent to the following Dirichlet problem:

$$\begin{cases} -\mathcal{L}V(x) = 0 & \text{for } x \in \mathcal{U} \\ V(x) = g(x) & \text{for } x \in \mathcal{U}^c \end{cases} \tag{2.30}$$

That is the equilibrium potential V is an harmonic function on $\mathcal{U}$ for the associated discrete Markov chain.

Proposition 2.5. *Let $(\mathcal{X}, c)$ be an electrical network and X its associated random walk with transition probabilities $p(x,y) = c(x,y)/\mu(x)$. Let $\mathcal{U} \subset \mathcal{X}$ and $g : \mathcal{U}^c \to \mathbb{R}$ a fixed potential on $\mathcal{U}^c$. Then the function $V : \mathcal{X} \to \mathbb{R}$ defined by*

$$V(x) := \mathbb{E}_x[g(X_{\tau_{\mathcal{U}^c}})] \tag{2.31}$$

is the unique equilibrium potential.

Proof. As we have observed above the equilibrium potential is an harmonic function that solves the Dirichlet problem (2.9). Then applying theorem (...) of the previous section the unique solution is the one defined in (2.10). □

As observed by A. Gaudillière in [3], if $(\mathcal{X}, c)$ is an electrical network then for each $\mathcal{U} \subset \mathcal{X}$ there is a unique set $\mathcal{H}_{\mathcal{U}}$ of harmonic functions on the associated random walk. But the functions in $\mathcal{H}_{\mathcal{U}}$ are harmonic for an infinite number of networks, since from different networks is possible to obtain the same random walk.
Actually the set of harmonic functions $\mathcal{H}_{\mathcal{U}}$ is associated to many more networks. Indeed the diagonal values $c(x,x)$ are irrelevant from an electrical point of view, but self-loops are possible according to the definition. Thus two networks that differ only in the diagonal values have the same set of harmonic potentials $\mathcal{H}_{\mathcal{U}}$, despite having associated different random walks.

2.3 Flows

We proceed by introducing the notion of flow, following the discussion in [3]. As we will see the flows are linked to the physical quantity previously defined not only by a physical point of view but also by the mathematical one.
Most of the results are quite simple in the case of finite state space, but without an effort under suitable assumptions it is possible to generalize them in the infinite case.

Definition 2.6. Any anti-symmetric real function on $\mathcal{E}$ is called *flow*, that is $\phi : \mathcal{E} \to \mathbb{R}$ such that $\phi(x,y) = -\phi(y,x)$.

For example given a potential f on $\mathcal{X}$ the current i that is constructed using Omh's law $i(x,y) = c(x,y)[f(x) - f(y)]$ is a flow and it is called *current flow*.
So any current associated to a potential is a flow, but viceversa a flow could not be the current associated to some potential.

Similarly to the lattice case of $\mathcal{X} = \mathbb{Z}^d$ the *divergence* of a flow is the function $\mathrm{div}(\phi) : \mathcal{X} \to \mathbb{R}$ defined by

$$\mathrm{div}(\phi)(x) = \sum_{e^- = x} \phi(e) \tag{2.32}$$

that is the sum of the flows from x towards its nearest neighbours.
If the network is finite, i.e. $|\mathcal{X}| < \infty$, then by the anti-symmetry of the flow we have that the *total divergence* of any flow ϕ on the network is

$$\sum_{x\in\mathcal{X}} \operatorname{div}(\phi)(x) = \sum_{\substack{x,y \\ (x,y)\in\mathcal{E}}} \phi(x,y) = 0 \tag{2.33}$$

since in the second sum both (x,y) and (y,x) are present.

The divergence of a current flow can also be rewritten by means of the infinitesimal generator $\mathcal{L}$ of random walk on the network

$$\operatorname{div}(i)(x) = \sum_{y:(x,y)\in\mathcal{E}} c(x,y)[f(x) - f(y)] = -\mu(x)\mathcal{L}f(x). \tag{2.34}$$

This gives an equivalent characterization of the harmonic potentials on $\mathcal{U} \subset \mathcal{E}$, that is a function $V : \mathcal{X} \to \mathbb{R}$ such that the current flow associated to it $\phi := -c\nabla V$ satisfies

$$\operatorname{div}(\phi)(x) = 0 \qquad \forall x \in \mathcal{U}. \tag{2.35}$$

Intuitively this means that for each $x \in \mathcal{U}$ the total incoming flux is equal to the total outgoing flux. A more explicit representation of the physical meaning comes from the application of the Stokes' Lemma.

Lemma 2.6 (Discrete Stokes' Lemma)**.** *For any flow ϕ and for any finite $K \subset \mathcal{X}$ it holds*

$$\sum_{e\in\partial K} \phi(e) = \sum_{x\in K} div(\phi)(x), \tag{2.36}$$

where the border of K is $\partial K := \{e \in \mathcal{E} : e^- \in K, e^+ \notin K\}$

Proof. We have

$$\sum_{x\in K} \operatorname{div}(\phi)(x) = \sum_{x\in K}\sum_{e^-=x} \phi(e) = \sum_{\substack{e\in\partial K \\ e^-\in K}} \phi(e) + \sum_{\substack{e\notin\partial K \\ e^-\in K}} \phi(e) \tag{2.37}$$

Now if $e \notin \partial K$ but $e^- \in K$, then both e^- and e^+ are in K. Then by symmetry also $-e \notin \partial K$ and $-e^- = e^+ \in K$. Thus e appears in the last summation if and only if $-e$ does. Hence by anti-symmetry of the flow $\phi(e) = -\phi(-e)$ and the second summation on the right hand side of (2.16) is 0. □

Corollary 2.7. *$V : \mathcal{X} \to \mathbb{R}$ is an harmonic potential on $\mathcal{U}$ if and only if its corresponding current flow $\phi := -c\nabla V$ has 0 flux through any finite cut-set ∂K, with $K \subset \mathcal{U}$.*

Given a flow ϕ for each $x \in \mathcal{X}$ the quantity $\mathrm{div}(\phi)(x)$ must be either $=, <$ or > 0. Thus we can divide the space into disjoint sets $A, B, \mathcal{X} \setminus (A \cup B)$, such that

$$\begin{aligned} \mathrm{div}(\phi)(a) &> 0 \qquad \forall a \in A, \\ \mathrm{div}(\phi)(b) &< 0 \qquad \forall b \in B, \\ \mathrm{div}(\phi)(x) &= 0 \qquad \forall x \notin A \cup B \end{aligned} \tag{2.38}$$

Definition 2.7. A flow ϕ is called *flow from A to B* if (2.17) holds. The elements of A are called *sources* whereas the element of B are called *sinks*. The *strength* of the flow is

$$|\phi| := \max \left\{ \sum_{a \in A} \mathrm{div}(\phi)(a); -\sum_{b \in B} \mathrm{div}(\phi)(b) \right\} \tag{2.39}$$

The flow is said to be *unitary* if $|\phi| = 1$; if the set B is empty we say ϕ is a flow from A *to infinity*.

By immediate application of Stokes' lemma we also see that if $\mathcal{X}$ is finite then $\sum_{a \in A} \mathrm{div}(\phi)(a) + \sum_{b \in B} \mathrm{div}(\phi)(b)$, i.e. the two quantities in (2.18) are equal.

Now we go back to the problem of identifying the equilibrium potentials on networks. Let $A, B \subset \mathcal{X}$ be disjoint subset such that

$$\mathbb{P}_x(\tau_{A \cup B} < \infty) = 1 \qquad \forall x \in \mathcal{X}, \tag{2.40}$$

where $\mathbb{P}$ stands for the law of the random walk X associated to the network. Condition (2.19) is clearly satisfied in the case $|\mathcal{X}| < \infty$ or more generally $|\mathcal{U}| < \infty$, where $\mathcal{U} = \mathcal{X} \setminus (A \cup B)$; also if we are considering an electrical network such condition holds, since the associated random walk X is ergodic.

Now we fix some constant values V_A and V_B on the sets A and B and we look for the solution to the Dirichlet problem

$$\begin{cases} \mathcal{L}V(x) = 0 & \text{for } x \in \mathcal{U} \\ V(x) = V_A & \text{for } x \in A \\ V(x) = V_B & \text{for } x \in B \end{cases} \tag{2.41}$$

If we call g the function such that $g(x) = V_A$ for $x \in A$, $g(x) = V_B$ for $x \in B$, then from Proposition (2.1) we know that the Kakutani's solution to the problem is

$$V(x) = \mathbb{E}_x[g(X_{\tau_{A \cup B}})] = V_A \mathbb{P}_x(\tau_A < \tau_B) + V_B \mathbb{P}_x(\tau_B < \tau_A) \tag{2.42}$$

that can be rewritten in a more compact way as

$$V(x) = V_B + (V_A - V_B)\mathbb{P}_x(\tau_A < \tau_B) \tag{2.43}$$

For $V_A = 1$ and $V_B = 0$ we obtain a very special solution that is denoted by $V_{A,B}$ whose value is

$$V_{A,B}(x) = \mathbb{P}_x(\tau_A < \tau_B) \tag{2.44}$$

The latter special case turns to be very useful. If i is the current flow

$$i = -c\nabla V \tag{2.45}$$

associated to the equilibrium potential which is solution of (2.20), then we have

$$\begin{aligned} i(x,y) &= -c(x,y)[V(y) - V(x)] \\ &= -c(x,y)[V_B + (V_A - V_B)\mathbb{P}_y(\tau_A < \tau_B) - V_B - (V_A - V_B)\mathbb{P}_x(\tau_A < \tau_B)] \\ &= -(V_A - V_B)c(x,y)[V_{A,B}(y) - V_{A,B}(x)] \end{aligned} \tag{2.46}$$

Thus changing the boundary values V_A and V_B the current flow does change only by a multiplicative factor.

The current flow in (2.24) coherently with our definition is a flow from A to B when $V_A > V_B$ or from B to A when $V_B > V_A$. Indeed if $x \in \mathcal{U}$ it holds $\mathrm{div}(i)(x) = 0$ and using the latter equation for any $a \in A$

$$\begin{aligned} \mathrm{div}(i)(a) &= \sum_{e^- = a} i(e) \\ &= -\sum_{y \in \mathcal{X}} (V_A - V_B)c(a,y)[V_{A,B}(y) - V_{A,B}(a)] \\ &= (V_A - V_B)\mu(a) \sum_{y \in \mathcal{X}} p(a,y)[\underbrace{\mathbb{P}_a(\tau_A < \tau_B)}_{=1} - \mathbb{P}_y(\tau_A < \tau_B)] \\ &= (V_A - V_B)\mu(a) \sum_{y \in \mathcal{X}} p(a,y)[\mathbb{P}_y(\tau_A > \tau_B)] \\ &= (V_A - V_B)\mu(a)\mathbb{P}_a(\tau_A^+ > \tau_B^+) \end{aligned} \tag{2.47}$$

where $\tau_K^+ := \min\{n \geq 1 : X_n \in K\}$ and in the latter equation we have used the Markov property. And analogously for $b \in B$ we get

$$\mathrm{div}(i)(b) = (V_B - V_A)\mu(b)\mathbb{P}_b(\tau_B^+ > \tau_A^+) \tag{2.48}$$

(Note: the divergence in A and in B is different from zero only on ∂A and on ∂B).

In order to compute the strength of the current notice that by the reversibility of X with respect to μ two paths, one the inverse of the other, starting with the stationary

measure have the same probability to occur. Thus we obtain

$$\begin{aligned}\sum_{a\in A}\mu(a)\mathbb{P}_a(\tau_A^+ > \tau_B^+) &= \sum_{a\in A}\sum_{n>0}\sum_{b\in B}\mu(a)\mathbb{P}_a(\tau_A^+ > \tau_B^+ = n, X_n = b)\\ &= \sum_{b\in B}\sum_{n>0}\sum_{a\in A}\mu(b)\mathbb{P}_b(\tau_B^+ > \tau_A^+ = n, X_n = a)\\ &= \sum_{b\in B}\mu(b)\mathbb{P}_b(\tau_B^+ > \tau_A^+)\end{aligned} \tag{2.49}$$

Definition 2.8. The *capacity* $C_{A,B}$ is the strength of the current flow associated to $V_{A,B}$, that is

$$C_{A,B} := \sum_{a\in A}\mu(a)\mathbb{P}_a(\tau_A^+ > \tau_B^+) = \sum_{b\in B}\mu(b)\mathbb{P}_b(\tau_B^+ > \tau_A^+) \tag{2.50}$$

Thus using (2.26) or (2.27) we see that

$$|i| = |V_A - V_B|C_{A,B}. \tag{2.51}$$

Furthermore if $C_{A,B}$ is finite (e.g. when $\mathcal{X}$ is finite) it is possible to define the flow

$$i_{A,B} := \frac{-c\nabla V_{A,B}}{C_{A,B}} \tag{2.52}$$

which is a unitary flow from A to B.

In chapter 4 of [3] Gaudillière provides a nice link between the condenser physics in the continuum and the electrical networks introduced above.
Following those considerations, if V is a potential and i the associated current flow, it is natural to define the *charge* in any $x \in \mathcal{X}$ as

$$q(x) := \mathrm{div}(i)(x) \tag{2.53}$$

Hence, from the previous considerations, if V is a potential such that i is a flow from A to B it holds

$$q(x) = \begin{cases} 0 & \text{if } x \notin A \cup B \\ (V_A - V_B)\mu(x)\mathbb{P}_x(\tau_A^+ > \tau_B^+) & \text{if } x \in A \\ (V_B - V_A)\mu(x)\mathbb{P}_x(\tau_B^+ > \tau_A^+) & \text{if } x \in B \end{cases} \tag{2.54}$$

and more generally if V is an harmonic function in $\mathcal{U}$ then each point of $\mathcal{U}$ has zero charge.

2.4 Green's function

Consider the electrical network $(\mathcal{X}, c)$ with the associated Markov chain X.

Definition 2.9. For any $B \subset \mathcal{X}$ and $x, y \in \mathcal{X}$ the *Green's function* is

$$G_B(x,y) := E_x\left[\sum_{n=0}^{\tau_B - 1} 1_{\{X(n)=y\}}\right], \tag{2.55}$$

where τ_B is the hitting time for the set B. Therefore the Green's function is the expected number of visits in y, starting from x, before hitting B.

Remark 1. Consider the case $x = y$. Let $N_B(y) := \sum_{n=0}^{\tau_B - 1} 1_{\{X(n)=y\}}$ be the random variable counting the number of visits in y before hitting B. If we call $\tau_y^+ = \inf\{n \geq 1 : X(n) = y\}$, the first return time in y, then, using the strong Markov property, we are able to write the probability distribution of $N_B(y)$ with respect to P_y as follows:

$$P_y(N_B(y) = k) = \left(P_y(\tau_y^+ < \tau_B)\right)^{k-1} P_y(\tau_y^+ > \tau_B) \tag{2.56}$$

that is a geometric distribution with success probability $p = P_y(\tau_y^+ > \tau_B)$. In particular, in the case $x = y$, the Green's function is the mean value of the geometric random variable $N_B(y)$ and is equal to $\frac{1}{p}$.

Note that the sum in (2.55) is almost surely finite. Then, exchanging the mean value with the sum, one gets

$$G_B(x,y) = \sum_{n \geq 0} P_x(X(n) = y \text{ and } n < \tau_B) \tag{2.57}$$

Then, from the reversibility of X, it holds the following symmetry condition:

$$\mu(x) G_B(x,y) = G_B(y,x)\mu(y). \tag{2.58}$$

If $A, B \subset \mathcal{X}$ satisfy (2.40) then the Green's function is linked to the equilibrium potential $V_{A,B}$. We define

$$L_{A,B} := \sup\{n \geq 0 : X(n) \in A \text{ and } n < \tau_B\} \tag{2.59}$$

with the convention $\sup\{\emptyset\} = -\infty$. Using the *last-exit decomposition* we have

$$\begin{aligned} V_{A,B}(x) &= P_x(\tau_A < \tau_B) = P_x(L_{A,B} \geq 0) \\ &= \sum_{n \geq 0} P_x(L_{A,B} = n) \\ &= \sum_{n \geq 0} \sum_{a \in A} P_x(X(n) = a, n < \tau_B) P_a(\tau_A^+ > \tau_B^+) \\ &= \sum_{a \in A} G_B(x,a) P_a(\tau_A^+ > \tau_B^+). \end{aligned} \tag{2.60}$$

And using (2.58) we obtain

$$V_{A,B}(x) = \sum_{a \in A} \frac{G_B(a,x)}{\mu(x)} \, \mu(a) P_a(\tau_A^+ > \tau_B^+) \tag{2.61}$$

where the factor $\mu(a)P_a(\tau_A^+ > \tau_B^+)$ is exactly the charge in a under the potential $V_{A,B}$.

