Universidad Nacional Autónoma de México

Posgrado Conjunto en Ciencias Matemáticas
UNAM-UMSNH

Stochastic Topology in the Curve Complex of a Surface

A thesis submitted in partial fulfillment for the
degree of **Master in Sciences**

by:

Ricardo Esteban Chávez Cáliz
ricardo.ch.cz@gmail.com
1150505J

Supervisor:

Dr. Noé Bárcenas Torres
barcenas@matmor.unam.mx
Maestría y Doctorado en Ciencias Matemáticas y de Especialización en Estadística Aplicada

MORELIA, MICH., May 2020

Universidad Michocana de
San Nicolás de Hidalgo

Posgrado Conjunto en Ciencias Matemáticas
UNAM-UMSNH

Stochastic Topology in the Curve Complex of a Surface

A thesis submitted in partial fulfillment for the
degree of **Master in Sciences**

by:

Ricardo Esteban Chávez Cáliz
ricardo.ch.cz@gmail.com
1150505J

Supervisor:

Dr. Noé Bárcenas Torres
barcenas@matmor.unam.mx

Maestría y Doctorado en ciencias Matemáticas y de la Especialización en Estadística Aplicada

MORELIA, MICH., May 2020

Fahre fort, übe nicht allein die Kunst, sondern dringe auch in ihr Inneres; sie verdient es, denn nur die Kunst und die Wissenschaft erhöhen den Menschen bis zur Gottheit.

Do not merely practice your art, but penetrate into its interior; it deserves that, because only art and science exalt man to divinity.

Ludwig van Beethoven

Acknowledgments

This work is the result of multiple efforts that pushed collaboration between different areas and institutions. I am fortunate enough to be part of this community, to which it is possible to give particular mention here to some of them.

First, this thesis would not have been possible without the support and patience of my principal supervisor, Dr. Noé Bárcenas Torres, from whom I have received some of the best advices for Math and Life.

I thank Dr. Jesus Hernandez Hernandez who introduced me to the world of rigid expansions and whose enthusiasm for collaboration planted the seed for this work. His follow up and expertise allowed to concrete the results.

I want to thank Dr. Octavio Arizmendi Echegaray for his valuable help in all the probability calculations and insights for random graphs.

I want to express my appreciation to the professors in CIMAT from whom I learned a statistics, probability and scientific computing.

Thanks to my committee members for all their helpful observations that refined the final work.

Financial support was provided by the project "Aplicación de ensamble de Baum-Connes: topología algebráica y geometría no conmutativa" with key IA100315. Also, I thank CONACYT foundational Research Grant 250747 and PAPIIT Grant IA 100119.

Last but not least, I want to thank all my family and friends for their support in this process, without them I am nothing.

Contents

Abstract

We propose the study of rigid expansions in graphs, a concept originally conceived in the curve graph of surface, in a stochastic context using the Radó graph and Erdös-Rényi model. We review the motivation to study this particular phenomenon and the feasibility to do it so through these probabilistic models. We also provide computational simulation with the corresponding optimizations for the processing.

Introduction

The curve graph $\Gamma(S)$ associated to a surface S appears naturally in the study of $Mod(S)$, the mapping class group of S, which is a central subject in contemporary mathematical research. We are interested in a rigidity concept of this graph; in general, the idea behind rigidity phenomena is to describe morphisms among objects using their structure.

The folkloric version of rigidity in the $Mod(S)$ context is that if we consider X and Y, under suitable conditions, then every homomorphism $Mod(X) \to Mod(Y)$ will be induced by a manipulation of the underlying surfaces.

Ivanov sketched in [1, Ivanov 97] the proof that every automorphism of $C(S)$ (the flag complex of $\Gamma(S)$), is induced by a self-homeomorphism of S. This argument is the favorite in the literature due to its simplicity and resemblance to proofs of other rigidity results.

A research line lead by Aramayona and Leininger propose the use of *rigid sets*, which can be interpreted as a subset *that allows to extend a local notion of rigidity to a global one*. In the aim of finding large rigid sets, in [2, Aramayona, Leininger 16] there's the proof of the existence of an increasing sequence of finite rigid sets that exhaust the curve graph. For this, they proposed a method called **rigid expansions**.

Rigidity in graphs is, regardless of its interpretation in the curve graph, an interesting phenomenon by it self. Due to the discrete nature of rigid expansions is reasonable to seek for a probabilistic approach; our goal is to address this particular path.

We want to answer the rather vague question: *How* ***common*** *is rigidity in graphs*, specifically by answering *how rigid expansions* ***usually*** *behave*. Also, the aim of the thesis is to review the feasibility of *studying the curve complex of a surface from a probabilistic point of view.*

Probabilistic models give formal meaning to words like "common" or "usually", we study rigidity phenomenon in this context and analyze the conditions under which these models fit the known properties of the curve graph.

In chapter one, we motivate the study of the curve graph and review the most important properties of it. Then, we introduce rigidity within the context of Graph theory.

In the second chapter, we propose the study of rigidity from the stochastic point of view through the Radó graph and the Erdös-Rényi model. In the aim to study the curve graph of a surface with a simple model, we justify that the genus of the surface cannot be finite. Thus, we end up with an asymptotic probabilistic analogue to the result due to Bering and Gaster, which asserts that the Radó graph embeds into the curve graph $C(S)$ of a surface S if and only if S has infinite genus.

Finally, we made a computational implementation of the algorithm to do rigid expansions. With the corresponding optimizations that the method require, we are able to take a closer look to rigidity phenomena.

Chapter 1

The curve graph of a surface

The study of surfaces in a strictly topological viewpoint has led us to forget significant information about them. A way to revert this is to attach a group to it, the **mapping class group** of the surface. It is denoted by $Mod(S)$ and encodes the *symmetries* of the surface. This group is defined as the set of isotopy classes of orientation-preserving homeomorphisms of S. In the first section of this chapter we give the formal definition of this group and establish the very important role of this concept in Mathematics.

The **curve complex** of the surface, denoted by $C(S)$ appears naturally in the study of $Mod(S)$. It is a simplicial complex that encodes intersection patterns of simple closed curves in S. We focus part of the discussion in the relationship between the algebraic structure of $Mod(S)$ and the combinatorial topology of S.

Many of the progress in understanding $Mod(S)$ has been possible by a well-known analogy among two very important classes of groups: arithmetic groups and mapping class groups. In this parallelism panorama arises the desire for an equivalent result to the Margulis Superrigidity for mapping class groups.

In the last section of this chapter we settle the bases to understand rigidity within a Graph theory context. An approach called *rigid expansions*, see [2, Aramayona, Leininger 16] and [3, Hernandez 19], allows us to build up subgraphs preserving the rigidity property and is compatible with stochastic tools.

Many results and definitions in this chapter where extracted from [4, Farb]. They are quite popular and equivalents can easily be found in the literature, however they are written here to establish nomenclature. Familiarity with basic concepts is assumed.

1.1 Mapping class group of a surface

We have the following fundamental, well-known result about surfaces.

Theorem 1.1.1 (Classification of surfaces)**.** *Any closed, connected, orientable surface is homeomorphic to the connect sum of a 2-dimensional sphere with $g \geq 0$ tori. Any compact, connected, orientable surface is obtained from a closed surface by removing $b \geq 0$ open disks with disjoint closures. Even more, the set of homeomorphism types of compact surfaces is in bijective correspondence with the set $\{(g, b) : g, b \geq 0\}$.*

We are so familiarized with this result that we usually forget what it is saying. It seems like, in the eyes of a topologist, there is nothing much interesting about surfaces, but this is because we are forgetting all the geometric information about them. $Mod(S)$ helps to recover this data, the magic happens when this group acts on the **Teichmüller space** of S, that is the space of hyperbolic metrics on S up to isotopy. A central result is that this action is properly discontinuous and the quotient space $M(S) = Teich(S)/Mod(S)$ is the **moduli space of Riemannian surfaces homeomorphic to** S. The space $M(S)$ is a essential object in mathematics and the group $Mod(S)$ encodes most of the topological features of $M(S)$.

$Mod(S)$, $Teich(S)$, and $M(S)$ can be found in a lot of different contexts in mathematics: hyperbolic geometry, algebraic geometry, combinatorial group theory, symplectic geometry, 3-manifold theory, dynamics and so on. The algebraic structure of $Mod(S)$, the geometry of $Teich(S)$, and the topology of $M(S)$ are just the strands which are used to weave the rich tapestry of the nature of the surface.

Before we continue, let's establish some nomenclature. The g in 1.1.1 is called the *genus* of the surface and the b is the number of *boundary components*. One way to obtain a non-compact surface from a compact one is to remove m points from the interior of it; in this case, we say that the resulting surface has m punctures. For now on, unless otherwise specified, we will be thinking in compact, connected, oriented surfaces that are possibly punctured (in this case they ceases to be compact). Therefore, we can specify the surfaces by the triplet (g, b, m). We will denote by $S_{g,m}$ a surface of genus g with m punctures and empty boundary; such a surface is homeomorphic to the interior of a compact surface with m boundary components. Also, for a closed surface of genus g, we will abbreviate $S_{g,0}$ as S_g and ∂S denote the (possibly disconnected) boundary of S.

There are a number of definitions for the mapping class group of a surface. We will be working with the following:

Definition 1.1.1.1. *Let S be a surface, the* ***mapping class group*** *of S, denoted by $Mod(S)$ is the following quotient:*

$$Mod(S) = Homeo^{+}(S)/Homeo_0(S)$$

where $Homeo^{+}(S)$ is the group of orientation-preserving, homeomorphisms of S, that are the identity on the boundary, this group can be endowed with the compact-open topology. $Homeo_0(S)$ is the subgroup formed by homeomorphisms of S which are isotopic to the identity, i.e. the connected component of the identity with this topology.

We could consider diffeomorphisms instead of homeomorphisms, or homotopy classes instead of isotopy classes; this will result in isomorphic groups, see [4, Farb, p. 41] for details in why we can do this. Summarizing, we can find the following variations in the definition of $Mod(S)$:

$$\begin{aligned} Mod(S) &= \pi_0(Homeo^{+}(S, \partial S)) \\ &\approx Homeo^{+}(S, \partial S)/homotopy \\ &\approx \pi_0(Diff^{+}(S, \partial S)) \end{aligned}$$

where $Diff^{+}(S, \partial S)$ is the group of orientation-preserving diffeomorphisms of S that are the identity on the boundary. It can be taken to be either smooth homotopy relative to the boundary or smooth isotopy relative to the boundary.

A lot of work had been made to describe the types of elements in $Mod(S)$. Thanks to the Thurston's classification theorem there is a characterization of the homeomorphisms of a compact orientable surface. This classification is useful to describe the curve graph which will be analyzed in the next section.

