Universidad de Cantabria

Facultad
de
Ciencias

Fenómenos de umbral en cadenas de Markov: dos ejemplos

Cutoff phenomena in Markov chains: two examples

Trabajo de Fin de Grado
para acceder al

GRADO EN MATEMÁTICAS

Autor: Jesús Ángel Pequeño Fernández
Director: Francisco Santos Leal
Septiembre - 2023

Contents

Agradecimientos

En primer lugar agradecer a mi director Paco, por su interés, atención y disponibilidad, su infinita paciencia y conocimiento, gracias por enseñarme lo que significa ser matemático.

Agradecer a la Universidad de Cantabria por permitirme estudiar este grado. Gracias a todos los profesores por vuestro trato cercano y pasión por las matemáticas, gracias por haber exigido tanto para poder aprender y disfrutar el estudio.

Gracias a mis padres y hermana por confiar en mí desde el principio, por ayudar en todo lo posible y un poco más, por apoyar en cada derrota y alegraros en cada pequeño logro. Me habéis demostrado que no hace falta estudiar matemáticas para ser los mejores compañeros de carrera, ese teorema lleva vuestro nombre. Sin vosotros esto no habría sido posible.

A mi familia, a los que están y a los que no, por vuestro interés todo este tiempo (en especial a mi abuelo por el interés diario) y por hacerme la persona que soy hoy.

A mis compañeros y amigos por las largas sesiones de estudio, por estar ahí y amenizar estos cuatro años a pesar de que el tiempo libre haya sido una función decreciente.

Gracias a todos.

Resumen

Una cadena de Markov es un proceso temporal aleatorio sin memoria. Es sabido que toda cadena de Markov irreducible y aperiódica en un conjunto finito tiene una única distribución estacionaria, a la cual converge cuando el tiempo tiende a infinito sea cual sea la distribución de probabilidad inicial.

El comportamiento asintótico de la cadena frecuentemente presenta un fenómeno de umbral: existe un valor concreto del tiempo en el cual la distribución de probabilidad pasa bruscamente de estar alejada de la estacionaria a ser prácticamente la estacionaria.

En este trabajo se introducen los conceptos básicos de cadenas de Markov y fenómenos de umbral y se estudian dos ejemplos de fenómenos de umbral en paseos aleatorios en grafos: paseos perezosos con sesgo en una cadena finita, y paseos en el modelo de configuración para grafos dirigidos.

Palabras clave: *grafo aleatorio, cadena de Markov, distribución estacionaria, modelo de configuración, fenómenos de umbral*

Abstract

A Markov chain is a random time process without memory. It is known that any irreducible and aperiodic Markov chain on a finite set has a unique stationary distribution, to which it converges when time tends to infinity whatever the initial probability distribution.

The asymptotic behaviour of the chain frequently exhibits a cutoff phenomenon: there is a particular value of time at which the probability distribution abruptly changes from being far from stationary to being almost stationary.

This paper introduces the basic concepts of Markov chains and cutoff phenomena and studies two examples of cutoff phenomena in random paths in graphs: lazy paths with laziness in a finite chain, and paths in the configuration model for directed graphs.

Key words: *random graph, Markov chain, stationary distribution, configuration model, cutoff phenomena*

Introducción y resumen

Los fenómenos de umbral aparecieron por primera vez en el año 1986 en un artículo de Persi Diaconis y David Aldous *"Shuffling cards and stopping times"* [8] donde, al estudiar el número de mezclas que hacían falta en un mazo de n cartas para que estuviese bien mezclado, se observó un comportamiento abrupto que hacía que el mazo pasase de no estar mezclado a estarlo prácticamente en cuestión de pocos pasos. Diaconis mencionó los fenómenos de umbral de manera más formal en parte de sus notas *"Group Representations in Probability and Statistics"* en 1992 [7] donde menciona como problema abierto en una de sus secciones el entender si el fenómeno de umbral suele suceder (pág 91 punto (8)). Sin embargo, no fue hasta 1996 cuando publicó otro artículo *"The cutoff phenomenon in finite Markov chains"* [6], donde definía estos fenómenos de umbral sobre cadenas de Markov en función de la distancia a su distribución estacionaria y daba una serie de ejemplos en los que tenía o no lugar este tipo de umbral.

En este trabajo, estudiaremos el fenómeno de cutoff en dos situaciones relacionadas con cadenas de Markov en grafos. Por un lado, en paseos aleatorios perezosos con sesgo en una línea, y por otro en paseos aleatorios en grafos dirigidos a su vez aleatorios, de acuerdo al modelo de configuración.

Nos hemos basado principalmente en dos fuentes: el libro de David A. Levin *"Markov Chains and Mixing Times"* [11] y el artículo de C. Bordenave, P. Caputo y J. Salez *"Random walk on sparse random digraphs"* [3]. Se ha utilizado el libro de Levin para una introducción y contextualización general de las cadenas de Markov y el fenómeno de umbral, así como el primer gran ejemplo del trabajo: El paseo perezoso y sesgado en una línea (sección 8.2.1 del libro, capítulo 3 de esta memoria). Por último, el artículo de C. Bordenave et al. utiliza el "configuration model" de Béla Bollobas [2], permitiendo dar el segundo gran ejemplo del trabajo: fenómenos de umbral en grafos dirigidos bajo el *"configuration model"* (capítulos 4 y 5 de la memoria).

Descripción de la memoria

El capítulo 1 consistirá en unos preliminares donde se enunciarán algunas definiciones, teoremas importantes y notación de teoría de grafos, probabilidad y notación asintótica, con el objetivo de agrupar prerrequisitos y dedicar el resto del trabajo al tema principal.

En el capítulo 2 comenzará el análisis de las cadenas de Markov. Definiremos lo que es una cadena de Markov, con las propiedades y elementos que la caracterizan como su matriz de transición o su grafo dirigido asociado. Posteriormente veremos el concepto de distribución estacionaria junto con el de variación total al equilibrio y se darán varios ejemplos, fundamental para comprender los fenómenos de umbral en cadenas de Markov.

A continuación estudiaremos la existencia y unicidad de distribución estacionaria. El Teorema 2.2.9 garantiza la existencia para toda cadena de Markov (finita y tiempo-homogénea), la Proposición 2.3.5 nos dice que esta es única si la cadena es irreducible (equivalentemente, si su grafo dirigido asociado es fuertemente conexo), y el Teorema 2.3.6 garantiza que si, además, la cadena es aperiódica, entonces la cadena converge siempre a la distribución estacionaria sea cual sea la distribución de probabilidad inicial. Cerraremos el capítulo definiendo el umbral de una sucesión de cadenas de Markov, así como una caracterización equivalente de esta propiedad.

En el capítulo 3 abordaremos el primer ejemplo importante del trabajo: la ventana de cutoff en el paseo sesgado y perezoso en una línea; es decir, consideramos el grafo formado por una cadena de aristas de longitud n y en él tenemos una cadena de Markov que consiste en moverse a la derecha con una cierta probabilidad $p/2$, a la izquierda con $(1-p)/2$, habiendo también una probabilidad $1/2$ de quedarse en el sitio. El teorema principal nos dice que esta cadena tiene un cutoff en tiempo $t = \beta^{-1}n$ y con una ventana del orden de $\sqrt{n}$, donde $\beta = p - 1/2$ es el sesgo (entendido como exceso de probabilidad de ir a la derecha en vez de a la izquierda).

Para demostrar el teorema principal del capítulo, relacionado con la ventana de cutoff, introduciremos el concepto de acoplamiento de dos cadenas de Markov. Esto permite estudiar propiedades de una cadena a partir de las de otra relacionada con ella, y será una técnica esencial que usaremos varias veces a lo largo de la prueba. La demostración se divide en dos partes. En la sección 3.1 demostramos varios resultados auxiliares que se necesitarán después pero cuyo enunciado y demostración no tiene que ver con el cutoff en sí, y en la sección 3.2 realizamos la prueba del teorema de cutoff, la cual a su vez está dividida en cuatro apartados.

El capítulo 4 consiste en la descripción del *"configuration model"* y de algunas de sus propiedades. En el modelo de configuración se fija un cierto número de vértices n así como sus grados de salida y de entrada, con la restricción obvia de que la suma de los grados de salida y los de entrada sea la misma, y luego se toma el grafo dirigido aleatorio que sale de dar probabilidad uniforme a todos los emparejamientos entre "salidas" y "entradas". El grafo resultante puede no ser simple (puede tener lazos y/o aristas múltiples). Este modelo fue introducido por Bollobas [2] para grafos no dirigidos. En este capítulo estudiamos algunas propiedades del configuration model para grafos dirigidos, como son el hecho de que localmente tienen estructura de árbol (Proposition 4.2.7) y, de manera más precisa, que el paseo aleatorio en el configuration model con gran probabilidad va a estar en nodos que tienen entornos con estructura de árbol (Proposition 4.3.1). Obsérvese que en este modelo tenemos superpuestos dos procesos aleatorios. Por un lado el configuration model produce un grafo aleatorio (con una certa distribución de probabilidad) y por otr lado en ese grafo tenemos la cadena de Markov consistente en hacer un paseo aleatorio.

En el último capítulo 5 del trabajo enunciamos el teorema fundamental (Teorema 5.1.1) que describe el umbral para paseos aleatorios en grafos bajo el modelo de configuración. La prueba consiste en dos grandes partes: Enunciamos los teoremas de forma similar pero trabajando con una distribución estacionaria aproximada para el cálculo de la distancia al equilibrio, para luego mostrar la equivalencia entre el resultado inicial y esta aproximación y por último demostrar ambos teoremas. En la última parte, por falta de espacio y debido a su complejidad, hay dos resultados técnicos que hemos enunciado sin demostración (Proposiciones 5.1.5 y 5.3.2, que son las Proposiciones 8 y 10 de [3]).

Introduction and summary

Cutoff phenomena first appeared in 1986 in a paper by Persi Diaconis and David Aldous *"Shuffling cards and stopping times"* [8] where, while studying the number of shuffles it took for a deck of n cards to be properly shuffled, they observed an abrupt behavior that caused the deck to go from unshuffled to shuffled in a matter of virtually a few steps. Diaconis mentioned cutoff phenomena more formally in part of his notes *"Group Representations in Probability and Statistics"* in 1992 [7] where he mentions as an open problem in one of his sections to understand whether cutoff phenomena usually happen (p 91 item (8)). However, it was not until 1996 when he published another article *"The cutoff phenomenon in finite Markov chains"* [6], where he defined the cutoff on Markov chains as a function of the distance to their stationary distribution and gave a number of examples in which this type of threshold did or did not occur.

In this paper, we will study the cutoff phenomenon in two situations related to Markov chains in graphs. On the one hand, in lazy random paths with bias on a line, and on the other hand in random paths in directed graphs random in turn random, according to the configuration model.

We have relied mainly on two sources: David A. Levin's book *"Markov Chains and Mixing Times"* [11] and the article by C. Bordenave, P. Caputo and J. Salez *"Random walk on sparse random digraphs"* [3]. Levin's book has been used for a general introduction and contextualization of Markov chains and the cutoff phenomenon, as well as the first major example of the work: The lazy and biased random walk on a line (section 8.2.1 of the book, chapter 3 of this paper). Finally, the paper by C. Bordenave et al. uses the configuration model of Béla Bollobas [2], allowing to give the second major example of the paper: threshold phenomena in directed graphs under the *"configuration model"* (chapters 4 and 5 of the paper).

Memory description

Chapter 1 will consist of some preliminaries where some definitions, important theorems and notation of graph theory, probability and asymptotic notation will be stated, with the aim of grouping prerequisites and dedicating the rest of the work to the main topic.

In chapter 2 we will begin the analysis of Markov chains. We will define what a Markov chain is, with the properties and elements that characterize it such as its transition matrix or its associated directed graph. Then we will see the concept of stationary distribution together with the concept of total variation distance from equilibrium and several examples will be given, fundamental to understand the threshold phenomena in Markov chains. Then we will study the existence and uniqueness of stationary distribution. The The-

orem 2.2.9 guarantees existence for every (finite and time-homogeneous) Markov chain, Proposition 2.3.5 tells us that it is unique if the chain is irreducible (equivalently, if its associated directed graph is strongly connected), and Theorem 2.3.6 guarantees that if, in addition, the chain is aperiodic, then the chain always converges to the stationary distribution regardless of the initial probability distribution. We will close the chapter by defining the cutoff of a succession of Markov chains, as well as an equivalent characterization of this property.

In chapter 3 we will address the first important example of this work: the cutoff window on the biased and lazy walk on a line; that is, we consider the graph formed by a chain of edges of length n and in it we have a Markov chain consisting of moving to the right with a certain probability $p/2$, to the left with $(1-p)/2$, there being also a probability $1/2$ of staying in place. The main theorem tells us that this chain has a cutoff in time $t = \beta^{-1}n$ and with a window of order $\sqrt{n}$, where $\beta = p - 1/2$ is the bias (understood as excess probability of going right instead of left).

To prove the main theorem of the chapter, related to the cutoff window, we will introduce the concept of coupling of two Markov chains. This allows us to study properties of one chain from those of another related chain, and will be an essential technique that we will use several times throughout the proof. The demonstration is divided into two parts. In section 3.1 we prove several auxiliary results that will be needed later but whose statement and proof has nothing to do with the cutoff itself, and in section 3.2 we show the proof of the cutoff theorem, which is divided in four sections.

The chapter 4 consists of the description of the configuration model and some of its properties. In the configuration model we fix a certain number of n vertices as well as their output and input degrees, with the obvious constraint that the sum of the output and input degrees be the same, and then take the random directed graph that comes out of giving uniform probability to all pairings between "outputs" and "inputs". The resulting graph may not be simple (it may have multiple loops and/or edges). This model was introduced by Bollobas [2] for undirected graphs. In this chapter we study some properties of the configuration model for directed graphs, such as the fact that they are locally tree-structured (Proposition 4.2.7) and, more precisely, that the random walk in the configuration model will most likely be on nodes that have tree-structured environments (Proposition 4.3.1). Note that in this model we have two overlapping random processes. On the one hand the configuration model produces a random graph (with a certain probability distribution) and on the other hand in that graph we have the Markov chain consisting of a random walk.

In the last chapter 5 of the paper we state the fundamental theorem (Theorem 5.1.1) describing the cutoff for random paths in graphs under the configuration model. The proof consists of two main parts: We state the theorems in a similar way but working with an approximate stationary distribution for the calculation of the distance to equilibrium, then show the equivalence between the initial result and this approximation and finally prove both theorems. In the last part, for lack of space and due to their complexity, there are two technical results that we have stated without proof (Propositions 8 and 10 of [3]).

Chapter 1

Basic definitions and results

1.1 Graph Theory

In order to understand Markov chains, it is necessary to have a solid knowledge of some underlying mathematical concepts. One of these main ideas is graphs, since they provide a visual representation of the elements and relations in a system. The concepts on this chapter have been extracted from [9].

Definition 1.1.1 (Graph)**.** A (finite) graph G is a triple $(V(G), E(G), \phi_G)$, consisting of a finite nonempty set of *vertices* $V(G)$, a set of *edges* $E(G)$ and an *incidence function* $\phi_G : E(G) \to P(V(G))$, that assigns each edge with a pair of vertices, i.e. $\phi_G(e) = \{x, y\}$ with $e \in E(G)$ and $x, y \in V$. If $\phi_G(a) = \{x\}$, we say that $e \in E(G)$ is a loop.

Definition 1.1.2 (Simple graph)**.** A graph G is simple if there are no loops and for any two vertices there is at most one edge that links them.

Definition 1.1.3 (Directed graph)**.** A directed graph or digraph is a triple $(V(G), E(G), \phi_G)$ where $V(G)$ and $E(G)$ are defined as on graph definition 1.1.1 and $\phi_G : E(G) \to V \times V$ an *incidence function* that associates an edge to an ordered pair of vertices.

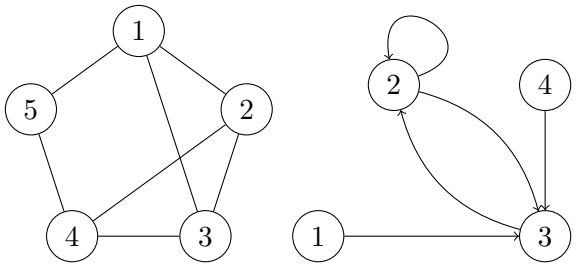


Figure 1.1: Simple graph and directed graph

Definition 1.1.4 (Strongly Connected Graph)**.** A directed graph is a *strongly connected graph* if for every pair of vertices $x, y \in V$ there exists a path from x to y.

1.2 Probability

Since Markov chains are stochastic processes, a great deal of proofs use random variables and their convergence. Some relevant definitions are:

Definition 1.2.1 (Convergence in Probability)**.** A sequence of real random variables $(X_0, X_1, \dots)$ converges to a random variable X *in probability* and denoted by $X_n \xrightarrow{\text{P}} X$ if

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0, \text{ for all } \epsilon > 0$$

Definition 1.2.2 (With high probability)**.** An event which depends on n occurs *with high probability* (W.H.P.) if its probability goes to 1 as n goes to infinity.