We conclude this section with a simple technique, that will be useful in the next chapter, to calculate the Green's function.
Let P, defined in (2.25), be the transition matrix of X. Observe that

$$\sum_{n \geq 0} P_x(X(n) = y \text{ and } n < \tau_B) = \sum_{n \geq 0} \left([P]_{B^c}\right)^n (x,y) \tag{2.62}$$

where $[P]_{B^c}$ is the restriction of the transition matrix to the set $\mathcal{X} \setminus B$. The latter is a matrix with all row summing to values strictly lower then 1. In particular $\left([P]_{B^c}\right)^n$ tends to the zero matrix as n goes to infinity and it is possible to apply the formula for the geometric series:

$$\sum_{n \geq 0} \left([P]_{B^c}\right)^n (x,y) = \left([I - P]_{B^c}\right)^{-1} (x,y). \tag{2.63}$$

Hence, from (2.57), we obtain

$$G_B(x,y) = \left([I - P]_{B^c}\right)^{-1} (x,y). \tag{2.64}$$

Chapter 3

Random spanning forests

3.1 Forest measure

The aim of this section is to introduce an appropriate measure for the set of all the forests spanning a graph.

Let $\mathcal{X}$ be a finite space with $|\mathcal{X}| = n$, and $w : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ a fixed weight function.
The weights $\{w(x, y) \in [0, \infty) : (x, y) \in \mathcal{X} \times \mathcal{X}\}$ induce the structure of oriented weighted graph on $\mathcal{X}$, by taking as oriented edges of the graph all the possible jumps within X, namely

$$\mathcal{E} = \{(x, y) \in \mathcal{X} \times \mathcal{X} : x \neq y, w(x, y) > 0\}. \tag{3.1}$$

Definition 3.1. Let $\mathcal{X}$ be a finite space, $w : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ a weight function, and $\mathcal{E}$ set of edges accordingly to (3.3). Then $\mathcal{G} = (\mathcal{X}, \mathcal{E}, w)$ is called *weighted oriented graph.*

A weighted oriented graph can be thought of as the graph associated to a Markov chain on the state space $\mathcal{X}$, with generator L defined as the linear operator

$$(Lf)(x) = \sum_{y \in \mathcal{X}} w(x, y)[f(y) - f(x)], \tag{3.2}$$

for arbitrary $f : \mathcal{X} \to \mathbb{R}$.

Now we would like to identify all the possible spanning forests in $\mathcal{G}$.
First we can construct an *$\mathcal{X}$-spanning unrooted forest*, that is any simple undirected graph without cycles obtained by the nodes in $\mathcal{X}$. Clearly each connected component of these forests are trees.
Next we can choose a node in each tree that is called root, so that we are able to direct all the edges in a tree towards its root. The structure obtained is thus an oriented graph ϕ.
If ϕ is a subgraph of $\mathcal{G}$, i.e. each oriented edge of ϕ belongs to $\mathcal{E}$, we will say ϕ is a *spanning rooted forest of* $\mathcal{G}$. From now on ϕ will be identify with the set of its edges, i.e. as a subset of $\mathcal{E}$.

Finally we denote by $\mathcal{F}$ the set of all the spanning rooted forests of $\mathcal{G}$. Note that $\emptyset \in \mathcal{F}$ is the trivial forest for which each node of $\mathcal{X}$ is a root of a trivial tree.

Given $\phi \in \mathcal{F}$ we can define the total weight $w(\phi)$ of the forest as the product of the weights of its edges

$$w(\phi) := \prod_{e \in \phi} w(e) \tag{3.3}$$

The set $\mathcal{F}$ is finite because $|\mathcal{X}| < \infty$. Thus we can introduce the forest probability measure on $\mathcal{F}$.

Definition 3.2. Let $\phi \in \mathcal{F}$ be a forest and denote by $\rho(\phi)$ the set of its roots. Fix $q > 0$ Then the measure $w_q(\phi) : \mathcal{F} \to [0, \infty)$

$$w_q(\phi) := q^{|\rho(\phi)|} \prod_{e \in \phi} w(e) = q^{|\rho(\phi)|} w(\phi) \tag{3.4}$$

is called *standard measure* of ϕ.
This measure is finite and, normalizing it with the partition function $Z(q) := \sum_{\phi \in \mathcal{F}} w_q(\phi)$, we get

$$\nu_q(\phi) := \frac{w_q(\phi)}{Z(q)}, \tag{3.5}$$

the *standard probability measure* of ϕ.

This notion can be generalized introducing a family of rates

$$\{q(x) \in [0, +\infty] : x \in \mathcal{X}\}, \tag{3.6}$$

one for each state in $\mathcal{X}$. We denote $Q(x, y) = q(x) 1_{\{x=y\}}$ the diagonal matrix with entries given by q and define the set $S = \{x \in \mathcal{X} : q = \infty\}$.

Definition 3.3. Let $\phi \in \mathcal{F}$ be a forest and $\rho(\phi)$ the set of its root. Given the set of killing rates $\{q(x) \in [0, \infty] : x \in \mathcal{X}\}$ the measure $w_Q(\phi) : \mathcal{F} \to [0, \infty)$

$$w_Q(\phi) := \prod_{x \in \rho(\phi) \setminus S} q(x) \quad \prod_{e \in \phi} w(e) \quad 1_{\{S \subset \rho(\phi)\}} \tag{3.7}$$

is the *generalized measure* of ϕ.
After introducing the partition function $Z(Q) := \sum_{\phi \in \mathcal{F}} w_Q(\phi)$, we get

$$\nu_Q := \frac{w_Q(\phi)}{Z(Q)} \tag{3.8}$$

the *generalized probability measure* of ϕ.

Observe that in the first product of (3.8) we are excluding the roots in S, i.e. such that $q(x) = \infty$. This implies that $w_Q(\phi) < \infty$. and means that the roots in S do not count for the forest weight.

We call *random spanning forest* a random variable Φ_q (or Φ_Q, in the generalized version) associated to the probability measure ν_q (or ν_Q) on $\mathcal{X}$.

3.2 Loop-erased trajectory

In this paragraph we would like to introduce the notion of loop-erased trajectory, giving a general construction and recovering the probability for one of this trajectories to be chosen among the others. Concretely we need to track a Markov process that starts from an arbitrary point in $\mathcal{Y}$ and stops when arrives into a subset $B \in \mathcal{Y}$, but deleting all the loops generated along the path.

Let $\mathcal{Y}$ be a finite space and Y a continuous time Markov chain on it, with transition rates $\{\alpha(y,z) \in [0,+\infty] : y,z \in \mathcal{Y}\}$. Then the infinitesimal generator $\mathcal{L}$ of Y is given by

$$(\mathcal{L}f)(y) = \sum_{z\in\mathcal{Y}} \alpha(y,z)[f(z) - f(y)] \tag{3.9}$$

for any arbitrary $f : \mathcal{Y} \to \mathbb{R}$. We assume that for each $y \in \mathcal{Y}$ there is at most one $z \in \mathcal{Y}$ such that $\alpha(y,z) = +\infty$, so that it cannot exist two different attractors for the same point.
Let

$$\alpha(y) := \sum_{z\in\mathcal{Y}\setminus\{y\}} \alpha(y,z). \tag{3.10}$$

For any point $y \in \mathcal{Y}$ there are two cases:

- $\alpha(y) = +\infty$: in this case there exists an attractor (because $\mathcal{Y}$ is finite), i.e. a node $z \in \mathcal{Y}$ such that $\alpha(y,z) = +\infty$. If the Markov process Y arrives in y then it is immediately forced to jump in z;

- $\alpha(y) < +\infty$: then if the Markov process Y is in y, it is free to move to the other nodes according to the weights $\alpha(y,\cdot)$.

Thus there is an important difference in terms of dynamics, a forced behaviour in the first case, a free behaviour in the second one. Then we are able to choose a subset $B \subset \mathcal{Y}$ such that

$$\{y : \alpha(y) = +\infty\} \subset B. \tag{3.11}$$

so that $\alpha(y) < \infty$ for any $y \in B^c = \mathcal{Y} \setminus B$. We assume that

$$P_y(T_B < \infty) = 1 \qquad \forall y \in \mathcal{Y} \tag{3.12}$$

where P_y is the the law of the process Y starting in y and $T_B = \inf\{t \geq 0 : Y_t \in B\}$ is the hitting time of B. This means that B is accessible for the process Y from any starting point.
From now on B will play the role of absorption subset for the Markov chain Y.

Take $y_0, ..., y_n \in \mathcal{Y}$ a list of ordered points (possibly repeated). We denote with $\xi = (y_0, ..., y_n)$ the oriented trajectory from y_0 to y_n passing neatly through the points y_i. Moreover if $\xi = (y_0, ..., y_n)$ is a trajectory we denote with $V(\xi) = \{y_0, ..., y_n\}$ the set of nodes it contains.
If some point is repeated in the trajectory we call *loop* the ordered list of points in between the point and its repetition, the latter included. For example if we have

$$\xi = (y_0, ..., y_i, ..., y_j, ..., y_l) \tag{3.13}$$

with $y_i = y_j$ then $\eta = (y_{i+1}, ..., y_j)$ is a loop. Note that if ξ has a loop then it's possible to "*loop-erase*" the trajectory, simply deleting from ξ the elements of its loop.

Definition 3.4. Let $y_0, ..., y_l \in \mathcal{Y}$ be different points, i.e. $y_i \neq y_j$ for any $i \neq j \in \{0, ..., l\}$, such that $y_0, ..., y_{l-1} \in B^c$ and $y_l \in B$. The trajectory of $l + 1$ points and length l $\gamma_B = (y_0, ..., y_l)$ is called *self-avoiding path* towards B.

By definition a self-avoiding path has not loops.

We are interested in defining a procedure to sample self-avoiding paths toward B starting from the Markov chain Y on $\mathcal{Y}$.

Definition 3.5. For $y_0 \in \mathcal{Y}$ let P_{y_0} be the law of the process Y starting from y_0. Let $B \subseteq \mathcal{Y}$ be any subset with assumptions (3.12), (3.13). Define a *random loop-erased trajectory* Γ_B obtained from Y under P_{y_0} in the following way:

1. start the process Y at the point y_0 and let it be the first term of the trajectory;
2. each time the process jumps to a new point y_i add it to the trajectory;
3. if after adding a point a loop has been created, delete this loop;
4. iterate until the process reaches B, and let Y_{T_B} be the last point of the trajectory.

From this procedure we have that Γ_B is a self-avoiding path. Note that if the starting point $y_0 \in B$ then $\Gamma_B = (y_0)$ is the trivial trajectory containing only one point.

In order to compute the probability that the above procedure realizes a given trajectory γ_B we need to use a discretization of the continuous-time Markov process Y absorbed in B.
Consider the $n \times n$ stochastic matrix $\hat{P}$, with $n = |\mathcal{Y}|$, defined by imposing

$$\hat{P}(y,z) = \begin{cases} \dfrac{\alpha(y,z)}{\overline{\alpha}} & \text{if } y \in B^c \text{ and } y \neq z, \\ 1 - \displaystyle\sum_{x \in \mathcal{Y} \setminus \{y\}} \hat{P}(y,x) & \text{if } y = z \in B^c, \\ \delta_{\{y=z\}} & \text{if } y \in B, \end{cases} \tag{3.14}$$

where we recall that $\alpha(y,z)$ for $y, z \in \mathcal{Y}$ are the transition rates for Y, whereas

$$\overline{\alpha} := \max_{y \in B^c} \alpha(y) \tag{3.15}$$

with $\alpha(y)$ defined as in (3.11). Notice that $\overline{\alpha} < \infty$ because the maximum is taken on the set B^c, that is such that $\alpha(y) < +\infty$ for any $y \in B^c$. Also $\overline{\alpha} \neq 0$ because $\alpha(y) > 0$ for all $y \in B^c$ (otherwise the existance of such point y would be in contraddiction with the assumption (3.13)).

By construction, we can look at the Markov process Y absorbed in B as the uniform Markov process $\hat{Y}_{N(t)}$, where $\hat{Y}$ is the discrete-time HMC on the state space $\mathcal{Y}$ associated to the transition matrix $\hat{P}$ and $\{N(t)\}$ is an HPP($\overline{\alpha}$). Namely, after an exponential time with rate $\overline{\alpha}$, the process will move according to the transition probabilities in (3.15).
Thus, for any arbitrary $f : \mathcal{Y} \to \mathbb{R}$ we have

$$(\mathcal{L}f)(y) = (\overline{\alpha}(\hat{P} - 1)f)(y) \quad \text{for } y \in B^c \tag{3.16}$$

and in particular

$$[\mathcal{L}]_{B^c} = \left[\overline{\alpha}(\hat{P} - 1)\right]_{B^c} \tag{3.17}$$

where $\mathcal{L}$ is the infinitesimal generator of Y in (3.10) and $[\,\cdot\,]_A$ denotes the restriction of a matrix to a subset $A \subset \mathcal{Y}$.

Finally, with these instruments and following the argument used by L. Avena and A. Gaudillière in [10], we are able to state and prove our goal proposition:

Proposition 3.1. *Let Y be a continuous-time Markov chain on the state space $\mathcal{Y}$ with infinitesimal generator $\mathcal{L}$ given by*

$$(\mathcal{L}f)(y) = \sum_{z \in \mathcal{Y}} \alpha(y,z)[f(z) - f(y)]. \tag{3.18}$$

Let $\gamma_B = (y_0, ..., y_l)$ be a self-avoiding path from y_0 to B, with $y_0, ..., y_{l-1} \in B^c$ distinct points, and $y_l \in B$. Then

$$P_{y_0}(\Gamma_B = \gamma_B) = \prod_{i=0}^{l-1} \alpha(y_i, y_{i+1}) \frac{\det_{B^c \setminus \{y_0,...,y_{l-1}\}}(-\mathcal{L})}{\det_{B^c}(-\mathcal{L})} \tag{3.19}$$

where $\det_A(\,\cdot\,) = \det([\,\cdot\,]_A)$ denotes the determinant of the matrix restricted to $A \subset \mathcal{Y}$.