1.1.1 Nielsen–Thurston classification

Given a homeomorphism $f : S \to S$, there is a map g isotopic to f such that at least one of the following statements holds:

- g is periodic, i.e. some power of g is the identity;
- g preserves some finite union of disjoint simple closed curves on S (in this case, g is called reducible); or
- g is pseudo-Anosov.

The definition of a **pseudo-Anosov map** relies on the notion of a measured foliation, a geometric structure on S. It consists of a singular foliation and a measure in the

transverse direction (i.e. that is constant in transverse arches). For the full definition of pseudo-Anosov elements and the proof of this theorem we can refer to [4, Farb, ch. 13].

The study of mapping class groups is a wide and challenging area. It is outside of the interests of this thesis to review the details and repercussions of this vastly field. Yet, there are a number of known properties of $Mod(S)$ that it would be nice to have in mind in further work, although different tools might be required.

- Finitely generated and presented
- It has a subgroup of finite index which doesn't have torsion.
- $Mod(S_{g,m}) \cong Out(\pi_1(S_{g,m}))$
- $H_1(Mod(S_{g,m}), \mathbb{Z}) = 1$ when $(g \geq 3, m = 0)$

1.2 Curve graph

1.2.1 Simple closed curves

Definition 1.2.0.1. *A **closed curve** in a surface S is a continuous map $\mathbb{S}^1 \to S$, is called **simple** if the map is injective. We will usually identify a closed curve with its image in S. A closed curve is called **essential** if it is not homotopic to a point, a puncture, or a boundary component.*

Among the adjectives that a curve can acquired we have the following:

- α is **separating**, if $S - \alpha$ has two components, otherwise it is called **non separating**.
- It is called **essential** if no component of $S - \alpha$ is a disk.
- It is **non-peripheral** if no component of $S - \alpha$ is an annulus.

We are interested in **essential** and **non-peripherial** curves, they will be assumed in this sense, unless otherwise specified.

The idea behind the construction of the curve graph is to stratify the set of homotopy classes of curves on a surface. For this to make sense we define the **geometric intersection number** between free homotopy classes a and b of simple closed curves in

a surface S. This is defined to be the minimal number of intersection points between a representative curve in the class a and a representative curve in the class b:

$$i(a,b) = min\{|\alpha \cap \beta| : \alpha \in a, \beta \in b\}$$

It is convenient to adopt a slight abuse of notation by writing $i(\alpha,\beta)$ for the intersection number between the homotopy classes of simple closed curves α and β. It is useful to think that this number can be calculate by finding representatives α and β that achieve the minimal intersection in their homotopy classes, so that $i(a,b) = |\alpha \cap \beta|$ (when this is the case, we say that α and β are in minimal position). Although the geometric intersection number is a useful and intuitive invariant it is not always easy to compute, whenever this is the case we can appeal to the algebraic intersection number. For a further discussion of this see [4, Farb]

1.2.2 The curve graph

Definition 1.2.0.2. *The* ***curve graph*** $\Gamma(S)$ *of a surface S is constructed by the following data:*

- ***Vertices****. There's a vertex in $\Gamma(S)$ for every isotopy class of essential closed curves in S.*
- ***Edges****. There's an edge between the corresponding vertices of isotopy classes a and b whenever $i(a,b) = 0$.*

Definition 1.2.0.3. *The* ***curve complex of the surface****, $C(S)$ is defined to be the flag complex of the curve graph just defined.*

1.3 Properties of the curve graph

The goal of this section is to enumerate known properties of the curve graph, we use this to establish the appropriate parameters in a probabilistic model. Notice that the construction of the curve complex is completely determined by the curve graph, hence the probabilistic models can work in the same sense. Lets keep in mind the following exceptional cases; they are responsible for the conditions stated in the hypothesis of the following theorems for g and n. For $S^2, S_{0,1}, S_{0,2}, S_{0,3}$ the curve graph is empty and for T^2, $S_{1,1}$ and $S_{0,4}$ is a countable disjoint union of points.

1.3.1 Cardinality of the number of vertices

Theorem 1.3.1. *If $g \geq 1$ or $n \geq 4$ then the set of vertices in $\Gamma(S_{g,n})$ is countably infinite.*

It is well known that for T^2 there is an explicit identification for the isotopy classes of essential curves with the rational numbers. In this case there aren't disjoint curves, the following figure can help to convince us of this fact.

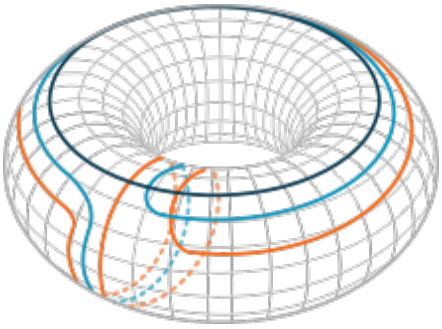

FIGURE 1.1: T^2 with representatives of typical elements of curves

This identification can be seen as the induction basis. The induction step over g comes from splitting the surface, by induction hypotheses none of the resulting surfaces can have a non-countable number of classes of curves.

1.3.2 Connectivity

Theorem 1.3.2. *If $3g + n \geq 5$, then $\Gamma(S_{g,n})$ is connected.*

To proof this theorem we can show that, for any two isotopy classes a and b of simple closed curves in $S_{g,n}$ exists a sequence of isotopy classes

$$a = c_1, \ldots, c_k = b$$

where $i(c_i, c_{i+1}) = 0$, this can be done proceeding by induction over $i(a, b)$. The full proof of this theorem can be found in [4, Farb, p. 93]

1.3.3 Locally infinite

Theorem 1.3.3. *If $3g + m \geq 5$, then $\Gamma(S_{g,n})$ is locally infinite.*

The idea behind the proof is that given any $\alpha \in \Gamma(S)$ we can construct a family of isotopy classes of curves which are disjoint to α. The following picture gives us an intuitive idea on how to do this whenever there are enough holes.

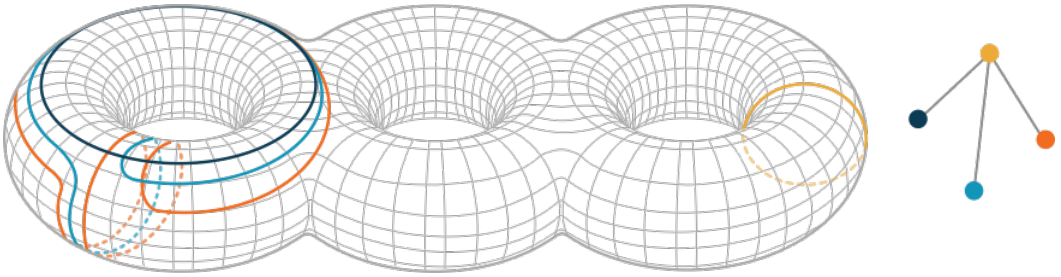

FIGURE 1.2: S_3 with typical representative curves which exemplify the idea behind locally infinity property

For the complete argument, let α be any simple closed curve on S, the surface $S-\alpha$, obtained by cutting S open along α, contains at least one connected component of Euler characteristic at most -2 (guaranteed by the $3g+m \geq 5$ condition). Such component contains infinitely many distinct homotopy classes of simple closed curves disjoint from α.

1.3.4 Clique number

A **clique** in a graph G is a complete subgraph of G. The clique number $cl(G)$ of a graph G is the maximum order of a clique of G.

Theorem 1.3.4. *If $3g+n \geq 5$, then the clique number of $\Gamma(S_{g,m})$ is $3g-3+m$.*

$3g-3+m$ is the number of curves in a pants decomposition of S, i.e. a maximal collection of disjoint, not freely homotopic, essential, simple closed curves which decompose S into $2g-2+m$ open subsurfaces homeomorphic to a thrice punctures sphere. For a full proof of this well-known fact refer to [5, Hatcher, Thurston 80]

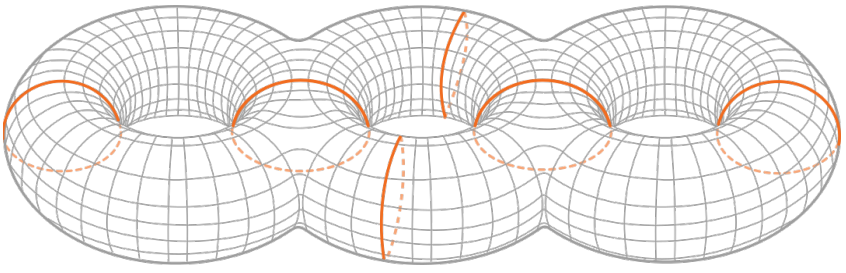

FIGURE 1.3: Exemplification of a pants decomposition of a surface

1.3.5 Infinite Diameter

Theorem 1.3.5. *If* $3g + m \geq 5$ *then* $diam(\Gamma(S)) = \infty$

The proof for this theorem relies on the fact that for any pseudo-Anosov element $h \in Mod(S)$, any $\gamma \in V(\Gamma(S))$ and any $k \in \mathbb{Z}$

$$d_C\Big(h^k(\gamma), \gamma\Big) \geq c|k|$$

This provide the infinite diameter property. For details, refer to [6, Masur, Minsky 99]

The curve graph and the curve complex are fundamental tools in the study of the surfaces. There are a number of known properties of them that it would be nice to have in mind to improve probabilistic models in further work.

1. $C(S)$ is hyperbolic
2. In the infinite case $diam(\Gamma(S)) = 2$
3. There's an isomorphism between $Mod(S)$ and $Aut(C(S))$, except when $(g, m) \in \{(1,2), (1,1), (2,0), (0,4)\}$

1.4 Rigidity in graphs

The intention of this section is to track down the motivation of rigid expansions and give the required definitions.

As mention in the introduction of the chapter, rigidity appears in the mapping class group context in light of its comparison with arithmetic groups. In [7, Aramayona, Souto 16] we can find a survey on the search of an analogue for the Margulis Superrigidity theorem. In this article, they provide three different perspectives: a Lie theoretical, a geometric and a folkloric one.

The Lie theoretic version states that every homomorphism $Mod(X) \to Mod(Y)$ is induced by a homomorphism between the associated groups of diffeomorphisms with compact support disjoint from the boundary $Diff_c(X) \to Diff_c(Y)$.

A direct formulation of geometric superrigidity cannot hold when the moduli space is endowed with any reasonable metric. However, there are ways to turn this around, saying that every (irreducible) homomorphism between mapping class groups induces a holomorphic map between the corresponding moduli spaces.

The folkloric version of Mostow and Margulis superrigidity claims that the only homomorphisms between lattices are the *"obvious ones"*, in the $Mod(S)$ context this will mean that if we consider X and Y, under suitable conditions, then every homomorphism $Mod(X) \to Mod(Y)$ will be induced by a manipulation of the underlying surfaces.