Theorem 1.2.3 (Central Limit Theorem)**.** *[1] Consider the random variable S_n as the sum of n random independent and identically distributed variables with mean μ and variance $\sigma^2 < \infty$. Define Z_n as a standarization of S_n:*

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Then:

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) := \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

where $\Phi(z)$ is called the "cumulative error function".

Theorem 1.2.4 (Boole's Inequality)**.** *For any finite or countable set of events, the probability that at least one events happens is upper bounded by the sum of the probability of the individual events. That is to say:*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

1.3 Asymptotic notation

When working with limits that tend to infinity, asymptotic analysis is used to describe limiting behaviour. The *Bachmann-Landauu notation* represent the behaviour of functions in a neighborhood in terms of other functions [12].

Definition 1.3.1 (Big O notation)**.** Let f be the real or complex function to be estimated and let g be the real-positive valued comparison function. $f(x) = O(g(x))$ if there exists a real number $M > 0$ and a real number x_0 such that:

$$|f(x)| \leq Mg(x) \quad \text{for all } x \geq x_0.$$

Definition 1.3.2 (Small o notation)**.** Let f be the real or complex function to be estimated and let g be the real-positive valued comparison function. $f(x) = o(g(x))$ if for every constant $\epsilon > 0$ there exists a real number x_0 such that:

$$|f(x)| \leq \epsilon g(x) \quad \text{for all } x \geq x_0.$$

Definition 1.3.3 (Big Θ notation)**.** Let f be the real or complex function to be estimated and let g be the real valued comparison function. $f(x) = \Theta(g(x))$ if for every constant $\epsilon > 0$ there exists two positive real numbers M_1 and M_2 and a real number x_0 such that:

$$M_1 g(x) \leq |f(x)| \leq M_2 g(x) \quad \text{for all } x \geq x_0.$$

Chapter 2

Markov Chains

Everything in this chapter is following the book [11] and with some proofs based on [10].

2.1 Definitions and basic properties

A sequence of finite or countable random variables $(X_0, X_1, ...)$ is a *discrete Markov chain* with *state space* $\mathcal{X}$ if it satisfies the Markov property:

$$\begin{aligned}&\forall n \in \mathbb{N} \text{ and } \forall i_1, i_2, ..., i_n \in \mathcal{X} : \\ &P(X_n = i_n | X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_n = i_n | X_{n-1} = i_{n-1}),\end{aligned}$$

where $\mathcal{X}$ is nothing but a set with every possible value that the random variables can have.

The interpretation is that we have a process that can be in different states (the elements of $\mathcal{X}$) and moves from one state to another as time passes, where time is discretized and represented by the index n. Having said that, the Markov chain property means that the probability of moving from a state i at time n to a state j at time $n+1$ (both in $\mathcal{X}$) does not depend on the previous steps to reach i.

From now on we assume that our Markov chains are *time-homogeneous* which means that the probability does not depend on n either. That is:

$$P(X_{n+1} = j | X_n = i) = P(X_n = j | X_{n-1} = i), \text{ for every } n.$$

We will also assume that $\mathcal{X}$ is finite.

This allows us to define a *transition matrix* P, a $|\mathcal{X}| \times |\mathcal{X}|$ matrix whose (i,j)-th entry (denoted by $P(i,j)$) is the probability

$$P(i,j) := P(X_1 = j | X_0 = i) = P(X_n = j | X_{n-1} = i) \tag{2.1}$$

The i-th row of P is the distribution $P(i, \cdot)$. P is stochastic:

$$\sum_{j \in \mathcal{X}} P(i,j) = 1 \text{ for all } i \in \mathcal{X} \text{ and } P(i,j) \geq 0 \text{ for all } i, j \in \mathcal{X}.$$

The directed graph of a Markov chain with transition matrix $P \in \mathbb{R}^{n \times n}$ is the directed graph with vertices $\{1, \ldots, n\}$ and with an edge (i,j) if and only if $P(i,j) \neq 0$.

As a Markov chain moves along the set $\mathcal{X}$, a question may appear: How can we know where are we going to be after t steps?

Let us define the initial distribution π_0 as the probability distribution of the Markov chain at time 0. i.e. for every state $i \in \mathcal{X}$ we express $\pi_0(i) = P(X_0 = i)$ as the probability that the chain starts at state i. In general, let us denote π_n as distribution of the Markov chain at time n, that is to say, $\pi_n(i) = P(X_n = i)\ \forall i \in \mathcal{X}$. By the Law of Total Probability:

$$\pi_{n+1}(j) = P(X_{n+1} = j) = \sum_{i \in \mathcal{X}} P((X_{n+1} = j) \cap (X_n = i)) =$$
$$= \sum_{i \in \mathcal{X}} P(X_{n+1} = j | X_n = i) P(X_n = i) = \sum_{i \in \mathcal{X}} P(i,j)\pi_n(i)$$

In matrix notation this is equivalent to:

$$\pi_{n+1} = \pi_n P$$

By induction we have that

$$\pi_n = \pi_0 P^n$$

Given an initial distribution π_0, we denote by $P_{\pi_0}(A)$ the probability of the event A having initial distribution π_0. Let j be a state in $\mathcal{X}$; the distribution having probability 1 of being at j is denoted δ_j but we abbreviate P_{δ_j} as P_j. A similar notation is used for the expectations E_{π_0}, E_j.

Consider $t \in \mathbb{N}$ and two vertices $i, j \in \mathcal{X}$. The probability of reaching j from i after exactly t steps is:

$$P_i(X_t = j) = (\delta_i P^t)(y) = P^t(i,j).$$

Definition 2.1.1 (Stationary distribution)**.** Let π be a distribution of the Markov chain at time t. π is a *stationary distribution* if the probability distribution of the Markov chain at any time $n \geq t$ equals π too; equivalently, if:

$$\pi = \pi P$$

It can be seen that this stationary distribution may be unique, have infinite possibilities or even not exist.

Definition 2.1.2 (Total variation)**.** The *total variation* between two distributions $\mu, \sigma : \mathcal{X} \longrightarrow \mathbb{R}$ is:

$$\|\mu - \sigma\|_{TV} = \max_{A \subset \mathcal{X}} |\mu(A) - \sigma(A)|$$

Proposition 2.1.3. *If the state space $\mathcal{X}$ is finite the total variation can be also expressed as*

$$\|\mu - \sigma\|_{TV} = \frac{1}{2} \sum_{i \in \mathcal{X}} |\mu(i) - \sigma(i)|$$

Proof. Let $B = \{i \in \mathcal{X} : \mu(i) \geq \sigma(i)\}$. Consider $A \subseteq \mathcal{X}$ any set of states. We can write $A = (A \cap B) \cup (A \cap B^c)$.

$$\mu(A) - \sigma(A) = \mu(A \cap B) + \mu(A \cap B^c) - \sigma(A \cap B) - \sigma(A \cap B^c) \leq \mu(A \cap B) - \sigma(A \cap B),$$

where the last inequality holds because if $i \in A \cap B^c \Longrightarrow \mu(i) - \sigma(i) < 0$.
Since $B = (A \cap B) \cup (B \backslash A)$ and having that $B \backslash A \subseteq B$,

$$\mu(A \cap B) - \sigma(A \cap B) \leq \mu(A \cap B) - \sigma(A \cap B) + \mu(B \backslash A) - \sigma(B \backslash A) = \mu(B) - \sigma(B).$$

Putting together both inequalities:

$$\mu(A) - \sigma(A) \leq \mu(B) - \sigma(B) \tag{2.2}$$

By a similar reasoning

$$\mu(B^c) - \sigma(B^c) \leq \mu(A) - \sigma(A).$$

Let us bound $|\mu(A) - \sigma(A)|$ by $\mu(B) - \sigma(B)$ as follows:

$$\begin{gathered}\mu(B) + \mu(B^c) = 1 = \sigma(B) + \sigma(B^c) \Longrightarrow \mu(B^c) - \sigma(B^c) = -(\mu(B) - \sigma(B)) \\ -(\mu(B) - \sigma(B)) \leq \mu(A) - \sigma(A) \leq \mu(B) - \sigma(B) \Longrightarrow \\ \Longrightarrow |\mu(A) - \sigma(A)| \leq \mu(B) - \sigma(B)\end{gathered}$$

As $\mu(A) - \sigma(A)$ is bounded by equation (2.2), the equality holds when $A = B$, so:

$$\begin{aligned}\|\mu - \sigma\|_{TV} =& \max_{A \subset \mathcal{X}} |\mu(A) - \sigma(A)| = \mu(B) - \sigma(B) = \\ =& \frac{1}{2} \left(\mu(B) - \sigma(B) + \sigma(B^c) - \mu(B^c)\right) = \\ =& \frac{1}{2} \left(\sum_{i \in B} |\mu(i) - \sigma(i)| + \sum_{i \in B^c} |\mu(i) - \sigma(i)|\right) = \\ =& \frac{1}{2} \sum_{i \in \mathcal{X}} |\mu(i) - \sigma(i)|\end{aligned}$$

□

In the next two sections we study conditions for the Markov chain to have a unique stationary distribution π.

A first approach to calculate this distribution is to solve the linear system

$$\pi(P - I) = 0,$$

This approach will lead to Theorem 2.3.5 showing that irreducible Markov chains have a unique stationary distribution.

An alternative to obtain π is to understand the stationary distribution as the limit of the distribution when time tends to infinity. Indeed, if the limit of P^n as n goes to infinity exists, then each row in the limit is a stationary distribution.

In Theorem 2.3.6 we show that this always happens for aperiodic irreducible Markov chains and, moreover, the convergence is exponentially fast.

Definition 2.1.4 ([11, Section 4.4], [3, Section 1])**.** For each i we define the *total variation distance from equilibrium at time t starting at i* as:

$$\mathcal{D}_i(t) := \|P^t(i, \cdot) - \pi\|_{TV}$$

We are mostly interested in the maximum over all i, the *maximum total variation distance from equilibrium at time t*

$$d(t) := \max_{i \in \mathcal{X}} \|P^t(i, \cdot) - \pi\|_{TV}$$

The main reason to look at this maximum is that it bounds the distance to equilibrium for *any* initial distribution, not only δ_i:

Proposition 2.1.5. *For any two distributions π_0 and π' and any $t \in \mathbb{N}$ we have*

$$\|\pi_0 P^t - \pi'\|_{TV} \leq \max_{i \in \mathcal{X}} \|P^t(i, \cdot) - \pi'\|_{TV}.$$

In particular for the stationary distribution π:

$$\|\pi_0 P^t - \pi\|_{TV} \leq d(t).$$

Proof. π_0 is a convex combination of the distributions δ_i, with coefficients $\lambda_i := \pi_0(i)$. Thinking of P^t as a linear map, this makes $\pi_0 P^t$ a convex combination of the distributions $P^t(i, \cdot) := \delta_i P^t$. That is:

$$\pi_0 P^t = \sum_{i \in \mathcal{X}} \lambda_i P^t(i, \cdot).$$

Convexity of the total variation [1] then implies $\|\pi_0 P^t - \pi'\|_{TV} \leq \max_{i \in \mathcal{X}} \|P^t(i, \cdot) - \pi'\|_{TV}$. □

An important result related with the distance to equilibrium is the fact that advancing the chain only gets closer to stationarity; that is to say

$$\|\pi_0 P^{t+1} - \pi\|_{TV} \leq \|\pi_0 P^t - \pi\|_{TV}.$$

This can be easily deduced from the following lemma:

Lemma 2.1.6. *Let P be the transition matrix of a Markov chain with state space $\mathcal{X}$, let μ and σ be two distributions on $\mathcal{X}$. We have that:*

$$\|\mu P - \sigma P\|_{TV} \leq \|\mu - \sigma\|_{TV}.$$

Proof. By the very definition of total variation distance

$$\begin{aligned}
\|\mu P - \sigma P\|_{TV} &= \frac{1}{2} \sum_{i \in \mathcal{X}} |\mu P(i) - \sigma P(i)| = \frac{1}{2} \sum_{i \in \mathcal{X}} \left| \sum_{j \in \mathcal{X}} P(j, i)[\mu(j) - \sigma(j)] \right| \leq \\
&\leq \frac{1}{2} \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X}} P(j, i) |\mu(j) - \sigma(j)| = \frac{1}{2} \sum_{j \in \mathcal{X}} |\mu(j) - \sigma(j)| \sum_{i \in \mathcal{X}} P(j, i) = \\
&= \frac{1}{2} \sum_{j \in \mathcal{X}} |\mu(j) - \sigma(j)| = \|\mu - \sigma\|_{TV}
\end{aligned}$$

□

[1] The total variation $\|\cdot\|_{TV}$ is a rescaling of the L_1-norm and for every norm and any convex combination $v = \sum_{i=1}^k \lambda_i v_i$ with $0 \leq \lambda_i \leq 1$ and $\sum_i \lambda_i = 1$ we have that

$$\|v - u\| = \left\| \sum_{i=1}^k \lambda_i v_i - u \right\| = \left\| \sum_{i=1}^k \lambda_i (v_i - u) \right\| \leq \sum_{i=1}^k \lambda_i \|v_i - u\| \leq \left(\sum_{i=1}^k \lambda_i \right) \max_j \|v_j - u\| = \max_j \|(v_j - u)\|.$$

Corollary 2.1.7. *For every initial distribution π_0 we have $\|\pi_0 P^{t+1} - \pi\|_{TV} \leq \|\pi_0 P^t - \pi\|_{TV}$. In particular, $d(t)$ is (weakly) decreasing with t.*

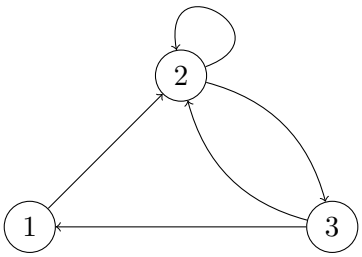
Proof. Apply the previous lemma with $\sigma = \pi$ and $\mu = \pi_0 P^t$ as follows:

$$\|\pi_0 P^{t+1} - \pi\|_{TV} = \|(\pi_0 P^t)P - \pi P\|_{TV} \leq \|\pi_0 P^t - \pi\|_{TV}. \qquad \square$$

Example 2.1.8. *Let us calculate the stationary distribution π in both of the previous ways. Consider the Markov chain with state space $\mathcal{X} = \{1, 2, 3\}$ and transition matrix:*

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

The graph of the chain can be represented as follows



- *Obtain π solving linear system:*

$$\begin{cases} \frac{1}{2}z & = x \\ x + \frac{1}{4}y + \frac{1}{2}z & = y \\ \frac{3}{4}y & = z \\ x + y + z & = 1 \end{cases}$$

  Where we obtain the following solution:

$$\pi = \begin{pmatrix} \frac{3}{17} & \frac{8}{17} & \frac{6}{17} \end{pmatrix} \approx (0.176 \;\; 0.471 \;\; 0.353)$$

- *Calculating the n-th power of P:*

  For $n = 5$:

$$\begin{pmatrix} 0.165 & 0.403 & 0.434 \\ 0.218 & 0.482 & 0.302 \\ 0.130 & 0.491 & 0.381 \end{pmatrix}$$

  For $n = 20$:

$$\begin{pmatrix} 0.176 & 0.471 & 0.353 \\ 0.176 & 0.471 & 0.353 \\ 0.176 & 0.471 & 0.353 \end{pmatrix}$$

  Which gives us a fairly good approximation of π.

2.2 Every Markov chain has a stationary distribution

In order to give the conditions for a Markov chain to have a unique stationary distribution, some previous definitions and propositions are required.
Let P be the transition matrix of a Markov chain with state space $\mathcal{X}$:

Definition 2.2.1 (Irreducible matrix)**.** P is a *irreducible matrix* if for any two states $i, j \in \mathcal{X}$ there exists a $t \in \mathbb{N}$ such that $P^t(i,j) > 0$, i.e., it is possible to get from any state to any other state using only transitions of positive probability.

Proposition 2.2.2. *A transition matrix $P \in \mathbb{R}^{n\times n}$ is irreducible $\iff$ its directed graph is strongly connected (for every $i, j \in V$ there exists a directed path from i to j)*

Proof. The proof is direct due to the very definition of P. □

In order to prove the existence of stationary distribution for any Markov chain we need to introduce some definitions and results related with the states in $\mathcal{X}$.

Let $i, j \in \mathcal{X}$; we say that j *is accessible from* i and denote it as $i \to j$ if there exists $r \in \mathbb{N}$ such that $P^r(i,j) > 0$. That is to say, if there is a path from i to j in the graph of the Markov chain. If i is also accessible from j (not necessarily with the same r), we say that i *communicates with* j and write $i \leftrightarrow j$. A state $i \in \mathcal{X}$ is *essential* if for every j such that $i \to j$ it is also true that $j \to i$. A state i is *inessential* if it is not essential.

Lemma 2.2.3. *If i is an essential state and $i \to j$, then j is essential.*

Proof. Let k be a reachable state from j. As $i \to j \to k$ and i is essential, then $k \to i$ and therefore $k \to j$. □

Communicating with defines an equivalence relation and we denote by $[i]$ the communicating class of i. Observe that communicating classes in the Markov chain are the same as strongly connected components in the graph.