Proof. We make use of $\hat{Y}$, the discretization of the Markov process Y absorbed in B constructed above. Consider the stopping times

$$\hat{T}^+_{y_0} := \inf\{k \geq 1 : \hat{Y}_k = y_0\} \quad \textit{first return time in } y_0 \tag{3.20}$$

$$\hat{T}_B := \inf\{k \geq 0 : \hat{Y}_k \in B\} \quad \textit{hitting time of } B \tag{3.21}$$

Since we are looking for the probability $P_{y_0}(\Gamma_B = \gamma_B)$, we can exploit the loop-erasing property of Γ_B appropriately.
Assume that $\hat{T}^+_{y_0} < \hat{T}_B$, i.e. the process $\hat{Y}$ returns in y_0 before hitting B: then, as soon the process returns in y_0, the loop in Γ_B (which consists of the entire trajectory except the initial point) is erased. Hence, thanks to the strong Markov property of $\hat{Y}$, the process starts anew with the same law P_{y_0}, and also the trajectory Γ_B restarts from the point y_0.
In other words, the fact that $\hat{T}^+_{y_0} < \hat{T}_B$ does not affect the P_{y_0} probability that $\Gamma_B = \gamma_B$:

$$P_{y_0}(\Gamma_B = \gamma_B | \hat{T}^+_{y_0} < \hat{T}_B) = P_{y_0}(\Gamma_B = \gamma_B). \tag{3.22}$$

From this identity, and noting that $P_{y_0}(\hat{T}^+_{y_0} = \hat{T}_B) = 0$, we get

$$\begin{aligned} P_{y_0}(\Gamma_B = \gamma_B) &= P_{y_0}(\Gamma_B = \gamma_B, \hat{T}^+_{y_0} < \hat{T}_B) + P_{y_0}(\Gamma_B = \gamma_B, \hat{T}^+_{y_0} > \hat{T}_B) \\ &= P_{y_0}(\Gamma_B = \gamma_B | \hat{T}^+_{y_0} < \hat{T}_B) P_{y_0}(\hat{T}^+_{y_0} < \hat{T}_B) + P_{y_0}(\Gamma_B = \gamma_B, \hat{T}^+_{y_0} > \hat{T}_B) \\ &= P_{y_0}(\Gamma_B = \gamma_B) P_{y_0}(\hat{T}^+_{y_0} < \hat{T}_B) + P_{y_0}(\Gamma_B = \gamma_B, \hat{T}^+_{y_0} > \hat{T}_B), \end{aligned}$$

and it follows

$$P_{y_0}(\Gamma_B = \gamma_B) = \frac{P_{y_0}(\Gamma_B = \gamma_B, \hat{T}^+_{y_0} > \hat{T}_B)}{1 - P_{y_0}(\hat{T}^+_{y_0} < \hat{T}_B)} = \frac{P_{y_0}(\Gamma_B = \gamma_B, \hat{T}^+_{y_0} > \hat{T}_B)}{P_{y_0}(\hat{T}^+_{y_0} > \hat{T}_B)}. \tag{3.23}$$

Consider now the random variable

$$N_{y_0} = \sum_{k=0}^{\infty} 1_{\{\hat{Y}_k = y_0\}} \tag{3.24}$$

counting the number of returns in y_0 for the random walk $\hat{Y}$. If the process starts in $\hat{Y}_0 = y_0 \in B^c$, by construction the probability that $\hat{Y}$ returns (for the first time) in

y_0 is $P_{y_0}(\hat{T}^+_{y_0} < \hat{T}_B)$. By the strong Markov property for $\hat{Y}$, we also know that this probability is the same at each return, because $P_{y_0}(\,\cdot\,)$ and $P_{y_0}(\,\cdot\,|\hat{T}^+_{y_0})$ have the same law. Thus

$$P_{y_0}(N_{y_0} = n) = \Big(P_{y_0}(\hat{T}^+_{y_0} < \hat{T}_B)\Big)^{n-1} P_{y_0}(\hat{T}^+_{y_0} > \hat{T}_B)$$
$$= \Big(1 - P_{y_0}(\hat{T}^+_{y_0} > \hat{T}_B)\Big)^{n-1} P_{y_0}(\hat{T}^+_{y_0} > \hat{T}_B) \quad (3.25)$$

i.e. it is the probability to return $n-1$ times in y_0 and then to reach the subset B: indeed once $\hat{Y}$ reaches B the process cannot come back to y_0. From (3.26) we get that under P_{y_0} the random variable N_{y_0} has geometric distribution with *success probability* $p = P_{y_0}(\hat{T}^+_{y_0} > \hat{T}_B)$.
Hence, we can express the mean number of returns in y_0 as the mean of a geometric distribution, that is

$$E_{y_0}[N_{y_0}] = \frac{1}{p} = \frac{1}{P_{y_0}(\hat{T}^+_{y_0} > \hat{T}_B)} \quad (3.26)$$

where the latter is the denominator of the r.h.s. of (3.24). Moreover we see that

$$E_{y_0}[N_{y_0}] = E_{y_0}\left[\sum_{k=0}^{\infty} 1_{\{\hat{Y}_k = y_0\}}\right] = \sum_{k=0}^{\infty} E_{y_0}\left[1_{\{\hat{Y}_k = y_0\}}\right] = \sum_{k=0}^{\infty} P_{y_0}(\hat{Y}_k = y_0) \quad (3.27)$$

hence

$$E_{y_0}[N_{y_0}] = \sum_{k=0}^{\infty} [\hat{P}]^k_{B^c}(y_0, y_0) = \Big([1 - \hat{P}]_{B^c}\Big)^{-1}(y_0, y_0) = \frac{\det_{B^c\setminus\{y_0\}}(1 - \hat{P})}{\det_{B^c}(1 - \hat{P})} \quad (3.28)$$

where we used the formula for the geometric series of a matrix and the Cramer's rule for the inverse of a matrix.

Furthermore, by the Markov property, the numerator in equation (3.24) can be written as

$$P_{y_0}(\Gamma_B = (y_0, ..., y_l), \hat{T}^+_{y_0} > \hat{T}_B) = \hat{P}(y_0, y_1) P_{y_1}(\Gamma_B = (y_1, ..., y_l), \hat{T}^+_{y_0} > \hat{T}_B)$$
$$= \hat{P}(y_0, y_1) P_{y_1}(\Gamma_{B\cup\{y_0\}} = (y_1, ..., y_l)), \quad (3.29)$$

where the latter equality is justified by the fact that if $\Gamma_B = (y_1, ..., y_l)$, and at the same time $\hat{T}^+_{y_0} > \hat{T}_B$, then Γ_B cannot make any (erased) loop containing y_0.

Finally from (3.24), summing up the results in (3.27), (3.29) and (3.30), we obtain a recursive formula for the probability in the statement of the proposition:

$$P_{y_0}(\Gamma_B = (y_0, ..., y_l)) = \hat{P}(y_0, y_1) \frac{\det_{B^c\setminus\{y_0\}}(1 - \hat{P})}{\det_{B^c}(1 - \hat{P})} P_{y_1}(\Gamma_{B\cup\{y_0\}} = (y_1, ..., y_l)). \quad (3.30)$$

Iterating this equation we notice that the last term is $P_{y_l}(\Gamma_{B\cup\{y_0,...,y_{l-1}\}} = (y_l)) = 1$ because the starting point y_l is contained in B. In conclusion, by iteration and with the notational convention $\{y_0, ..., y_j\} = \emptyset$ for $j < 0$, we get

$$\begin{aligned}
P_{y_0}(\Gamma_B = \gamma_B) &= \prod_{i=0}^{l-1} \hat{P}(y_i, y_{i+1}) \frac{\det_{B^c\setminus\{y_0,...,y_i\}}(1-\hat{P})}{\det_{B^c\setminus\{y_0,...,y_{i-1}\}}(1-\hat{P})} \\
&= \frac{\det_{B^c\setminus\{y_0,...,y_{l-1}\}}(1-\hat{P})}{\det_{B^c}(1-\hat{P})} \prod_{i=0}^{l-1} \hat{P}(y_i, y_{i+1}) \\
(*) \quad &= \frac{\det_{B^c\setminus\{y_0,...,y_{l-1}\}}(-\mathcal{L}/\overline{\alpha})}{\det_{B^c}(-\mathcal{L}/\overline{\alpha})} \prod_{i=0}^{l-1} \frac{\alpha(y_i, y_{i+1})}{\overline{\alpha}} \\
&= \overline{\alpha}^l \frac{\det_{B^c\setminus\{y_0,...,y_{l-1}\}}(-\mathcal{L})}{\det_{B^c}(-\mathcal{L})} \frac{1}{\overline{\alpha}^l} \prod_{i=0}^{l-1} \alpha(y_i, y_{i+1}) \\
&= \frac{\det_{B^c\setminus\{y_0,...,y_{l-1}\}}(-\mathcal{L})}{\det_{B^c}(-\mathcal{L})} \prod_{i=0}^{l-1} \alpha(y_i, y_{i+1}),
\end{aligned}$$

where in $(*)$ we used the identity (3.18) and the transition probabilities of the Markov matrix in (3.15).

□

3.3 Wilson's algorithm

It looks difficult to sample the standard and generalized probability measures in (3.6) and (3.9) and to explicitly compute their partition function. The purpose of this section is to introduce an algorithm, due to D. Wilson and J. Propp [9], that allows us to achieve this goal.

Recall, from section 3.1, that X is an irreducible Markov process on $\mathcal{X}$ defined by its generator L in (3.2).
We now consider an extended process $\overline{X}$ on $\overline{\mathcal{X}} = \mathcal{X}\cup\{\Delta\}$ where $\Delta \notin \mathcal{X}$ is an absorbing state: this can be obtained by leaving unchanged the dynamics inside $\mathcal{X}$ and by adding some killing rates $\{q(x) \in [0, +\infty] : x \in \mathcal{X}\}$, each one corresponding to a Poisson process with rate $q(x)$ associated to the transition from $x \in \mathcal{X}$ to Δ.

The infinitesimal generator of $\overline{X}$ is given by the linear operator

$$\mathcal{L}f(x) = \begin{cases} Lf(x) + q(x)[f(\Delta) - f(x)] & \text{if } x \neq \Delta \\ 0 & \text{if } x = \Delta \end{cases} \tag{3.31}$$

or, equivalently, in matrix form, by

$$\mathcal{L} = (\alpha(x,y))_{x,y\in\overline{\mathcal{X}}} = \left(\begin{array}{ccc|c} & & & q(x_1) \\ & L-Q & & \vdots \\ & & & q(x_n) \\ \hline 0 & \cdots & 0 & 0 \end{array}\right) \tag{3.32}$$

where we recall that $Q(x,y) = q(x)1_{\{x=y\}}$. In particular we notice that

$$\alpha(x,y) = \begin{cases} w(x,y) & \text{if } x\in\mathcal{X}, y\in\mathcal{X}, x\neq y \\ q(x) & \text{if } x\in\mathcal{X}, y=\Delta \end{cases} \tag{3.33}$$

Moreover

$$[-\mathcal{L}]_{\mathcal{X}} = Q - L \qquad \text{and} \qquad [-\mathcal{L}]_A = [Q-L]_A \quad \forall A\subset\mathcal{X}. \tag{3.34}$$

Definition 3.6. We define the *Wilson's Algorithm* as follows:

1. start the process $\overline{X}$ from any $x_1\in\mathcal{X}$ until it reaches Δ, i.e. perform a random loop-erased trajectory Γ_Δ starting from x_1 under $\mathcal{L}$. Call γ_1 this self-avoiding path;

2. if $V(\gamma_1) = \overline{\mathcal{X}}$ stop; otherwise choose any point $x_2\in\mathcal{X}\setminus V(\gamma_1)$ and perform a loop-erased trajectory $\Gamma_{V(\gamma_1)}$ starting from x_2. Call γ_2 this self-avoiding path;

3. if $\cup_{i=1,2}V(\gamma_i) = \overline{\mathcal{X}}$ stop; otherwise choose any point $x_3\in\mathcal{X}\setminus\cup_{i=1,2}V(\gamma_i)$ and perform a loop-erased trajectory $\Gamma_{\cup_{i=1,2}V(\gamma_i)}$ starting from x_3. Call γ_3 this self-avoiding path;

4. iterate this procedure until $\overline{\mathcal{X}}$ is completely covered.

Let $\mathcal{T}_{\overline{\mathcal{X}}}$ be the set of the spanning oriented (towards the root) trees on $\overline{\mathcal{X}}$ rooted at Δ. By construction, the Wilson's algorithm produces an element $\tau\in\mathcal{T}_{\overline{\mathcal{X}}}$ in finite time, because the irreducibility of $\overline{X}$ ensures that, for any $B\subset\overline{\mathcal{X}}$, $\inf\{t\geq 0 : \overline{X}_t\in B\} < \infty$ almost surely.
Let $\mathcal{W}$ denote the random element of $\mathcal{T}_{\overline{\mathcal{X}}}$ obtained from the Wilson's algorithm. Then it holds the following.

Proposition 3.2. *Fix $\tau\in\mathcal{T}_{\overline{\mathcal{X}}}$, a spanning tree rooted at Δ. Then*

$$\mathbb{P}(\mathcal{W}=\tau) = \frac{\displaystyle\prod_{x\in\partial_-\rho(\tau)\setminus S} q(x) \prod_{e\in\tau\setminus\partial\rho(\tau)} w(e)}{\det_{\mathcal{X}\setminus S}(-\mathcal{L})} 1_{\{S\subset\partial_-\rho(\tau)\}}. \tag{3.35}$$

where $\partial\rho(\tau) = \{e\in\tau : e^+ = \Delta\}$ is the set of edges in τ pointing to the root and $\partial_-\rho(\tau) = \cup_{e\in\partial\rho(\tau)}\{e^-\}$.

Proof. The proof follows as a corollary of Proposition 3.1 from the following observations.

First notice that the Wilson algorithm can only produce trees τ in $\mathcal{T}_{\overline{\mathcal{X}}}$ such that $S \subset \partial_-\rho(\tau)$. Indeed, whenever the process $\overline{X}$ reaches a node $s \in S$ it immediately jumps to Δ because $q(s) = +\infty$: thus $s \in \partial_-\rho(\tau)$. In particular $(s, \Delta) \in \tau$ for any $s \in S$.
From now on we thus consider τ such that $S \subset \partial_-\rho(\tau)$.

Pick any arbitrary $x_1 \in \mathcal{X} \setminus B_1$ with $B_1 = \{\Delta\} \cup S$. By definition of tree, there exists a unique oriented path of minimum length from x_1 to some element in B_1 for the fixed tree τ; we call it

$$\gamma_1 = (x_1^{(0)}, x_1^{(1)}, ..., x_1^{(l_1)}), \tag{3.36}$$

with $x_1^{(0)} = x_1$. Notice that it could be either $x_1^{(l_1)} = \Delta$ or $x_1^{(l_1)} \in S$, but in both cases the knowledge of γ_1 is sufficient to completely determine the path from x_1 to the root Δ, thanks to the observation above.

Thus, following the Wilson algorithm, the probability to be generate γ_1 is equal to $P_{x_1}(\Gamma_{B_1} = \gamma_1)$, where Γ denotes the random loop-erased trajectory obtained from $\overline{X}$. From (3.20) we can compute this probability using the generator of $\overline{X}$ defined in (3.33):

$$P_{x_1}(\Gamma_{B_1} = \gamma_1) = \frac{\det_{B_1^c \setminus V(\gamma_1)}(-\mathcal{L})}{\det_{B_1^c}(-\mathcal{L})} \prod_{i=0}^{l_1-1} \alpha(x_1^{(i)}, x_1^{(i+1)}) \tag{3.37}$$

Now let $B_2 = B_1 \cup V(\gamma_1)$. If $B_2 = \mathcal{X}$ the algorithm stops, otherwise pick any $x_2 \in \mathcal{X} \setminus B_2$. Again, following Wilson algorithm, the unique oriented path of minimum length from x_2 to some element in B_2 is denoted by $\gamma_2 = (x_2^{(0)}, ..., x_2^{(l_2)})$, with $x_2^{(0)} = x_2$. γ_2 has a probability to occur given by $P_{x_2}(\Gamma_{B_2} = \gamma_2)$. Notice that as before the knowledge of γ_2 is sufficient to determine the entire path from x_2 to Δ, because if $x_2^{(l_2)} \neq \Delta$ then the portion of path from $x_2^{(l_2)}$ to Δ has been detected yet in the previous step of the algorithm.
From (3.20) we have

$$P_{x_2}(\Gamma_{B_2} = \gamma_2) = \frac{\det_{B_2^c \setminus V(\gamma_2)}(-\mathcal{L})}{\det_{B_2^c}(-\mathcal{L})} \prod_{i=0}^{l_2-1} \alpha(x_2^{(i)}, x_2^{(i+1)}). \tag{3.38}$$

We may iterate this procedure until we find some point $x_n \in \mathcal{X} \setminus B_n$ such that $B_n \cup$

$V(\gamma_n) = \mathcal{X}$. Since $B_n^c \setminus V(\gamma_n) = \emptyset$, as consequence of Proposition 3.1 it holds

$$\begin{aligned} P_{x_n}(\Gamma_{B_n} = \gamma_n) &= \frac{\det_{B_n^c \setminus V(\gamma_n)}(-\mathcal{L})}{\det_{B_n^c}(-\mathcal{L})} \prod_{i=0}^{l_n-1} \alpha(x_n^{(i)}, x_n^{(i+1)}) \\ &= \frac{1}{\det_{B_n^c}(-\mathcal{L})} \prod_{i=0}^{l_n-1} \alpha(x_n^{(i)}, x_n^{(i+1)}). \end{aligned} \tag{3.39}$$

Finally, we have $P(\mathcal{W} = \tau) = \prod_{k=1}^n P_{x_k}(\Gamma_{B_k} = \gamma_k)$ because the loop-erased trajectories are independent and their union forms the entire tree τ. Moreover, since by definition $B_k^c \setminus V(\gamma_k) = B_{k+1}^c$ for any $k = 1, ..., n-1$, it follows

$$\begin{aligned} P(\mathcal{W} = \tau) &= \prod_{k=1}^n P_{x_k}(\Gamma_{B_k} = \gamma_k) \\ &= \prod_{k=1}^n \left[\frac{\det_{B_k^c \setminus V(\gamma_k)}(-\mathcal{L})}{\det_{B_k^c}(-\mathcal{L})} \prod_{i=0}^{l_k-1} \alpha(x_k^{(i)}, x_k^{(i+1)}) \right] \\ &= \frac{1}{\det_{B_1^c}(-\mathcal{L})} \prod_{k=1}^n \prod_{i=0}^{l_k-1} \alpha(x_k^{(i)}, x_k^{(i+1)}) \end{aligned} \tag{3.40}$$

By construction the terms in the double product are distinct and each one is biunivocally associated to an edge in τ (except for the edges connecting the elements in S with Δ which are not counted). In particular using the expression of $\alpha(x, y)$ in (3.34) we can split the product over the two sets $\{e \in \partial\rho(\tau) : e^- \notin S\}$ and $\{e \in \tau \setminus \partial\rho(\tau)\}$:

$$\mathbb{P}(\mathcal{W} = \tau) = \frac{\displaystyle\prod_{e \in \partial\rho(\tau) : e^- \notin S} q(x) \prod_{e \in \tau \setminus \partial\rho(\tau)} w(e)}{\det_{B_1^c}(-\mathcal{L})} \tag{3.41}$$

which is equal to the claim of the theorem, as $B_1^c = \mathcal{X} \setminus S$.