A result due to Ivanov [1, Ivanov 97], Korkmaz [8, Korkmaz 99] and Luo [9, Luo 00], asserts that, excluding few well-understood cases, the curve complexes are simplicially rigid. This means that the group $Aut(C(S))$ of simplicial automorphisms of $C(S)$ is isomorphic to the extended mapping class group. This result is sometimes interpreted as a proof that **the automorphisms of the curve complex are all geometric**.

In the aim to generalize this result to broader types of simplicial self-maps, in [2, Aramayona, Leininger 12] was introduced the concept of **rigid sets** in the curve complex. Let S be a surface different from $S_{1,2}$, $Y \subset C(S)$ is called **rigid** if for every locally injective simplicial map $\Phi : Y \to C(S)$ there exists $h \in Mod*(S)$ with $h|_Y = \Phi$, unique up to the pointwise stabilizer of Y in $Mod^*(S)$.

Later in [10, Aramayona, Leininger 16] a method for enlarging rigid subgraphs is presented and the proof that for almost all surfaces of finite topological type, there exists an **increasing sequence of finite rigid sets** that exhaust the curve graph of which has trivial pointwise stabilizer in $Mod^*(S_g, n)$.

In [3, Hernández 19] there is the proof of a similar result to Aramayona and Leininger's. The method, called **rigid expansions**, allowed to obtain new results concerning edge-preserving maps.

Losing the big picture that rigidity represents, we can land this vastly journey in the following graph theory definitions.

Definition 1.4.0.1. *Let Γ be a simplicial graph and $H < \Gamma$ a vertex-induced subgraph. A function $f : y \to \Gamma$ is **locally injective** if $f|_{star(v)}$ is injective for all $v \in V(y)$.*

Note 1.4.1. *Remember, $star(v)$ is the vertex-induced subgraph with vertices $\{v\} \cup N(v)$ (v plus its neighborhood).*

Definition 1.4.1.1. *$H < \Gamma$ is **rigid** if every locally injective function defined in H can be extended to an automorphism of Γ.*

A vertex $v \in V$ in a graph is called to be uniquely determined by $A \subset V(G)$, denoted $v = \langle A \rangle$, if v is the unique common neighbor of every element of B, i.e.

$$\{v\} = \bigcap_{w \in B} lnk(w)$$

Definition 1.4.1.2. *The first rigid expansion of $Y \subset \Gamma$, denoted by Y^1, is the vertex-induced subgraph whose vertices are*

$$V(Y) \cup \{v \in V(\Gamma) : \exists A \subset V(Y) \text{ where } v = \langle A \rangle\}$$

We also define $Y^0 = Y$ and, inductively, $Y^k = (Y^{k-1})^1$.

Recalling that in Proposition 3.5 in [2, Aramayona, Leininger 16], Aramayona and Leininger prove that if $Y \subset C(S)$ is a rigid set, then so is Y^r for all $r \geq 0$. So this method in fact preserve the desired property.

It would be nice to have conditions which determine whether a subgraph is rigid or not. So far we know that there aren't non-trivial necessary conditions to check rigidity, i.e. other than connectivity there's not much else.

With this definitions we can proceed to settled a probabilistic model so that we can analyze the rigidity concept in graphs from a stochastic point of view. With the reviewed properties of the curve graph we can determine the feasibility of studying the curve graph through simple models.

Chapter 2

Rigidity in random graphs

The use of the probabilistic method in discrete mathematics has become a prominent idea in the area in recent times. It provides existence proofs where objects have certain desirable properties and the construction of explicit examples is challenging. This has been just the beginning of the use of probabilistic tools within a deterministic context.

Complex topological spaces arise quite natural in a lot of scientific contexts. Probability theory implements different approaches to model those spaces; even in complex configurations, it can be possible by doing approximations, to study topological invariants. In this sense, stochastic topology can be thought as a tool for topology in the same sense as statistical mechanics is used to study a macroscopic physical system when the classical mechanics finds these systems very complicated to solve.

Stochastic topology finds its early motivation in applied problems. Nevertheless, in recent articles it has been used to provide a deeper insight for theoretical questions. For example, with probabilistic analogs of very classical topology conjectures, like Whitehead's Asphericity Conjecture [11, Costa, Faber 15].

Probability theory can help us understand the ubiquity of certain mathematical phenomena. For example, *many* simplicial complexes and posets which arise from combinatorial constructions are homotopy equivalent to a wedge of spheres, or that hiperbolicity is *common* in random groups. With Probability theory we can give formal meaning to these expressions.

In this chapter we review rigid expansions in simple probabilistic models. Afterwards, we analyze the feasibility of modeling the curve graph using these proposals.

Familiarity with basic concepts in probability theory such as probability spaces, random variables, and basic theorems are assumed.

2.1 Models for random graphs

2.1.1 The Radó graph

Let $0 < p < 1$ be fixed, $\mathcal{G}(\mathbb{N}, p)$ is the probability space which consists of all graphs with vertex set $\mathbb{N}$, whose edges are chosen independently with probability p. In other words, a random graph $G \in \mathcal{G}(\mathbb{N}, p)$ is a collection $(X_{ij}) = \{X_{ij} : 1 \leq i < j\}$ of independent $Bernoulli(p)$ r.v., where a pair ij is an edge of G if and only if $X_{ij} = 1$.

Erdös and Rényi proved in [12, Erdös, Rényi 63], that every countably infinite random graph is isomorphic to the **Radó graph**. A construction of this graph can be done using binary numbers; identify the vertices of the graph with the natural numbers and then every edge appears between vertices x and y in the graph (assuming $x < y$) whenever the x-th bit of the binary representation of y is nonzero. This means, for example, that all odd-numbered vertices will be neighbors of vertex 0, and that the larger neighbors of vertex 1 are all vertices with numbers congruent to 2 or 3 mod 4.

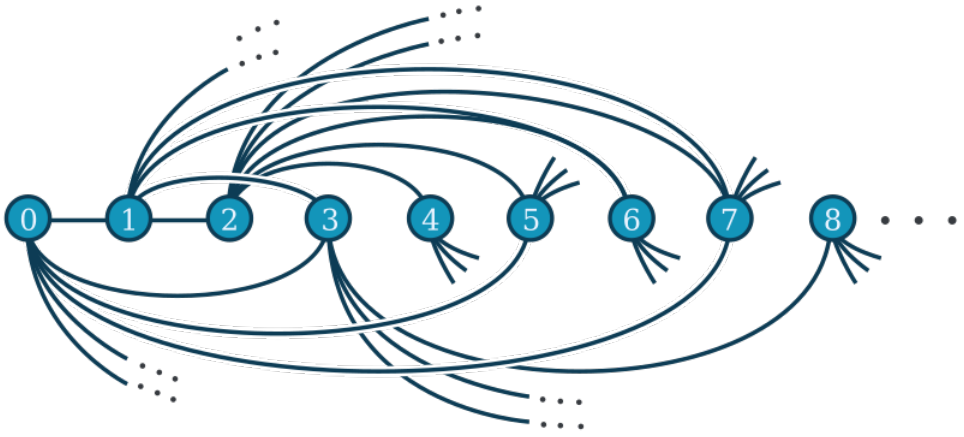

FIGURE 2.1: Binary construction of the Radó graph

2.1.2 Erdös-Rényi model

The Erdös-Rényi model for random graphs is the finite version of the Radó graph. In this model the parameter p is usually taken as a function of n. This provides, unlike the past model, a variety of graphs when n tends to infinity.

Definition 2.1.0.1. *Denote by $\mathcal{G}(n, p)$ to the probability space formed by all the graphs of n vertices and probability measure*

$$\mathbb{P}\Big(G \in \mathcal{G}(n, p)\Big) = p^k (1-p)^{\binom{n}{2}-k}$$

where k is the number of edges in G, the σ-algebra is given by the power set.

Note: There is a variation of the model, where we rather choose randomly exactly m edges among the $\binom{n}{2}$ possible.

We can also think this model like $\binom{n}{2}$ i.i.d. $Bernoulli(p)$ that represent the edges. From this, we can immediately get some properties of the degree of a vertex v.

- The probability that a given vertex v has degree k is given by

$$b(k; n-1, p) = \binom{n-1}{k} \cdot p^k \cdot (p-1)^{n-k-1}$$

- The expected degree is $(n-1) \cdot p$
- The variance of this degree is $(n-1) \cdot (1-p) \cdot p$

The degree distribution can be helpful to do optimizations in the rigid expansions algorithms. This is outlined in the next chapter.

2.2 Rigid expansions

We will focus in the rigidity calculations for the finite case. Then, we will analyze when n tends to infinite. For this, we need the calculations of the probability that the following events occur.

1. Given v a vertex and A_m a subset of vertices of size m, $v = \langle A_m \rangle$. Let's call E_m to this event.
2. A_k generates a rigid expansion.
3. A_k generates a rigid expansion with s new elements.

For the calculations concerning the first event, take a look to following figure

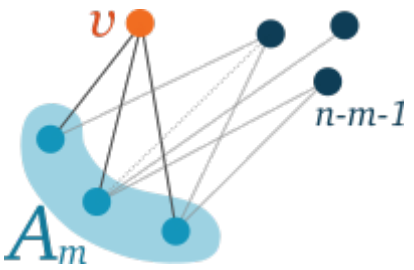

FIGURE 2.2: Probability of uniquely determined vertices

If $\langle A_m \rangle = v$, there is a edge between v and every vertex in A, and none of the remaining $n - m - 1$ vertices is also connected to every vertex in A, i.e.

$$\mathbb{P}(E_m) := \mathbb{P}(\langle A_m \rangle = v) = p^m(1 - p^m)^{n-m-1}$$

Using the `networkx` library in `python` we reproduce the following experiment:

Uniquely determined vertex experiment. Let n, p, m be fixed.

1. Generate an Erdös-Rényi graph $G \in \mathcal{G}(n, p)$ with labeled vertices.
2. Excluding the n-th vertex, take a random subset of vertices of size m.
3. Verify if this subset uniquely determine the n-th vertex.

In the next chapter we explain how to generate random graphs for the first step of the experiment. To simplify the process we took, without loosing generality, the last vertex as a particular element of the experiment.

Fixing different values for n and p is possible to compute the empiric probability for each possible value of m.