Lemma 2.2.4. *Every Markov chain with finite $\mathcal{X}$ has at least one essential class. [11]*

Proof. Let us build a sequence of states $(i_0, i_1, ...)$. Fixed i_0, for $m \geq 1$, if i_m is essential we stop due to Lemma 2.2.3, as we have an essential class made up of every state that communicates with i_m. If i_m is not essential, we choose i_{m+1} such that $i_m \to i_{m+1}$ but $i_{m+1} \not\to i_m$. There can not be repeated states, because if $i_{m+1} = i_l$ for $l < m$, then $i_{m+1} \to i_m$ as $i_l \to i_m$.
As $|\mathcal{X}| < \infty$, we will reach an essential state at some point, concluding the proof.
□

Lemma 2.2.5. *In every directed graph $G = (V, E)$ without sinks, there exists a subset $W \subset V$ such that:*

- $G\big|_W$ *is strongly connected.*
- *W has no edge that leaves W.*

Proof. Consider a Markov chain with graph G. For example, we can give equal probability to all the edges that go out from each $i \in V$. That is, we take $P(i,j) = 1/\deg^+(i)$, where $\deg^+(i)$ denotes the out-degree of i.

By Lemma 2.2.4, there exists an essential class $W \subset V$. Being a class implies the induced subgraph is strongly connected and being essential implies that W can not be left once you reach it. □

Corollary 2.2.6. *Knowing that every irreducible chain has stationary distribution, so does the reducible chains.*

Proof. Consider the induced directed graph G from the transition matrix P of the reducible chain. By Lemma 2.2.5 consider W as the strongly connected component of G. Reordering rows of the transition matrix we obtain

$$\begin{pmatrix} W & 0 \\ * & * \end{pmatrix}$$

where the upper left quadrant is the transition matrix of the Markov chain defined as on proof of Lemma 2.2.5. The stationary distribution of the reducible chain is a vector made of the stationary distribution of W π_W, and $|V \backslash W|$ zeros:

$$\pi = (\pi_w \, 0 \ldots 0),$$

and it is easily shown that

$$\pi P = \pi$$

□

Definition 2.2.7 (Hitting time)**.** For $i \in \mathcal{X}$, the *hitting time* of i is the first time the chain visits i.

$$\tau_i := min\{t \geq 0 : X_t = i\}$$

If we exclude the case in which we start at i we can define

$$\tau_i^+ := min\{t \geq 1 : X_t = i\}$$

Lemma 2.2.8. *Given a finite irreducible Markov chain, for every $k \in \mathcal{X}$ we have that:*

- $P_k\{\tau_k^+ < \infty\} = 1$
- $E_k(\tau_k^+) < \infty$

Proof. For every state $i \in \mathcal{X}$ we denote by l_i as the minimum number of steps to reach k from i in the graph of the Markov chain, and by $p_i := P(X_{l_i} = k | X_0 = i)$, which is always greater than 0 since the chain is irreducible. Consider

$$L := \max_{i \in \mathcal{X}} \{l_i\}.$$

By the definition of τ_k^+, for every initial distribution π_0 we have that

$$P_{\pi_0}(\tau_k^+ \leq L) \geq \min p_i.$$

Letting $\epsilon = \min_{i \in V} p_i > 0$ and iterating multiples of L steps:

$$P_{\pi_0}(\tau_k^+ > L) \leq 1 - \epsilon$$
$$P_{\pi_0}(\tau_k^+ > 2L) = P_{\pi_0}(\tau_k^+ > L) \cdot P_{P^L \pi_0}(\tau_k^+ > L) \leq (1-\epsilon)^2$$

and, in general

$$P_{\pi_0}(\tau_k^+ > mL) \leq (1-\epsilon)^m, \quad \forall m \in \mathbb{N}. \tag{2.3}$$

We prove the first point of the Lemma taking $\pi_0 = \delta_k$ and making m tend to infinity:

$$\lim_{m \to \infty} P_k(\tau_k^+ > mL) \leq \lim_{m \to \infty} (1-\epsilon)^m = 0$$

In order to prove the second result

$$\begin{aligned} E_k(\tau_k^+) &= \sum_{t=1}^{\infty} t P_k(\tau_k^+ = t) = \\ &= \sum_{t=1}^{L} t P_k(\tau_k^+ = t) + \sum_{t=L+1}^{2L} t P_k(\tau_k^+ = t) + \sum_{t=2L+1}^{3L} t P_k(\tau_k^+ = t) + \cdots \leq \\ &\leq L^2 P_k(\tau_k^+ > 0) + 2L^2 P_k(\tau_k^+ > L) + 3L^2 P_k(\tau_k^+ > 2L) + \cdots = \\ &= L^2 \sum_{m=0}^{\infty} (m+1)(1-\epsilon)^m < \infty. \end{aligned}$$

The last inequality (saying that $\sum_{m=0}^{\infty}(m+1)(1-\epsilon)^m$ is convergent) follows for example from *D'Alambert ratio test* (also known as *ratio test*): in a series of positive terms, if the ratio between consecutive terms has limit strictly smaller than 1 then the series converges. In our case, calling $a_m = (m+1)(1-\epsilon)^m$:

$$C = \lim_{m \to \infty} \frac{a_{m+1}}{a_m} = \lim_{m \to \infty} \frac{(m+2)(1-\epsilon)^{m+1}}{(m+1)(1-\epsilon)^m} = 1 - \epsilon < 1. \qquad \square$$

Theorem 2.2.9. *Every (finite, time-homogeneous) Markov chain has some stationary distribution.*

Proof. By Corollary 2.2.6 we assume without loss of generality that the Markov chain is irreducible.

The proof is based on Lemma 2.2.8.
Let

$$\begin{aligned} \tilde{\pi}(j) :=& E_k(\text{number of visits to } j \text{ before returning to } k) = \\ =& \sum_{t=0}^{\infty} P_k\{X_t = j \wedge \tau_k^+ > t\} = \sum_{t=0}^{\infty} P_k\{X_t = j \wedge \tau_k^+ \geq t+1\}. \end{aligned} \tag{2.4}$$

The idea of the proof is to show that the vector $\tilde{\pi}$ normalized so that $\sum_{i \in \mathcal{X}} \tilde{\pi}(i) = 1$ is a stationary distribution.

Recall that $\tilde{\pi}$ is a stationary distribution if $\tilde{\pi} P = \tilde{\pi}$, $\sum_{i \in \mathcal{X}} \tilde{\pi}(i) = 1$ and $\tilde{\pi}(i) \geq 0$ $\forall i \in V$. In this case we need to prove the first two equalities as the third is granted by the very definition of $\tilde{\pi}$.

Observe that $\tilde{\pi}(k) = 1$ and that

$$\sum_{i \in \mathcal{X}} \tilde{\pi}(i) = E_k(\tau_k^+),$$

which gives us the fact that $\tilde{\pi}(i) \leq \infty \ \forall i \in \mathcal{X}$ by Lemma 2.2.8. Let us show that $\tilde{\pi}$ is stationary:

$$\sum_{i \in \mathcal{X}} \tilde{\pi}(i) P(i,j) = \sum_{i \in \mathcal{X}} \sum_{t=0}^{\infty} P_k\{X_t = i \wedge \tau_k^+ > t\} P(i,j). \tag{2.5}$$

We can write the last term as

$$P_k\{X_t = i \wedge \tau_k^+ > t\} P(i,j) = P_k\{X_t = i \wedge X_{t+1} = j \wedge \tau_k^+ \geq t+1\}.$$

When substituting on (2.5) we can get rid of $X_t = i$ due to the first summation hence we obtain

$$\sum_{i \in \mathcal{X}} \tilde{\pi} P(i,j) = \sum_{t=0}^{\infty} P_k\{X_{t+1} = j \wedge \tau_k^+ \geq t+1\} = \sum_{t=1}^{\infty} P_k\{X_t = j, \tau_k^+ \geq t\}.$$

By the definition of $\tilde{\pi}$ (2.4)

$$\begin{aligned}\sum_{t=1}^{\infty} P_k\{X_t = j, \tau_k^+ \geq t\} =& \tilde{\pi}(j) - P_k\{X_0 = j \wedge \tau_k^+ \geq 0\} + \sum_{t=1}^{\infty} P_k\{X_t = j \wedge \tau_k^+ = t\} = \\ =& \tilde{\pi}(j) - P_k\{X_0 = j\} + P_k\{X_{\tau_k^+} = j\}\end{aligned}$$

If we show that $P_k\{X_0 = j\} = P_k\{X_{\tau_k} = j\}$ the first part of the proof is done. We can consider two cases:

- $j = k$: As the initial distrubution is δ_k, the first term is equal to 1 and by the definition of τ_k^+, $P_k\{X_{\tau_k^+} = j\} = 1$.
- $j \neq k$: Again due to the initial distribution, the first term is equals to 0, and so is the second.

This proves that $\tilde{\pi} P = \tilde{\pi}$. We now normalize by

$$\sum_{i \in \mathcal{X}} \tilde{\pi}(i) = E_k(\tau_k^+)$$

obtaining

$$\pi(i) = \frac{\tilde{\pi}(i)}{E_z(\tau_k^+)}$$

□

Remark 1. *The theorem can also be proved using Brouwer fixed-point theorem:*

Theorem 2.2.10 (Brouwer's fixed-point theorem)**.** *Every continuous function from $\mathcal{B}^d \longrightarrow \mathcal{B}^d$ has some fixed point.*

The proof of the theorem for any $d \in \mathbb{N}$ goes beyond the scope of this work, although throughout the mathematics degree various proofs for $d = 2$ have been studied, with both topological and complex analysis arguments.

Second proof of Theorem 2.2.9. This proof is based on the homeomorphism between Δ^{n-1} and $\mathcal{B}^{n-1}$, $\Delta^{n-1} \cong \mathcal{B}^{n-1}$, where

$$\Delta^{n-1} = \{(x_1, ..., x_n) \Big| \sum_{i=1}^{n} = 1 \wedge x_i \geq 0 \forall i\}$$

is called the *unit simplex* or equivalently *probability simplex*.
We can now interpret the transpose of P as a continuous function from $\Delta^{n-1} \longrightarrow \Delta^{n-1}$ and by Brouwer's fixed-point theorem, there exists a vector $v \in \mathbb{R}^n$ such that $P^T v = v$. This vector v is the stationary distribution we are looking for. □

If we do not consider a finite Markov chain (Markov chain with finite state space), the existence of stationary distribution is not granted:

Given the state space $\mathcal{X} = \mathbb{N}$, consider the transition matrix P as the infinite matrix with $P(i, i+1) = 1$ for every $i \in \mathcal{X}$.

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Let π_0 be a distribution. If π_0 is a stationary distribution $\pi_0 P = \pi_0$, then we have that $\pi_0(j) = \pi_0(j+1)\ \forall j > 0$. This implies that π_0 is a constant vector, but since the state space is infinite, there is no infinite vector such that $\sum_{i=1}^{\infty} \pi_0(i) = 1$ and $\pi_0(i) = c$ for some $c \in \mathbb{R}$.

2.3 Convergence to the stationary distribution

In Theorem 2.3.6 we are going to show that any Markov chain with irreducible, aperiodic transition matrix P on a finite state space $\mathcal{X}$ converges to its unique stationary distribution π:

$$\forall (i, j) \in \mathcal{X}^2, \ \ P^t(i, j) \ \xrightarrow{n \to \infty} \ \pi(j)$$

By aperiodic we mean the following:

Definition 2.3.1 (Period. Aperiodic chain)**.** Let $i \in \mathcal{X}$ and $T(i) = \{t \geq 1 : P^t(i, i) > 0\}$ the set of times when it is possible to return to i. The *period* of the state i is $g.c.d(T(i))$. A chain is *aperiodic* if all states have period 1.

Proposition 2.3.2. *If P is aperiodic and irreducible, there exist $r_0 \in \mathbb{N}$ such that $P^r(i, j) > 0\ \forall i, j \in \mathcal{X}$ and for every $r \geq r_0$.*

Proof. Let $i \in \mathcal{X}$. As P is aperiodic, $g.c.d(T(i)) = 1$. For $l, m \in T(i)$, since $P^{l+m}(i,i) \geq P^l(i,i)P^m(i,i) > 0$ we have $l+m \in T(i)$. Then it has to exist $t(i) \in \mathbb{N}$ such that for every $t \geq t(i)$, $t \in T(i)$.
Since P is irreducible, for every $j \in \mathcal{X}$, let $r(i,j)$ be an integer such that $P^r(i,j) > 0$. Consider $t'(i) := t(i) + \max_{j\in\mathcal{X}} r(i,j)$, we have that $P^t(i,j) > 0 \ \forall j$ and $\forall t \geq t(i)$. Given that i was fixed, if $t \geq \max_{i\in\mathcal{X}}(t'(i))$, $P^t(i,j) > 0$ for every $i, j \in \mathcal{X}$. □

Definition 2.3.3 (Harmonic function)**.** Consider $\mathcal{X}$ a finite set and P a transition matrix. A function $h : \mathcal{X} \to \mathbb{R}$ is *harmonic* at $i \in \mathcal{X}$ if:

$$h(i) = \sum_{j\in\mathcal{X}} P(i,j)h(j)$$

If h is harmonic at every state, that is, if $Ph = h$ we simply say that h is harmonic on $\mathcal{X}$.

Observe that a harmonic function of P is nothing but a right eigenvector with eigenvalue 1, while a stationary distribution is a (nonnegative, normalized to 1) left eigenvector with that eigenvalue.

Lemma 2.3.4. *If P is an irreducible transition matrix then the harmonic functions on $\mathcal{X}$ are exactly the constant functions.*

Proof. Since every row of P has sum equal to 1, any constant vector is harmonic.

For the converse, let h be harmonic and let $i_0 \in \mathcal{X}$ be such that $h(i_0)$ reaches its maximum (which exists because $\mathcal{X}$ is finite). Let us denote $M = h(i_0)$. Since h is harmonic:

$$h(i_0) = M = (Ph)(i_0) = \sum_{j\in\mathcal{X}} P(i_0,j)h(j)$$

Let $k \in \mathcal{X}$ be an arbitrary state:

- If $P(i_0,k) > 0$,

$$M = \sum_{j\in\mathcal{X}} P(i_0,j)h(j) = P(i_0,k)h(k) + \sum_{\substack{j\in\mathcal{X}\\ j\neq k}} P(i_0,j)h(j) \leq$$
$$\leq P(i_0,k)h(k) + (1 - P(i_0,k))M$$

  That is to say
$$M \leq P(i_0,k)h(k) + (1 - P(i_0,k))M,$$
  which implies
$$M \leq h(k).$$

- If $P(i_0,k) = 0$, as P is irreducible, there exists $n_k \in \mathbb{N}$ such that $P^{n_k}(i_0,k) > 0$. As $P^{n_k}h = h$, reasoning in the same way as in the first case, we obtain that $M \leq h(k)$.

Since for every state k in $\mathcal{X}$, $M \leq h(k)$, we conclude that $M = h(i_0) = h(k)$, finishing the proof. □

Proposition 2.3.5. *If P is irreducible then it has a unique stationary distribution.*

Proof. We already saw in Theorem 2.2.9 that a stationary distribution π exists, so we only need to proof uniqueness.

The proof is based on Lemma 2.3.4:
Since P is irreducible, the constant vectors are the only vectors with $Ph = h$, that is, the only vectors with $(P - I)h = 0$. That is, $\dim(\ker(P - I)) = 1$. By rank-nullity theorem,

$$rank(P - I) = |\mathcal{X}| - 1 = rank(P^t - I).$$

Using rank-nullity theorem again, $rank(ker(P^t - I)) = 1$, which means that the equation

$$(P^t - I)v^t = 0$$

has a 1-dimensional space of solutions. This equation is equivalent to $vP = v$. By hypothesis, as $\pi P = \pi$, every other solution must be of the form $v = \lambda\pi$ but as the sum of v must be 1, $\lambda = 1 \Longrightarrow v = \pi$. □

The reciprocal of 2.3.5 is not true: Consider a directed graph with only two vertices, one directed edge 12 and one loop on 2. The transition matrix is

$$P = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

whose stationary distribution is $\pi = (0, 1)$ and unique, but P is reducible, as you can not leave 2 once you get there.