Note that the random choice of the points $x_1, ..., x_n$ used by the Wilson's algorithm is irrelevant, since the result is independent from them.

□

Remark 2. In the construction of the Wilson's algorithm we have introduced the absorbing state Δ, and this let us define the extended process $\overline{X}$.
Anyway, in the case of uniform killing rate $q > 0$, the dynamics can also be described in a different way: indeed consider the loop-erased random walk (LERW) obtained from X in the sense of Definition 3.5, that is erasing loops as soon as they appear, with target subset $B = \emptyset$. Thus the Wilson's algorithm is obtained by running sequentially the LERW and stopping it at independent times T_q, which are exponential random variables of parameter q. For more details see [11].

3.4 Partition function

The set $\mathcal{F}$ of spanning rooted forests in $\mathcal{X}$ and the set $\mathcal{T}_{\overline{\mathcal{X}}}$ of spanning trees rooted at Δ are in natural bijection.
Indeed, if $\phi \in \mathcal{F}$, then we may uniquely define $\tau(\phi) \in \mathcal{T}_{\overline{\mathcal{X}}}$ by adding to ϕ all the edges connecting the roots of ϕ to Δ, that is the set of edges $\{(e^-, e^+) : e^- \in \rho(\phi), e^+ = \Delta\}$. Conversely, if $\tau \in \mathcal{T}_{\overline{\mathcal{X}}}$, we obtain an unique element $\phi(\tau) \in \mathcal{F}$ by removing from τ all the edges containing the root Δ.

Thanks to the latter observation we are able to sample the measure (3.9) via Wilson's algorithm.

Theorem 3.3. *Let $\{q(x) \in [0, +\infty] : x \in \mathcal{X}\}$ be a collection of killing rates and $w_Q, Z(Q), \nu_Q$ the associated forest measure, partition function, forest probability measure on $\mathcal{F}$ as in definition 3.3. For any $\phi \in \mathcal{F}$, with $\rho(\phi)$ denoting the set of the roots, let $\tau(\phi) \in \mathcal{T}_{\overline{\mathcal{X}}}$ be the oriented spanning tree inside $\overline{\mathcal{X}}$ rooted at Δ and such that $\tau(\phi) = \phi \cup \{(e^-, e^+) : e^- \in \rho(\phi), e^+ = \Delta\}$. Then*

$$Z(Q) = \det_{\mathcal{X}\setminus S}(Q - L) \tag{3.42}$$

and

$$\nu_Q(\phi) = P(\mathcal{W} = \tau(\phi)) \tag{3.43}$$

Proof. We recall that by definition

$$w_Q(\phi) := \prod_{x \in \rho(\phi)\setminus S} q(x) \quad \prod_{e \in \phi} w(e) \quad 1_{\{S \subset \rho(\phi)\}}. \tag{3.44}$$

Note that we can rewrite the two sets indexing the products as

$$\{x \in \rho(\phi) \setminus S\} = \{x \in \partial_- \rho(\tau(\phi)) \setminus S\}$$

so we obtain that

$$\mathbb{P}(\mathcal{W} = \tau) = \frac{\displaystyle\prod_{x \in \partial_- \rho(\tau)\setminus S} q(x) \quad \prod_{e \in \tau \setminus \partial\rho(\tau)} w(e)}{\det_{\mathcal{X}\setminus S}(-\mathcal{L})} 1_{\{S \subset \partial_- \rho(\tau)\}} = \frac{w_Q(\phi)}{\det_{\mathcal{X}\setminus S}(-\mathcal{L})}. \tag{3.45}$$

Now the denominator in the latter equation is independent from ϕ, thus in order to obtain $Z(Q)$ we just need to sum over $\phi \in \mathcal{F}$ and use the fact that the Wilson's algorithm may only produce trees $\tau(\phi) \in \mathcal{T}_{\overline{\mathcal{X}}}$ for some $\phi \in \mathcal{F}$. Then

$$Z(Q) = \sum_{\phi \in \mathcal{F}} w_Q(\phi) = \det_{\mathcal{X}\setminus S}(-\mathcal{L}) \sum_{\phi \in \mathcal{F}} \mathbb{P}(\mathcal{W} = \tau(\phi)) = \det_{\mathcal{X}\setminus S}(-\mathcal{L}) \tag{3.46}$$

and to conclude, using observation (3.35), $\det_{\mathcal{X}\setminus S}(-\mathcal{L}) = \det_{\mathcal{X}\setminus S}(Q - L)$.

Finally from (3.46)

$$\nu_Q(\phi) = \frac{w_Q(\phi)}{Z(Q)} = \mathbb{P}(\mathcal{W} = \tau(\phi)). \tag{3.47}$$

□

Corollary 3.4. *If $q > 0$ is an uniform killing rate the partition function for the standard forest measure is*

$$Z(q) = q \prod_{i=1}^{n-1} (q + \lambda_i) \tag{3.48}$$

where λ_i's, for $i = 0, ..., n-1$, are the eigenvalues of $-L$ ordered by non-decreasing real part and $\lambda_0 = 0$.

Proof. Applying Theorem 3.3 to the case $q(x) \equiv q > 0$, we have $S = \emptyset$ and $Q = qI$. Then

$$Z(q) = \det_{\mathcal{X}}(Q - L) = \det(q1 - L) = \chi_L(q) \tag{3.49}$$

where $\chi_L(q)$ is the characteristic polynomial of L. Hence, factorizing it, we have $Z(q) = \prod_{i=0}^{n-1}(q + \lambda_i)$.
At last $\lambda_0 = 0$ because the infinitesimal generator matrix L is singular.

□

The latter result leads to another important corollary for the standard case with killing rates $q(x) \equiv q > 0$, because it allows to describe the probability distribution of the root number of Φ_q. Recall that Φ_q is a random variable taking values in $\mathcal{F}$ such that $\mathbb{P}(\Phi_q = \phi) = \nu_q(\phi)$.

Corollary 3.5. *Let $q > 0$ be an uniform killing rate and assume $-L$ has real spectrum $Sp(-L) = \{\lambda_0, ..., \lambda_{n-1}\}$. Let $|\rho(\Phi_q)|$ be the random variable counting the number of roots of Φ_q. Then*

$$|\rho(\Phi_q)| \sim \sum_{i=0}^{n-1} \mathfrak{B}\left(\frac{q}{q + \lambda_i}\right) \tag{3.50}$$

the sum of n independent Bernoulli random variables with parameters $\frac{q}{q+\lambda_i}$.

Proof. For fixed $k \geq 0$, we aim to evaluate $\mathbb{P}(|\rho(\Phi_q)| = k)$, that can be expressed as a fraction as follows

$$\mathbb{P}(|\rho(\Phi_q)| = k) = \frac{1}{Z(q)} \sum_{\phi : |\rho(\phi)| = k} w_q(\phi) = \frac{Z_k(q)}{Z(q)}, \tag{3.51}$$

where the numerator corresponds to the total weight of the forests in $\mathcal{F}$ having k roots.

Observe that, by definition, $Z(q) = \sum_{\phi\in\mathcal{F}} q^{|\rho(\phi)|} w(\phi)$. If we consider the partition function as a polynomial of q, we deduce that $Z_k(q)$ is equal to the term of degree k in $Z(q)$.

Besides, from Corollary 3.4, the partition function is $Z(q) = \prod_{i=0}^{n-1}(q+\lambda_i)$, product of monomials. Hence the coefficient of degree k of such polynomial is simply given by the sum of the products of all the possible combinations of $n-k$ elements in the set $\{\lambda_0, ..., \lambda_{n-1}\}$:

$$Z_k(q) = q^k \sum_{I\in\mathcal{P}[n-k]} \prod_{i\in I} \lambda_i \tag{3.52}$$

where $\mathcal{P}[m]$ denotes the set of all possible choices of m elements in $\{0, 1, ..., n-1\}$. Summing up and rearranging the terms, we obtain

$$\begin{aligned}
\mathbb{P}(|\rho(\Phi_q)| = k) = \frac{Z_k(q)}{Z(q)} &= \frac{q^k \displaystyle\sum_{I\in\mathcal{P}[n-k]} \prod_{i\in I} \lambda_i}{\displaystyle\prod_{i\in\{0,...,n-1\}} (q+\lambda_i)} \\
&= \sum_{I\in\mathcal{P}[n-k]} \left[\prod_{i\in I} \left(\frac{\lambda_i}{q+\lambda_i}\right)\right] \left[\prod_{j\notin I} \left(\frac{q}{q+\lambda_j}\right)\right] \\
&= \sum_{J\in\mathcal{P}[k]} \left[\prod_{j\in J} \left(\frac{q}{q+\lambda_j}\right)\right] \left[\prod_{i\notin J} \left(1-\frac{q}{q+\lambda_i}\right)\right]
\end{aligned} \tag{3.53}$$

that is equivalent to the claim. □

Remark 3. Observe that any forest $\phi \in \mathcal{F}$ must contain at least one and at most $n = |\mathcal{X}|$ trees (or roots). This is consistent with Corollary 3.6 because, as we know, $\lambda_0 = 0$ is an eigenvalue of $-L$ and then the first Bernoulli random variable is $\mathfrak{B}\left(\frac{q}{q+\lambda_0}\right) = \mathfrak{B}(1) = 1$ almost surely. So the statement (3.51) can be rewritten more explicitly as

$$|\rho(\Phi_q)| \sim 1 + \sum_{i=1}^{n-1} \mathfrak{B}\left(\frac{q}{q+\lambda_i}\right) \tag{3.54}$$

Remark 4. In Corollary 3.6 we used the hypothesis that the eigenvalues of $-L$ are real. In fact it is possible to remove such condition and obtain the same exact result of equation (3.54). However the outcome is not the sum of Bernoulli random variables, as λ_i may be complex numbers.
A clever way to deal with this problem is provided by L. Avena and A. Gaudillière

in [12] introducing new random variables with *complex Bernoulli law*. Since L is a real matrix, also its characteristic polynomial is real, and consequently the eigenvalues must be real or complex conjugate. Then if we denote $p_i := \frac{q}{q+\lambda_i}$ we observe that

$$\overline{p_i} = \overline{\left(\frac{q}{q+\lambda_i}\right)} = \frac{q}{q+\overline{\lambda_i}}. \tag{3.55}$$

Now, for simplicity, we define the partition of $I = \{1, 2, ..., n-1\}$ into three disjoint sets, $I = I_0 \cup I_+ \cup I_-$, where

$$I_0 = \{i \in I : Im(\lambda_i) = 0\}, \quad I_+ = \{i \in I : Im(\lambda_i) > 0\}, \quad I_- = \{i \in I : Im(\lambda_i) < 0\}. \tag{3.56}$$

The terms in the sum of (3.54) with index is in I_0 are still contributing as Bernoulli with parameter p_i. The remaining terms can be coupled pairwise in a single random variable $\mathfrak{C}(p_i)$, with $i \in I_+$, having the following distribution:

$$\begin{aligned} &\mathbb{P}(\mathfrak{C}(p_i) = 2) = p_i\overline{p_i}, \\ &\mathbb{P}(\mathfrak{C}(p_i) = 1) = p_i(1-\overline{p_i}) + \overline{p_i}(1-p_i), \\ &\mathbb{P}(\mathfrak{C}(p_i) = 0) = (1-p_i)(1-\overline{p_i}) \end{aligned} \tag{3.57}$$

(and the right terms are all real values between 0 and 1, thanks to the fact that the eigenvalues have non-negative real part). In conclusion the statement of the latter corollary, without the hypothesis of real spectrum, is

$$|\rho(\Phi_q)| = 1 + \sum_{i \in I_0} \mathfrak{B}\left(\frac{q}{q+\lambda_i}\right) + \sum_{i \in I_+} \mathfrak{C}\left(\frac{q}{q+\lambda_i}\right). \tag{3.58}$$

Moreover, it is possible to compute the mean value and the variance of the random variable $|\rho(\Phi_q)|$. The straightforward way is using the explicit formula (3.55) (or (3.59) in the complex case) and observing that the random variables in the sum are independent. But there is also another more interesting way that only involves the knowledge of the partition function.

Proposition 3.6. *Let $q > 0$. Then*

$$\mathbb{E}[|\rho(\Phi_q)|] = \sum_{i=0}^{n-1} \frac{q}{q+\lambda_i}, \tag{3.59}$$

$$Var(|\rho(\Phi_q)|) = \sum_{i=0}^{n-1} \frac{q}{q+\lambda_i} - \left(\frac{q}{q+\lambda_i}\right)^2. \tag{3.60}$$

Proof. Recall that Φ_q is a random variable taking values in $\mathcal{F}$ and with law ν_q defined in (3.6). Thus if $O(\Phi_q)$ is an observable its mean value is given by

$$\mathbb{E}(O(\Phi_q)) = \sum_{\phi\in\mathcal{F}} O(\phi)\nu_q(\phi) = \frac{\sum_{\phi\in\mathcal{F}} O(\phi)q^{|\rho(\phi)|}w(\phi)}{\sum_{\phi\in\mathcal{F}} q^{|\rho(\phi)|}w(\phi)} \tag{3.61}$$

where the denominator is the partition function $Z(q)$. In particular if we differentiate the logarithm of $Z(q)$ with respect to q we obtain

$$\frac{\partial}{\partial q}\log(Z(q)) = \frac{1}{Z(q)}\frac{\partial}{\partial q}Z(q) = \frac{1}{Z(q)}\sum_{\phi\in\mathcal{F}}\frac{\partial}{\partial q}\left(q^{|\rho(\phi)|}w(\phi)\right) = \\ \frac{1}{Z(q)}\sum_{\phi\in\mathcal{F}}|\rho(\phi)|q^{|\rho(\phi)|-1}w(\phi) = \frac{1}{q}\mathbb{E}[|\rho(\Phi_q)|]. \tag{3.62}$$

Then, using the result of Corollary 3.5 we have

$$\mathbb{E}[|\rho(\Phi_q)|] = q\frac{\partial}{\partial q}\log(Z(q)) = q\frac{\partial}{\partial q}\log\left(\prod_{i=0}^{n-1}(q+\lambda_i)\right) = q\sum_{i=0}^{n-1}\frac{1}{q+\lambda_i}. \tag{3.63}$$

For what concerns the variance, observe that

$$\left(q\frac{\partial}{\partial q}\right)^2\log(Z(q)) = q\frac{\partial}{\partial q}\left[\frac{1}{Z(q)}\sum_{\phi\in\mathcal{F}}|\rho(\phi)|q^{|\rho(\phi)|}w(\phi)\right] \\ = \frac{\sum_{\phi\in\mathcal{F}}|\rho(\phi)|^2q^{|\rho(\phi)|}w(\phi)}{Z(q)} - \left(\frac{\sum_{\phi\in\mathcal{F}}|\rho(\phi)|q^{|\rho(\phi)|}w(\phi)}{Z(q)}\right)^2 \\ = \mathbb{E}[|\rho(\Phi_q)|^2] - \mathbb{E}[|\rho(\Phi_q)|]^2 = \mathrm{Var}(|\rho(\Phi_q)|) \tag{3.64}$$

so the claim follows easily differentiating the right hand side of (3.64) and multiplying by q.