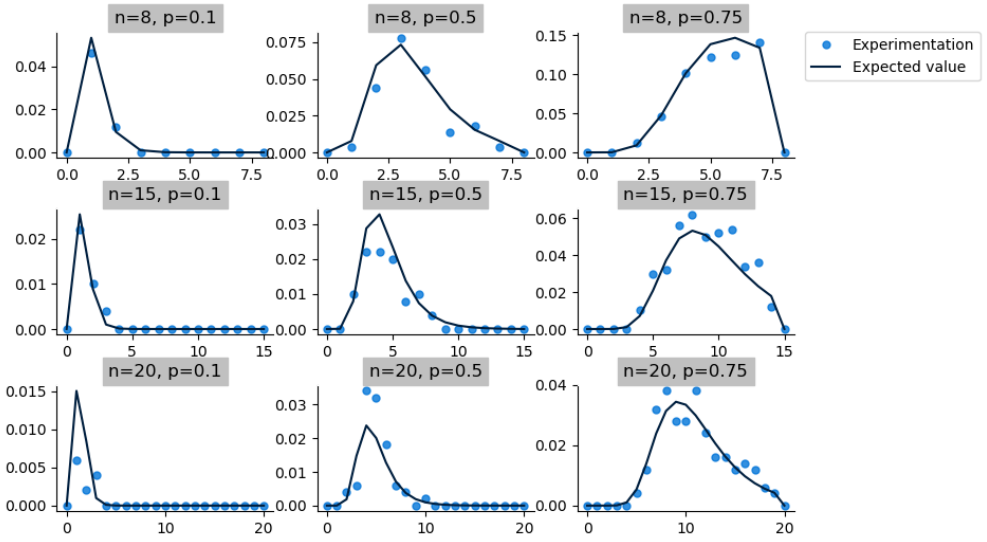

FIGURE 2.3: Theoretical and empirical probabilities of uniquely determined a vertex. For different values of n and p varying among all the possible values of m

For these estimations we calculated the empiric probability by repeating this experiment 500 times and counting the number of times when the n-th vertex was uniquely determined by the random set.

Notice that there are certain values of m more *effective* than others, in the sense that, depending on the parameters of the model, is more likely that a subset of certain size uniquely determine a vertex. This can be used, as described in the next chapter, to do optimizations in the simulations.

In this table appear the supremum of absolute differences between hypothesized and empirical probability $\forall m \in \{1, 2, \dots n\}$ for the different values of n and p.

p \ n	8	15	20
0.1	$9.01E-03$	$1.20E-02$	$8.09E-03$
0.1	$9.01E-03$	$1.20E-02$	$8.09E-03$
0.1	$9.01E-03$	$1.20E-02$	$8.09E-03$

TABLE 2.1: Supremum of absolute differences between hypothesized and empirical probabilities

For the second event, if A_k does not generate a rigid expansion is because none of the subsets of A_k uniquely determined a vertex outside of it. We have:

$$\mathbb{P}(A_k \text{ generates a rigid expansion}) = 1 - \prod_{m=1}^{k} (\rho_{m,k})^{\binom{k}{m}}$$

where $\rho_{m,k} = \Big(1 - \mathbb{P}(E_m)\Big)^{n-k}$.

Just as before, we reproduce the following experiment:

Rigid expansion experiment. Let n, p, k be fixed.

1. Generate an Erdös-Rényi graph $G \in \mathcal{G}(n, p)$.
2. Take a random subset of vertices of size k.
3. Verify if this set generates a rigid expansion.

Notice that the third step is a critical point of the experiment; we must verify among **all the possible subsets** of A_k. In the next chapter we explain the optimizations that needed to be done.

Again, we repeated this experiment 500 times and calculated the empiric probability that a random set generates a rigid expansion.

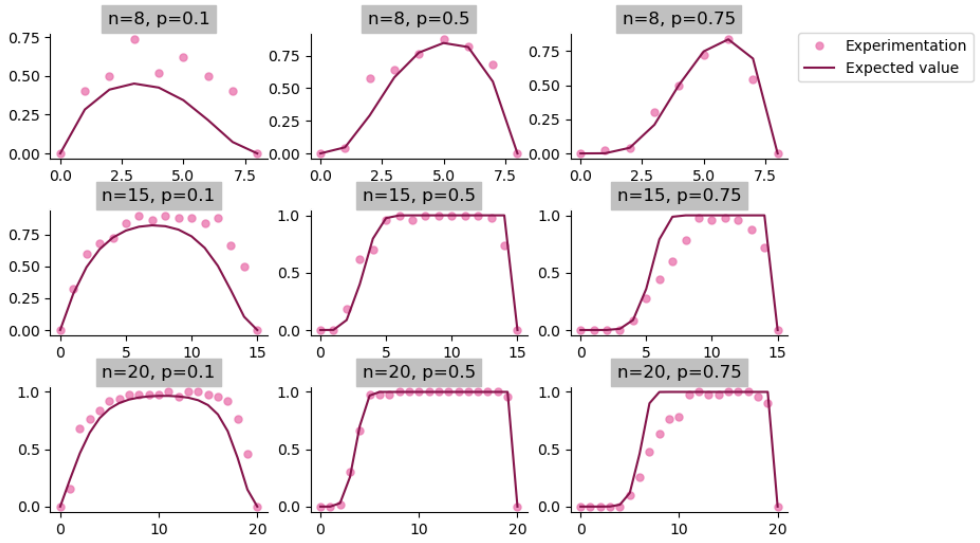

FIGURE 2.4: Theoretical and empirical probabilities of expanding A_k. For different values of n and p varying among all the possible values of k

In the following table appear the supremum of absolute differences between hypothesized and empirical probabilities $\forall k \in \{1, 2, \dots n\}$ for the different values of n and p.

p \ n	8	15	20
0.1	$9.01E-03$	$1.20E-02$	$8.09E-03$
0.1	$9.01E-03$	$1.20E-02$	$8.09E-03$
0.1	$9.01E-03$	$1.20E-02$	$8.09E-03$

TABLE 2.2: Supremum of absolute differences between hypothesized and empirical probabilities

The calculations for the last question are helpful if we want to approximate the sequence of rigid expansions of A_k by a Markov chain. Consider $\{0, 1, \dots n\}$ as the states space of the Markov chain with transition matrix given by:

$$a_{k,k+s} = \mathbb{P}(A_k \text{ generates a rigid expansion by } s \text{ elements})$$

Notice that the deterministic process stops once a iteration fails to add new vertices. In our stochastic approximation a new $G \in \mathcal{G}(n,p)$ is considered for each step, hence, it is allowed to *"have extra tries to expand"*.

This probability calculations are more difficult to obtain. To start understanding this phenomenon we can simulate with our computational tools the following experiment:

Increase size by a rigid expansions experiment
Let n, p, k be fixed.

1. Generate an Erdös-Rényi graph $G \in \mathcal{G}(n, p)$.
2. Take A_k a random set of k vertices.
3. Produce the first rigid expansion from the graph spanned by A_k
4. Return the size of the expanded subgraph

This experiment yields a random variable which depends on n, p and k. Fixing n and p we obtained a sample of size 50 for every possible value of k. Using the resulting histogram as an empirical density function we obtain the following figure. It graphically describes the nature of the transition matrix of a sequence of rigid expansions.

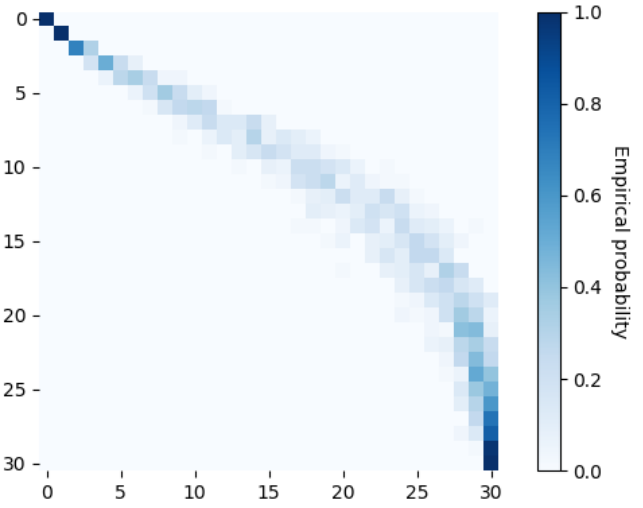

FIGURE 2.5: Empirical transition probability matrix

2.3 Radó graph as a model for the curve graph

In chapter one we outlined the properties of the curve graph associated to a surface, thus the proposed models should at least guarantee the following properties:

1. Countably infinite number of vertices
2. Connectedness

3. Locally infinite
4. Clique number $3g - 3 + m$
5. Infinite diameter

The Radó graph satisfy that every finite or countably infinite graph is an induced subgraph of it [13, Cameron 97]. The the clique number property of $C(S)$ implies that, if S is a surface of finite genus, is not possible that the Radó graph is embedded into $C(S)$. Even more, a result by Bering and Gaster [14, Bering, Gaster 17] state that the converse is also valid.

Theorem 2.3.1. *The random graph embeds into the curve graph $C(S)$ of a surface S if and only if S has infinite genus.*

Therefore, if we want to study the curve graph of a **surface of finite genus** using the Radó graph, we have to think it as a subgraph of it. A simple approach to do it is to take a random subset of vertices of a graph $G \in \mathcal{G}(\mathbb{N}, p)$ and then consider the vertex induced subgraph. It turns out that for a.e. $G \in \mathcal{G}(\mathbb{N}, p)$ the sequence $cl(G_n)$ is almost entirely determined.

Theorem 2.3.2. *For a.e. $G \in \mathcal{G}(\mathbb{N}, p)$ there is a constant $m_0 = m_0(G)$ such that if $n \geq m_o$ and $n'_r \leq n \leq n_{r+1}$, then $cl(G_n) = r$.*

The theorem states that if r is fixed and finite, the number of vertices must be finite as well, the proof can be find in [15, Bollobás p. 284]. Therefore, **it is not possible to obtain a curve graph by an uniform selection of vertices**.

Notice that the **clique number property is not generic** at all, unlike the others listed above, the clique number is the only property which actually depends on the genus of the surface. Therefore, is not a surprise that this property is highly restrictive in the plan of setting a generic model.

Prohibited configurations appear often in the literature, for example in [16, Alcazar 15] they want to ensure that a random graph does not have cycles, implying that the clique number is 2. A discrete MCMC algorithm was used to sample uniformly random trees of size n, with the `generate algorithm`. It produces a maximal tree of any not directed graph with n vertices uniformly among all the possible ones. In the appendix we summarize the results of this method, read [17, Broder 89] for the full analysis.

There's a chance of finding an analog of the generate algorithm to ensure a fixed clique number, but the scope of this work is to study rigid expansions in a simple probabilistic model. In this spirit, it remains to examine the plausibility of the Erdös-Rényi model and do an asymptotic analysis.