The condition of irreducibility of P on Proposition 2.3.5 is necessary for uniqueness, as if P is reducible, we may have infinite stationary distributions: Consider the $n \times n$ transition matrix

$$P = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

P is reducible since $P^r = P \ \forall r \in \mathbb{N}$ and for every $\pi \in \mathbb{R}^n$, $\pi P = \pi$

Theorem 2.3.6 (Convergence Theorem)**.** *Suppose P is irreducible and aperiodic with stationary distribution π. Then there exist constants $\alpha \in (0, 1)$ and $C > 0$ such that*

$$\max_{i \in \mathcal{X}} \|P^t(i, \cdot) - \pi\|_{TV} \leq C\alpha^t.$$

Proof. Since P is aperiodic and irreducible, Proposition 2.3.2 ensures that there exists r such that the matrix P^r has only positive entries. For ϵ small enough

$$P^r(i,j) \geq \epsilon\pi(j) \quad \text{for every } i,j \in \mathcal{X}.$$

Let $\theta = 1-\epsilon$ and Π be the $|\mathcal{X}| \times |\mathcal{X}|$ matrix with every row equal to π. The following equation defines a stochastic matrix Q:

$$P^r = (1-\theta)\Pi + \theta Q \tag{2.6}$$

In order to prove the theorem we need to show by induction on k that

$$P^{rk} = (1-\theta^k)\Pi + \theta^k Q \tag{2.7}$$

For $k=1$, it is obvious due to (2.6). Assuming that it holds for $k=n$, let us prove it for $k=n+1$:

$$\begin{aligned} P^{r(n+1)} &= [(1-\theta^n)\Pi + \theta^n Q]P^r \\ &= [1-\theta^n]\Pi P^r + (1-\theta)\theta^n Q^n \Pi + \theta^{n+1}Q^{n+1} \end{aligned}$$

Since $\pi P = \pi$, $\pi P^m = (\pi P)P^{m-1} = \cdots = \pi$ for every $m \in \mathbb{N}$, we have that $\Pi P^m = \Pi$. The same occurs for Q^n:

$$(Q^n\Pi)(i,j) = \sum_{k\in\mathcal{X}} Q^n(i,k)\Pi(i,j) = \pi(j)\sum_{k\in\mathcal{X}} Q^n(i,k) = \pi(j) = \Pi(i,j),$$

by the fact that Q is stochastic. Hence:

$$P^{r(n+1)} = [(1-\theta^{n+1})\Pi + \theta^{n+1}Q^{n+1},$$

concluding that the equality (2.7) holds for every k. Multiplying by P^l with $0 \leq l < r$ and reordering terms

$$P^{rk+l} - \Pi = \theta^k(Q^kP^l - \Pi).$$

To conclude the proof we focus on the i-th row of the resulting matrices. If we sum the absolute values of the row and divide by 2, we can use the Proposition 2.1.3. The right term can be bounded by θ^k as $\|Q^kP^l(i,\cdot) - \pi\|_{TV}$ is at most 1. Therefore

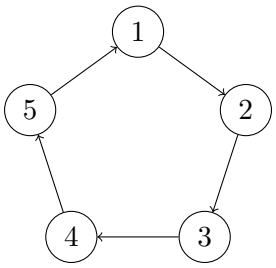
$$\|P^{rk+l}(i,\cdot) - \pi\|_{TV} \leq \theta^k$$

Taking $t = rk+l$, $\alpha = \theta^{1/r} \in (0,1)$ and $C = \alpha^{-r} = 1/\theta > 1$ we finish the proof in the following manner:

$$\|P^{rk+l}(i,\cdot) - \pi\|_{TV} \leq \theta^k = \alpha^{rk} \leq \alpha^{rk}\alpha^{l-r} = \alpha^{-r}\alpha^{rk+l} = C\alpha^t.$$

□

Remark that we need the aperiodicity of P, as otherwise the theorem would not be true:



Example 2.3.7. *Consider a Markov with state space* $\mathcal{X} = \{1, 2, 3, 4, 5\}$*, transition matrix*

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and with the following cycle directed graph:

Consider π *a distribution on the state space* $\mathcal{X}$*. In order for* π *to be stationary it must be a solution of the following system of equations*

$$\begin{cases} \pi(1) = \pi(2) \\ \pi(2) = \pi(3) \\ \cdots \\ \pi(5) = \pi(1) \\ \sum_{i=1}^{5} \pi(i) = 1 \end{cases}$$

hence

$$\pi(1) = \pi(2) = \cdots = \pi(5) = \frac{1}{5}$$

P *is not aperiodic, as g.c.d*$(T(i)) = 5$ *for every state on* $\mathcal{X}$*. Let us show that there does not exist* α *and* C *defined as on Theorem 2.3.6: For every row of the matrix* P^t*,*

$$\|P^t(i, \cdot) - \pi\|_{TV} = \frac{1}{2}\left(\left|1 - \frac{1}{5}\right| + \left|-\frac{1}{5}\right| + \left|-\frac{1}{5}\right| + \left|-\frac{1}{5}\right| + \left|-\frac{1}{5}\right|\right) = \frac{4}{5}.$$

For every pair C_0, α_0*, since* C_0 *is constant and* α_0^t *tends to 0 as* $t \to \infty$*,* $\frac{4}{5} > C_0\alpha_0^t$ *for every* $t \geq t_0$*, which contradicts the Convergence Theorem.*

2.4 Cutoff phenomena

When understanding total variation distance from equilibrium new concepts appear:

Remember that we denote by $d(t)$ the maximum total variation distance at time t:

$$d(t) := \max_{i \in \mathcal{X}} \|P^t(i, \cdot) - \pi\|_{TV}$$

We define the *variation distance mixing time* or simply *mixing time* as

$$t_{mix}(\epsilon) := \min\{t \in \mathbb{N} : d(t) \leq \epsilon\}.$$

It is convenient to speak of the mixing time without referring to an ϵ. For this we consider:

$$t_{mix} := t_{mix}(1/4)$$

For each $n \in \mathbb{N}$, let $\mathcal{X}_n$ be a finite state space. Let $X^{(n)} = (X_0^{(n)}, X_1^{(n)}, X_2^{(n)}, \dots)$ be a Markov Chain with state space $\mathcal{X}_n$. The mixing time for the n-th chain is denoted by $t_{mix}^{(n)}(\epsilon)$.

Definition 2.4.1 (Cutoff)**.** The sequence of chains $X^{(0)}, X^{(1)}, X^{(2)}, \dots$ has a *cutoff* if for every $\epsilon \in (0,1)$

$$\lim_{n\to\infty} \frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1-\epsilon)} = 1 \tag{2.8}$$

Lemma 2.4.2. *A sequence of Markov chains has a cutoff if and only if*

$$\lim_{n\to\infty} d^{(n)}(ct_{mix}^{(n)}) = \begin{cases} 1 & \text{if } 0 < c < 1 \\ 0 & \text{if } c > 1 \end{cases} \tag{2.9}$$

Proof. Assume that (2.9) holds and let $\epsilon > 0$ be fixed.

Given $\gamma \in (0,1)$, (2.9) implies

$$\begin{cases} \lim_{n\to\infty} d^{(n)}\big((1+\gamma)t_{mix}^{(n)}\big) = 0 \\ \lim_{n\to\infty} d^{(n)}\big((1-\gamma)t_{mix}^{(n)}\big) = 1 \end{cases}$$

From the first equation we get that for n large enough $d^{(n)}\big((1+\gamma)t_{mix}^{(n)}\big) < \epsilon$, which implies that

$$t_{mix}^{(n)}(\epsilon) \le (1+\gamma)t_{mix}^{(n)} \tag{2.10}$$

relying on the very definition of $t_{mix}^{(n)}(\epsilon)$. Similarly, the second equation says that for n large enough $d^{(n)}\big((1-\gamma)t_{mix}^{(n)}\big) > 1-\epsilon$, which implies that

$$t_{mix}^{(n)}(1-\epsilon) \ge (1-\gamma)t_{mix}^{(n)}. \tag{2.11}$$

Equations (2.10) and (2.11) give

$$\frac{t_{mix}^{(n)}(\epsilon)}{t_{mix}^{(n)}(1-\epsilon)} \le \frac{1+\gamma}{1-\gamma}$$

Making γ tend to 0 we have (2.8).

Conversely, assume that (2.8) holds. Given $\gamma > 0$, for any $\epsilon > 0$ and for n large enough, $t_{mix}^{(n)}(\epsilon) \le (1+\gamma)t_{mix}^{(n)}$. Indeed, if $\epsilon > 1/4$ then

$$t_{mix}^{(n)}(\epsilon) \le t_{mix}^{(n)} \le (1+\gamma)t_{mix}^{(n)};$$

and if $\epsilon < 1/4$ then

$$t_{mix}^{(n)}(\epsilon) < (1+\gamma)t_{mix}^{(n)}(1-\epsilon) \le (1+\gamma)t_{mix}^{(n)},$$

where the first inequality follows from (2.8), for n large enough.

Hence $\lim_{n\to\infty} d^{(n)}\big((1+\gamma)t_{mix}^{(n)}\big) \leq \epsilon$. Since this holds for every ϵ,

$$\lim_{n\to\infty} d^{(n)}\big((1+\gamma)t_{mix}^{(n)}\big) = 0.$$

The same arguments give $t_{mix}^{(n)}(1-\epsilon) \geq (1-\gamma)t_{mix}^{(n)}$ for n large enough $\lim_{n\to\infty} d^{(n)}\big((1-\gamma)t_{mix}^{(n)}\big) \geq 1-\epsilon$. As this also holds for every ϵ,

$$\lim_{n\to\infty} d^{(n)}\big((1-\gamma)t_{mix}^{(n)}\big) = 1$$

□

Remark 2. *Observe that this proof would work the same taking $t_{mix} := t_{mix}(r)$ for any $r \in (0, 1/2)$, instead of the choice $r = 1/4$.*

Definition 2.4.3 (Cutoff window)**.** Let $(w_n)_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ (equivalently, we can consider $w : \mathbb{N} \to (0,\infty)$ a function). A sequence of Markov chains has a cutoff with a window of size $O(w_n)$ if $w_n = o(t_{mix}^{(n)})$ and

$$\lim_{\alpha\to\infty} \liminf_{n\to\infty} d^{(n)}(t_{mix}^{(n)} - \alpha w_n) = 1$$
$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} d^{(n)}(t_{mix}^{(n)} + \alpha w_n) = 0$$

To understand better the concept of cutoff, we dedicate the following chapter to analyze an example taken from [11, Section 18].

Chapter 3

The lazy biased random walk on a chain

3.1 Lazy biased random walk on a line segment

Consider the lazy nearest-neighbor random walk with bias β on the interval $\mathcal{X}_n = \{0, 1, \ldots, n\}$. When at an interior vertex, the walk remains in its current position with probability $1/2$, moves one step to the right with probability $p/2$ with $p \in (1/2, 1)$ and one step to the left with probability $q/2$ with $q = 1 - p$. When at an end vertex, the walk remains in its current position with probability $1/2$ and moves to the interior vertex with probability $1/2$. The bias β is defined as $\beta := (p - q)/2 = p - 1/2$.

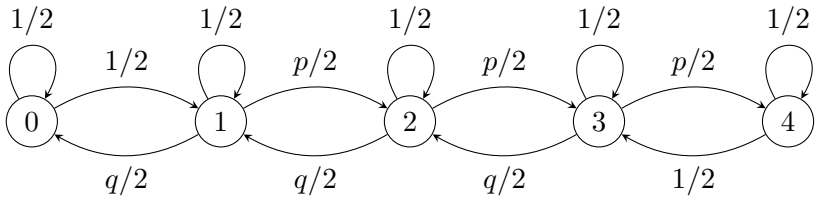


Figure 3.1: Biased random walk with five elements

Theorem 3.1.1. *For each $n \in \mathbb{N}$, let $(X_t^{(n)})_{t\in\mathbb{N}}$ be the lazy random walk with bias $\beta = p - 1/2$ on $\mathcal{X}_n = \{0, 1, 2, \ldots, n\}$. Then, there exists a constant $c(\beta) > 0$ that depends on β only, such that for every $\alpha \in \mathbb{R}$*

$$\lim_{n\to\infty} d^{(n)}(t_n(\alpha)) = \Phi(-c(\beta)\alpha),$$

where $t_n(\alpha) = \beta^{-1}n + \alpha\sqrt{n}$ and Φ denotes the cumulative normal distribution function.

Corollary 3.1.2. *The lazy random walks $(X_t^{(n)})_{t\in\mathbb{N}}$ with bias $\beta = p - 1/2$ on $\mathcal{X}_n = \{0, 1, 2, \ldots, n\}$ have a cutoff at $t = \beta^{-1}n$ with a window of size $O(\sqrt{n})$.*

Let us here prove the corollary from the theorem. The proof of the theorem itself will occupy the rest of this chapter.

Proof. By the definition of cutoff window, we need to show that

$$\lim_{\alpha\to\infty}\liminf_{n\to\infty} d^{(n)}(t_{mix}^{(n)} - \alpha w_n) = 1$$
$$\lim_{\alpha\to\infty}\limsup_{n\to\infty} d^{(n)}(t_{mix}^{(n)} + \alpha w_n) = 0,$$

where the first equality can be rewritten as

$$\lim_{\alpha\to-\infty}\liminf_{n\to\infty} d^{(n)}(t_{mix}^{(n)} + \alpha w_n) = 1.$$

Assuming Theorem 3.1.1, we can distinguish two cases depending on α:

- If $\alpha < 0$
$$\lim_{\alpha\to-\infty}\lim_{n\to\infty} d^{(n)}(\beta^{-1}n + \alpha\sqrt{n}) = \lim_{\alpha\to-\infty}\Phi(-c(\beta)\alpha),$$
which tends to 1 by properties of the cumulative error function combined with the fact that $c(\beta) > 0$.

- If $\alpha > 0$, we have
$$\lim_{\alpha\to\infty}\lim_{n\to\infty} d^{(n)}(\beta^{-1}n + \alpha\sqrt{n}) = \lim_{\alpha\to\infty}\Phi(-c(\beta)\alpha),$$
which tends to 0 due to a same reasoning as on the equality above.

□

Along the proof we use several times the idea of coupling two or more Markov chains (extracted from Chapter 5 of [11]), with the following definition:

Definition 3.1.3 (Coupling)**.** [[11, Chapter 5]] Let $\{X_t\}$ and $\{Y_t\}$ be two Markov chains with state spaces $\mathcal{X}$ and $\mathcal{Y}$ respectively.

A *coupling* of the two chains is any Markov chain $\{(X_t, Y_t)\}$ on $\mathcal{X}\times\mathcal{Y}$ such that forgetting X_t (respectively Y_t) the chain equals Y_t (respectively X_t).

We always assume that the coupling is *Markovian*:

$$P(X_{t+1} = i'|X_t = i, Y_t = j) = P(X_{t+1} = i'|X_t = i),$$
$$P(Y_{t+1} = j'|X_t = i, Y_t = j) = P(Y_{t+1} = j'|Y_t = j).$$

That is, the evolution of one of the variables is independent of where the other one is.

One result that we need about Markovian coupling is:

Corollary 3.1.4 ([11, Chapter 5])**.** *Let P be a transition matrix on $\mathcal{X}$ and suppose that for each pair of states $i, j \in \mathcal{X}$ there is a Markovian coupling (X_t, Y_t) with $X_0 = i$ and $Y_0 = j$. For each coupling, let τ_{couple} be the first time s where $X_s = Y_s$. Then*

$$d(t) \le \max_{i,j\in\mathcal{X}} P_{i,j}(\tau_{couple} > t) \tag{3.1}$$

We also need the following properties about $X^{(n)}$. Let $\theta = q/p$, which we assume to be smaller than 1.

Lemma 3.1.5. *In the lazy random walk $X^{(n)}$ we have*

$$P_1(\textit{hit } 0 \textit{ before } n) = \frac{\theta - \theta^n}{1 - \theta^n}$$

Proof. Remember that $P_1(\cdot)$ denotes the probability of something when the Markov chain starts with $X_0^{(n)} = 1$. That is, we want to compute the probability that the random walk starting at 1 passes through 0 before passing through n.

Our first remark is that this depends only on the trajectory of the walk and not on the time it takes to move, so the probability is the same in $X^{(n)}$ as in the usual biased random walk that moves to the left with probabilty q and to the right with probability p.

We are going to prove the statement by induction on n, starting with $n = 2$:

- For $n = 2$:

  $$P_1(\text{hit 0 before 2}) = q = \frac{\theta}{1-\theta} = \frac{\theta - \theta^2}{1 - \theta^2},$$

  where we use that $\theta = q/p$ and $q + p = 1$ imply that p and q are described by

  $$q = \frac{\theta}{1+\theta}, \qquad p = \frac{1}{1+\theta}$$

- We now assume that the formula holds for $n - 1$. We claim that:

  $$P_1(\text{hit 0 before } n) = q + p\, P_2(\text{hit 1 before } n) P_1(\text{hit 0 before } n)$$

  The explanation of this formula is as follows: Starting at 1 we either move to 0 with probability q (which gives the first summand) or to 2, with probability p. From 2 we need to go to 1 and then to 0, both without hitting n.

  This leads to

  $$P_1(\text{hit 0 before } n) = q \cdot \frac{1}{1 - p \cdot P_2(\text{hit 1 before } n)} \tag{3.2}$$

  If we denote by $A_n := P_1(\text{hit 0 before } n)$ then we have

  $$A_{n-1} = P_1(\text{hit 0 before } n-1) = P_2(\text{hit 1 before } n),$$

  since the Markov chain in $\{0, \ldots, n-1\}$ has the same probabilities as in $\{1, \ldots, n\}$.