□

As we have seen, Markov chains and weighted graphs are two side of the same coin. Moreover the results of this section show how the study of random spanning forests is strictly related to the spectral analysis of the infinitesimal generator of the random walk associated to the graph.

Example. Consider a finite homogeneous complete graph $\mathcal{K}_N$, where $N \geq 1$ is the number of vertices, with constant edge weights $w > 0$. The Laplacian assumes the

following form:

$$L = \begin{pmatrix} * & w & \cdots & w \\ w & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & w \\ w & \cdots & w & * \end{pmatrix} \tag{3.65}$$

where the element on the diagonal is $* = -(N-1)w$. Our aim is to find the eigenvalues for L.

First we notice that $\gamma_1 = 0$ is an eigenvalue, since the matrix L is singular. Indeed the relative eigenspace has dimension 1 and it is

$$V_1 = \langle\, v_1 \,\rangle, \qquad \text{with} \qquad v_1 = e_1 + e_2 + ... + e_n \tag{3.66}$$

(here $E = \{e_1, ..., e_n\}$ denotes the canonical basis).

In order to find the other eigenvalues, we just need to exploit the symmetry of the problem, and observe that subtracting the value $\lambda_2 = -Nw$ on the diagonal we obtain:

$$L - \lambda_2 I = \begin{pmatrix} w & \cdots & w \\ \vdots & \ddots & \vdots \\ w & \cdots & w \end{pmatrix}. \tag{3.67}$$

This matrix has obviously rank 1, since all its columns (or rows) are identical. The eigenspace relative to λ_2 then is:

$$V_2 = \langle\, v_2, v_3, ..., v_n \,\rangle, \qquad \text{with} \qquad v_k = e_1 - e_k \quad \text{for } k \geq 2. \tag{3.68}$$

Hence the matrix L has the eigenvalue λ_1 with multiplicity 1, and the eigenvalue λ_2 with multiplicity $n-1$. The characteristic polynomial for L is:

$$\chi_L(q) = q(q + Nw)^{n-1} \tag{3.69}$$

As a consequence of Corollary 3.5, the root number distribution for a random spanning forest over $\mathcal{K}_N$, with killing rate $q > 0$, is

$$|\rho(\Phi_q)| \sim 1 + \sum_{i=1}^{n-1} \mathfrak{B}\left(\frac{q}{q + Nw}\right) \tag{3.70}$$

which is the sum of independent and identically distributed Bernoulli random variables.

In general it is a hard task to compute explicitly the eigenvalues of the infinitesimal generator for a specific Markov processes.

3.5 Root process

The aim of this section is to prove that the random set $\rho(\Phi_Q)$ is a determinantal process. We begin with some definition.

Definition 3.7. Let Λ be a locally compact *Polish* space, i.e. a separable and completely metrizable topological space. A *point process* ξ is a random integer-valued positive Radon measure on Λ. When ξ almost surely assigns at most value 1 to the singletons the point process is said to be *simple*.

In other words, a point process can be interpreted as a random counting measure; moreover a simple point process can be identified with a random discrete subset of Λ, and $\xi(D)$ represents the number of points of this random set that fall in $D \subset \Lambda$.

Definition 3.8. The *joint intensities* of a point process ξ with respect to the measure μ are functions (if any exist) $\rho_k : \Lambda^k \to [0, \infty)$, for $k \geq 1$, such that for any family of mutually disjoint subsets $D_1, ..., D_k$ of Λ

$$\mathbb{E}\left[\prod_{i=1}^{k} \xi(D_i)\right] = \int_{\prod_i D_i} \rho_k(x_1, ..., x_k)\, d\mu(x_1) \cdots d\mu(x_k). \tag{3.71}$$

Definition 3.9. A point process ξ on Λ is said to be a *determinantal process* with kernel K if it is simple and its joint intensities satisfy

$$\rho_k(x_1, ..., x_k) = \det_I(K) \tag{3.72}$$

with $I = \{x_1, ..., x_k\}$ for every $k \geq 1$ and $x_1, ..., x_k \in \Lambda$.

In [13] can be found more details about point processes and some important examples of determinantal processes.

We state now the important result claimed in the beginning of this section.
Consider the extended Markov process $\overline{X}$ defined on $\overline{\mathcal{X}} = \mathcal{X} \cup \Delta$ defined in (3.32) by its infinitesimal generator $\mathcal{L}$. Up to time $T_\Delta = \inf\{t \geq 0 : X_t = \Delta\}$, the process $\overline{X}$ can be coupled with the Markov process X in $\mathcal{X}$, with infinitesimal generator L given in (3.2), stopped at rate $q(x)$ in each $x \in \mathcal{X}$. We denote with T_Q this stopping time. In the uniform case, $q(x) \equiv q > 0$, the time T_Q (denoted simply T_q) has exponential distribution of parameter q.
By construction, the point X_{T_Q} is the last node of $\mathcal{X}$ visited by $\overline{X}$ before going directly in Δ.

Theorem 3.7. *The random set $\rho(\Phi_Q)$ is a determinantal process with kernel K_Q defined by*

$$K_Q(x, y) = P_x(X_{T_Q} = y), \qquad x, y \in \mathcal{X}. \tag{3.73}$$

That is, for any $A \subset \mathcal{X}$

$$\mathbb{P}(A \subset \rho(\Phi_Q)) = \det_A(K_Q). \tag{3.74}$$

In order to prove the theorem, it is useful to introduce the trace process. Assume that Y is a generic Markov chain on $\mathcal{Y}$ with generator

$$(\mathcal{L}f)(y) = \sum_{z \in \mathcal{Y}} \alpha(y,z)[f(z) - f(y)]. \tag{3.75}$$

As usual let $\overline{\alpha} = \max_y \alpha(y)$ (assuming it is finite) and construct the discrete process $\hat{Y}$ on $\mathcal{Y}$ with transition matrix $\hat{P}$ and such that

$$\mathcal{L} = \overline{\alpha}(\hat{P} - 1_{\mathcal{Y}}). \tag{3.76}$$

Vice versa, Y is obtained from $\hat{Y}$ updating the position of the process after independent times distributed as exponential of parameter $\overline{\alpha}$.

If we fix $A \subset \mathcal{Y}$ we can build a new Markov process $\hat{Y}^A$, with state space A, which is the trace of the process $\hat{Y}$ on A, i.e. it captures the trajectory of $\hat{Y}$ only inside A ignoring what happens in A^c. More precisely the transition matrix of the process $\hat{Y}^A$ is $\hat{P}^A$ defined by

$$\hat{P}^A(x,y) = P_x\left(\hat{Y}_{\hat{T}_A^+} = y\right), \qquad x, y \in A \tag{3.77}$$

with $\hat{T}_A^+ = \inf\{n \geq 1 : \hat{Y}_n \in A\}$ is the first return time of $\hat{Y}$ in A.

Finally, going back to continuous time, we may define the continuous process Y^A whose dynamics is given by $\hat{Y}^A$ but whose jumps are performed after independent exponential times of parameter $\overline{\alpha}$. In other words it is a process with infinitesimal generator

$$\mathcal{L}^A = \overline{\alpha}(\hat{P}^A - 1_A). \tag{3.78}$$

Notation: if $B, C \subset \mathcal{Y}$, $[\mathcal{L}]_B$ is the restriction of the matrix $\mathcal{L}$ on the set B whereas $[\mathcal{L}]_{B,C}$ is the operator containing only the transition rates from B to C of the matrix $\mathcal{L}$.
With this notation it holds:

Proposition 3.8. *Let Y be a Markov process on $\mathcal{Y}$ with infinitesimal generator $\mathcal{L}$. Fix $A \subset \mathcal{Y}$ and let Y^A be the trace process of Y on A with infinitesimal generator $\mathcal{L}^A$. Then $\mathcal{L}^A$ is the Schur complement of $[\mathcal{L}]_{A^c}$ in $\mathcal{L}$:*

$$\mathcal{L}^A = [\mathcal{L}]_A - [\mathcal{L}]_{A,A^c}[\mathcal{L}]_{A^c}^{-1}[\mathcal{L}]_{A^c,A}. \tag{3.79}$$

Proof. We start computing $\hat{P}^A$. From (3.72), if $\hat{Y}_0 = x$, then the contribution to $\hat{P}^A(x,y)$ comes from two different possible cases:

- the next jump of $\hat{Y}$ is performed towards y with some positive probability;

- in the next jump $\hat{Y}$ leaves A, it continues moving outside A for some time and finally it jumps from A^c to y with positive probability.

Due to this observation it holds

$$\hat{P}^A(x,y) = \hat{P}(x,y) + \sum_{z,z'} \hat{P}(x,z) \left(\sum_{k\geq 0} [\hat{P}]^k_{A^c}(z,z') \right) \hat{P}(z',y). \tag{3.80}$$

The term inside the parenthesis is a geometric series, thus

$$\hat{P}^A(x,y) = [\hat{P}]_A(x,y) + \sum_{z,z'} \hat{P}(x,z) \left(1_{A^c} - [\hat{P}]_{A_c} \right)^{-1} (z,z') \hat{P}(z',y) \tag{3.81}$$

From (3.76), subtracting 1_A on both sides and multiplying by $\overline{\alpha}$, we get

$$\overline{\alpha}(\hat{P}^A - 1_A) = \overline{\alpha}[\hat{P} - 1]_A - \overline{\alpha}[\hat{P} - 1]_{A,A^c} \left(\overline{\alpha}[\hat{P} - 1]_{A_c} \right)^{-1} \overline{\alpha}[\hat{P} - 1]_{A^c,A} \tag{3.82}$$

that is the claim. □

If $B \subset \mathcal{Y}$ is an absorbing set, with $A \subset B^c$, the same computations holds using the sub-Markovian generator $[\mathcal{L}]_{B^c}$.

For $x, y \in A$ the Green's function of Y and Y^A coincide

$$G_B(x,y) = G^A_B(x,y) \tag{3.83}$$

because the mean local time in y starting from x before hitting B is the same for Y and for Y^A.

Before proving Theorem 3.10 we show a lemma that let us express the kernel K_Q in (3.68) in terms of the Green's function.

Lemma 3.9. *It holds*

$$K_Q(x,y) = \begin{cases} G_Q(x,y)q(y) & \text{if } x,y \notin S \\ P_x(X_{T_Q} = X_{T_S} = y) & \text{if } y \in S, x \notin S \\ 1_{\{x=y\}} & \text{if } x \in S \end{cases} \tag{3.84}$$

where $S = \{x \in \mathcal{X} : q(x) = \infty\}$ *and, for* $x, y \notin S$*, the Green's function is*

$$G_Q(x,y) = E_x[l_y(T_Q)] = [Q - L]^{-1}_{\mathcal{X}\setminus S}(x,y) \tag{3.85}$$

the mean local time spent in y *by the process* X *starting in* x *up to the stopping time* T_Q*.*

Proof. The only non-trivial case is when $x, y \notin S$, otherwise the restatement comes directly from the definition of K_Q. Set $\mathcal{Y} = \overline{\mathcal{X}}$ and $Y = \overline{X}$ the extended process, and consider the discrete process $\hat{Y}$ associated to it in the usual way, such that the generator of Y and the transition matrix of $\hat{Y}$ are linked by equation (3.71).
Let $\hat{T}_\Delta$ be the discrete hitting time of the absorbing state Δ. Describing K_Q in terms of the discrete dynamics, it holds

$$\begin{aligned} K_Q(x,y) = P_x(X_{T_Q} = y) &= \sum_{k\geq 1} P_x(\hat{Y}_{k-1} = y, \hat{T}_\Delta = k) \\ &= \sum_{k\geq 1} P_x(\hat{Y}_{k-1} = y, \hat{T}_\Delta \geq k)\hat{P}(y,\Delta). \end{aligned} \tag{3.86}$$

The second factor in the last term is independent of k and its value is $\frac{q(y)}{\overline{\alpha}}$. The remaining term in the sum is exactly $E_x[l_y(\hat{T}_\Delta)]$, the mean local discrete time spent by $\hat{Y}$ in y, i.e. the number of visits in y before the process gets absorbed.
Note that $E_x[l_y(T_Q)] = \frac{E_x[l_y(\hat{T}_\Delta)]}{\overline{\alpha}}$ because the jumps of Y occur after independent exponential times with parameter $\overline{\alpha}$, thus, each time Y visits y, on average it remains there for an exponential time of mean value $\frac{1}{\overline{\alpha}}$.
Then

$$K_Q(x,y) = E_x[l_y(\hat{T}_\Delta)]\frac{q(y)}{\overline{\alpha}} = E_x[l_y(T_Q)]q(y). \tag{3.87}$$

Lastly, observe that

$$E_x[l_y(\hat{T}_\Delta)] = \sum_{k\geq 0}[\hat{P}]^k_{\mathcal{X}\setminus S}(x,y) = ([1-\hat{P}]_{\mathcal{X}\setminus S})^{-1}(x,y). \tag{3.88}$$

Hence dividing both terms by $\overline{\alpha}$ and recalling that $[-\mathcal{L}]_{\mathcal{X}} = Q - L$, we obtain equation (3.81).

□

Proof of Theorem 3.10. We follow the probabilistic proof provided in [10]. The key idea is to construct two different absorbing states for the Wilson's algorithm. Indeed the algorithm works also in the case of more than one absorbing state, producing a spanning oriented forest instead of a spanning oriented tree.

For simplicity assume $S = \emptyset$.
Fix $A \subset \mathcal{X}$ and let Δ_A, Δ_{A^c} be different absorbing states accessible respectively only from A or A^c. More precisely, if $q(x)$ are the killing rates for the process X, we set $\mathcal{Y} = \mathcal{X} \cup \{\Delta_A, \Delta_{A^c}\}$ and define the Markov process Y via its generator

$$(\mathcal{L}f)(x) = \begin{cases} (Lf)(x) + q(x)[f(\Delta_A)1_{\{x\in A\}} + f(\Delta_{A^c})1_{\{x\in A^c\}} - f(x)], & x \in \mathcal{X} \\ 0, & x \in \{\Delta_A, \Delta_{A^c}\} \end{cases} \tag{3.89}$$

similarly to what we did in (3.32), with L being the infinitesimal generator of X. Also set $A^* = A \cup \{\Delta_A, \Delta_{A^c}\}$ and consider the process Y^{A^*} that traces Y inside A^*.

It is possible to run the Wilson's algorithm on the two different spaces:

- $\overline{\mathcal{X}} = \mathcal{X} \cup \Delta$, with absorbing state Δ and generator as in (3.32). In this case we denote the random outcome of the algorithm with $\mathcal{W}$;
- A^*, with absorbing states Δ_A, Δ_{A^c} and generator $\mathcal{L}^{A^*}$ (that is the generator of the trace process). In this case we denote the random outcome of the algorithm with $\mathcal{W}^*$.