2.4 Erdös-Rényi as a model for the curve graph

2.4.1 Connectivity

Theorem 2.4.1. *Let $\omega(n)$ be a function that tends to infinity arbitrarily slow as n tends to infinity*

- *If $p \geq \frac{log(n)+\omega(n)}{n}$ then*

$$\lim_{n\to\infty} \mathbb{P}(G \in \mathcal{G}(n,p) \text{ is connected}) = 1$$

- *If $p \leq \frac{log(n)-\omega(n)}{n}$ then*

$$\lim_{n\to\infty} \mathbb{P}(G \in \mathcal{G}(n,p) \text{ is disconnected}) = 1$$

Here $\omega(n)$ represents the control over the convergence, in other words, the uncertainty. The theorem is just saying that in order to reduce $\omega(n)$ you must increase the size of the graph.

This theorem can be proven by first showing that for a large n almost all graphs consist of a single connected component and k isolated points, the theorem follows from a counting argument. The complete proof can be found in [18, Erdös-Rényi, p. 59].

2.4.2 Locally infinite

To fulfill this property we need to take $p(n)$ in a way that the expected degree grows along with the vertices, i.e $np \to \infty$. Notice that the conditions for the connectivity threshold are more than enough to ensure this. The full argument is stated in the following theorem:

Theorem 2.4.2. *Let $G \in \mathcal{G}(n,p)$ and $D_n \sim b(k; n-1, p)$ the random variable that describes the degree of a vertex in G. Take $p(n) = \frac{\epsilon}{n^a}$ with fixed $\epsilon, a > 0$ then:*

- *If $a \geq 1$ then $\{D_n\} \xrightarrow{p} c$, where c is a finite constant.*
- *If $a < 1$ then $\lim_{n\to\infty} \mathbb{P}(D_n \text{ is finite}) = 0$*

Proof. When $a = 1$ this is the Binomial's Poisson approximation.

For $a > 1$ take $k = 0$ so we have:

$$\lim_{n\to\infty} \mathbb{P}(D_n = 0) = \lim_{n\to\infty} \left(1 - \frac{\epsilon}{n^a}\right)^n = \lim_{n\to\infty} exp\left(ln\left(1 - \frac{\epsilon}{n^a}\right)^n\right) = \lim_{n\to\infty} exp\left(n \cdot ln\left(1 - \frac{\epsilon}{n^a}\right)\right)$$

If $f(n) = n \cdot ln\left(1 - \frac{\epsilon}{n^a}\right)$, then $\lim_{n\to\infty} f(n) = \lim_{n\to\infty} \frac{ln\left(1 - \frac{\epsilon}{n^a}\right)}{\frac{1}{n}}$. Using L'ôpital's rule for limits we obtain:

$$\lim_{n\to\infty} f(n) = \lim_{n\to\infty} \frac{\frac{1}{\left(1-\frac{\epsilon}{n^a}\right)} \cdot \left(\epsilon a n^{-a-1}\right)}{-1 \cdot n^{-2}} = \lim_{n\to\infty} \frac{\frac{n^a}{n^a-\epsilon} \cdot \left(\epsilon a n^{-a-1}\right) \cdot n^2}{-1} = \lim_{n\to\infty} -\frac{\epsilon a n}{n^a - \epsilon} = \lim_{n\to\infty} -\frac{\epsilon a}{a n^{a-1}}$$

from here, if $a > 1$ then $\lim_{n\to\infty} f(n) = 0$, if $a < 1$ then $\lim_{n\to\infty} f(n) = -\infty$. So we have

$$\lim_{n\to\infty} \mathbb{P}(D_n = 0) = \begin{cases} \lim_{n\to\infty} e^{f(n)} = 1, & \text{if } a > 1 \\ \lim_{n\to\infty} e^{f(n)} = 0, & \text{if } a < 1 \end{cases}$$

This conclude the first part of the theorem.

For $a < 1$ and $k > 0$:

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\Big(D_n = k\Big) = & \lim_{n\to\infty} \binom{n}{k} \cdot \left(\frac{\epsilon}{n^a}\right)^k \cdot \left(1 - \frac{\epsilon}{n^a}\right)^{n-k} \\ = & \lim_{n\to\infty} C_k \cdot n^k \cdot \frac{\left(\frac{\epsilon}{n^a}\right)^k}{\left(\frac{n^a-\epsilon}{n^a}\right)^k} \cdot \left(1 - \frac{\epsilon}{n^a}\right)^n \\ = & C_k \lim_{n\to\infty} n^k \cdot \left(\frac{1}{n^a - \epsilon}\right)^k \cdot \left(1 - \frac{\epsilon}{n^a}\right)^n \\ = & C_k \lim_{n\to\infty} n^s \cdot \left(1 - \frac{\epsilon}{n^a}\right)^n \end{aligned}$$

where $s = k(1 - a) > 0$, hence $\lim_{n\to\infty} \mathbb{P}\Big(D_n = k\Big) = 0, \forall k > 0$

□

So, when $a < 1$ we obtain the locally infinite property and this condition is always satisfied in the connectivity threshold.

Consider $X_k = X_k(G)$, the random variable that describes the number of vertices of degree k in a graph G. The following theorem gives a complete description of the degree distribution.

Theorem 2.4.3. *Let $\epsilon > 0$ be fixed, $\epsilon n^{-3/2} \leq p = p(n) \leq 1 - \epsilon n^{-3/2}$, let $k = k(n)$ be a natural number and set $\lambda_k = \lambda_k(n) = \mathbb{E}(X_k) = n \cdot b(k; n-1, p)$. Then the following assertions hold.*

- *If $\lim_{n\to\infty} \lambda_k(n) = 0$, then $\lim_{n\to\infty} \mathbb{P}(X_k = 0) = 1$.*

- *If* $\lim_{n\to\infty} \lambda_k(n) = \infty$, *then* $\lim_{n\to\infty} \mathbb{P}(X_k > t) = 1$ *for every fixed* t.
- *If* $0 < \lim_{n\to\infty} \lambda_k(n) < \lim_{n\to\infty} \lambda_k(n) < \infty$, *then* X_k *has asymptotically Poisson distribution with mean* λ_k*:*

$$P(X_k = r) \sim e^{\lambda_k} \cdot \lambda_k^r / r!$$

for every fixed r.

The $\epsilon n^{-3/2} \leq p = p(n) \leq 1 - \epsilon n^{-3/2}$ hypothesis is to rule out when we consider a loose upper bound on the expected degree of X_k, if $pn^2 \to \infty$ then almost every $G \in \mathcal{G}(n,p)$ consist of independent edges and isolated vertices.

The first assertion comes directly from Markov's inequality and implies that if k is a finite fixed number and $\lim \lambda_k(n) = 0$ then a.a.s. there are no vertices of degree k. In the second case there are an infinite number of vertices with degree k. In the third case we can describe explicitly the degree distribution. The complete arguments for proving this theorem can be found in [15, Bollobás, p. 61].

2.4.3 Clique number

As outlined for the Radó graph model, the clique number is a highly restrictive property, for the Erdös Rényi model it's not different. Let X_r be the random variable that counts the number of $r-$cliques in a graph G, we are looking for a threshold were:

$$\lim_{n\to\infty} \mathbb{E}(X_r) > 0 \text{ and } \lim_{n\to\infty} \mathbb{E}(X_{r+1}) = 0$$

But this is not possible if r is a fixed finite number. The closest we can get is stated in the following result:

Theorem 2.4.4. *Let* $r = r(n) = O(n^{1/3})$ *and let* $p = p(n)$, $0 < p < 1$, *be such that*

$$\binom{n}{r} p^{\binom{r}{2}} \to \infty \text{ and } \binom{n}{r+1} p^{\binom{r+1}{2}} \to 0$$

Then a.e $G \in \mathcal{G}(n,p)$ *has clique number* r.

This theorem can be proven using the calculations for $\mathbb{E}(X_r)$ and a first moment argument. A full proof can be find in [15, Bollobás, p. 290].

Given that $r(n)$ must grow along with n and $r = 3g - 3 + m$, this implies that the curve graph corresponds to a surface with infinite genus or with an infinite number of punctures.

2.4.4 Diameter

The diameter of a graph G, denoted by $diam(G)$, is the maximal distance between pairs of vertices of G.

If we want to model a surface of finite genus we must ensure infinite diameter, there are a number of theorems that describe the conditions under which this can be achieved [15, Bollobás, p. 259].

Following the past result we now must guarantee that the diameter is equal to 2. The idea is to have an analogue of Bering and Gaster result (Theorem 2.3.1).

Theorem 2.4.5. *Let d be a fixed integer d, if*

$$\frac{(pn)^{d-1}}{n} \to 0 \text{ and } \frac{(pn)^{d}}{n} \to \infty$$

then, a.a.s $G \in \mathcal{G}(n,p)$ has diameter d

The proof of this theorem is due to Klee and Larman and can be find in [19, Klee, Larman 81]. For $d = 2$ this means we are looking that $p(n) \to 0$ and $p^2\dot{n} \to \infty$, i.e $p(n) = \frac{f(n)}{n^{1/2}}$ where $f(n) \in o(n^{1/2})$ and $f(n) \to \infty$.

Using the expression for $p(n)$ in 2.4.2, we must have $a < \frac{1}{2}$ to ensure diameter 2.

To conclude, the following result states that Erdös-Rényi graphs, that asymptotically approach the Radó graph, naturally satisfy the condition for the diameter. This give us some insight about the interconnection between the object of study and the models when an asymptotic context is taken; the diameter condition of the curve graph appears naturally when modeling a surface of infinite genus.

Theorem 2.4.6. *If p is taken fixed $G(n,p)$ has diameter 2 with high probability*

Proof. Let X_n be random variable that counts the number of vertex pairs in a graph in $\mathcal{G}(n,p)$ with no common neighbors. By Markov's inequality we have that

$$\begin{aligned} \mathbb{P}(X_n \geq 1) \leq \mathbb{E}(X_n) &= \binom{n}{2} \cdot \mathbb{P}(\text{Two vertices doesn't have common neighbors}) \\ &= \binom{n}{2}(1-p^2)^{n-2} \end{aligned}$$

where $\lim_{n\to\infty} \binom{n}{2}(1-p^2)^{n-2} = 0$ □

2.5 Conclusions

Theorems 2.4.1, 2.4.2 and 2.4.5 give the thresholds where each of the properties of the curve graph of infinite genus surface are satisfied. Theorem 2.4.4 describe the asymptotic behaviour of the clique number.

In summary, if $\epsilon, a > 0$ are fixed real numbers and $\omega(n)$ a function that tends to infinity arbitrary slow, we have:

1. $p \geq \frac{log(n)+\omega(n)}{n} \implies$ a.a.s $G \in \mathcal{G}(np)$ is connected
2. $p = \frac{\epsilon}{n^a}$ with $a < 1 \implies$ a.a.s $G \in \mathcal{G}(np)$ is locally infinite
3. $p = (n) = \frac{f(n)}{n^{1/2}}$ with $f(n) \in o(n^{1/2})$ and $f(n) \to \infty$, particularly if $p(n) = \frac{\epsilon}{n^a}$ $a < \frac{1}{2}$ $\implies$ a.a.s $G \in \mathcal{G}(np)$ have diameter 2
4. $r = r(n) = O(n^{1/3})$ with $\binom{n}{r} p^{\binom{r}{2}} \to \infty$ and $\binom{n}{r+1} p^{\binom{r+1}{2}} \to 0$ *implies* a.a.s $G \in \mathcal{G}(n,p)$ has clique number r.