  Hence (3.2) can be expressed as:

  $$A_n = q \cdot \frac{1}{1 - p \cdot A_{n-1}} = \frac{1}{1/q - p/q \cdot A_{n-1}} \tag{3.3}$$

  Returning to equation (3.3) and applying induction hypothesis

$$A_n = \frac{1}{\dfrac{1+\theta}{\theta} - \dfrac{\theta - \theta^{n-1}}{\theta(1-\theta^{n-1})}} = \frac{\theta}{\dfrac{(1+\theta)(1-\theta^{n-1}) - (\theta - \theta^{n-1})}{(1-\theta^{n-1})}} = \frac{\theta(1-\theta^{n-1})}{1-\theta^n} = \frac{\theta - \theta^n}{1-\theta^n}$$

□

Lemma 3.1.6. *Let $h \in \mathbb{N}$ and let $\pi^{(n)}$ denote the stationary distribution of $X^{(n)}$. Then*

$$\pi^{(n)}(n-k) \leq \theta^{k-1}$$

for every $k = \{0, \ldots, n\}$. In particular,

$$\pi^{(n)}(\{0, \ldots, n-h\}) \leq \frac{\theta^{h-1}}{1-\theta}.$$

Proof. The second formula follows from the first one by summing the geometric series:

$$\pi^{(n)}(\{0, \ldots, n-h\}) = \sum_{k=h}^{n} \pi^{(n)}(n-k) \leq \sum_{k=h}^{n} \theta^{k-1} = \frac{\theta^{h-1} - \theta^{n}}{1-\theta} \leq \frac{\theta^{h-1}}{1-\theta}$$

Let us first show that for computing the stationary distribution we can also forget the laziness factor of the random walk. That is, we can consider the random walk that moves to the right with probability p and to the left with probability q when at an interior vertex, and that moves with probability one to the closest interior vertex when at an end vertex. For $n = 3$, the transition matrix of $X^{(3)}$ is

$$P^{(3)} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ q/2 & 1/2 & p/2 & 0 \\ 0 & q/2 & 1/2 & p/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Getting rid of the laziness:

$$P^{(3)}_{nolazy} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

For this, let $\pi^{(n)}$ be the stationary distribution of $X^{(n)}$ (that is, $\pi^{(n)}P^{(n)} = \pi^{(n)}$), where $P^{(n)}$ is the transition matrix of the lazy walk. Then, denoting $P^{(n)}_{nolazy}$ the transition matrix of the non-lazy walk we have:

$$P^{(n)} = \frac{I + P^{(n)}_{nolazy}}{2}$$

Multiplying on both sides by $\pi^{(n)}$:

$$\pi^{(n)} = \pi^{(n)}P^{(n)} = \pi^{(n)}\left(\frac{I + P^{(n)}_{nolazy}}{2}\right) = \frac{1}{2}\pi^{(n)} + \frac{1}{2}\pi^{(n)}P^{(n)}_{nolazy},$$

so $\pi^{(n)} = \pi^{(n)}P^{(n)}_{nolazy}$, which implies that $\pi^{(n)}$ is also the stationary distribution of $P^{(n)}_{nolazy}$.

Considering that $\pi^{(n)}$ is the unique solution of the linear system $\pi^{(n)}P^{(n)} = \pi^{(n)}$, we have the following explicit solution:

$$\begin{aligned}
\pi^{(n)}(0) &= q\pi^{(n)}(1) \\
\pi^{(n)}(1) = \pi^{(n)}(0) + q\pi^{(n)}(2) = q\left(\pi^{(n)}(1) + \pi^{(n)}(2)\right) \Longrightarrow \quad & \pi^{(n)}(1) = \frac{q}{p}\pi^{(n)}(2) \\
\pi^{(n)}(2) = p\pi^{(n)}(1) + q\pi^{(n)}(3) = q\left(\pi^{(n)}(2) + q\pi^{(n)}(3)\right) \Longrightarrow \quad & \pi^{(n)}(2) = \frac{q}{p}\pi^{(n)}(3) \\
&\vdots \\
&\pi^{(n)}(n-2) = \frac{q}{p}\pi^{(n)}(n-1) \\
&\pi^{(n)}(n-1) = \frac{1}{p}\pi^{(n)}(n)
\end{aligned}$$

In particular,

$$\pi^{(n)}(i-1) = \theta\pi^{(n)}(i), \quad \text{for } i = 2, \ldots, n-1$$

which gives

$$\pi^{(n)}(n-k) = \theta^{k-1}\pi^{(n)}(n-1) \leq \theta^{k-1}, \quad \text{for } k = 1, \ldots, n-1,$$

as we wanted to show. For $k = 0$ the statement is trivial and for $k = n$ we use

$$\pi^{(n)}(n-n) = \pi^{(n)}(0) = q\pi^{(n)}(1) \leq \theta\pi^{(n)}(1) \leq \theta^{n-1}.$$

□

3.2 Proof of the cutoff Theorem 3.1.1

The proof consists in proving the equality by demonstrating the upper bound and the lower bound separately.

Upper bound. Part 1. First of all, we are going to prove that if $\tau_n := \min\{t \in \mathbb{N} : X_t = n\}$, then:

$$\limsup_{n\to\infty} P_0(\tau_n > t_n(\alpha)) \leq \Phi(-c(\beta)\alpha). \tag{3.4}$$

Let $(S_t)_{t\in\mathbb{N}}$ be a lazy biased random walk with $\mathcal{X} = \mathbb{Z}$. We can write

$$S_t = \widetilde{S}_1 + \widetilde{S}_2 + \cdots + \widetilde{S}_t,$$

where each $\widetilde{S}_i$ is a random variable which represents a step and can be equal to 1 with probability $p/2$, -1 with probability $q/2$ and 0 with probability $1/2$. We can obtain the expected value for each of these random variables as

$$E\left[\widetilde{S}_1\right] = 1 \cdot \frac{p}{2} - 1 \cdot \frac{q}{2} + 0 \cdot \frac{1}{2} = \frac{p-q}{2} = \beta$$

We can calculate the expected value of S_t starting at $x_0^{(n)} = k$ using the linearity of the expected value:

$$E_k\left[S_t\right] = k + E\left[\widetilde{S}_1\right] + E\left[\widetilde{S}_2\right] + \cdots + E\left[\widetilde{S}_t\right] = k + \beta t \tag{3.5}$$

We now couple with (S_t) the following Markov chain (X_t) with state space $\mathcal{X}_+ := \{0, 1, 2, \dots\}$. We define $X_0 = S_0$ and

$$X_{t+1} = \begin{cases} 1 & \text{if } X_t = 0 \text{ and } S_{t+1} - S_t = -1 \\ X_t + S_{t+1} - S_t & \text{otherwise.} \end{cases}$$

Observe that (assuming $k \le n$), for every $t \le \tau_n$ we have that $X_t = X_t^{(n)}$.

Let us to prove that for every $t \le \tau_n$, we have that $X_t^{(n)} \ge S_t$. For this we use induction:

- $X_0^{(n)} = S_0$ by the very definition of the coupling.
- Assume that it holds for $t \le \tau_n$. Let us show that it holds for $t + 1 \le \tau_n$:

$$X_{t+1}^{(n)} = X_t^{(n)} + S_{t+1} - S_t \ge S_{t+1} \iff X_t^{(n)} - S_t \ge 0 \iff X_t^{(n)} \ge S_t,$$

  which is true due to induction hypothesis.

By equation (3.5) $E_0 S_{t_n(\alpha)} = t_n(\alpha)\beta = n + \alpha\beta\sqrt{n}$ and

$$P_0(S_{t_n(\alpha)} < n) = P_0\left(\frac{S_{t_n(\alpha)} - ES_{t_n(\alpha)}}{\sqrt{t_n(\alpha)v}} < \frac{-\alpha\beta\sqrt{n}}{\sqrt{t_n(\alpha)v}}\right),$$

where $v = 1/2 - \beta^2$. By the Central Limit Theorem 1.2.3

$$\limsup_{n\to\infty} P_0(S_{t_n(\alpha)} < n) = \Phi(-c(\beta)\alpha) \tag{3.6}$$

with $c(\beta) = \beta^{3/2}\sqrt{v}$.

Since $X_t^{(n)} = X_t \ge S_t$ for $t \le \tau_n$,

$$P_0\left(\max_{0\le s\le t_n(\alpha)} X_s^{(n)} < n\right) \le P_0\left(\max_{0\le s\le t_n(\alpha)} S_s < n\right) \le P_0(S_{t_n(\alpha)} < n). \tag{3.7}$$

By the definition of τ_n

$$P_0(\tau_n > t_n(\alpha)) = P_0\left(\max_{0\le s\le t_n(\alpha)} X_s^{(n)} < n\right).$$

Combining this with (3.7) and (3.6) we obtain the desired (3.4).

Upper bound. Part 2 We now show that

$$\lim_{n\to\infty} d^{(n)}(t_n(\alpha)) \le \limsup_{n\to\infty} P_0(\tau_n > t_n(\alpha)). \tag{3.8}$$

We couple two copies of $(X_t^{(n)})$ as follows: Consider the Markov chain $\{(X_t, Y_t)\}$ with state space $\mathcal{X}_n \times \mathcal{X}_n$ constructed as follows. At each time t we chose which particle (X or Y) to move, each with probability $1/2$. Once we have decided which particle moves, if the particle is at an interior state it takes a $+1$ step with probability p and a -1 step with

probability q. If the particle is at an end state (i.e. at 0 or n) it moves to the adjacent vertex with probability 1.

Observe that if we forget one of the particles, the Markov chain followed by the other particle is exactly the original $(X_t^{(n)})$). That is, this is a Markovian coupling according to the Definition 3.1.3.

The coupled Markov chains may have different initial distributions. We write $P_{i,j}(\cdot)$ with $i, j \in \mathcal{X}_n$ for the probability of some event in the coupled Markov chain assuming that the first particle starts at i and the second at j.

We call *meeting time* of the coupled chain the first time when the two particles are at the same position. We denote it τ_{couple}.

If $i < j$, the time when the particles meet τ_{couple} is upper bounded by the time it takes for the first particle to hit n. From this we deduce

$$P_{i,j}(\tau_{couple} > t_n(\alpha)) \leq P_0(\tau_n > t_n(\alpha)) \tag{3.9}$$

Combining this with Corollary 3.1.4 we have

$$\limsup_{n\to\infty} d_n(t_n(\alpha)) \leq \Phi(-c(\beta)\alpha)$$

Lower bound. Part 1. Consider $\theta = (q/p)$. In this first step we are going to show that

$$\forall h \in [n], \qquad \limsup_{n\to\infty} P_0(X_{t_n(\alpha)} > n - h) \leq 1 - \Phi(-c(\beta)\alpha) + \theta^{h-1}. \tag{3.10}$$

Consider the coupled chains X_t and S_t of the upper bound, and couple with them a third Markov chain $\widetilde{X}_t$ defined as follows: We define $\widetilde{X}_0 = 1$ and

$$\widetilde{X}_{t+1} = \begin{cases} 1 & \text{if } \widetilde{X}_t = 1 \text{ and } S_{t+1} - S_t = -1 \\ \widetilde{X}_t + S_{t+1} - S_t & \text{otherwise.} \end{cases}$$

Observe that $\widetilde{X}_t$ stays always in $\{1, 2, 3, \dots\}$. However, we interpret it as a lazy biased random walk on $\{0, 1, \dots\}$ that starts at 1 and which "restarts from 1" whenever it should hit 0. When this happens (that is, when $\widetilde{X}_t = 1$ and $S_{t+1} - S_t = -1$) we say that $\widetilde{X}_{t+1}$ "returns to zero".

We have the following properties:

1. $\widetilde{X}_t - X_t \in \{0, 1\}$ for all t. In fact, this difference changes from 1 to 0 and from 0 to 1 everytime that $\widetilde{X}_t$ returns to 0.

   Indeed, let
   $$t_0 = \min\{t \geq 1 : X_t = 0 \wedge \widetilde{X}_t = 1 \wedge S_{t+1} - S_t = -1\}$$
   be the first time when $\widetilde{X}_t$ returns to 0. By the definition of X_t and $\widetilde{X}_t$, $X_t - \widetilde{X}_t = 1$ for every $t \leq t_0$.

   At time $t_0 + 1$ we have that $X_{t_0+1} = 1 = \widetilde{X}_{t_0+1}$ and from there we will have that $X_t = \widetilde{X}_t$ until the time
   $$t_1 := \min\{t > t_0 : X_t = \widetilde{X}_t = 1 \wedge S_{t+1} - S_t = -1\}.$$
   At time t_{t_1+1} we have $X_{t_1+1} = 0$ and $\widetilde{X}_{t_1+1} = 1$, so we are again at the initial situation.

2. In particular, $\widetilde{X}_t \geq X_t$ for all t.

3. $\widetilde{X}_t - S_t = h + 1$ where h is the number of times that $\widetilde{X}_t$ has returned to zero before time t.

   Indeed, at $t = 0$ we have $\tilde{X}_0 = 1 = S_0 + 1$, and by definition of $\tilde{X}$ the difference $\widetilde{X}_t - S_t$ stays constant unless $\tilde{X}$ returns to zero, and when this happens it increases by 1.

As a result, we have that

$$P_0(X_t > x) \leq P_0(\widetilde{X}_t > x), \tag{3.11}$$

for every $x \geq 0$ and that

$$P_0(\widetilde{X}_t - S_t \geq h) = P_0(\text{at least } h-1 \text{ returns of } (\widetilde{X}_t) \text{ to } 0)$$

By Lemma 3.1.5 the chance that a random walk on $\mathbb{N}$ starting on 1, hits 0 before n is $1 - (1-\theta)/(1-\theta^n)$.

Letting $n \to \infty$ we can deduce that the probability of hitting 0 is θ. Hence

$$P_0(\text{at least } h-1 \text{ returns of } (\widetilde{X}_t) \text{ to } 0) = \theta^{h-1} \tag{3.12}$$

By (3.11) and Boole's Inequality 1.2.4

$$\begin{aligned} P_0(X_{t_n(\alpha)} > n-h) &\leq P_0(\widetilde{X}_{t_n(\alpha)} > n-h) = \\ &= P_0(\widetilde{X}_{t_n(\alpha)} + S_{t_n(\alpha)} - S_{t_n(\alpha)} > n - 2h + h) \leq \\ &\leq P_0((\widetilde{X}_{t_n(\alpha)} - S_{t_n(\alpha)} > n-2h) \vee (S_{t_n(\alpha)} > h)) \leq \\ &\leq P_0(S_{t_n(\alpha)} > n-2h) + P_0(X_{t_n(\alpha)} - S_{t_n(\alpha)} \geq h) \end{aligned}$$

Combining the inequality with (3.12) we obtain

$$P_0(X_{t_n(\alpha)} > n-h) \leq P_0(\widetilde{X}_{t_n(\alpha)} > n-h) \leq P_0(S_{t_n(\alpha)} > n-2h) + \theta^{h-1}. \tag{3.13}$$

As shown on the first part of the upper bound by the Central Limit Theorem 2.3.6, for every $h \in \mathbb{N}$

$$\lim_{n\to\infty} P(S_{t_n(\alpha)} > n-2h) = 1 - \Phi(-c(\beta)\alpha). \tag{3.14}$$

We prove (3.10) joining (3.13) with (3.14).

Lower bound. Part 2

Observe that for every subset $S \subset [n]$ we have

$$d_n(t_n(\alpha)) \geq \pi^{(n)}(S) - P_0(X_{t_n(\alpha)} \in S).$$

Taking $S = \{n-h+1, \ldots, n\}$, Lemma 3.1.6 says that $\forall h \in [n]$,

$$\pi^{(n)}(S) \geq 1 - \frac{\theta^{h-1}}{1-\theta}.$$

On the other hand in part 1 we have shown that

$$\limsup_{n\to\infty} P_0(X_{t_n(\alpha)} \in S) \leq 1 - \Phi(-c(\beta)\alpha) + \theta^{h-1}. \tag{3.15}$$

Hence

$$\liminf_{n\to\infty} d_n(t_n(\alpha)) \geq 1 - \frac{\theta^{h-1}}{1-\theta} - (1 - \Phi(-c(\beta)\alpha) + \theta^{h-1}) = \Phi(-c(\beta)\alpha) - \frac{\theta^{h-1}}{1-\theta} - \theta^{h-1}.$$

Since this holds for any h, we can take h going to infinity (for example $h = n/2$) and get

$$\liminf_{n\to\infty} d_n(t_n(\alpha)) \geq \Phi(-c(\beta)\alpha).$$

Chapter 4

The configuration model and its Markov chain

4.1 Definition and properties

The main goal of the remaining chapters is to understand and describe the main results in the paper [3].

In order to build a directed multigraph (directed graph which allows multiple edges) with prescribed degree sequences we start with a vertex set $V = \{1, ..., n\}$ and two sequences of natural numbers $(d_i^-)_{1\leq i\leq n}$ and $(d_i^+)_{1\leq i\leq n}$, with equal sum m which represent the in-degrees and out-degrees of the n vertices. We equip each vertex $i \in V$ with a set E_i^+ of d_i^+ (out-degree) *tails* and a set E_i^- of d_i^- (in-degree) *heads*. We then build the *environment*: a bijection $\omega : \bigcup_i E_i^+ \to \bigcup_i E_i^-$, where $\omega(e) = f$ means arc ef from vertex e to vertex f.

In what follows we denote by Δ the maximum among all the d_i^- and d_i^+ and by δ the minimum, and we assume $\delta \geq 2$ and $\Delta = O(1)$.

As ω is a bijection between two sets of size m, there are $m!$ possible environments.

Definition 4.1.1 (Configuration model)**.** We call *configuration model* the random multigraph obtained giving equal probability to each of the $m!$ possible environments.

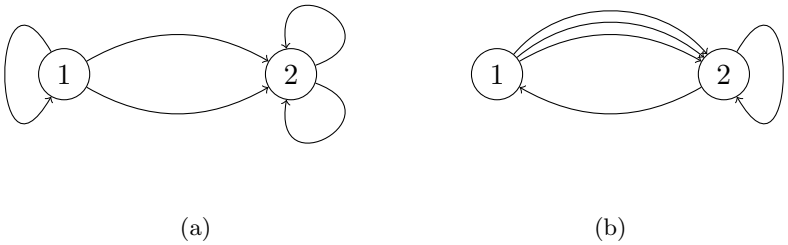
The configuration model was introduced by B. Bollobas for undirected graphs [2].