Now define

$$D = \{y \in \overline{\mathcal{X}} : (x, y) \in \mathcal{W}, x \in A\} \tag{3.90}$$

$$D^* = \{y \in A^* : (x, y) \in \mathcal{W}^*, x \in A\} \tag{3.91}$$

that are the set of ending points of the edges starting from A after running the Wilson's algorithm. Observe that by construction

$$\mathbb{P}(A \subset \rho(\Phi_Q)) = \mathbb{P}(D = \{\Delta\}) = \mathbb{P}(D^* = \{\Delta_A\}), \tag{3.92}$$

because if the elements of A are roots of the random spanning forest, then all the edges starting from A need to end inside the absorbing state Δ, in the first case, or Δ_A in the second one.

Now, the last term of (3.88) is easy to compute using Proposition 3.2. Indeed the probability of the event $D^* = \{\Delta_A\}$ is the probability that for the trace process all the paths of length 1 from A to Δ_A occur (and all these events are independent from each other). Then

$$\mathbb{P}(D^* = \{\Delta_A\}) = \frac{\prod_{a \in A} q(a)}{\det_A(-\mathcal{L}^{A^*})} = \frac{\det_A(Q)}{\det_A(-\mathcal{L}^{A^*})} = \det_A(Q) \det\left(\left(\left[-\mathcal{L}^{A^*}\right]_A\right)^{-1}\right) \tag{3.93}$$

where the last equality follows recalling that $\det_A(M) = \det([M]_A)$, and from the rule for the determinant of the inverse matrix.

Now denote by $G^{A^*}_{\{\Delta_A, \Delta_{A^c}\}}$ the Greeen's function of the process Y^{A^*} stopped in the absorbing states Δ_A and Δ_{A^c}. Recall that $G^{A^*}_{\{\Delta_A, \Delta_{A^c}\}}(x, y)$, for $x, y \in A$, is the local time spent in y by Y^{A^*}, starting from x and before being absorbed. Then $G^{A^*}_{\{\Delta_A, \Delta_{A^c}\}} = \left(\left[-\mathcal{L}^{A^*}\right]_A\right)^{-1}$, the matrix appearing in the latter equation above. Moreover observe that by construction, for any $x, y \in A$,

$$G^{A^*}_{\{\Delta_A, \Delta_{A^c}\}}(x, y) = G^{\mathcal{Y}}_{\{\Delta_A, \Delta_{A^c}\}}(x, y) = G^{\overline{\mathcal{X}}}_{\{\Delta\}}(x, y) \tag{3.94}$$

and note that $G^{\overline{\mathcal{X}}}_{\{\Delta\}}(x,y) = (Q-L)^{-1}(x,y)$ for $x, y \in \mathcal{X}$. Hence, summing up equations (3.88), (3.89) and using (3.81), we obtain

$$\mathbb{P}(A \subset \rho(\Phi_Q)) = \det_A((Q-L)^{-1}) \det_A(Q). \tag{3.95}$$

Finally, since Q is a diagonal matrix, it holds

$$\det_A((Q-L)^{-1}) \det_A(Q) = \det_A((Q-L)^{-1}Q)$$

and by the application of Proposition 3.11 we obtain $\mathbb{P}(A \subset \rho(\Phi_Q)) = \det_A(K_Q)$.

□

Chapter 4

Analysis of loop-erased partitions

4.1 Interaction potential

Given a graph $\mathcal{G} = (\mathcal{X}, \mathcal{E}, w)$ any spanning forest $\phi \in \mathcal{F}$ induces a partition of $\mathcal{X}$ into disjoint sets, each one containing the vertices of the different trees in the forest. We call this partition $\Pi(\phi)$.
As seen in section 3.4, it is possible to sample a random spanning forest using the Wilson's Algorithm. This fact suggests the following definition.

Definition 4.1. Let $q > 0$. A *loop-erased partition* of $\mathcal{X}$ is a random variable Π_q with law

$$\mathbb{P}(\Pi_q = \Pi_m) = \mu_q(\Pi_m) = \frac{q^m \sum_{\phi:\Pi(\phi)=\Pi_m} w(\phi)}{Z(q)} \tag{4.1}$$

where Π_m is a partition of $\mathcal{X}$, $w(\phi)$ and $Z(q)$ are defined in section 3.1.

Remark 5. Comparing (3.6) and (4.1) we observe that

$$\mu_q(\Pi_m) = \sum_{\phi:\Pi(\phi)=\Pi_m} \nu_q(\phi), \tag{4.2}$$

that is the probability to choose the partition Π_m is equal to the probability that the Wilson's algorithm produces a forest ϕ which induces such partition.

From the latter observation it is clear that the measure μ_q can be sampled using the Wilson's Algorithm. According to definition 3.6, in order to use the algorithm it should be necessary to introduce the absorbed process $\overline{X}$ with infinitesimal generator (3.32) on the extended state space $\overline{\mathcal{X}} = \mathcal{X} \cup \{\Delta\}$.
However in the case of uniform killing rate $q > 0$, recalling remark 3.3, the absorption time of the process occurs after a time τ_q distributed as an exponential random variable of parameter q. Then, if we consider the loop-erased random walk (LERW) obtained running the process X on $\mathcal{X}$ and erasing the loops as soon as they appear, it is possible to describe the algorithm in the following way:

1. run a LERW starting from any vertex in $\mathcal{X}$ up to a time τ_q; call γ_1 this self-avoiding path

2. run a LERW starting from any vertex in $\mathcal{X} \setminus \gamma_1$ until $\min\{\tau_q, \tau_{\gamma_1}\}$, where τ_{γ_1} is the hitting time for γ_1; call γ_2 the union of γ_1 with the new self-avoiding path

3. iterate step 2 with γ_{i+1} in place of γ_i until all vertex of $\mathcal{X}$ are contained in γ_{i+1}.

As observed in [11] by L. Avena, P. Milanesi, M. Quattropani, the algorithm shows the tendency of the partition to *cluster* in the same tree the points that with high probability can be visited by the random walk X on the time scale τ_q: this makes the study of *metastable-like regions* interesting.

The clustering analysis can be performed by studying two-points correlations associated to μ_q.

Definition 4.2. Let $q > 0$ and $\mathcal{G} = (\mathcal{X}, \mathcal{E}, w)$ a weighted graph. Fix $x, y \in \mathcal{X}$. We call *pairwise interaction potential* the probability

$$U_q(x, y) := \mathbb{P}(x \text{ and } y \text{ are in different blocks of } \Pi_q). \tag{4.3}$$

Since the starting points of the algorithm can be chosen arbitrarily, we may describe explicitly the quantity in (4.3)

$$U_q(x, y) = \sum_{\gamma} P_x(\Gamma_\Delta = \gamma) P_y(\tau_{\gamma \setminus \{\Delta\}} > \tau_{\{\Delta\}}) \tag{4.4}$$

where $\tau_{\{\Delta\}} = \inf\{t \geq 0 : \overline{X}_t = \Delta\}$, $\tau_{\gamma\setminus\{\Delta\}} = \inf\{t \geq 0 : \overline{X}_t \in \gamma \setminus \{\Delta\}\}$ and the sum runs over all possible loop-erased paths γ starting at x.

Recall from the previous chapter that P_x is the law of the random walk Y on $\mathcal{Y}$ starting from x. If Γ_B denotes a random loop-erased trajectory obtained from Y and stopped at $B \subset \mathcal{Y}$, we saw in proposition 3.1 how to compute the probability $P_x(\Gamma_B = \gamma)$. In order to apply this result to the Wilson's algorithm it was sufficient to apply it using $Y = \overline{X}$, $\mathcal{Y} = \overline{\mathcal{X}}$ and appropriately changing the target set B at each iteration of the algorithm.

The pairwise interaction potential strongly depends on the geometry of the graph and thus it is difficult to characterize. In the next sections, we present some results achieved in [11] in simple geometries.

4.2 Mean-field model

The easiest geometry for a graph is the mean-field model. This model involves a complete graph denoted by $\mathcal{K}_N$, $N \geq 1$ being the number of vertices. The set of vertices of $\mathcal{K}_N$ is $[N] = \{1, 2, ..., N\}$ and $\mathcal{E}$, the set of edges, contains all possible pairs of elements in $[N]$. The term *mean field* indicates that each edge is associated to a constant weight $w > 0$.

Theorem 4.1. *Fix $N \geq 1$ and $q > 0$. Consider the graph $\mathcal{K}_N$ with constant edge weight $w > 0$. Then for any $x \neq y \in [N]$*

$$U_q^{(N)}(x,y) \equiv U_q^{(N)} = \sum_{n\geq 1} \frac{q}{q+Nw} \left(\frac{Nw}{q+Nw}\right)^{n-1} \prod_{k=2}^{n} \left(1 - \frac{k}{N}\right). \tag{4.5}$$

Proof. Consider the absorbed process $\overline{X}$ on the state space $\overline{\mathcal{X}} = [N] \cup \{\Delta\}$ associated to the weighted graph. In particular the infinitesimal generator of $\overline{X}$ is

$$\mathcal{L} = (\alpha(x,y))_{x,y\in\overline{\mathcal{X}}} = \left(\begin{array}{ccc|c} & L - qI & & \begin{matrix} q \\ \vdots \\ q \end{matrix} \\ \hline 0 & \cdots & 0 & 0 \end{array}\right), \qquad \text{where} \qquad L = \begin{pmatrix} * & w & \cdots & w \\ w & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & w \\ w & \cdots & w & * \end{pmatrix} \tag{4.6}$$

is an $N \times N$ matrix with diagonal entries $* = -(N-1)w$.

We use the representation of the interaction potential given in (4.4). Fix a loop-erased trajectory $\gamma = \{x_0, x_1, ..., x_{l-1}, x_l\}$ of length $1 \leq l \leq N$, with $x_0 = x, x_l = \Delta$. Applying proposition 3.1 one gets

$$P_x(\Gamma_{\{\Delta\}} = \gamma) = \prod_{i=0}^{l-1} \alpha(x_i, x_{i+1}) \frac{\det_{\overline{\mathcal{X}}\setminus\{x_0,...,x_{l-1},\Delta\}}(-\mathcal{L})}{\det_{\overline{\mathcal{X}}\setminus\{\Delta\}}(-\mathcal{L})} = qw^{l-1} \frac{\det_A(qI - L)}{\det(qI - L)} \tag{4.7}$$

where $A := [N] \setminus \{x_0, ..., x_{l-1}\}$. The terms in the fraction are equal to the characteristic polynomials of the matrices $[L]_A$ and L respectively. Applying the same reasoning of the last example of section 3.4, the eigenvalues of $[L]_A$ and L are

$$Spec([L]_A) = \{-lw, \underbrace{-Nw, ..., -Nw}_{N-l-1}\}, \qquad Spec(L) = \{0, \underbrace{-Nw, ..., -Nw}_{N-1}\}. \tag{4.8}$$

Then

$$P_x(\Gamma_{\{\Delta\}} = \gamma) = qw^{l-1} \frac{(q+lw)(q+Nw)^{N-l-1}}{q(q+Nw)^{N-1}} = w^{l-1} \frac{q+lw}{(q+Nw)^l}. \tag{4.9}$$

Next consider the stopping times $\tau_{\gamma\setminus\{\Delta\}}$ and $\tau_{\{\Delta\}}$ for the random walk with law given by P_y. The transition rates for the process $\overline{X}$ are time-independent until it reaches the absorbing state: this means that, starting from y, there are l Poisson processes with rate w (one for each element in $\gamma \setminus \{\Delta\}$) and a Poisson process with rate q competing with each other.
Then

$$P_y(\tau_{\{\Delta\}} < \tau_{\gamma\setminus\{\Delta\}}) = \begin{cases} \frac{q}{q+lw} & \text{if } y \notin \gamma \\ 0 & \text{if } y \in \gamma \end{cases} \tag{4.10}$$

Since the terms in (4.9) and in (4.10) only depend on the length l of the loop-erased path, to conclude the proof it remains to count how many paths of length l do not contain y.
The case $l = 1$ is trivial, since γ is the path that contains just x and Δ.
Also $l = N$ is trivial, because in that case all the elements of $[N]$ must be in γ, so $y \in \gamma$.
If $2 \leq l \leq N-1$, the paths such that $y \notin \gamma$ are obtained by choosing neatly the points $x_1, x_2, ..., x_{l-1}$ other than y. In other words, it is necessary to select neatly $l-1$ elements from $N-2$ different possible choices: then the number of paths is $\frac{(N-2)!}{(N-l-1)!} = (N-2)(N-3)\cdots(N-l)$.

Summing up we obtain

$$\begin{aligned} U_q^{(N)}(x,y) &= \sum_{\gamma} P_x(\Gamma_\Delta = \gamma) P_y(\tau_{\gamma\setminus\{\Delta\}} > \tau_{\{\Delta\}}) \\ &= \sum_{l=1}^{N-1} w^{l-1} \frac{q+lw}{(q+Nw)^l} \frac{q}{q+lw} \prod_{i=2}^{l} (N-i) \\ &= \sum_{l=1}^{N-1} \frac{q}{q+Nw} \left(\frac{w}{q+Nw}\right)^{l-1} N^{l-1} \prod_{i=2}^{l} \left(1 - \frac{i}{N}\right) \\ &= \sum_{l=1}^{N-1} \frac{q}{q+Nw} \left(\frac{Nw}{q+Nw}\right)^{l-1} \prod_{i=2}^{l} \left(1 - \frac{i}{N}\right) \end{aligned} \tag{4.11}$$

with the convention that the product term when $l = 1$ is equal to 1.

□

Note that result (4.5) has an interesting interpretation. Consider a random variable T_q with geometric distribution of parameter $\tilde{q} = \frac{q}{q+Nw}$ so that

$$\mathbb{P}(T_q = k) = \tilde{q}(1-\tilde{q})^{k-1} \tag{4.12}$$

If we set

$$f(k,N) := \prod_{i=2}^{k} \left(1 - \frac{i}{N}\right) \tag{4.13}$$

then the interaction potential can be rewritten as

$$\begin{aligned} U_q^{(N)} &= \sum_{k=1}^{N-1} \frac{q}{q+Nw} \left(\frac{Nw}{q+Nw}\right)^{k-1} \prod_{i=2}^{k} \left(1 - \frac{i}{N}\right) \\ &= \sum_{k=1}^{N-1} \tilde{q}(1-\tilde{q})^{k-1} f(k,N) \\ &= \mathbb{E}\left[f(T_q, N)\right], \end{aligned} \tag{4.14}$$

where the last equality follows by observing that $f(k,N)=0$ for all $k \geq N$.

Remark 6. In the mean-field hypothesis, $U_q^{(N)}(x,y)$ does not depend on the choice of the pair (x,y). This is a consequence of the total symmetry of the system.

In the large N limit, for the mean-field model, it is also possible to determine the law for the interaction potential, when the parameter q scales with the square root of N.

Proposition 4.2. *In the hypothesis of Theorem 1, if $q = z\sqrt{N}w$, for fixed $z > 0$, then*

$$U_q := \lim_{N\to\infty} U_q^{(N)} = \sqrt{2\pi} z e^{\frac{z^2}{2}} \mathbb{P}(Z > z) \tag{4.15}$$

where Z is a standard Gaussian random variable.