For the simplified form of $p(n) = \frac{\epsilon}{n^a}$ we can plot the following thresholds:

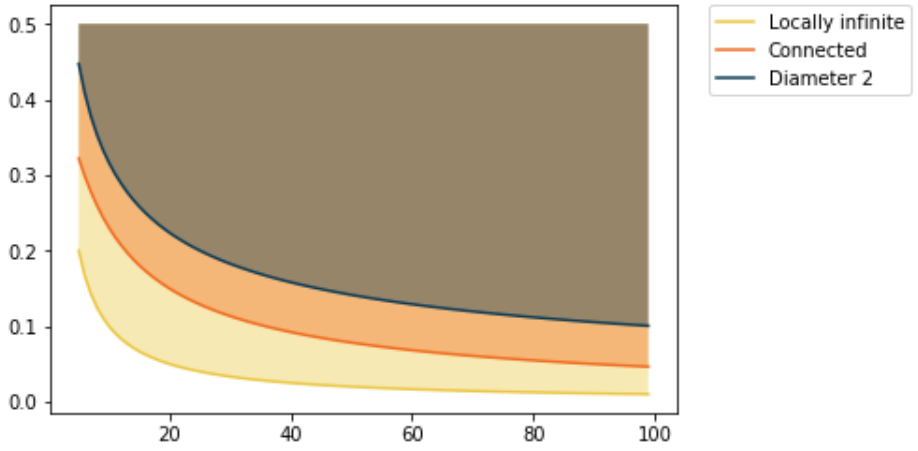

FIGURE 2.6: Thresholds for the properties of the curve complex

Some particular properties of the random graphs, such as a fixed clique number, will be linked to the asymptotic behaviour of the vertices. Meanwhile some conditions, like diameter 2, are strong enough that other properties come along.

The clique number property is highly restrictive when setting a generic model, thus a different techniques must be implemented. The problem with this approach is that we'll end up with a model so complicated that it won't be suitable for studying further phenomena.

There is a lot of progress in the study of random clique complexes that can help understand better the curve complex for infinite genus surfaces, for example in [20, Khale,09] we find the following result:

Theorem 2.5.1. *If $p = n^a$, with $a < -1/k$ or $a > -1/2k+1$, then the k-th homology group of $X(G(n,p))$ is almost always vanishing, and if $-1/k < a < -1/(k+1)$, then it is almost always nonvanishing*

Although interpretation might not be direct, the thresholds are within those defined for $C(S)$.

Studying a concept from the stochastic perspective could be helpful to understand its particularities. Also, the challenges to explain a phenomena in a specific framework, although it can be defined in a general manner.

From the Gaster and Bering result, we can also conclude that some objects, with apparently very particular constructions, end up being very much a like generic objects.

Using Erdös-Rényi model can bring insightful results of the asymptotic behavior of certain properties that in the Radó graph are given immediately.

Understanding the probabilistic nature of anything, in which computational operations are involved, can be helpful when optimizing procedures. This is extended in the next chapter.

Chapter 3

Computational experimentation

Nowadays Scientific Computing is one the most important tools that we have for Stochastic means. It becomes crucial when the problem cannot be solved by traditional experimentation or theoretical means. There are a number or reasons why this might happen, for example whenever experimentation may be dangerous, to expensive or time-consuming.

In this chapter we use the outlined probabilistic material to accomplish an efficient implementation of rigidity phenomenon in graphs. We describe our results with their technical difficulties and the actions taken to endure them.

All the computational experimentation was developed in `python`. We used `NetworkX` library to create and modify graphs.

3.1 Simulating Erdös-Rényi random graphs

There is a direct algorithm to obtain a graph in $G(n,p)$, it simulates a $Bernoulli(p)$ r.v. for each of the $\frac{n(n-1)}{2}$ possible edges. Thus, it runs in $O(n^2)$ time.

It is possible to execute faster algorithms for small values of p. It runs in $O(n+m)$ time, where m is the expected number of edges, which equals to $\frac{pn(n-1)}{2}$. This is the one that we use for generating all the graphs in our executions. Details in performance and accuracy can be found in [21, Batagelj, Brandes 05].

Visual aid is helpful while writing the code for the experimentation. In Figure 3.1 appear a set of graphs obtained with the built in algorithms for Erdös–Rényi graphs, fixing $n = 10$ an varying the parameter p.

Figure 3.2 show the execution times varying n

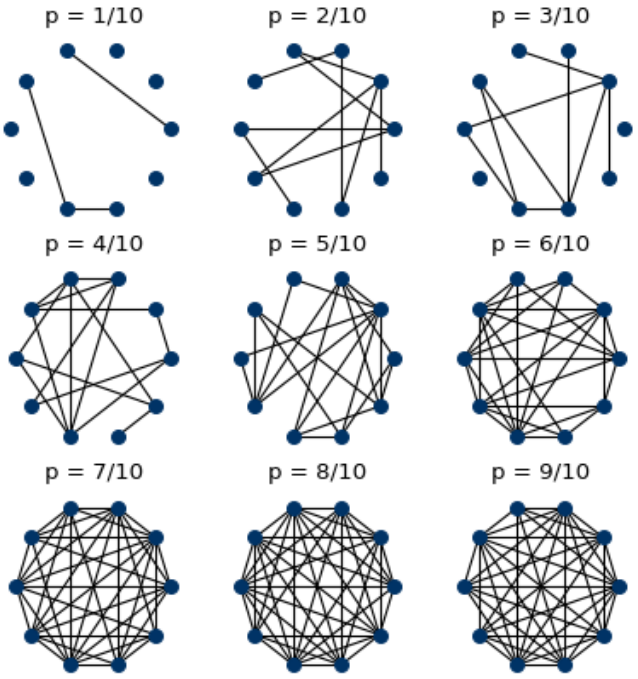

FIGURE 3.1: Erdös-Rényi random graphs with n fixed and varying p

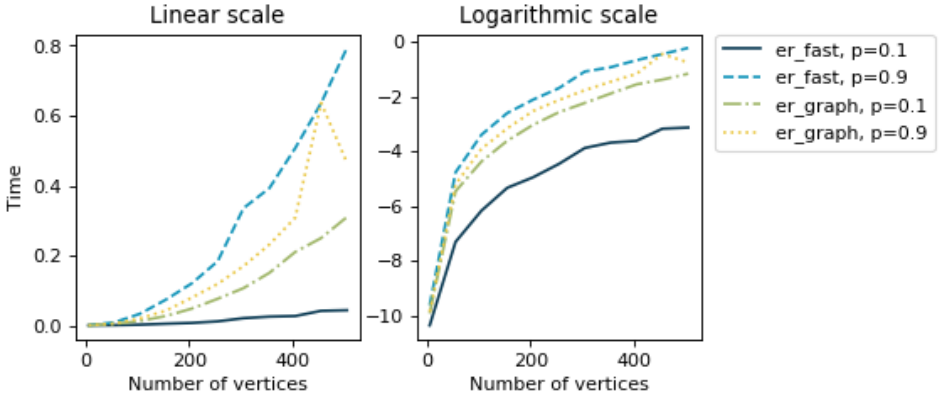

FIGURE 3.2: Execution times varying n. Both normal and logarithmic scale appear in the figure

3.2 Rigid expansions algorithms

A priory, the algorithm to determine a rigid expansion is supposed to be executed in a large amount of time. As the definition let us see, it depends on the size of G, and exponentially on the size of A; it must check among all the possible subsets of A, that's 2^k verifications, where $|A| = k$. Thus, it is important to do some optimizations and evaluate when they have more impact in the expected execution time according to the parameters taken.

The following is the straightforward algorithm for rigid expansions.

Rigid Expansions Algorithm

`Input:` `Random graph` G `(dictionary),`
`set of vertices` A `(array).`

`Output:` `Set of vertices obtained after expanding` A `(array)`

1. Initialize N as empty (the set of new vertices)
2. For every B, subset of A:
   If $\bigcap_{b \in B} N(b) = v$ and $v \notin A \cup N$:
   Add v to N
3. If N is not empty:
   Replace A by $A \cup N$ and return to step 1.
   Otherwise:
   Return A

To optimize memory in step two the iterations were indexed by `generators`.

For time-execution optimizations, we implemented the following:

1. **Consideration of isolated vertices and leaves.** None of the isolated vertices in G have any influence in rigid expansions, so they should not be consider. Also, whenever A contains a leave is convenient to ignore them; the unique neighbor of a leave, which we will call *petioles* should be automatically added in the first expansion an then it does not contribute to uniquely determine new vertices. This means that the input should be replace with:

$$A' = A \cup \{u : \exists x, N(x) = \{u\}\} - \{v : deg(v) \leq 1\}$$

   and add them again by the end of the expansions. This will be particularly helpful for small values of p.

2. **Relative size of A.** In Step 2, if A is big enough is faster to check if a vertex outside of A can be uniquely determinate by a subset of A. This can reduce dramatically the execution time when p is small; it reduce the size of revisions by

taking only the *effective* part of A, this is convenient to do whenever

$$k \cdot log(2) > log(n-k) + (kp) \cdot log(2)$$

where k is the size of A

3. **Restriction to effective subsets**. Calculations in Chapter 2 showed that some subsets are more likely to generate rigid expansions than others. This depend on the parameters of the space and the size of the subsets. If we restrict verifications to these effective subsets we can reduce the number of verifications.

With these optimizations we obtain the following algorithm

Optimized Rigid Expansions Algorithm

`Input:` `Random graph` G `(dictionary),`
`set of vertices` A `(array),`
n`(int) and` p`(float)`

`Output:` `Set of vertices obtained after expanding` A `(array)`

1. Remove isolated vertices and replace A with A'

2. Calculate the range of effective subsets.
   If $k \cdot log(2) > log(n-k) + (kp) \cdot log(2)$:
   For every $v \in V - A$:
   Take $C = A \cap N(v)$ and for every B, effective subset of C:
   If $\bigcap_{b \in B} N(b) = v$ and $v \notin A \cup N$:
   Add v to N

   Otherwise:
   For every B, effective subset of A:
   If $\bigcap_{b \in B} N(b) = v$ and $v \notin A \cup N$:
   Add v to N

3. If N is not empty:
   Replace A with $A \cup N$, initialize N as empty and return to step 3.
   Otherwise:
   Return $A \cup \{v : deg(v) \leq 1\}$

3.3 Time execution comparison

The task of finding the first rigid expansion of $A \subset V$ of size k in $G \in \mathcal{G}(n,p)$ depends on n, p, k. To keep track of the enhancements implemented we measured the execution time varying the parameters.