The fact that we consider $\delta \geq 2$ is that the random multigraph, denoted by G, is strongly connected with high probability (see [5]) which combined with Propositions 2.2.2 and 2.3.5 implies that the stationary distribution is unique with high probability.

An important observation needs to be made: different environments can produce the same graph, so that the number of possible graphs is smaller than $m!$, and different graphs may arise with different probabilities.

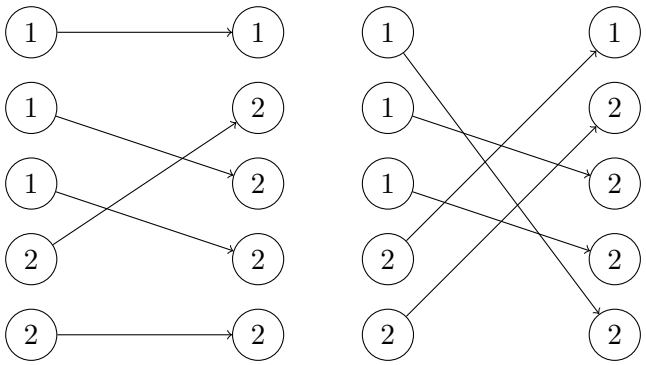
Example 4.1.2. *Let us consider the in-degree sequence* $(1, 4)$ *and the out-degree* $(3, 2)$*, with* $m = 5$*. There are two possible directed graphs, depicted in Figure 4.1. They arise, respectively with probability 6/10 and 4/10, as shown in Example 4.1.3.*

Each environment in Figure 4.2 matches with a different graph in Figure 4.1 but both environments have the same probability $\frac{1}{5!}$



(a) (b)

Figure 4.1: The two possible graphs with the given sequences



(a) (b)

Figure 4.2: Two possible environments

Example 4.1.3. *Suppose we have two vertices 1 and 2 with out-degrees* (d_1^+, d_2^+) *and in-degrees* (d_1^-, d_2^-)*. Without loss of generality suppose that* $\min\{d_1^+, d_2^+, d_1^-, d_2^-\} = d_1^-$*. Then the number of loops at vertex one can be any number between zero and and* d_1^-*, and the number of loops at vertex 1 completely determines the rest of the graph:*
Denote by i *the number of loops on vertex 1 and by* j *the number of loops on vertex 2. The edges 12 and 21 appear with multiplicity* $d_1^+ - i$ *and* $d_2^+ - j$ *respectively. Moreover* i *and* j *are dependent since*

$$d_1^+ - i = d_2^- - j,$$

hence

$$i - j = d_1^+ - d_2^- = d_1^- - d_2^+$$

We want to compute the probability of having each of these possible $d_1^- + 1$ *graphs. The probability that a graph has* i *loops on vertex 1 is:*

$$P(i \text{ loops vertex } 1) = \frac{\binom{d_1^+}{i}\binom{d_2^+}{d_1^- - i}}{\binom{m}{d_1^-}} \tag{4.1}$$

The approach to obtaining the formula 4.1 is as follows: $\binom{m}{d_1^-}$ *represents the number of tail sets that can be joined to the* d_1^- *heads at vertex one. Each of such sets appears with*

the same probability, so the probability of having i loops is the fraction of those sets that consist of exactly i tails from 1 (that is, loops at vertex 1) and $d_1^- - i$ tails from 2 (that is, edges from 2 to 1). These two numbers give the numerator $\binom{d_1^+}{i}\binom{d_2^+}{d_1^- - i}$.

Remark 3. *In the configuration model we say that something happens* with high probability *if the environments for which it happens have probability going to 1 as n goes to infinity for every choice of the degree sequences (as long as every degree is at least two and at most Δ; the upper bound Δ is considered bounded and not depending on n).*

For example, the biased random walk on a chain studied in the previous chapter is a possible environment in the configuration model. Indeed, if, say, $p = 3/5$ and $q = 2/5$ we can mimic the lazy biased chain letting $d_i^+ = d_i^- = 10$ and putting it each vertex i three edges going to the right, two edges going to the left, and five loops.

However, this "environment" will appear with probability going to zero as $n \to \infty$, so the cutoff theorems of the previous chapter are not contradicting the ones in this and the next chapter.

4.2 Sequential generation and local tree structure

One important property of the environment ω is that it can be generated sequentially by repeating m times (where m is the number of edges) the following steps:

1. An unmatched tail e is selected arbitrarily.
2. An unmatched head f is chosen uniformly at random.
3. e is matched with f so that $\omega(e) = f$

It is important to notice that in step 1 we do not need to choose the tail e uniformly at random. Any rule for step 1 produces the uniform environment.

Theorem 4.2.1. *Given a sequence of in-degrees and out-degrees (with respective sum m), any random graph built with this algorithm has the same probability as in the configuration model.*

Proof. By definition, in the configuration model each environment ω has the same probability, equal to $\frac{1}{m!}$. We want to show that the sequential algorithm gives this same probability to them.

For this, let ω be an environment, that is to say, a bijection between the m tails and m heads, and let us calculate the probability of getting that ω with the sequential algorithm:

The first unmatched tail is chosen arbitrarily from the set of m tails. There are m possible heads to build an edge, and we choose one uniformly at random, so the probability of choosing precisely the one that is in ω is $1/m$. We select another unmatched tail and match it with one of the $m-1$ remaining heads. The probability of choosing the one in ω is $1/(m-1)$. We repeat this until there are no tails/heads left. As each of these experiments of choosing head to match a tail is independent, the probability of obtaining the resulting environment is:

$$\frac{1}{m} \cdot \frac{1}{m-1} \cdots \frac{1}{1} = \frac{1}{m!}$$

□

During the sequence of matching tail to heads we can say that a *collision* occurs whenever a head, whose endpoint i was alive (some tail $e \in E_i^+$ or head $f \in E_i^-$ had previously been picked), gets chosen. When the k-th head gets chosen, since less than $2k$ vertices are alive, (equality holds only when there are no collisions), less than $2\Delta k$ of the $m-k+1$ possible choices for the head can cause a collision, as at most every alive vertex has Δ heads. Since the head is chosen uniformly at random, the probability that the k-th arc results in a collision is less than

$$\frac{2\Delta k}{m-k+1}. \tag{4.2}$$

In the next result, for two random variables A and B with integer values we say that B stochastically dominates A if $P(A \geq k) \leq P(B \geq k)$ for all k.

Lemma 4.2.2. *Let $1 \leq k \leq m$, and assume $2\Delta k \leq m-k+1$. Then, in any sequential generation of the environment for the configuration model the number Z_k of collisions caused by the first k arcs is stochastically dominated by a $\mathrm{Bin}(k, \frac{2\Delta k}{m-k+1})$ random variable.*

Proof. We want to prove that

$$P(Z_k \geq i) \leq P(Binomial\left(k, \frac{2\Delta k}{m-k+1}\right) \geq i)$$

Suppose that the statement is true for $k-1$. Let us prove that it is true for k:

$$\begin{aligned} P(Z_k \geq i) &= P\left((Z_{k-1} \geq i) \cup (Z_k = i | Z_{k-1} = i+1)\right) = \\ &= P(Z_{k-1} \geq i) + P(Z_k = i | Z_{k-1} = i+1) \leq \end{aligned}$$

Applying induction hypothesis and the inequality 4.2

$$\begin{aligned} &\leq P(\mathrm{Bin}\left(k-1, \frac{2\Delta(k-1)}{m-(k-1)+1}\right) \geq i) + \frac{2\Delta k}{m-k+1} \leq \\ &\leq P\Big(Bin\left(k-1, \frac{2\Delta k}{m-k+1}\right) \geq i) + \\ &\qquad + P(\mathrm{Bin}\left(k, \frac{2\Delta k}{m-k+1}\right) = i \Big| \mathrm{Bin}\left(k-1, \frac{2\Delta k}{m-k+1}\right) = i-1\Big) = \\ &= P(\mathrm{Bin}\left(k, \frac{2\Delta k}{m-k+1}\right) \geq i) \end{aligned}$$

□

In particular:

Corollary 4.2.3. *In any sequential generation of the configuration model the number Z_k of collisions caused by the first k arcs satisfies:*

$$P(Z_k \geq 1) \leq \frac{2\Delta k^2}{m-k+1} \text{ and } P(Z_k \geq 2) \leq \frac{2\Delta^2 k^4}{(m-k+1)^2}$$

In order to prove the first inequality we will use the next theorem:

Theorem 4.2.4 (Bernouilli's inequality)**.** *Let $x \in \mathbb{R}$ such that $1+x > 0$ and $n \in \mathbb{N}$. Then $(1+x)^n \geq 1+nx$.*

Proof. Let us prove the result by induction on n:
Let $n = 1$. It is obvious that $1 + x \geq 1 + x$.
Suppose that the statement is true for $n = k-1$ and let us prove that it is true for $n = k$:

$$(1+x)^k = (1+x)^{k-1} \cdot (1+x)$$

Using induction hypothesis, and the fact that $n \geq 1$ and $x^2 \geq 0\ \forall x \in \mathbb{R}$, we conclude the prove:

$$(1+x)^{k-1} \cdot (1+x) \geq (1+(k-1)x) \cdot (1+x) =$$
$$= 1 + x + (k-1)x + (k-1)x^2 = 1 + kx + (k-1)x^2 \geq 1 + kx.$$

Observe that the first inequality needs the hypothesis $1 + x > 0$. □

Corollary 4.2.5. *Let $k \in \mathbb{N}$ and $p \in [0,1]$. Then $P(Binomial(k,p) \geq 1) \leq kp$ and $P(Binomial(k,p) \geq 2) \leq k^2p^2$.*

Proof. The first inequality is easily proved thanks to Bernouilli's inequality 4.2.4:

$$P(Binomial(k,p) \geq 1) = 1 - P(Binomial(k,p) = 0) = 1 - \binom{k}{0} \cdot p^0 \cdot (1-p)^k =$$
$$= 1 - (1-p)^k \leq 1 - (1-kp) = kp$$

For the second inequality:

$$P(Binomial(k,p) \geq 2) = 1 - P(Binomial(k,p) = 1) - P(Binomial(k,p) = 0)$$

We are not using Bernouilli's inequality in this step, as we would obtain a looser bound.

$$= 1 - \binom{k}{1} p(1-p)^{k-1} - \binom{k}{0}(1-p)^k = 1 - kp(1-p)^{k-1} - (1-p)^k =$$
$$= 1 - (1-p)^{k-1}(kp + (1-p)) = 1 - (1-p)^{k-1}((k-1)p + 1).$$

This time we will employ the Bernouilli's inequality in order to conclude the proof:

$$1 - (1-p)^{k-1}((k-1)p+1) \leq 1 - (1-(k-1)p)(1+(k-1)p) = 1 - (1-(k-1)^2p^2) =$$
$$= 1 - 1 + (k-1)^2p^2 \leq k^2p^2$$

□

Proof of Corollary 4.2.3. The proof is based on the stochastic dominance displayed on Lemma 4.2.2 and Corollary 4.2.5 and goes as follows:
For the first inequality:

$$P(Z_k \geq 1) \leq P(Binomial\left(k, \frac{2\Delta k}{m-k+1}\right) \geq 1) \leq \frac{2\Delta k^2}{m-k+1}$$

For the second inequality we reason the same way:

$$P(Z_k \geq 2) \leq P(Binomial\left(k, \frac{2\Delta k}{m-k+1}\right) \geq 2) \leq \frac{4\Delta^2 k^4}{(m-k+1)^2}$$

□

Definition 4.2.6 (Forward ball)**.** The *forward ball* of radius t around a vertex $i \in V$ is the subgraph $\mathcal{B}^+(i,t) \subseteq G$ induced by the directed paths of length t from i. We can generate it sequentially choosing first those tails e from vertices at the lowest distance from i until the minimal distance is greater than t.

Proposition 4.2.7. *Let $h = \lfloor \frac{\ln n}{10 \ln \Delta} \rfloor$. Then, with high probability, G is locally tree-like in the following sense:*

$$\forall i \in V, \mathcal{B}^+(i, 2h) \text{ is either a directed tree, or a directed tree with an extra arc.}$$

Proof. Consider the configuration model built sequentially from a certain vertex i in a breath-first search manner: first choose the neighbors of i, then the neighbors of the neighbors, and so on.

Let k be the number of edges that have been selected when we finish constructing the forward ball $\mathcal{B}^+(i, 2h)$. We have

$$k \leq \Delta + \cdots + \Delta^{2h} \leq 2\Delta^{2h} \leq 2\Delta^{\frac{\ln n}{5 \ln \Delta}} = 2n^{\frac{\ln \Delta}{5 \ln \Delta}} = 2n^{\frac{1}{5}},$$

where in the first inequality we use that $\Delta \geq 2$ and n the first equality that $a^{\ln b} = b^{\ln a}$.

By Corollary 4.2.3 we have that

$$P(Z_k \geq 2) \leq \frac{2\Delta^2 k^4}{(m-k+1)^2} \leq \frac{32\Delta^2 n^{4/5}}{(2n - 2n^{1/5} + 1)^2}$$

Thus, for each vertex i the probability of having more than one collision in the forward ball of i is bounded above by $O(n^{-6/5})$. The probability that this happens for one of the n vertices is bounded by the sum of the individual probabilities, that is, by $O(n^{-1/5})$. Hence, with high probability the forward boalls of all vertices have at most one collison. □

4.3 The uniform random walk in the configuration model

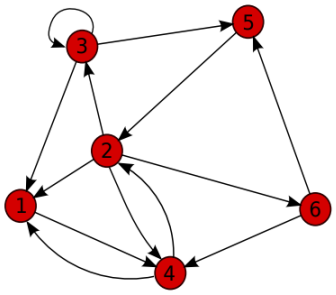
A random walk on the directed graph G of the configuration model is now defined as the discrete-time Markov chain with state space V and transition matrix

$$P(i,j) = \frac{1}{d_i^+} card\left\{ e \in E_i^+ : \omega(e) \in E_j^- \right\}.$$

That is, from each vertex $i \in V$ we take with equal probability $1/d_i^+$ any of the d_i^+ edges that have i as tail.

A *directed path* of length t from $i \in V$ to $j \in V$ is a sequence of arcs $\mathfrak{p} = (e_1 f_1, ..., e_t f_t)$ with $e_k \in E_{i_{k-1}}^+$, $f_k \in E_{i_k}^-$ for some sequence of vertices $\{i_k\}_k \in V$ such that $i_0 = i$ and $i_t = j$, and where the weight of the path is defined by

$$\mathbf{w}(\mathfrak{p}) = \frac{1}{d_{i_0}^+ ... d_{i_{t-1}}^+} \leq \frac{1}{\delta^t} \leq \frac{1}{2^t}.$$



$\mathcal{X} = \{1, \ldots, 6\}$

0	0	0	1	0	0
$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$
$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0
$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
0	1	0	0	0	0
0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0

Figure 4.3: Directed graph with its transition matrix [4]

Let P_{ij}^t be the set of all directed paths of length t from $i \in V$ to $j \in V$. The probability that a walk starting from i reaches j at time t (in t steps) can be expressed as:

$$P^t(i,j) = \sum_{\mathfrak{p} \in P_{ij}^t} \mathbf{w}(\mathfrak{p})$$

Let $V_\star$ denote the set of vertices $i \in V$ such that $\mathcal{B}^+(i,h)$ is a directed tree. Combining Proposition 4.2.7 with the fact that all out-degrees are at least 2, we can prove the following proposition:

Proposition 4.3.1. *With high probability,*

$$\forall i \in V, \forall \ell \in \mathbb{N}, \qquad P^\ell(i, V \backslash V_\star) \leq 2^{-min(\ell,h)}$$

Proof. Recall that $V \backslash V_\star$ are the vertices from where you can have a collision in h steps. By Proposition 4.2.7 with high probability we can suppose that $\mathcal{B}^+(i,2h)$ has at most one collision for every vertex $i \in V$. We distinguish the case $\ell \geq h$ from $\ell < h$:

- $\ell \geq h$: Let $\widetilde{V}_i^{\ell-h}$ be the set of vertices reachable from i in $\ell - h$ steps. By hypothesis $\mathcal{B}^+(j,2h)$ has at most one collision. We can bound $P^\ell(i, V \backslash V_\star)$ as:

$$P^\ell(i, V \backslash V_\star) = \sum_{j \in \widetilde{V}_i^{\ell-h}} P^{\ell-h}(i,j) \cdot P^h(j, V \backslash V_\star) \leq \max_{j \in \widetilde{V}_i^{\ell-h}} P^h(j, V \backslash V_\star),$$

  so we can bound the right term. If $\mathcal{B}^+(j,2h)$ has no collision, then $P^h(j, V \backslash V_\star) = 0$. If there is exactly one collision two cases may occur:

  - If the collision occurs in $k > h$ steps, then only one of the more than 2^h possible walks from j leads to a vertex from which we reach a collision: From j we reach a vertex x, which is in $V \backslash V_\star$, since we can reach a collision from x in $k - h$ steps. Hence, $P^h(j, V \backslash V_\star)$ is $\leq 2^{-h}$.
  - If the collision occurs in $k \leq h$ steps, then there are two subcases: (a) the head of the collision lies in the (unique) path from j that leads to the tail of the collision. Then, a cycle is formed in $\mathcal{B}^+(j,2h)$, and it is the only cycle because

it is the only collision. In this case, the only path leading from j to a vertex in $V \setminus V_\star$ is the path that first goes from j to the collision and then stays in the cycle. This is one among the more than 2^{-h} possible paths, so its probability is less than 2^{-h}. (b) the head of the collision does not lie in the path from j to the collision. In this case all paths of length h from j end in vertices of $V_\star$ and $P^h(j, V \setminus V_\star)$ is zero.