Proof. Following the observation above, we consider the interaction potential $U_q^{(N)}$ as the mean value of $f(T_q, N)$ where T_q is a geometric random variable of parameter $\tilde{q} = \frac{q}{q+Nw}$.
We start approximating $f(k,N)$. Using the Taylor series for the logarithm we get

$$\begin{aligned}
\ln f(k,N) &= \ln \prod_{i=2}^{k} \left(1 - \frac{i}{N}\right) = \sum_{i=2}^{k} \ln\left(1 - \frac{i}{N}\right) = -\sum_{i=2}^{k} \frac{i}{N} + O\left(\frac{i^2}{N^2}\right) \\
&= -\frac{1}{N}\left[\frac{k(k+1)}{2} - 1\right] + kO\left(\frac{k^2}{N^2}\right) \\
&= -\frac{1}{N}\frac{k^2+k-2}{2} + O\left(\frac{k^3}{N^2}\right) \\
&= -\frac{k^2}{2N} + O\left(\frac{k}{N} + \frac{k^3}{N^2}\right) =: -\frac{k^2}{2N} + c_N(k).
\end{aligned} \tag{4.16}$$

Then

$$U_q^{(N)} = \mathbb{E}\left[f(T_q, N)\right] = \mathbb{E}\left[e^{-\frac{T_q^2}{2N} + c_N(T_q)}\right] \tag{4.17}$$

By hypothesis $q = z\sqrt{N}w$. Since $T_q \sim Geo(\tilde{q})$ with $\tilde{q} = \frac{z\sqrt{N}w}{z\sqrt{N}w+Nw} \xrightarrow[N\to\infty]{} \frac{z}{\sqrt{N}}$, then $\frac{T_q}{\sqrt{N}}$ converges in distribution to $Y \sim Exp(z)$, an exponential random variable of parameter z. In particular it holds

$$\lim_{N\to\infty} \left| \mathbb{E}\left[e^{-\frac{T_q^2}{2N}}\right] - \mathbb{E}\left[e^{-\frac{Y^2}{2}}\right] \right| = 0 \tag{4.18}$$

Moreover set $Z \sim \mathcal{N}(0,1)$ a random variable distributed as a standard Gaussian. Using a simple substitution inside the integral, we have

$$\begin{aligned}\mathbb{E}\Big[e^{-\frac{Y^2}{2}}\Big] &= \int_0^\infty e^{-\frac{y^2}{2}} z e^{-zy}\, dy \\ &= z e^{\frac{z^2}{2}} \int_0^\infty e^{-\frac{1}{2}(y+z)^2}\, dy \\ &= z e^{\frac{z^2}{2}} \int_z^\infty e^{-\frac{t^2}{2}}\, dt \\ &= \sqrt{2\pi} z e^{\frac{z^2}{2}} \mathbb{P}(Z > z).\end{aligned} \tag{4.19}$$

Summing up we obtain

$$\begin{aligned}\Big|U_q^{(N)} - \sqrt{2\pi} z e^{\frac{z^2}{2}} \mathbb{P}(Z>z)\Big| &= \left|\mathbb{E}\Big[e^{-\frac{T_q^2}{2N}+c_N(T_q)}\Big] - \mathbb{E}\Big[e^{-\frac{Y^2}{2}}\Big]\right| \\ &\le \left|\mathbb{E}\Big[e^{-\frac{T_q^2}{2N}+c_N(T_q)}\Big] - \mathbb{E}\Big[e^{-\frac{T_q^2}{2N}}\Big]\right| + o(1) \\ &\le \left|\mathbb{E}\Big[e^{-\frac{T_q^2}{2N}+c_N(T_q)}\Big] - \sum_{k=0}^{M} \mathbb{P}(T_q=k) e^{-\frac{k^2}{2N}+c_N(k)}\right| \\ &\quad + \left|\sum_{k=0}^{M} \mathbb{P}(T_q=k) e^{-\frac{k^2}{2N}+c_N(k)} - \mathbb{E}\Big[e^{-\frac{T_q^2}{2N}}\Big]\right| + o(1) \\ &\le \sum_{k=M+1}^{\infty} \mathbb{P}(T_q=k) + \sum_{k=0}^{M} \mathbb{P}(T_q=k)\Big[e^{-\frac{k^2}{2N}+c_N(k)} - e^{-\frac{k^2}{2N}}\Big] + o(1)\end{aligned}$$

with arbitrary $M \in \mathbb{N}$. In particular, by choosing $M = \lfloor N^\delta \rfloor$ with $\delta \in (\frac{1}{2}, \frac{2}{3})$, both sums are $o(1)$, thus the proof is complete.

□

4.3 Mean-field-community model

Next we consider a slightly more complicated geometry with non-homogeneous edge weights. Let $\mathcal{K}_{2N}$ be a complete graph, where the set of vertices is $[2N] = \{1, 2, ..., 2N\}$ and $\mathcal{E}$ is the set of edges. Consider the weight function $w : \mathcal{E} \to [0, +\infty)$ defined as

$$w(e) = \begin{cases} w_1 & \text{if } e_-, e_+ \in [N] \text{ or } e_-, e_+ \in [2N] \setminus [N] \\ w_2 & \text{otherwise} \end{cases} \tag{4.20}$$

where w_1, w_2 are positive real numbers. This weight function produces a community structure (of the two communities $[N]$ and $[2N] \setminus [N]$).

We aim to compute the interaction potential $U_q^{(N)}(x,y)$, but thanks to the symmetry of the system, such value will turn out to be invariant for any choice of elements x, y belonging to the same community and for any choice of elements x, y that are in different communities. Therefore we introduce the following notation.

- We call "*out*" the case in which x and y belong to different communities. In such situation we denote the interaction potential as $U_q^{(N)}(out)$.

- Instead we call "*in*" the case in which x and y are inside the same community. In this circumstance we denote the interaction potential as $U_q^{(N)}(in)$.

Theorem 4.3. *Fix $N \geq 1$ and $q > 0$. Consider the graph $\mathcal{K}_{2N}$ with edge weight as in (4.21). Let T_q be a geometric random variable with success parameter*

$$\alpha := \frac{q}{q + N(w_1 + w_2)}. \tag{4.21}$$

Let $\tilde{X}$ be a discrete Markov chain with state space $\{1,2\}$ and transition matrix

$$\tilde{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}, \qquad p = \frac{w_1}{w_1 + w_2}; \tag{4.22}$$

denote by $\ell(t) := \sum_{s<t} 1_{\{\tilde{X}_s = 1\}}$ the local time in state 1 spent by $\tilde{X}$ up to time t, and by $\tilde{\mathbb{P}}_1$ the path measure of $\tilde{X}$ starting from state 1.
Fix $x, y \in [2N]$. Then

$$U_q^{(N)}(out) = \sum_{n\geq 1} \mathbb{P}(T_q = n) \sum_{k=1}^{n} \tilde{\mathbb{P}}_1(\ell(n) = k)\ \beta_{out}(n,k) \prod_{i=1}^{k-1}\left(1 - \frac{i}{N}\right) \prod_{j=1}^{n-k}\left(1 - \frac{j}{N}\right), \tag{4.23}$$

$$U_q^{(N)}(in) = \sum_{n\geq 1} \mathbb{P}(T_q = n) \sum_{k=1}^{n} \tilde{\mathbb{P}}_1(\ell(n) = k)\ \beta_{in}(n,k) \prod_{i=1}^{k}\left(1 - \frac{i}{N}\right) \prod_{j=1}^{n-k-1}\left(1 - \frac{j}{N}\right) \tag{4.24}$$

where

$$\beta_\star(n,k) := \begin{cases} \frac{q+2Nw_2+(w_1-w_2)k}{q+2Nw_2} & \text{if } \star = out, \\ \frac{q+2Nw_2+(w_1-w_2)(n-k)}{q+2Nw_2} & \text{if } \star = in \end{cases} \tag{4.25}$$

Proof. For simplicity we denote the communities

$$V_1 = [N], \qquad , V_2 = [2N] \setminus [N], \qquad V = V_1 \cup V_2,$$

with $|V_1| = |V_2| = N$. We split the proof in three parts.

1. The loop-erased trajectory starting from x.
Consider the absorbed process $\overline{X}$ on $\overline{\mathcal{X}} = V \cup \{\Delta\}$ associated to the weighted graph, i.e. with infinitesimal generator

$$\mathcal{L} = (\alpha(x,y))_{x,y\in\overline{\mathcal{X}}} = \left(\begin{array}{ccc|c} & & & q \\ & L - qI & & \vdots \\ & & & q \\ \hline 0 & \cdots & 0 & 0 \end{array}\right) \tag{4.26}$$

where the matrix

$$L = \left(\begin{array}{ccc|ccc} d & \cdots & w_1 & w_2 & \cdots & w_2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w_1 & \cdots & d & w_2 & \cdots & w_2 \\ \hline w_2 & \cdots & w_2 & d & \cdots & w_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w_2 & \cdots & w_2 & w_1 & \cdots & d \end{array}\right), \tag{4.27}$$

with $d = -(N-1)w_1 - Nw_2$, is such that each block has size $N \times N$.

Now fix a self-avoiding trajectory γ of length k in $\overline{\mathcal{X}}$, starting from x and ending in Δ. Without loss of generality, thanks to the symmetry of the problem, we may assume $x \in V_1$. We set

$$\gamma_1 = \gamma \cup V_1, \qquad \gamma_2 = \gamma \cup V_2, \tag{4.28}$$

and clearly $\gamma = \gamma_1 \cup \gamma_2 \cup \{\Delta\}$. Applying proposition 3.1 one gets

$$P_x(\Gamma_{\{\Delta\}} = \gamma) = \prod_{i=0}^{l-1} \alpha(x_i, x_{i+1}) \frac{\det_{\overline{\mathcal{X}}\setminus\gamma}(-\mathcal{L})}{\det_V(-\mathcal{L})} = \prod_{i=0}^{l-1} \alpha(x_i, x_{i+1}) \frac{\det_{V\setminus\gamma}(qI - L)}{\det(qI - L)} \tag{4.29}$$

The nodes in the trajectory γ can be picked both in V_1 and in V_2. To determine exactly what are the coefficients $\alpha(x_i, x_{i+1})$, it is necessary to exploit the symmetry of the problem. Indeed we need to know:

- the number of nodes in the trajectory that are inside V_1 and inside V_2, i.e. the numbers $k_1 = |\lambda_1|$ and $k_2 = |\lambda_2|$;
- the number of transition from V_1 to V_2, namely j_1, and from V_2 to V_1, namely j_2 performed by the trajectory γ.

Notice that j_1 and j_2 must be less or equal to $\min\{k_1, k_2\}$, because the number of jumps from one community to the other cannot be larger than the number of nodes

contained by each community. Moreover j_1 and j_2 must differ, at most, by 1, since there cannot be two consecutive jumps from V_1 to V_2, or viceversa. Also the hypothesis $x \in V_1$ ensures that $j_1 \geq j_2$: thus $j_2 = j_1 - 1$ or $j_2 = j_1$.
Then:

- $j_1 + j_2$ of the α's in (4.25) have value w_2 (they are jumps between different communities);
- $k_1 + k_2 - (j_1 + j_2) - 1$ of the α's have value w_1 (they are jumps inside the same community);
- 1 of the α's has value q (the last jump to the absorption node).

Hence

$$P_x(\Gamma_{\{\Delta\}} = \gamma) = w_1^{k_1+k_2-j_1-j_2-1} w_2^{j_1+j_2} q \frac{\det_{V\setminus\gamma}(qI - L)}{\det(qI - L)} \tag{4.30}$$

Observe that

$$\det_{V\setminus\gamma}(qI - L) = \det_{V\setminus\{1,\dots,k_1,N+1,\dots,N+k_2\}}(qI - L)$$

thanks to the structure of the matrix L.
Both term in numerator and denominator are the characteristic polynomials of the matrix $\hat{L} := [L]_{V\setminus\gamma}$ and L.
The matrix $\hat{L}$ is

$$\hat{L} = \left(\begin{array}{ccc|ccc} d & \cdots & w_1 & w_2 & \cdots & w_2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w_1 & \cdots & d & w_2 & \cdots & w_2 \\ \hline w_2 & \cdots & w_2 & d & \cdots & w_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ w_2 & \cdots & w_2 & w_1 & \cdots & d \end{array}\right) = \begin{pmatrix} D_1 & B \\ B^T & D_2 \end{pmatrix} \tag{4.31}$$

where $d = -(N-1)w_1 - Nw_2$, D_1 is a square block of size $N - k_1$ and D_2 is a square block of size $N - k_2$.
Now notice that:

- adding a diagonal matrix of value $N(w_1 + w_2)$ to L we obtain a matrix of rank 2. Thus $\lambda_0 = -N(w_1 + w_2)$ is eigenvalue of L with multiplicity $2N - 2$;
- similarly, adding a diagonal matrix of value $N(w_1 + w_2)$ to $\hat{L}$ we obtain a matrix of rank 2. Thus $\hat{\lambda_0} = -N(w_1 + w_2)$ is eigenvalue of $\hat{L}$ with multiplicity $2N - k_1 - k_2 - 2$.

The remaining two eigenvalues, for both L and $\hat{L}$, can be computed in the following way. Assume that L has eigenvectors of the form $v = (x_1, ..., x_1, x_2, ..., x_2)^T$, where the upper component and the lower one have length N.
Analogously assume that $\hat{L}$ has eigenvectors of the form $\hat{v} = (x_1, ..., x_1, x_2, ..., x_2)^T$, where the upper component has length $N - k_1$ and the lower one has length $N - k_2$.
Then the problems $Lv = \lambda v$ and $\hat{L}\hat{v} = \hat{\lambda}\hat{v}$ become:

$$Lv = \lambda v \quad \longrightarrow \quad \begin{pmatrix} d + (N-1)w_1 & Nw_2 \\ Nw_2 & d + (N-1)w_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{4.32}$$

$$\hat{L}\hat{v} = \hat{\lambda}\hat{v} \quad \longrightarrow \quad \begin{pmatrix} d + (N-k_1-1)w_1 & (N-k_2)w_2 \\ (N-k_1)w_2 & d + (N-k_2-1)w_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \hat{\lambda} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{4.33}$$

By direct computations the eigenvalues of the problem (4.33) are $\lambda_1 = 0$ and $\lambda_2 = -Nw_2$, whereas the eigenvalues of the problem (4.34) (which depends on the values k_1 and k_2) are

$$\hat{\lambda}_i(k_1, k_2) = -\frac{1}{2}\left[w_1(k_1 + k_2) + 2Nw_2 + (-1)^i\sqrt{w_1^2(k_1 - k_2)^2 + 4(N - k_1)(N - k_2)w_2^2}\right] \tag{4.34}$$

with $i = 1, 2$.

Hence

$$\det_{V\setminus\gamma}(qI - L) = (q - \hat{\lambda}_0)^{2N-k_1-k_2-2}(q - \hat{\lambda}_1(k_1, k_2))(q - \hat{\lambda}_2(k_1, k_2)) \tag{4.35}$$

and

$$\det(qI - L) = (q - \lambda_0)^{2N-2}(q - \lambda_1)(q - \lambda_2). \tag{4.36}$$

Therefore we obtain

$$P_x(\Gamma_{\{\Delta\}} = \gamma) = w_1^{k_1+k_2-j_1-j_2-1}w_2^{j_1+j_2}q\ \theta(k_1, k_2) \tag{4.37}$$

with

$$\theta(k_1, k_2) := \frac{(q - \hat{\lambda}_1(k_1, k_2))(q - \hat{\lambda}_2(k_1, k_2))}{q(q + 2Nw_2)(q + N(w_1 + w_2))^{k_1+k_1}} \tag{4.38}$$

We conclude this first part of the proof by observing again that the quantity $P_x(\Gamma_{\{\Delta\}} = \gamma)$ only depends on the variables k_1, k_2, j_1, j_2.

2. The absorbed trajectory starting from y.
Let γ be a self-avoiding trajectory starting from x and ending in Δ and fix y either in V_1 or in V_2. Now we want to characterize the quantity $P_y(\tau_{\{\Delta\}} < \tau_{\gamma\setminus\{\Delta\}})$.