For each collection of n, p and k we calculated the mean execution time for 30 different rigid expansions, with and without optimizations. We took k in some proportion of n, explicitly: 1/4, 1/2 and 3/4.

Having a larger n impacts heavily when executing the non-optimized algorithm and even in certain thresholds the enhancement algorithms still take to much time. Also, considering the nature of rigid expansions and that we must execute multiple tests for each collection of parameters, n is fixed to be small, 20 in this case.

Results are presented in Figure 3.3

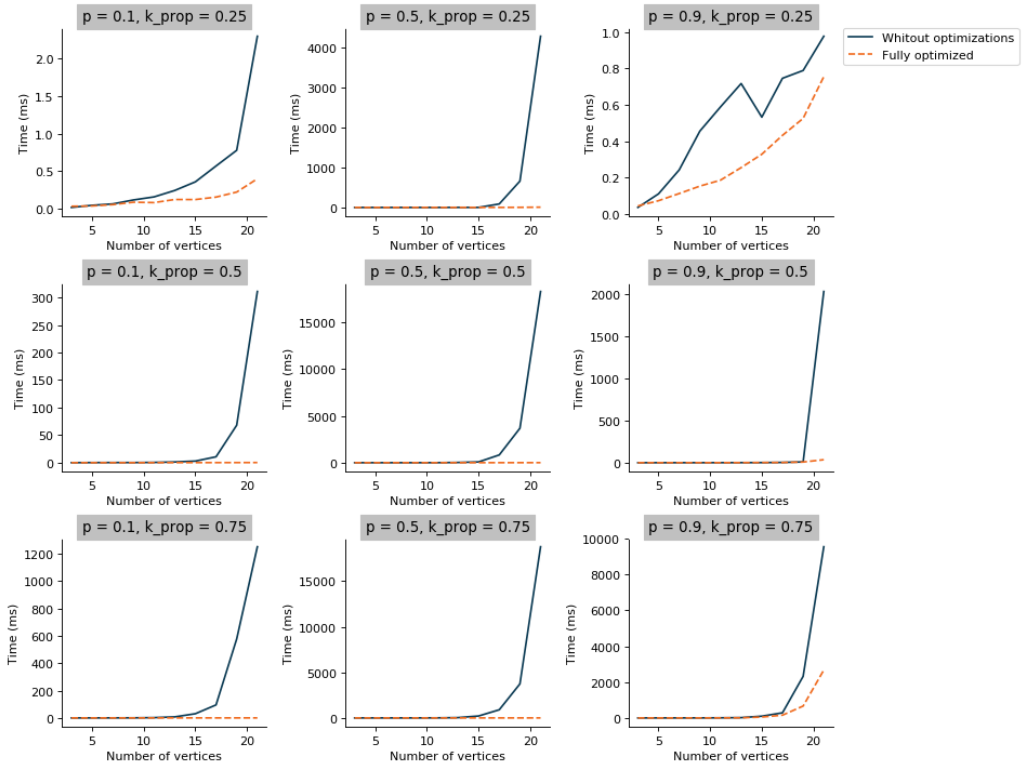

FIGURE 3.3: Mean execution time varying n, p and k for 30 rigid expansions. Measured in ms

Notice that this enhancements have an important impact in reducing the execution time. The results presented in Chapter 2 were obtained with these optimizations, so we can also conclude that accuracy is not compromised.

We expect to see an exponential behaviour for the execution time. To have a better comparison we can use the logarithmic scale.

Results in log scale are presented in Figure 3.4

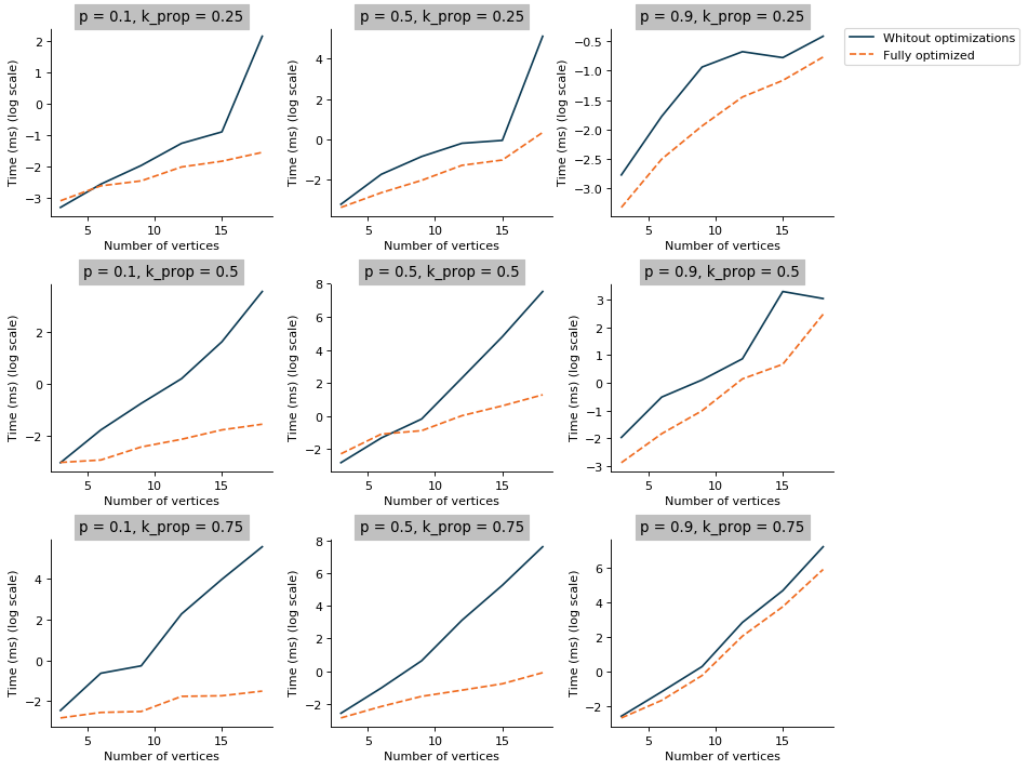

FIGURE 3.4: Mean execution time varying n, p and k for 30 rigid expansions. Measured in ms. Log scale

Notice that optimizations have more impact for lower values of p, this sounds reasonable giving that the firsts proposed optimizations are explicitly helpful for sparse graphs.

Increasing k has a big impact in performance that corresponds with still having to search among an exponentially bigger number of subsets. Even using the second and third optimization we are still obtaining the same exponential behaviour.

Further optimizations can be performed, such as excluding the parts of G that are not connected with vertices of A, nevertheless this will only will have impact in sparse graphs as well.

3.4 Conclusions

Probabilistic optimization and the understanding the combinatorial nature of rigid expansions allowed us to execute experiments for larger graphs that otherwise could not

have been performed.

In general, the use of probability theory in computational has demonstrated to be powerful and efficient. Algorithms based in random sampling provide state-of-the-art techniques due to their great degree of flexibility and reliability.

To name a few:

- The PageRank algorithm was the first method used by Google to order search results [22, Page 99]. It outputs a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page.
- In motion planning, the use of Rapidly-exploring random trees (RRTs) are one of the most successful algorithms [16, Alcazar 15]. Problems in motion planning consist on finding a collision-free path that connects an initial configuration of geometric bodies to a final goal configuration. A RRT is a rooted tree that grows from a starting configuration by using random samples from the search space.

But it also works in the opposite direction, the use of computational tools can bring value for theoretical means. For theoretical means it has allowed, for example, to verify whether the established conditions in a probabilistic model are sharp enough ([23, Aronshtam 13]).

Bottom line, the use of computational tools can be very helpful to understand a topic even in the most theoretical contexts.

Appendix A

The generate algorithm

In this section we describe the algorithm due to [17, Broder 89]. Given a not directed graph G with n vertices it produces a maximal tree of G sampled uniformly among all the possibles. For almost every graph the expected executed time of the algorithm is $O(n \cdot log(n))$ for each tree and O(n^3) in the worst cases.

One of the first algorithms published for this problem has execution time $O(n^5)$. It is based on the fact that the total number of directed trees in a graph can be explicitly calculated through a determinant of $n \times n$ size. The algorithm consider the edges of the graph labeled from 1 to m, each maximal tree is label by the set of its edges. This induces a lexicographic order in the set of trees and the same tree can be find calculation at most m determinants. Further improvements by [24, Colbourn 89] reduce the number of calculations, thus reducing the execution time to $O(n^3)$ or $O(L(n))$, where $L(n)$ is the execution time of multiplying matrices of size $n \times n$, but the new algorithms turn out to be far more complicated.

For an stochastic approach, consider a particle that moves among vertices in a graph. At each step it moves, choosing uniformly random, from the current vertex to a neighbor of it. This stochastic process is a Markov chain called **Random Walk**.

Generate Algorithm

`Input:` `Graph` G `(dictionary),`

`Output:` `Maximal tree` T `(dictionary)`

1. Choose a random vertex s of G (uniformly).
2. Simulate a simple random walk in G. It stops when every vertex gets visited.
3. For each i in $V - s$ collect the edge (j, i), the first entrance corresponds to the vertex where the particle was before it visited for the first time the vertex i. Let T be the collection of such edges.
4. Return T.

T is a maximal tree because it contains $|V|-1$ edges; it has an edge for every vertex in G except for s, and by construction it does not contains cycles.

The `Generate` algorithm is based in a simulation of Markov chains in the space of interest. In this case, the Markov chain has a stationary distribution $\pi_i = d_i / \sum_{j \in V} d_j$ where d_i is the degree of the vertex i. The pounded digraph associated to this chain $G_M = (V, E')$, is obtained by replacing each edge $\{i, j\} \in A$ by two directed edges; (i, j) with weight $1/d_i$ and (j, i) with weight $1/d_j$. The justification that the algorithm actually provides a method to sample with uniform distribution is summarized in the next three results, their proofs can be found in [17, Broder 89].

Let $\mathcal{T}_i(G_M)$ be the family of maximal directed trees of G_M with root i, when the root is not under consideration it will be denoted simply by $\mathcal{T}(G_M)$.

Theorem A.0.1. *Let M be a irreducible Markov chain in n states with stationary distribution $\pi_1, \ldots, \pi_n$. Let G_M be the weighted digraph associated to M. Then*

$$\pi_i = \frac{\sum_{T \in \mathcal{T}_i(G_M)} \omega(T)}{\sum_{T \in \mathcal{T}(G_M)} \omega(T)}$$

where $\omega(T) = \prod_{a \in A(T)} \omega(a)$, this means that the weight of the a directed tree is defined as the product of the weight of the edges of the tree.