- $\ell \leq h$: In this case, instead of working with $\mathcal{B}^+(j, 2h)$ for the vertices $j \in \widetilde{V}_i^{\ell-h}$ we use a similar reasoning directly with the ball $\mathcal{B}^+(i, h+\ell) \subseteq \mathcal{B}^+(i, 2h)$. This ball has at most one collision and the same case distinction gives us bounds of 2^{-h} or $2^{-\ell}$ depending on whether the collision happens after or before h steps. Since in this case $2^{-h} \leq 2^{-\ell} = 2^{-min(\ell,h)}$ we are done.

□

Chapter 5

Cutoff in the configuration model

5.1 Statement of the cutoff theorem and idea of the proof

In this chapter we study the cutoff behaviour of the configuration model. First, we will show that the mixing time of a random walk under the configuration model is logarithmic in the number n of nodes. For this, let us define

$$t_\star := \frac{\ln n}{\mu}$$

where μ is constant depending on the input data and defined as

$$\mu := \frac{1}{m}\sum_{i=1}^{n} d_i^- \ln d_i^+$$

Since $\ln \delta \leq \ln d_i^+ \leq \ln \Delta$ and $m = \sum_i d_i^- = \sum_i d_i^+$ we have

$$\ln \delta \leq \mu \leq \ln \Delta.$$

Recall that $\mathcal{D}_i(t)$ denotes the total variation distance from equilibrium at time t:

$$\mathcal{D}_i(t) = \|P^t(i,\cdot) + \pi\|_{TV}$$

With this we can state the main result in this chapter, saying that the distance from equilibrium exhibits a *cutoff* at time $t_\star$.

Theorem 5.1.1. *For $t = \lambda t_\star + o(t_\star)$ with $\lambda > 0$, we have*

$$\lambda < 1 \Longrightarrow \min_{i\in V} \mathcal{D}_i(t) \xrightarrow[n\to\infty]{P} 1,$$
$$\lambda > 1 \Longrightarrow \max_{i\in V} \mathcal{D}_i(t) \xrightarrow[n\to\infty]{P} 0$$

To prove this theorem, instead of working with the stationary distribution π we use as an approximation the following one. Denote $\pi_h^- := \pi_0^- P^h$ where π_0^- is the in-degree distribution, that is, $\pi_0^-(j) := \frac{d_j^-}{m}$ for every $j \in V$.

When h is sufficiently big π_h^- is close to π. In what follows we fix h to be the same as in Proposition 4.2.7. That is to say:

$$h := \left\lfloor \frac{\ln n}{10 \ln \Delta} \right\rfloor .$$

The following statement is the analogue to Theorem 5.1.1 with the distribution π_h^- instead of the stationary distribution π:

Theorem 5.1.2. *For h as above and letting*

$$\widetilde{\mathcal{D}}_i(t) := \|P^t(i,\cdot) - \pi_h^-\|_{TV},$$

and with $t = \lambda t_\star + o(t_\star)$ and $\lambda > 0$, we have

$$\lambda < 1 \Longrightarrow \min_{i \in V} \widetilde{\mathcal{D}}_i(t) \xrightarrow[n\to\infty]{P} 1,$$
$$\lambda > 1 \Longrightarrow \max_{i \in V} \widetilde{\mathcal{D}}_i(t) \xrightarrow[n\to\infty]{P} 0$$

The proof of Theorem 5.1.2 occupies most of the rest of this chapter, but before starting the proof let us see how it implies Theorem 5.1.1.

Corollary 5.1.3. *With $\widetilde{\mathcal{D}}$ defined as in Theorem 5.1.2 we have that*

$$\sup_{i \in V, t \in \mathbb{N}} |\widetilde{\mathcal{D}}_i(t) - \mathcal{D}_i(t)| \xrightarrow[n\to\infty]{P} 0.$$

Proof. By Theorem 5.1.2 we have in particular that for (say) $t \geq 2t_\star$

$$\max_{i \in V} \|P^t(i,\cdot) - \pi_h^-\|_{TV} \xrightarrow[n\to\infty]{\mathrm{P}} 0.$$

By convexity of the total variation (Proposition 2.1.5), this is true for any starting distribution, that is, putting $\pi_0 P^t$ instead of $P^t(i,\cdot)$, for an arbitrary π_0. In particular, if we consider $\pi_0 = \pi$, since $\pi P^t = \pi$, we have

$$\|\pi - \pi_h^-\|_{TV} \xrightarrow[n\to\infty]{\mathrm{P}} 0.$$

By the triangle inequality we have

$$\sup_{i \in V,\, t \in \mathbb{N}} |\widetilde{\mathcal{D}}_i(t) - \mathcal{D}_i(t)| \xrightarrow[n\to\infty]{\mathrm{P}} 0.$$

□

That is, Theorem 5.1.1 is equivalent to Theorem 5.1.2.

To prove Theorem 5.1.2 we introduce the following notation and state Propositions 5.1.4 and 5.1.5 which will be proved in the next sections.

Recall that we denote P_{ij}^t the set of paths of length t from node i to node j, and that for each $\mathfrak{p} \in P_{ij}^t$ we denote $\mathbf{w}(\mathfrak{p})$ the weight of the path, which is the probability that

starting at i the next t steps follow the path $\mathfrak{p}$. With this notation we now define for each vertex i, number of steps t, and threshold weight θ

$$\mathcal{Q}_{i,t}(\theta) := \sum_{j\in V}\sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p})1_{\mathbf{w}(\mathfrak{p})>\theta}.$$

That is, $\mathcal{Q}_{i,t}(\theta)$ equals the sum of weights of all paths of length t and weight greater than θ starting at i. Observe that $\mathcal{Q}_{i,t}(\theta)$ is a decreasing function of θ with $\mathcal{Q}_{i,t}(0) = 1$ and $\mathcal{Q}_{i,t}(1) = 0$.

With this, Theorem 5.1.2 decomposes in the following two steps:

Proposition 5.1.4 ([3, Prop. 7])**.** *For any $t = t(n)$:*

$$\forall i \in V, \qquad \widetilde{\mathcal{D}}_i(t) \geq \mathcal{Q}_{i,t}\left(\frac{\ln^3 n}{n}\right) - o_P(1),$$
$$\max_{i\in V}\widetilde{\mathcal{D}}_i(t) \leq \max_{i\in V}\mathcal{Q}_{i,t}\left(\frac{1}{n\ln^3 n}\right) + o_P(1).$$

where $o_P(1)$ is a term that converges to 0 in probability as n tends to infinity.

Proposition 5.1.5 ([3, Prop. 8])**.** *Consider $t = \Theta(\ln n)$ and θ depending arbitarily on n:*

1. *If $\frac{\mu t+\ln\theta}{\sqrt{t}} \longrightarrow \infty$ as $n \longrightarrow \infty$, then*

$$\max_{i\in V}\mathcal{Q}_{i,t}(\theta) \xrightarrow[n\to\infty]{P} 0.$$

2. *If $\frac{\mu t+\ln\theta}{\sqrt{t}} \to -\infty$ as $n \longrightarrow \infty$, then*

$$\min_{i\in V}\mathcal{Q}_{i,t}(\theta) \xrightarrow[n\to\infty]{P} 1.$$

Proof of Theorem 5.1.2. Let us consider

$$t = \lambda\frac{\ln n}{\mu} + o\left(\frac{\ln n}{\mu}\right) = \Theta(\ln n)$$

- If $\lambda < 1$, we are going to lower bound $\min_{i\in V}\widetilde{\mathcal{D}}_i(t)$ by 1, to prove the first implication of Theorem 5.1.2:

  Take $\theta = \frac{\ln^3 n}{n}$,

$$\lim_{n\to\infty}\frac{\mu t+\ln\theta}{\sqrt{t}} = \lim_{n\to\infty}\frac{\mu\lambda\frac{\ln n}{\mu} + o(\ln n) + \ln\left(\frac{\ln^3 n}{n}\right)}{\sqrt{\lambda\frac{\ln n}{\mu} + o\left(\frac{\ln n}{\mu}\right)}} =$$
$$= \lim_{n\to\infty}\frac{\lambda\ln n + o(\ln n) + \ln(\ln^3 n) - \ln n}{\sqrt{\lambda\frac{\ln n}{\mu} + o\left(\frac{\ln n}{\mu}\right)}}$$

  The denominator is greater than 0 and since $\lambda < 1$, for n sufficiently large the numerator is smaller than 0. These implies that the limit tends to $-\infty$ as $n \to \infty$. This limit implies that we can use the second result of Proposition 5.1.5 which combined with the lower bound on Proposition 5.1.4 leads to the wanted implication.

- For $\lambda > 1$ we establish an upper bound for $\max_{i\in V} \widetilde{\mathcal{D}}_i(t)$ of 0. Take $\theta = \frac{1}{n\ln^3 n}$

$$\lim_{n\to\infty} \frac{\mu t + \ln\theta}{\sqrt{t}} = \lim_{n\to\infty} \frac{\mu\lambda\frac{\ln n}{\mu} + o(\ln n) + \ln\left(\frac{1}{n\ln^3 n}\right)}{\sqrt{\lambda\frac{\ln n}{\mu} + o\left(\frac{\ln n}{\mu}\right)}} =$$
$$= \lim_{n\to\infty} \frac{\lambda\ln n + o(\ln n) - \ln n - \ln(\ln^3 n)}{\sqrt{\lambda\frac{\ln n}{\mu} + o\left(\frac{\ln n}{\mu}\right)}}$$

  In this case, for n large enough the numerator is < 0 since $\lambda > 1$, hence the numerator tends to infinity as n tends to infinity. Combining the first result of Proposition 5.1.5 with the upper bound on Proposition 5.1.4 we conclude the proof of the theorem.

□

So, we have a proof of Theorem 5.1.2 except for the fact that we need to prove the two propositions 5.1.4 and 5.1.5. For lack of space we are not going to say anything about the proof of Proposition 5.1.5 and we are going to give an incomplete one for Proposition 5.1.4. Still, this "incomplete" proof will occupy the rest of this chapter. The missing part is that in Section 5.3 we state without proof the very technical Proposition 5.3.2.

5.2 Proof of the lower-bound in Proposition 5.1.4

Proof of the lower-bound. Consider ω an environment, i, j two vertices of V, $\widetilde{\pi}$ a probability distribution on V, $t \in \mathbb{N}$ and $\theta \in (0,1)$. As

$$P^t(i,j) = \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})\le\theta} + \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})>\theta}, \tag{5.1}$$

we deduce that

$$P^t(i,j) \ge \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})\le\theta}. \tag{5.2}$$

When equality holds on 5.2, then

$$\widetilde{\pi}(j) - \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})\le\theta} \le [\widetilde{\pi}(j) - P^t(i,j)]_+$$

where $[\widetilde{\pi}(j) - P^t(i,j)]_+ = \max_{j\in V}\left(\widetilde{\pi}(j) - P^t(i,j), 0\right)$. If the inequality 5.2 is strict, there must exist $\mathfrak{p} \in P_{ij}^t$ such that $\mathbf{w}(\mathfrak{p}) > \theta$, which implies that

$$\widetilde{\pi}(j) - \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})\le\theta} \le \widetilde{\pi}(j) = \widetilde{\pi}(j)\, 1_{P^t(i,j)>\theta}$$

due to $1_{P^t(i,j)>\theta} = 1$ by the definition of $P^t(i,j)$ and because of the path from i to j whose weight exceeds θ.

In either case

$$\widetilde{\pi}(j) - \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})\leq\theta} \leq [\widetilde{\pi}(j) - P^t(i,j)]_+ + \widetilde{\pi}(j) 1_{P^t(i,j)>\theta}$$

By Equation 5.1

$$\widetilde{\pi}(j) - P^t(i,j) + \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})>\theta} \leq [\widetilde{\pi}(j) - P^t(i,j)]_+ + \widetilde{\pi}(j) 1_{P^t(i,j)>\theta} \tag{5.3}$$

We are now going to sum over all j, obtaining the following equalities

$$\sum_{j\in V} \widetilde{\pi}(j) = \sum_{j\in V} P^t(i,j) = 1$$
$$\sum_{j\in V} \sum_{\mathfrak{p}\in P_{ij}^t} \mathbf{w}(\mathfrak{p}) 1_{\mathbf{w}(\mathfrak{p})} = \mathcal{Q}_{i,t}(\theta)$$
$$\sum_{j\in V} [\widetilde{\pi}(j) - P^t(i,j)]_+ = \|\widetilde{\pi} - P^t(i,\cdot)\|_{TV},$$

where in the last equality we use that $[\widetilde{\pi}(j) - P^t(i,j)]_+$ is nonzero if and only if $\widetilde{\pi}(j) \geq P^t(i,j)$. Hence,

$$\mathcal{Q}_{i,t}(\theta) \leq \|\widetilde{\pi} - P^t(i,\cdot)\|_{TV} + \sum_{j\in V} \widetilde{\pi}(j) 1_{P^t(i,j)>\theta}. \tag{5.4}$$

For the last term we use the Cauchy-Schwartz inequality. We regard the sum as the dot product of the following two vectors:

- $\widetilde{\pi}$: The probability distribution on V.
- $v_{P^t(i,j)>\theta}$: a vector in $\{0,1\}^{|V|}$, where each position j is 1 if $P^t(i,j) > \theta$ for every $j \in V$. At most $\frac{1}{\theta}$ positions are 1 otherwise the sum of the i-th row of the matrix P^t would be greater than 1.

Hence, Cauchy-Schwartz gives

$$\sum_{j\in V} \widetilde{\pi}(j) 1_{P^t(i,j)>\theta} \leq \|\widetilde{\pi}\|_2 \cdot \|v_{P^t(i,j)>\theta}\|_2$$

As we said before, there can be at most $\frac{1}{\theta}$ non zero terms on $v_{P^t(i,j)>\theta}$. Then

$$\sum_{j\in V} \widetilde{\pi}(j) 1_{P^t(i,j)>\theta} \leq \sqrt{\sum_{j\in V} \widetilde{\pi}(j)^2} \sqrt{\frac{1}{\theta}} = \sqrt{\frac{1}{\theta} \sum_{j\in V} \widetilde{\pi}(j)^2}$$

Applying this bound on Equation 5.4

$$\mathcal{Q}_{i,t}(\theta) \leq \|\widetilde{\pi} - P^t(i,\cdot)\|_{TV} + \sqrt{\frac{1}{\theta} \sum_{j\in V} \widetilde{\pi}(j)^2} \tag{5.5}$$

We now need to show that the last term is $o_P(1)$ when we take $\widetilde{\pi} = \pi_h^-$ and $\theta = \frac{\ln^3 n}{n}$. In fact, let us prove that a stronger equality holds:

$$E\left[\sum_{j\in V}(\pi_h^-)^2(j)\right] = O\left(\frac{\ln^2 n}{n}\right)$$

Since $\pi_h^- = \pi_0^- P^h$, we can interpret the left term as

$$\sum_{j\in V} P_{\pi_0^-}(X_h(j) = Y_h(j)) = P_{\pi_0^-}(X_h = Y_h),$$

where $(X_k)_{0\leq k\leq h}$ and $(Y_k)_{0\leq k\leq h}$ are two independent walks starting with the same distribution π_0^-. In order to build $(X_k)_{0\leq k\leq h}$, we choose X_0 according to π_0^-. The following steps can be repeated for every $k \geq 1$:

- We choose a tail e from the vertex X_{k-1} uniformly at random.
- If e is unmatched, we choose a head f uniformly at random.
- Let X_k be the endpoint of $\omega(e)$.

We generate by a similar process $(Y_k)_{0\leq k\leq h}$. At most $2h$ arcs are formed. $X_h = Y_h$ may occur either because $X_0 = Y_0$, which has probability $\frac{\Delta}{m}$, or due to a collision, which has probability bounded from above by $\frac{2\Delta(2h)^2}{m-2h+1}$ according to Lemma 4.2.2. Substituting h by its value $\lfloor \frac{\ln n}{10\ln\Delta} \rfloor$, we obtain an upper bound of the order of $O\left(\frac{\ln^2 n}{n}\right)$. Considering $\theta = \frac{\ln^3 n}{n}$ and taking minimum value for every $i \in V$ on 5.5 we have

$$\min_{i\in V} \mathcal{Q}_{i,t}\left(\frac{\ln^3 n}{n}\right) - o_P(1) \leq \min_{i\in V} \|\widetilde{\pi} - P^t(i,\cdot)\|_{TV}$$

□

5.3 Proof of the upper-bound in Proposition 5.1.4 (Sketch)

For lack of space, and because this proof is quite complicated (six pages in [3]) we only give a sketch of the proof of the upper bound. What we do not include is the proof of Proposition 5.3.2 (Proposition 10 in [3]).