First observe that we can reduce the problem in the following way. Instead of using the Markov process$\overline{X}$ we define a new Markov process $\overline{Y}$, with values in the state space $\{1,2,3,4\}$ corresponding to the sets $\{V_1\setminus\gamma_1, V_2\setminus\gamma_2, \gamma_1\cup\gamma_2, \Delta\}$, whose transition matrix is given by

$$\overline{P} = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \tag{4.39}$$

where

$$Q := \frac{1}{d^*}\begin{pmatrix} (N-k_1-1)w_1 & (N-k_2)w_2 \\ (N-k_1)w_2 & (N-k_2-1)w_1 \end{pmatrix}, \qquad R := \frac{1}{d^*}\begin{pmatrix} k_1w_1+k_2w_2 & q \\ k_2w_1+k_1w_2 & q \end{pmatrix},$$

with $d^* = (N-1)w_1 + Nw_2 + q$ the normalization constant for the first two rows of $\overline{P}$. In this new setting, calling T_{abs} the hitting time of the absorbing set $\{3,4\}$, we have that

$$P_y(\tau_{\{\Delta\}} < \tau_{\gamma\setminus\{\Delta\}}) = \begin{cases} \mathbb{P}_1(\overline{Y}(T_{abs}) = 4) & \text{if } y \in V_1 \\ \mathbb{P}_2(\overline{Y}(T_{abs}) = 4) & \text{if } y \in V_2 \end{cases} \tag{4.40}$$

The quantities on the right side of (4.35) can be computed directly:

$$\begin{aligned} \mathbb{P}_1(\overline{Y}(T_{abs}) = 4) &= \sum_{k=0}^{\infty} \overline{P}^k(1,1)\frac{q}{d^*} + \sum_{k=0}^{\infty} \overline{P}^k(1,2)\frac{q}{d^*} \\ &= \frac{1}{d^*}\left(\sum_{k=0}^{\infty} Q^k\right)\begin{pmatrix} q \\ q \end{pmatrix}(1) \\ &= \frac{1}{d^*}(I-Q)^{-1}\begin{pmatrix} q \\ q \end{pmatrix}(1) \end{aligned} \tag{4.41}$$

and analogously for the other case. In particular, defining the vector

$$v^* := \frac{1}{d^*}(I-Q)^{-1}\begin{pmatrix} q \\ q \end{pmatrix} \tag{4.42}$$

we have that

$$\mathbb{P}_1(\overline{Y}(T_{abs}) = 4) = v^*(1), \qquad \mathbb{P}_2(\overline{Y}(T_{abs}) = 4) = v^*(2). \tag{4.43}$$

By direct computation the vector v^* is

$$v^* = \frac{q}{c}\begin{pmatrix} q + k_2(w_1-w_2) + 2w_2N \\ q + k_1(w_1-w_2) + 2w_2N \end{pmatrix} \tag{4.44}$$

with

$$c := (q+k_1w_1)(q+k_2w_1) + Nw_2(2q+kw_1) + w_2^2(Nk - k_1k_2). \tag{4.45}$$

Lastly observe that, assuming as done before $x \in V_1$, the case $y \in V_1$ corresponds to the *intra-community* case, whereas $y \in V_2$ is the *inter-community* case. This let us distinguish two possible scenarios for the interaction potential: $U_q^{(N)}(in)$ if x, y are in the same community, or $U_q^{(N)}(out)$ if x, y belong to different communities.

3. Counting the number of paths γ.
As we have seen, for γ fixed, the quantities $P_x(\Gamma_{\{\Delta\}} = \gamma)$ and $P_y(\tau_{\{\Delta\}} < \tau_{\gamma\setminus\{\Delta\}})$ only depend on the variables k_1, k_2, j_1, j_2. The question now is: how many paths share the same k's and j's values?

Fix the 4 parameters k_1, k_2, j_1, j_2. Again, assume $x \in V_1$ and let us distinguish the case $y \in V_1$ or $y \in V_2$.
We focus on the case $y \in V_2$. In order to select a path with such parameters we have to:

- Choose a neat sequence of length k_1 (without repetitions) of elements inside V_1, corresponding to the nodes in γ_1. Being x fixed, we can choose this sequence in $(N-1)_{k_1-1}$ different ways.

- Choose a neat sequence of length k_2 (without repetitions) of elements inside V_2, corresponding to the nodes in γ_2. Being $y \in V_2$, if we exclude the trivial case $y \in \gamma$, we can choose such sequence in $(N-1)_{k_2}$ different ways;

- Choose when the jumps j_1 and j_2 occur. Remind that the if path starts in $x \in V_1$, we only have the two possibilities $j_2 = j_1$ or $j_2 = j_1 - 1$:

  - if $j_2 = j_1$, the last node of γ_1 must jump to the absorbing state Δ, whereas from the last node of γ_2 there must be a jump to γ_1;
  - if $j_2 = j_1 - 1$, it is the opposite;

  hence the possible choices are $\binom{k_1-1}{f_1(j_1,j_2)}\binom{k_2-1}{f_2(j_1,j_2)}$ where $f_1(j_1, j_2) := j_1 - 1_{j_1 \neq j_2}$ and $f_2(j_1, j_2) := j_1 - 1_{j_1 = j_2}$.

With a similar reasoning we can count the possible choices in the case $y \in V_1$.

Conclusion
Summing up, distinguishing the two cases $y \in V_2$ (inter-community) or $y \in V_1$ (intra-

community), we have

$$U_q^{(N)}(out) = \sum_{k_1=1}^{N}\sum_{k_2=0}^{N-1}(N-1)_{k_1-1}(N-1)_{k_2}\,\theta(k_1,k_2)\,v^*(2)\,q\cdot \\ \cdot\sum_{j_1=0}^{\min(k_1,k_2)}\sum_{j_2=j_1-1}^{j_1}\binom{k_1-1}{f_1(j_1,j_2)}\binom{k_2-1}{f_2(j_1,j_2)}w_1^{k_1+k_2-1-j_1-j_2}w_2^{j_1+j_2}, \quad (4.46)$$

$$U_q^{(N)}(in) = \sum_{k_1=1}^{N-1}\sum_{k_2=0}^{N}(N-2)_{k_1-1}(N)_{k_2}\,\theta(k_1,k_2)\,v^*(1)\,q\cdot \\ \cdot\sum_{j_1=0}^{\min(k_1,k_2)}\sum_{j_2=j_1-1}^{j_1}\binom{k_1-1}{f_1(j_1,j_2)}\binom{k_2-1}{f_2(j_1,j_2)}w_1^{k_1+k_2-1-j_1-j_2}w_2^{j_1+j_2}. \quad (4.47)$$

We conclude the proof simplifying the quantities $U_q^{(N)}(out)$ and $U_q^{(N)}(in)$.

Consider a Markov chain $\tilde{X}$ on the state space $\{1,2\}$ with transition matrix

$$\tilde{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}, \qquad p = \frac{w_1}{w_1+w_2}. \quad (4.48)$$

Denoting by $\ell(t) := \sum_{s<t} 1_{\{\tilde{X}_s=1\}}$ the local time in state 1 spent by $\tilde{X}$ up to time t and by $\tilde{\mathbb{P}}_1$ the path measure of $\tilde{X}$ starting at 1, it is easy to observe that

$$\tilde{\mathbb{P}}_1(\ell(k_1+k_2)=k_1) = \sum_{j_1=0}^{\min(k_1,k_2)}\sum_{j_2=j_1-1}^{j_1}\binom{k_1-1}{f_1(j_1,j_2)}\binom{k_2-1}{f_2(j_1,j_2)}\cdot \\ \cdot\left(\frac{w_1}{w_1+w_2}\right)^{k_1+k_2-1-j_1-j_2}\left(\frac{w_2}{w_1+w_2}\right)^{j_1+j_2}, \quad (4.49)$$

as the right side of equation (4.49) is exactly the total weight of all the possible trajectories of $\tilde{X}$ such that $\ell(k_1+k_2)=k_1$.
Hence, in (4.46) and in (4.46) we can replace the j's summation terms with the quantity: $(w_1+w_2)^{k_1+k_2-1}\tilde{\mathbb{P}}_1(\ell(k_1+k_2)=k_1)$.

Notice that the terms $\theta(k_1,k_2)$ and v^* can be simplified: indeed, after explicit computations, it results that the second degree polynomial $(q-\hat{\lambda_1}(k_1,k_2))(q-\hat{\lambda_2}(k_1,k_2))$ in the numerator of $\theta(k_1,k_2)$ coincides with the factor c contained in v^*.

Hence, reassembling wisely the pieces and multiplying and dividing by $N^{k_1+k_2-1}$, the interaction potentials assume the following forms:

$$U_q^{(N)}(out) = \sum_{k_1=1}^{N} \sum_{k_2=0}^{N-1} \frac{qN^{k_1+k_2-1}}{(q+N(w_1+w_2))^{k_1+k_2}} (w_1+w_2)^{k_1+k_2-1} \tilde{\mathbb{P}}_1(\ell(k_1+k_2)=k_1) \cdot \\ \cdot \frac{q+k_1(w_1-w_2)+2Nw_2}{q+2Nw_2} N^{-(k_1+k_2-1)}(N-1)_{k_1-1}(N-1)_{k_2} \quad (4.50)$$

$$U_q^{(N)}(in) = \sum_{k_1=1}^{N-1} \sum_{k_2=0}^{N} \frac{qN^{k_1+k_2-1}}{(q+N(w_1+w_2))^{k_1+k_2}} (w_1+w_2)^{k_1+k_2-1} \tilde{\mathbb{P}}_1(\ell(k_1+k_2)=k_1) \cdot \\ \cdot \frac{q+k_2(w_1-w_2)+2Nw_2}{q+2Nw_2} N^{-(k_1+k_2-1)}(N-2)_{k_1-1}(N)_{k_2} \quad (4.51)$$

Now we recognize inside (4.50) and (4.51) the geometric term

$$\begin{aligned} \frac{qN^{k_1+k_2-1}}{(q+N(w_1+w_2))^{k_1+k_2}} (w_1+w_2)^{k_1+k_2-1} &= \frac{q}{q+N(w_1+w_2)} \left(\frac{N(w_1+w_2)}{q+N(w_1+w_2)} \right)^{k_1+k_2-1} \\ &= \mathbb{P}(T_q = k_1+k_2), \end{aligned} \quad (4.52)$$

where T_q is a geometric random variable with success probability $\alpha := \frac{q}{q+N(w_1+w_2)}$. Lastly, we can rewrite as follows

$$\begin{aligned} N^{-(k_1+k_2-1)}(N-1)_{k_1-1}(N-1)_{k_2} &= \frac{1}{N^{k_1+k_2-1}} \prod_{i=1}^{k_1-1}(N-i) \prod_{j=1}^{k_2}(N-j) \\ &= \prod_{i=1}^{k_1-1}\left(1-\frac{i}{N}\right) \prod_{j=1}^{k_2}\left(1-\frac{j}{N}\right) \\ N^{-(k_1+k_2-1)}(N-2)_{k_1-1}(N)_{k_2} &= \frac{1}{N^{k_1+k_2-1}} \prod_{i=2}^{k_1}(N-i) \prod_{j=0}^{k_2-1}(N-j) \\ &= \prod_{i=2}^{k_1}\left(1-\frac{i}{N}\right) \prod_{j=0}^{k_2-1}\left(1-\frac{j}{N}\right) \end{aligned} \quad (4.53)$$

Finally, using the rewrites in (4.52) and (4.53), we obtain

$$U_q^{(N)}(out) = \sum_{k_1=1}^{N}\sum_{k_2=0}^{N-1}\mathbb{P}(T_q = k_1+k_2)\tilde{\mathbb{P}}_1(\ell(k_1+k_2)=k_1)\cdot \\ \cdot\ \frac{k_1(w_1-w_2)+q+2Nw_2}{q+2Nw_2}\prod_{i=1}^{k_1-1}\left(1-\frac{i}{N}\right)\prod_{j=1}^{k_2}\left(1-\frac{j}{N}\right), \quad (4.54)$$

and

$$U_q^{(N)}(in) = \sum_{k_1=1}^{N-1}\sum_{k_2=0}^{N}\mathbb{P}(T_q = k_1+k_2)\tilde{\mathbb{P}}_1(\ell(k_1+k_2)=k_1)\cdot \\ \cdot\ \frac{k_2(w_1-w_2)+q+2Nw_2}{q+2Nw_2}\prod_{i=2}^{k_1}\left(1-\frac{i}{N}\right)\prod_{j=0}^{k_2-1}\left(1-\frac{j}{N}\right), \quad (4.55)$$

which is equivalent to the statement of the theorem. □

We conclude pointing out the similarities and the differences between the mean-field model (MF) and the mean-field-community model (MFC).

(I) Both in MF and MFC, inside the interaction potential appears a *geometric term*

$$\mathbb{P}(T_q = n), \quad (4.56)$$

where T_q is a geometric random variable with success parameter

$$\alpha^{(MF)} = \frac{q}{q+Nw}, \qquad \text{or} \qquad \alpha^{(MFC)} = \frac{q}{q+N(w_1+w_2)}. \quad (4.57)$$

(II) Unlike the first model, in the MFC the pairwise potential contains the summation over k of the term

$$\tilde{\mathbb{P}}_1(\ell(n)=k)\beta_{out}(n,k)g_{out}(n,k), \qquad \text{or} \qquad \tilde{\mathbb{P}}_1(\ell(n)=k)\beta_{in}(n,k)g_{in}(n,k),$$

where β_{out}, β_{in} are defined in the statement of Theorem 4.3, whereas

$$g_{out}(n,k) := \prod_{i=1}^{k-1}\left(1-\frac{i}{N}\right)\prod_{j=1}^{n-k}\left(1-\frac{j}{N}\right), \\ g_{in}(n,k) := \prod_{i=1}^{k}\left(1-\frac{i}{N}\right)\prod_{j=1}^{n-k-1}\left(1-\frac{j}{N}\right). \quad (4.58)$$

The factors $g_{out}(n,k)$ and $g_{in}(n,k)$ can be considered *entropic terms*, relative to the total number of possible configurations. Note that in MF was present the factor $g(n) := \prod_{i=2}^{n} \left(1 - \frac{i}{N}\right)$.
The values β_{out} and β_{in} can be interpreted as weights, and they disappear in the specific case $w_1 = w_2$.
Lastly, notice that, summing over k, we are taking the expectation of the function $\beta_{out} g_{out}(n,k)$ or $\beta_{in} g_{in}(n,k)$ with respect to the local time spent in state 1 of the *coarse-grained* random walk $\tilde{X}$.

Bibliography

[1] Pierre Brémaud, *Markov Chains - Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer (1998)

[2] Søren Feodor Nielsen, *Continuous-time homogeneous Markov chains*, http://web.math.ku.dk/ susanne/kursusstokproc/

[3] Alexandre Gaudillière, *Condenser physics applied to Markov chains - A brief introduction to potential theory* (2009).

[4] Sandro Salsa, *Partial Differential Equations in Action - From Modelling to Theory*, Springer (2015).

[5] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer (1987).

[6] Shizuo Kakutani, *Two-dimensional Brownian motion and harmonic functions*, Proc. Imp. Acad. 20, no. 10, 706-714 (1944).

[7] Bernt Øksendal, *Stochastic Differential Equations - An Introduction with Applications*, Springer (2003).

[8] David A. Levin, Yuval Peres, Elizabeth L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society (2009).

[9] James G. Propp, David B. Wilson, *How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph*, Journal of Algorithms 27 170-217 (1998).

[10] Luca Avena, Alexandre Gaudillière, *Random spanning forests, Markov matrix spectra and well distributed points*, arXiv:1310.1723v4 [math.PR] (2016)

[11] Luca Avena, Paolo Milanesi, Matteo Quattropani, *Loop-erased partitioning of a graph*, arXiv:1906.03858v2 [math.PR] (2019).

[12] Luca Avena, Alexandre Gaudillière, *Two applications of random spanning forests*, J. Theor. Probab. 31, 1975-2004 (2018).

[13] J. Ben Hough, Manjunath Krishnapur, Yuval Peres, Bálint Virág, *Determinantal processes and indipendence*, Probab. Surveys Volume 3, 206-229 (2006).

[14] David B. Wilson, *Generating random spanning trees more quickly than the cover time*, Proceedings of the twenty-eight annual acm symposium on the theory of computing, 296-303 (1996).

[15] Russell Lyons, Yuval Peres, *Probability on Trees and Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, Volume 42, Cambridge University Press (2016).

[16] Luca Avena, Fabienne Castell, Alexandre Gaudillière, Clothilde Mélot, *Random forests and networks analysis*, J. Stat. Phys. 173, 985-1027 (2018).

[17] Luca Avena, Fabienne Castell, Alexandre Gaudillière, Clothilde Mélot, *Intertwining wavelets or Multiresolution analysis on graphs through random forests*, ACHA DOI:10.1016/j.acha2018.09.006 (2018).

[18] Yusuf Y. Pilavci, Pierre-Olivier Amblard, Simon Barthelmé, Nicolas Tremblay, *Smoothing graph signals via random spanning forests*, arXiv:1910.07963v1[cs.DM] (2017).

[19] Philippe Marchal, *Loop-erased random walks, spanning trees and hamiltonian cycles*, Elect. Comm. in Probab. 5, 39–50 (2000).