We define the (*forward tree*) at time t, F_t as follows: Let I_t be the set of states visited before time $t+1$. For every $i \in I_t$, let $p(i, t)$ the first time that the state i was visited. The

root of the tree F_t is $\{(X_{p(i,t)}, X_{p(i,t)-1}) | i \in I_t - X_0\}$, where $(X_t)_{t\in\mathbb{N}}$ corresponds to the Markov chain given by the random walk. In other words F_t is constructed by overlapping the first entrance at each state with inverted orientation. Clearly F_t is a directed tree with root where each edge points from the leaves to the root.

Let C be the *covering time*, i.e. the first time that all the states where visited. Clearly for $t \geq C$ the tree F_t is a directed maximal tree and $F_t = F_C$. Note that with the past definition,the random walk $\{X_t\}$ in the vertices of G_M induces a Markov chain $\{F_t\}$ in the space of all directed trees of G_M, it is called forward trees chain.

For this chain every no maximal tree is a transitive state and every maximal tree is an absorbent state. Even more, the next theorem establish the distribution of F_C.

Theorem A.0.2. *With the same notation and conditions of the past theorem. Let F_C be the forward tree in time C. Then, for any maximal directed tree with root T of G_M we have*

$$\mathbb{P}(F_C = T) = \frac{\prod_{(i,j)\in T} P_{i,j}}{\sum_{T\in T(G_M)} \prod_{(i,j)\in T'} P_{i,j}}$$

Corollary A.0.2.1. *(Proof of the `Generate` algorithm) Let M be a simple random walk in a connected non-directed graph $G = (V, E)$, staring from a vertex s, G_M the directed graph associated to M and covering time C for G starting from the stationary distribution, we have F_C without considering direction is a maximal tree of G with random uniform distribution among all the maximal possible trees of G.*

This algorithm can be implemented using `Python`. Fixing G as the complete graph K_n it was possible, using the generate algorithm, to sample uniformly from the set of maximal trees with n vertices. In the figure A.1 appear a set of trees obtained with this method, which are drawn using the function `draw_random` of the `NetworX` library.

The expected execution time of the algorithm per tree is equal to $\mathbb{E}(C_s)$. It is known that for the connected graph $\mathbb{E}(C_v) = O(n^3)$, nevertheless in [25, Broder, Andrei 89] there is a proof that if the transition matrix of a random walk have the second greater eigenvalue bounded away from 1, then the expected covering time is only $O(n \cdot log(n))$. Almost every graph in the Erdös-Rényi model satisfy this condition when $p > \frac{c \cdot log(n)}{n}$, in particular when $p = \frac{1}{2}$ and for almost every d-regular graphs [26, Friedman 89]. In the figure A.2 appear the results of the execution time of the implemented algorithm.

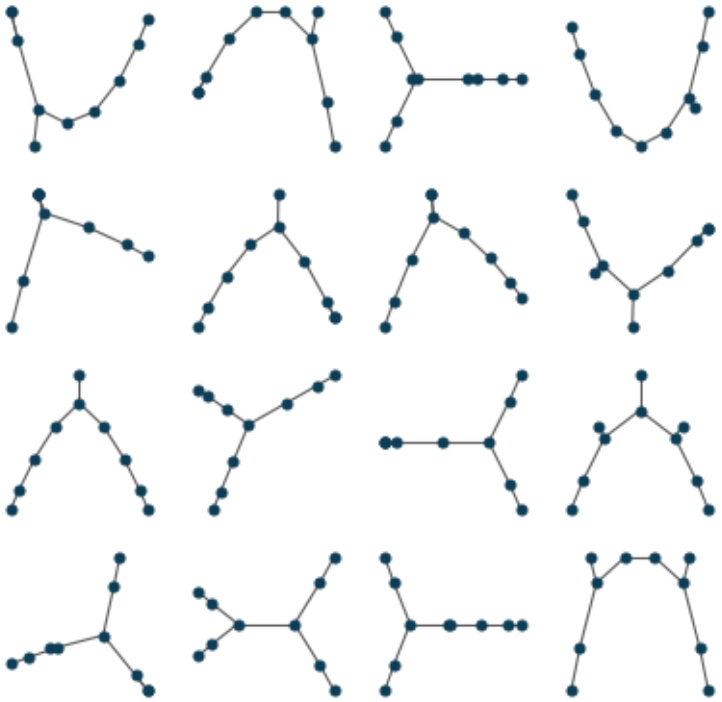

FIGURE A.1: Maximal tree chosen randomly with uniform distribution among all the possible ones in complete graph (10) of vertices.

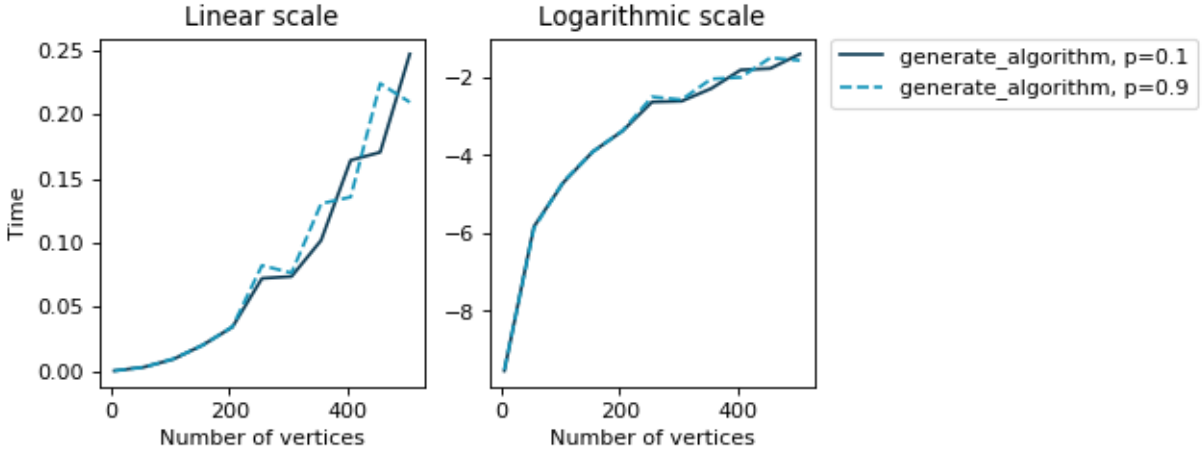

FIGURE A.2: Execution time of the algorithm in seconds varying the size of the tree. It appears the normal and the logarithmic scale

Bibliography

[1] N. V. Ivanov, "Automorphism of complexes of curves and of Teichmüller spaces," *Internat. Math. Res. Notices*, no. 14, pp. 651–666, 1997.

[2] J. Aramayona and C. J. Leininger, "Finite rigid sets in curve complexes," *J Topology Anal*, vol. 5, no. 2, 2012.

[3] J. Hernández Hernández, "Exhaustion of the curve graph via rigid expansions," *Glasg. Math. J.*, vol. 61, no. 1, pp. 195–230, 2019.

[4] B. Farb and D. Margalit, *A primer on mapping class groups*, vol. 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.

[5] A. Hatcher and W. Thurston, "A presentation for the mapping class group of a closed orientable surface," *Topology*, vol. 19, no. 3, pp. 221–237, 1980.

[6] H. A. Masur and Y. N. Minsky, "Geometry of the complex of curves. I. Hyperbolicity," *Invent. Math.*, vol. 138, no. 1, pp. 103–149, 1999.

[7] J. Aramayona and J. Souto, "Rigidity phenomena in the mapping class group," in *Handbook of Teichmüller theory. Vol. VI*, vol. 27 of *IRMA Lect. Math. Theor. Phys.*, pp. 131–165, Eur. Math. Soc., Zürich, 2016.

[8] M. Korkmaz, "Automorphisms of complexes of curves on punctured spheres and on punctured tori," *Topology Appl.*, vol. 95, no. 2, pp. 85–111, 1999.

[9] F. Luo, "Automorphisms of the complex of curves," *Topology*, vol. 39, no. 2, pp. 283–298, 2000.

[10] J. Aramayona and C. J. Leininger, "Exhausting curve complexes by finite rigid sets," *Pacific J. Math.*, vol. 282, no. 2, pp. 257–283, 2016.

[11] A. Costa and M. Farber, "Large random simplicial complexes, II; the fundamental group," *J. Topol. Anal.*, vol. 9, no. 3, pp. 441–483, 2017.

[12] P. Erdős and A. Rényi, "Asymmetric graphs," *Acta Math. Acad. Sci. Hungar.*, vol. 14, pp. 295–315, 1963.

[13] P. J. Cameron, "The random graph," in *The mathematics of Paul Erdős, II*, vol. 14 of *Algorithms Combin.*, pp. 333–351, Springer, Berlin, 1997.

[14] E. A. Bering, IV and J. Gaster, "The random graph embeds in the curve graph of any infinite genus surface," *New York J. Math.*, vol. 23, pp. 59–66, 2017.

[15] B. Bollobás, *Random graphs*, vol. 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second ed., 2001.

[16] V. Alcázar, S. Fernández, D. Borrajo, and M. Veloso, "Using random sampling trees for automated planning," *AI Commun.*, vol. 28, no. 4, pp. 665–681, 2015.

[17] A. Broder, "Generating random spanning trees," in *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, SFCS '89, (Washington, DC, USA), pp. 442–447, IEEE Computer Society, 1989.

[18] P. Erdős and A. Rényi, "On random graphs. I," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.

[19] V. Klee and D. Larman, "Diameters of random graphs," *Canadian J. Math.*, vol. 33, no. 3, pp. 618–640, 1981.

[20] M. Kahle, "Topology of random clique complexes," *Discrete Mathematics*, vol. 309, no. 6, pp. 1658–1671, 2009.

[21] V. Batagelj and U. Brandes, "Efficient generation of large random networks," *Phys. Rev. E*, vol. 71, p. 036113, Mar 2005.

[22] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[23] L. Aronshtam, N. Linial, T. Ł uczak, and R. Meshulam, "Collapsibility and vanishing of top homology in random simplicial complexes," *Discrete Comput. Geom.*, vol. 49, no. 2, pp. 317–334, 2013.

[24] C. J. Colbourn, R. P. J. Day, and L. D. Nel, "Unranking and ranking spanning trees of a graph," *J. Algorithms*, vol. 10, no. 2, pp. 271–286, 1989.

[25] A. Z. Broder and A. R. Karlin, "Bounds on the cover time," *Journal of Theoretical Probability*, vol. 2, pp. 101–120, Jan 1989.

[26] J. Friedman, J. Kahn, and E. Szemerédi, "On the second eigenvalue of random regular graphs," in *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, (New York, NY, USA), pp. 587–598, ACM, 1989.