In order to calculate $\mathcal{Q}_{i,t}(\theta)$, we are going to estimate $P^t(i,j)$ considering only certain paths, defined as follows. Recall that we have defined

$$h = \left\lfloor \frac{\ln n}{10\ln\Delta} \right\rfloor.$$

Definition 5.3.1 (Nice path)**.** A path $\mathfrak{p}$ starting at $i \in V$ of length t is a *nice path* if it satisfies:

- $\mathbf{w}(\mathfrak{p}) \leq \frac{1}{n\ln^2 n}$.

- The first $t-h$ steps are contained in a tree $\mathcal{T}_i \subseteq \mathcal{B}^+(i, t-h) \subset G$, constructed sequentially as explained in Section 4.2, with a certain rule for step 1 that we are not going to specify (see details in Section 6.2 of [3]).
- Let j be the vertex after $t-h$ steps in the path and k the end of the path. Then, the last h steps in the path are the only path of length at most h from j to k.

Remark that

$$P^t(i,j) = \sum_{\mathfrak{p} \in P^t_{ij}} \mathbf{w}(\mathfrak{p}). \tag{5.6}$$

We denote by $P_0^t(i,j)$ the sum 5.6 restricted only to nice paths, i.e

$$P_0^t(i,j) = \sum \{\mathbf{w}(\mathfrak{p}) : \mathfrak{p} \text{ nice path of length } t \text{ from } i \text{ to } j\}.$$

Proposition 5.3.2 ([3, Proposition 10])**.** *Let $\epsilon > 0$ and $t = t_\star + o(t_\star)$. Then with high probability:*

1. *Every pair of vertices $i, j \in V$ satisfies*

$$P_0^t(i,j) \leq (1+\epsilon)\pi_h^-(j) + \frac{\epsilon}{|V|} \tag{5.7}$$

2. *Every vertex $i \in V_\star$ satisfies $1 - P_0^t(i,V) \leq \mathcal{Q}_{i,t}\left(\frac{1}{n \ln^2 n}\right) + \epsilon$*

We then have the following:

Lemma 5.3.3. *Let $\epsilon > 0$ and $t = t_\star + o(t_\star)$. Then with high probability:*

$$\|\pi_h^- - P^t(i,\cdot)\|_{TV} \leq 1 - P_0^t(i,V) + 2\epsilon.$$

Proof.

$$\begin{aligned}
\|\pi_h^- - P^t(i,\cdot)\|_{TV} &= \sum_{j \in V} (\pi_h^-(j) - P^t(i,j))_+ \\
&\leq \sum_{j \in V} \left(\pi_h^-(j)(1+\epsilon) + \frac{\epsilon}{|V|} - P_0^t(i,j)\right)_+, \\
&= \sum_{j \in V} \left(\pi_h^-(j)(1+\epsilon) + \frac{\epsilon}{|V|} - P_0^t(i,j)\right),
\end{aligned}$$

where the first inequality follows from $P_0^t(i,j) \leq P^t(i,j)$ and $\pi_h^-(j) \leq \pi_h^-(j)(1+\epsilon) + \frac{\epsilon}{|V|}$, and the second inequality from (5.7) in part (1) of Proposition 5.3.2.

We can split the summation in two obtaining the following result:

$$\begin{aligned}
&\sum_{j \in V} \left(\pi_h^-(j)(1+\epsilon) + \frac{\epsilon}{|V|} - P_0^t(i,j)\right)_+ = -P_0^t(i,V) + \sum_{j \in V} \left(\pi_h^-(j)(1+\epsilon) + \frac{\epsilon}{|V|}\right) \leq \\
&\leq -P_0^t(i,V) + \sum_{j \in V} \left(\epsilon\left(\pi_h^-(j) + \frac{1}{|V|}\right) + \pi_h^-(j)\right) = 1 - P_0^t(i,V) + 2\epsilon
\end{aligned}$$

Where the last equality comes from the fact that

$$\sum_{j\in V}\pi_h^-(j)=1=\sum_{j\in V}\frac{1}{|V|}$$ □

Corollary 5.3.4. *For every t and for every $i\in V_\star$:*

$$\|\pi_h^- - P^t(i,\cdot)\|_{TV}\le \mathcal{Q}_{i,t}\left(\frac{1}{n\ln^2 n}\right)+o_P(1).$$

Proof. Assume first that $t=t_\star+o(t_\star)$. Then by Lemma 5.3.3 and Proposition 5.3.2 for every ϵ we have that with high probability

$$\|\pi_h^- - P^t(i,\cdot)\|_{TV}\le 1-P_0^t(i,V)+2\epsilon\le \mathcal{Q}_{i,t}\left(\frac{1}{n\ln^2 n}\right)+3\epsilon\le \mathcal{Q}_{i,t}\left(\frac{1}{n\ln^2 n}\right)+o_P(1).$$

To extend this to arbitrary t we use Proposition 5.1.5 with $\theta=\frac{1}{n\ln^2 n}$, distinguishing whether $t>t_\star+o(t_\star)$ or $t<t_\star+o(t_\star)$.

If $t<t_\star+o(t_\star)$ then we can assume $t<t_\star-{t_\star}^{2/3}$. Let $t'=t_\star-{t_\star}^{2/3}$. We have:

$$\begin{aligned}\lim_{n\to\infty}\frac{\mu t'+\ln\theta}{\sqrt{t'}}&=\lim_{n\to\infty}\frac{\mu\left(\frac{\ln n}{\mu}-\left(\frac{\ln n}{\mu}\right)^{2/3}\right)+\ln\left(\frac{1}{n\ln^2 n}\right)}{\sqrt{\frac{\ln n}{\mu}-\left(\frac{\ln n}{\mu}\right)^{2/3}}}=\\&=\lim_{n\to\infty}\frac{\ln n-\mu^{1/3}\ln^{2/3}n-\ln n-\ln\ln^2 n}{\sqrt{\frac{\ln n}{\mu}-\left(\frac{\ln n}{\mu}\right)^{2/3}}}=\\&=\lim_{n\to\infty}-\frac{-\mu^{1/3}\ln^{2/3}n-\ln\ln^2 n}{\sqrt{\frac{\ln n}{\mu}+\left(\frac{\ln n}{\mu}\right)^{2/3}}}=\lim_{n\to\infty}-\mu^{5/6}\ln^{1/6}n=-\infty\end{aligned}$$

Hence, part 1 of Proposition 5.1.5 ensures that $\mathcal{Q}_{i,t'}\left(\frac{1}{n\ln^2 n}\right)\ge 1-o_P(1)$. Since $\mathcal{Q}_{i,t}$ is non-increasing with t we have that $\mathcal{Q}_{i,t}\left(\frac{1}{n\ln^2 n}\right)\ge 1-o_P(1)$ for every $t\le t_\star-{t_\star}^{2/3}$ (in particular whenever $t<t_\star+o(t_\star)$) and then the result follows from $\|\pi_h^- - P^t(i,\cdot)\|_{TV}\le 1$.

If $t>t_\star+o(t_\star)$ then we can assume $t>t_\star+{t_\star}^{2/3}$. Let $t'=t_\star+{t_\star}^{2/3}$. Then, we have:

$$\begin{aligned}\lim_{n\to\infty}\frac{\mu t'+\ln\theta}{\sqrt{t'}}&=\lim_{n\to\infty}\frac{\mu\left(\frac{\ln n}{\mu}+\left(\frac{\ln n}{\mu}\right)^{2/3}\right)+\ln\left(\frac{1}{n\ln^2 n}\right)}{\sqrt{\frac{\ln n}{\mu}+\left(\frac{\ln n}{\mu}\right)^{2/3}}}=\\&=\lim_{n\to\infty}\frac{\ln n+\mu^{1/3}\ln^{2/3}n-\ln n-\ln\ln^2 n}{\sqrt{\frac{\ln n}{\mu}+\left(\frac{\ln n}{\mu}\right)^{2/3}}}=\\&=\lim_{n\to\infty}\frac{\mu^{1/3}\ln^{2/3}n-\ln\ln^2 n}{\sqrt{\frac{\ln n}{\mu}+\left(\frac{\ln n}{\mu}\right)^{2/3}}}=\lim_{n\to\infty}\mu^{5/6}\ln^{1/6}n=+\infty,\end{aligned}$$

so part 2 of Proposition 5.1.5 states that $\mathcal{Q}_{i,t'}\left(\frac{1}{n\ln^2 n}\right) \leq o_P(1)$. Observe that Proposition 2.1.5, together with the triangular inequality, implies

$$\|\pi_h^- - P^t(i,\cdot)\|_{TV} \leq \max_{j\in V} \|P^t(j,\cdot) - P^t(i,\cdot)\|_{TV} \leq 2\max_{j\in V} \|\pi - P^t(j,\cdot)\|_{TV},$$

$$\|\pi - P^t(i,\cdot)\|_{TV} \leq \max_{j\in V} \|P^t(j,\cdot) - P^t(i,\cdot)\|_{TV} \leq 2\max_{j\in V} \|\pi_h^- - P^t(j,\cdot)\|_{TV}.$$

Since $\|\pi - P^t(j,\cdot)\|_{TV}$ is non-increasing with t (Corollary 2.1.7) we have that for $t > t_\star + o(t_\star)$

$$\begin{aligned}
\|\pi_h^- - P^t(j,\cdot)\|_{TV} &\leq 2\max_{j\in V} \|\pi - P^t(j,\cdot)\|_{TV} \\
&\leq 2\max_{j\in V} \|\pi - P^{t'}(i,\cdot)\|_{TV} \\
&\leq 4\max_{j\in V} \|\pi_h^- - P^{t'}(j,\cdot)\|_{TV} \\
&\leq 4\max_{j\in V} \mathcal{Q}_{j,t'}\left(\frac{1}{n\ln^2 n}\right) + o_P(1) = o_P(1)
\end{aligned}$$

where the last inequality uses the case $t' = t_\star + o(t_\star)$ shown at the beginning of the proof. □

Lemma 5.3.5. *For every $s \in \mathbb{N}$, $\mathcal{Q}_{i,t-s}(\theta) \leq Q_{i,t}(\theta\Delta^{-s})$*

Proof. Consider $\mathfrak{p}$ a walk from $i \in V$ to $j \in V$ in $t-s$ steps with weight $\mathbf{w}(\mathfrak{p}) > \theta$. In s steps from j, the weight will be at least $\theta\Delta^{-s}$. □

Lemma 5.3.6. *W.h.p., for every $i \in V$ and for $s < h$ we have that*

$$\widetilde{\mathcal{D}}_i(t) \leq \max_{i\in V_\star} \widetilde{\mathcal{D}}_i(t-s) + 2^{1-s}.$$

Proof. We clearly have that

$$\widetilde{\mathcal{D}}_i(t) = \widetilde{\mathcal{D}}_{P^s(i,\cdot)}(t-s),$$

where we denote

$$\widetilde{\mathcal{D}}_\pi(t) := \|\pi P^t - \pi_h^-\|_{TV}$$

the distance to π_h^- after the Markov chain runs for t steps starting with a probability distribution π.

Now, by Proposition 4.3.1, we have that w.h.p.,

$$\lambda := P^s(i, V_\star) \geq 1 - 2^{-s}, \qquad 1 - \lambda = P^s(i, V \setminus V_\star) \leq 2^{-s}.$$

For each subset W of V let $P^s(i,\cdot)_W$ denote the probability distribution $P^s(i,\cdot)$ conditioned to be in W. We have that

$$P^s(i,\cdot) = \lambda P^s(i,\cdot)_{V_\star} + (1-\lambda)P^s(i,\cdot)_{V\setminus V_\star}.$$

Since $||\cdot||_{TV}$ is a metric on the space of probability distributions, we have that

$$\begin{aligned}\widetilde{\mathcal{D}}_{P^s(i,\cdot)}(t-s) =& \|P^s(i,\cdot)P^{t-s} - \pi_h^-\|_{TV} \\ \leq& \|P^s(i,\cdot)_{V_\star}P^{t-s} - \pi_h^-\|_{TV} + \|P^s(i,\cdot)_{V_\star}P^{t-s} - P^s(i,\cdot)P^{t-s}\|_{TV} \\ \leq& \|P^s(i,\cdot)_{V_\star}P^{t-s} - \pi_h^-\|_{TV} + (1-\lambda) \\ \leq& \|P^s(i,\cdot)_{V_\star}P^{t-s} - \pi_h^-\|_{TV} + 2^{-s} \\ \leq& \max_{i\in V_\star} \|\delta_i P^{t-s} - \pi_h^-\|_{TV} + 2^{-s} \\ =& \max_{i\in V_\star} \|P^{t-s}(i,\cdot) - \pi_h^-\|_{TV} + 2^{-s} = \max_{i\in V_\star} \widetilde{\mathcal{D}}_i(t-s).\end{aligned}$$

Here, the last inequality uses the same convexity argument of Proposition 2.1.5 but taking into account that the distribution $P^s(i,\cdot)_{V_\star}$ has support contained in $V_\star$, hence it is a convex combination of the distributions $\{\delta_i\}_{i\in V_\star}$. □

Proof of the upper bound in Proposition 5.1.4. Consider $s = \lfloor \frac{\ln\ln n}{\ln \Delta} \rfloor$. Observe that

$$\Delta^s \leq \Delta^{\frac{\ln\ln n}{\ln\Delta}} = e^{\ln\ln n} = \ln n.$$

On the other hand,

$$\Delta^h \simeq \Delta^{\frac{\ln n}{10\ln\Delta}} = n^{1/10},$$

so that, for large n, we have $s < h$.

Then, for every $i \in V_\star$ we have

$$\begin{aligned}\widetilde{\mathcal{D}}_i(t-s) + o_P(1) =& \|\pi_h^- - P^{t-s}(i,\cdot)\|_{TV} + o_P(1) \\ \leq& \mathcal{Q}_{i,t-s}\left(\frac{1}{n\ln^2 n}\right) + o_P(1) \\ \leq& \mathcal{Q}_{i,t}\left(\frac{1}{n\ln^3 n}\right) + o_P(1),\end{aligned}$$

where the first equality is the definition of $\widetilde{\mathcal{D}}$, the next inequality is Corollary 5.3.4 and the last inequality is Lemma 5.3.5, taking into account that $\Delta^s \leq \ln n$. Then we have

$$\begin{aligned}\max_{i\in V} \widetilde{\mathcal{D}}_i(t) \leq& \max_{i\in V_\star} \widetilde{\mathcal{D}}_i(t-s) + o_P(1) \\ \leq& \max_{i\in V_\star} \mathcal{Q}_{i,t}\left(\frac{1}{n\ln^3 n}\right) + o_P(1) \\ \leq& \max_{i\in V} \mathcal{Q}_{i,t}\left(\frac{1}{n\ln^3 n}\right) + o_P(1),\end{aligned}$$

where the first inequality if Lemma 5.3.6 (since $2^{1-s} \in o(1)$), the second one is what we have shown above, and the third one is obvious. □

Bibliography

[1] Jimena Blatoitta, Pablo Delieutraz. Teorema Central del Limite. *Universidad de Buenos Aires* (2004). https://www.u-cursos.cl/ingenieria/2009/2/MA3401/1/material_docente/bajar?id_material=260765

[2] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled graphs, European Journal on Combinatorics 1 (1980) 311–316. http://dx.doi.org/10/ggktcq

[3] Charles Bordenave, Pietro Caputo, Justin Salez. Random walk in sparse random digraphs. *Probab. Theory Relat. Fields* 170 (2018), 933–960. https://doi.org/10.1007/s00440-017-0796-7

[4] Charles Bordenave, Pietro Caputo, Justin Salez. Random walk in sparse random digraphs. Lecture at *Université Paris Diderot* (2016).

[5] C. Cooper, A. Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence, Comb. Probab. Comput 13 (3) (2004) 319—337. https://www.math.cmu.edu/~af1p/Texfiles/dirgiant.pdf

[6] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* 93 (1996), 1659-1664. https://www.pnas.org/doi/pdf/10.1073/pnas.93.4.1659

[7] Persi Diaconis. Group representations in probability and statistics. IMS Lecture Notes Monogr. Ser., 11: 198pp. (1988). DOI: 10.1214/lnms/1215467407

[8] Persi Diaconis, David Aldous. Shuffling cards and stopping times. *The American Mathematical Monthly* 93, no. 5 (1986), 333–348. https://www.jstor.org/stable/2323590

[9] M. Fioravanti. Matemática discreta: Teoría de Grafos *Universidad de Cantabria* (2021). https://personales.unican.es/fioravam/MatDiscreta/Mat-Discreta_3ra-parte_nov2021.pdf

[10] Ari Friedman. Convergence theorem for finite Markov chains. *University of Chicago.* Disponible en https://math.uchicago.edu/~may/REU2017/REUPapers/Freedman.pdf

[11] David Levin, Yuval Peres. Markov Chains and Mixing Times, second edition. *University of Oregon.* Disponible en https://pages.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf

[12] Patrick Tantalo. CSE 102: Introduction to Analysis of Algorithms. *University of California, Santa Cruz*, 2021 [fecha de consulta: Marzo del 2023]. Disponible en `https://classes.soe.ucsc.edu/cse102/Fall21/Handouts/AsymptoticGrowth.pdf`